Unit - I

Chapter 1

MARKETING RESEARCH

Objective:

The objective of this chapter is to understand:

- The meaning of Marketing Research
- The difference between basic and applied research
- The various classifications of Marketing Research
- The scope of Marketing Research
- The various methods of Marketing Research

Definition of Research

Research always starts with a question to which we seek an answer using scientific methods. We define the question as a 'Problem".

Research is often described as an active, diligent, and systematic process of inquiry aimed at discovering, interpreting and revising facts.

The word *research* is derived from the French language; its literal meaning is 'to investigate thoroughly'.

Undertaking research is basically applying scientific methods to find solution to a problem. It is a systematic and explorative study carried out to analyse and apply various solutions to a defined problem.

Research can be classified into two broad categories:

1. Basic Research and
2. Applied Research

Basic research

Basic research is also called *fundamental* or *pure* research. As the name itself refers, Basic Research is of basic nature which is not carried out in response to a problem. It is more educative, towards understanding the fundamentals and aim at expanding the knowledge base of an individual or organisation. It does not have any commercial potential.

Applied research

Applied Research on the otherhand is carried out to seek alternate solutions for a problem at hand. Applied research is done to solve specific, practical questions; its primary aim is not to gain knowledge. It specifies possible outcomes of each of the alternatives and its commercial implications.

Applied research can be carried out by academic or industrial institutions. Often, an academic instituion such as a university will have a specific applied research program funded by an industrial partner interested in that program. Electronics, informatics, computer science, process engineering and drug design are some of the common areas of applied research.

Applied research can further be divided into:

1. **Problem-solving research**: It involves research oriented towards a crucial problem facing the organisation which may be issue specific.
   Ex: How do we improve the communication skills of our employees?
2. **Problem-oriented research**: The research is oriented towards a crucial problem facing the organisation. It is undertaken inside the organisation or by an external consultant on its behalf. This research is conceptual in nature and newer innovative techniques of problem-solving are applied.
   Ex: How to improve the production yield from machine X using modern techniques?

Activity 1:

Classify the following on the basis of basic research or aplied research:

1. *Research carried out to understand the disease Typhoid.*
2. *Research carried out to understand the methods to improve the productivity of people working on Machine X.*
3. *Research carried out to study the impact of absenteism on productivity.*
4. *Understanding a new software programme which has been launched.*

Defining Marketing Research:

Marketing research (also called **consumer research**) is a form of business research. The field of marketing research as a statistical science was pioneered by Arthur Nielsen with the founding of the ACNielsen Company in 1923.

Marketing research is a systematic and objective study of problems pertaining to the marketing of goods and services. It is applicable to any area of marketing.

Research is the only tool an organization has to keep in contact with its external operating environment. In order to be proactive and change with the environment simple questions need to be asked:

- What are the customer needs and how are they changing? How to meet these changing needs? What do the customers think about existing products or services? What more are they looking at?
- What are the competitors doing to retain customers in this environment? Are their strategies exceeding or influencing yours? What should you do to be more competitive?
- How are macro and micro environmental factors influencing your organisation? How will you react t this environment?

Authors have defined Marketing Research in many ways:

- Kotler (1999) defines marketing research as 'systematic problem analysis, model-building and fact-finding for the purpose of improved decision-making and control in the marketing of goods and services'.
- The American Marketing Association (AMA, 1961) defines it as 'the systematic gathering, recording and analyzing of data relating to the marketing of goods and services'.
- Green and Tull have defined marketing research as the systematic and objective search for and analysis of information relevant to the identification and solution of any problem in the field of marketing.

The aim of marketing management is to satisfy the needs of the consumer. Marketing research helps in achieving this. Marketing research is a systematic and logical way of assessing ways of satisfying customer needs.

According to all the above definitions, Marketing Research starts by stating the problem or the issue to be investigated; indicate what kind of information is required to resolve the problem; identify where and how to get it; specify the

methodology for analyzing the research findings; sum up the research findings and then suggest the best solution for marketing decision making.

Scope of marketing research:

Marketing research can be used in:

- **Product Management:** One of the major scope of marketing research is to manage the current products and new products. In product management Marketing Research is helpful in
    - **Competitive Intelligence –** To understand the competitive product stretegy.
    - **Prelaunch strategy for new products**
    - **Test Marketing –** To monitor the performance of the brand by launching in a select area and then taking it across the country. In other words it is a small-scale product launch used to determine the likely acceptance of the product when it is introduced into a wider market.
    - **Concept testing** - to test the acceptance of a concept by target consumers.
- **Sales analyis**: Marketing research is used to study the sales trend and make suitable strategies when required. It is used to
    - Assess market potential
    - Estimation of demand for a product
    - Market share estimation
    - Study seasonal variation for a product
    - Market segmentation studies
    - Estimate size of the market
    - Need analysis to find out where the product fits in
- **Corporate Research**: Marketing Research is used to analyse the corporate effectiveness. Some examples are:

    - Assessing the image of the company
    - Knowledge of the company activities
- **Advertising Research**: Advertising is an arena in which Marketing Research is extensively used. Some scope are:
    - Readership feedbacks – Mainly carried out for newspapers and magazines
    - Advertising Recall – To assess the recall of telivision or other advertising and thereby assess its effectiveness.
- **Syndicated Research**: This is compiled by agencies on a regular basis and sold to organisations on subscription basis.

All of these forms of marketing research can be classified as either ***problem-identification research*** or as ***problem-solving research***.

A similar distinction exists between *exploratory research* and *conclusive research*.

- **Exploratory** research provides insights into and comprehension of an issue or situation. It should draw definitive conclusions only with extreme caution.
- **Conclusive research** draws conclusions: the results of the study can be generalized to the whole population.

Research can also be:

- **Primary Marketing Research**: It is research conducted by an organisation for its own purpose which addresses its requirements. It is generally expensive but is specific and objective to the organisation's requirement.
- **Secondary Marketing Research**: This is used if the organisation is considering extending its business into new markets or adding new services or product lines. This type of research is based on information obtained from studies previously performed by government agencies,

chambers of commerce, trade associations and other organizations. This also includes Census Bureau information.

In other terms this is research published previously and usually by someone else. Secondary research costs less than primary research, but seldom comes in a form that exactly meets the needs of the researcher. It can cater to anyone who wishes to use the data.

This data can be found in local libraries or on the Web, but books and business publications, as well as magazines and newspapers, are also great sources.

Hence, Primary research delivers more specific results than secondary research, which is an especially important while launching a new product or service. In addition, primary research is usually based on statistical methodologies that involve sampling as small as 1 percent of a target market. This tiny representative sample can give an accurate representation of a particular market.

With the advance in technology a lot of software have been developed which help in primary market research online and offline thereby making analysis and interpretation easier.

The ideal way to conduct Marketing Research is to do secondary research first and then do the primary research for the data not available form secondary sources.

Hence, secondary research lays the groundwork and primary research helps fill in the gaps. By using both types of market research, organisations get a better picture of their market and have the information they need to make important business decisions.

Activity 2:

List down areas in marketing where Marketing Research would be helpful:

……
……
……

Marketing Research Methods

Methodologically, marketing research uses four types of research designs, namely:

- **Qualitative marketing research** – This is generally used for exploratory purposes. The data collected is qualitative and focuses on people's opinions and attitudes towards a product or service.
  The respondents are generally few in number and the findings cannot be generalised tot eh whole population. No statistical methods are generally applied.
  Ex: Focus groups, In-depth interviews, and Projective techniques
- **Quantitative marketing research** – This is generally used to draw conclusions for a specific problem. It tests a specific hypothesis and uses random sampling techniques so as to infer from the sample to the population. It involves a large number of respondents and analysis is carried out using statistical techniques.
  Ex: Surveys and Questionnaires
- **Observational techniques** - The researcher observes social phenomena in their natural setting and draws conclusion from the same. The observations can occur **cross-sectionally** (observations made at one time) or **longitudinally** (observations occur over several time-periods)
  Ex: Product-use analysis and computer cookie tracing
- **Experimental techniques** – Here, the researcher creates a quasi-artificial environment to try to control spurious factors, then manipulates at least one of the variables to get an answer to a research
  Ex: Test marketing and Purchase laboratories

More than one research designs could be used at a time. They may start with secondary research to get background information, then conduct a focus group (qualitative research design) to explore the issues. Finally they might do a full nation-wide survey (quantitative research design) in order to devise specific recommendations for the client organisation.

Difference between Market Research and Marketing Research

Generally Market Research and Marketing Research are confused to be the same. But there is a clear distinction between the both.

***Market Research*:** Market Research involves researching a specific industry or market.

Ex: Researching the automobile industry to discover the number of competitors and their market share.

Marketing Research: Marketing Research analyses a given marketing opportunity or problem, defines the research and data collection methods required to deal with the problem or take advantage of the opportunity, through to the implementation of the project. It is a more systematic method which aims to discover the root cause for a specific problem within an organisation and put forward solutions to that problem.

Ex: Research carried out to analyze and find solution for increasing turnover in an organisation.

Summary:

The meaning of research is to investigate thoroughly. It can be divided into basic and applied research. Basic research is the pure research which is more educative. Applied research is carried out to seek alternate solution for a problem.

Applied research can further be classified as problem-solving research and problem-oriented research depending upon the research problem at hand.

Marketing research is a systematic and objective study of problems pertaining to the marketing of goods and services. It is applicable to any area of marketing. For ex. Product management, sales, advertising research, etc.

Marketing Research can be Primary Market Research or Secondary Market research depending on the data source used. It can be qualitative or quantitative research depending upon the nature of the research.

Marketing Research is different from Market Research wherein the former is oriented towards solving marketing problems and the latter is market related.

Questions:

Answer the following questions:

1. What is Marketing Research?
2. Define Marketing Research as stated by various authors.

3. Why should you conduct Marketing Research?
4. Differentiate between basic research and applied research.
5. List down the scope of Marketing Research.
6. What is the difference between problem-solving research and problem-oriented research?
7. Give an example of situations where you will use exploratory research.
8. Enumerate the differences between primary and secondary marketing research.
9. Give examples of Qualitative and Quantitative Marketing Research.
10. Is there any difference between Market Research and Marketing Research? Explain

Chapter 2

THE MARKETING RESEARCH PROCESS

Objectives:

- To understand the Marketing Research Process.
- To learn in detail about the various steps in the Marketing Research Process

The Marketing Research Process

As we saw earlier Marketing Research is very much essential to make strategic decisions which are important for the growth of the organisation. It helps in making the right decisions systematically using statistical methods.

Marketing Research reduces the uncertainty in the decision-making process and increase the probability and magnitude of success if conducted in a systematic, analytical, and objective manner. Marketing research by itself does not arrive at marketing decisions, nor does it guarantee that the organization will be successful in marketing its products. It is only a tool which helps in the decision making process.

The Marketing Research Process involves a number of inter-related activities which have bearing on each other. Once the need for Marketing Research has been established, broadly it involves the steps as depicted in Figure 1 below:

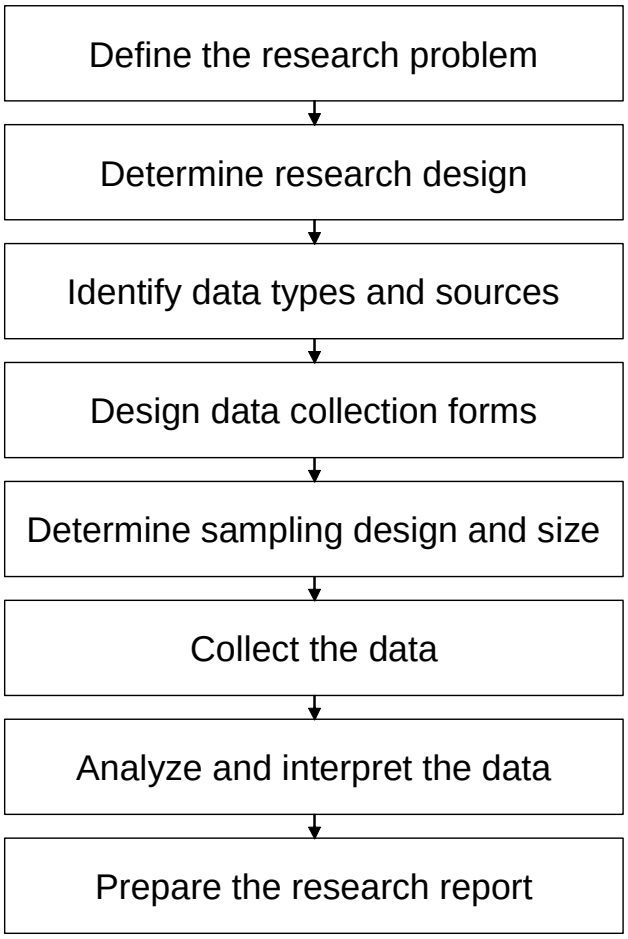


Figure1: The steps in Marketing Research Process

Let us now know in detail about the various steps involved in the Marketing Research Process.

1. Define the research problem

The first step in Marketing is to define the research problem. A problem well defined is half-solved. If a problem is poorly defined, a good research design cannot be developed.

The decision problem faced by the organisation must be translated into a market research problem in the form of questions. These questions must define the information that is required to make the decision and how this information can be obtained. This way, the decision problem gets translated into a research problem.

For example, a decision problem may be whether to launch a new product. The corresponding research problem might be to assess whether the market would accept the new product.

In order to define the problem more precisely, an exploratory research can be carried out. Survey of secondary data, pilot studies or experience surveys are some of the popular methods.

2. Determine research design

The research design specifies the method and procedure for conducting a particular study.

As studied already, marketing research and hence the research designs can be classified into one of three categories

- Exploratory research
- Descriptive research
- Causal research

This classification is based on the objective of the research. In some cases the research will fall into one of these categories, but in other cases different phases of the same research project will fall into different categories.

Problems are formulated clearly in exploratory research. It aims at clarifying concepts, gathering explanations, gaining insight, eliminating impractical ideas,

and forming hypotheses. Exploratory research can be performed using a literature search, surveying certain people about their experiences, focus groups, and case studies. During the survey, exploratory research studies would not try to acquire a representative sample, but rather, seek to interview those who are knowledgeable and who might be able to provide insight concerning the relationship among variables. Case studies can include contrasting situations or benchmarking against an organization known for its excellence. Exploratory research may develop hypotheses, but it does not seek to test them. Exploratory research is characterized by its flexibility.

A descriptive study is undertaken when the researcher wants to know the characteristics of certain groups such as age, sex, educational level, income, occupation, etc. Descriptive research is more rigid than exploratory research and seeks to describe users of a product, determine the proportion of the population that uses a product, or predict future demand for a product. Descriptive research should define questions, people surveyed, and the method of analysis prior to beginning data collection. In other words, the who, what, where, when, why, and how aspects of the research should be defined. Such preparation allows one the opportunity to make any required changes before the costly process of data collection has begun.

There are two basic types of descriptive research: longitudinal studies and cross-sectional studies. Longitudinal studies are time series analyses that make repeated measurements of the same individuals, thus allowing one to monitor behavior such as brand-switching. However, longitudinal studies are not necessarily representative since many people may refuse to participate because of the commitment required. Cross-sectional studies sample the population to make measurements at a specific point in time. A special type of cross-sectional analysis is a cohort analysis, which tracks an aggregate of individuals who

experience the same event within the same time interval over time. Cohort analyses are useful for long-term forecasting of product demand.

Causal research seeks to find cause and effect relationships between variables. It accomplishes this goal through laboratory and field experiments.

3. Identify data types and sources

The next step is to determine the sources of data to be used. The researcher has to decide whether to go for primary data or secondary data. Sometimes a combination of both is used.

Before going through the time and expense of collecting primary data, one should check for secondary data that previously may have been collected for other purposes but that can be used in the immediate study. Secondary data may be internal to the firm, such as sales invoices and warranty cards, or may be external to the firm such as published data or commercially available data. The government census is a valuable source of secondary data.

Secondary data has the advantage of saving time and reducing data gathering costs. The disadvantages are that the data may not fit the problem perfectly and that the accuracy may be more difficult to verify for secondary data than for primary data.

Many a time the secondary data might have to be supplemented by primary data originated specifically for the study at hand. Some common types of primary data are:

- Demographic and socioeconomic characteristics
- Psychological and lifestyle characteristics
- Attitudes and opinions
- Awareness and knowledge - for example, brand awareness
- Intentions - for example, purchase intentions. While useful, intentions are not a reliable indication of actual future behavior.
- Motivation - a person's motives are more stable than his/her behavior, so motive is a better predictor of future behavior than is past behavior.
- Behavior

Primary data can be obtained by communication or by observation. Communication involves questioning respondents either verbally or in writing. This method is versatile, since one needs to only ask for the information; however, the response may not be accurate. Communication usually is quicker and cheaper than observation. Observation involves the recording of actions and is performed by either a person or some mechanical or electronic device. Observation is less versatile than communication since some attributes of a person may not be readily observable, such as attitudes, awareness, knowledge, intentions, and motivation. Observation also might take longer since observers may have to wait for appropriate events to occur, though observation using scanner data might be quicker and more cost effective. Observation typically is more accurate than communication.

Personal interviews have an interviewer bias that mail-in questionnaires do not have. For example, in a personal interview the respondent's perception of the interviewer may affect the responses.

4. Design data collection forms

Once it has been decided to obtain primary data, the mode of collection needs to be decided. Two methods are available for data collection:

1. Observational methods
2. Survey methods

Observational methods: As the name itself suggests, the data are collected through observation. An observer observes and records the data faithfully and accurately. This may be suitable in case of some studies but is not useful to observe attitudes, opinions, motivations and other intangible states of mind.

Also in this method, the data collected is non-reactive, as it does not involve the respondent.

Surveys: It is one of the most common methods of collecting data for primary marketing research. Surveys can be:

- Personal: The information is sought through personal interviews. A questionnaire is prepared and administered to the respondent during the interview. This is a detailed method of collecting information.
- Telephonic: This is suitable if limited information is sought in a fixed time frame.
- Mail: Here, the questionnaire is sent out in mail and the response is sought. Timely response cannot be sought in this method as there is no control over the survey. All the people to whom the mail was sent may not respond.

Sometimes a combination of two or more methods may be used. Whatever be the method, a structured questionnaire is required to be used. The questionnaire is an important tool for gathering primary data. Poorly constructed questions can result in large errors and invalidate the research data, so significant effort should be put into the questionnaire design. The questionnaire should be tested thoroughly prior to conducting the survey.

5. Determine sampling design and size

A sampling plan is a very important part of the research process. The marketing researcher has to decide whether it will be a sample survey or a census. Definitely a sample survey has its distinct merits.

The population from which the sample has to be drawn has to be well defined. A broad choice is to be made between probability sampling and non-probability sampling. The sample design is then chosen depending on the suitability and the availability of the sample frame.

The size of the sample chosen is based on statistical methods. This is well defined and also reproduces the characteristics of the population. In practice, however, this objective is never completely attained on account of the occurrence of two types of errors – errors due to bias in the selection and sapling errors.

6. Collect the data

The next step is to collect the data for which the research process has been spelled out. The interviewing and the supervision of field work should be looked into. One of the most difficult tasks is interviewing for marketing research. Many a time the respondents may not part with crucial information unless approached with tact and intelligence. Supervision of field work is important to ensure timely and proper completion of the field survey.

If this is not carried out properly, then there results an interview error which may be detrimental to marketing research.

7. Analyze and interpret the data

The next step is to analyze the data that has been collected from the field survey. The raw data is transformed into the right format. First, it is edited so that errors can be corrected or omitted. The data is then coded; this procedure converts the edited raw data into numbers or symbols. A codebook is created to document how the data is coded. Finally, the data is tabulated to count the number of samples falling into various categories.

Simple tabulations count the occurrences of each variable independently of the other variables. Cross tabulations, also known as contingency tables or cross tabs, treats two or more variables simultaneously.

Cross tabulation is the most commonly utilized data analysis method in marketing research. Many studies take the analysis no further than cross tabulation.

Once the tabulation is done, the following analysis can be carried out.

- **Conjoint Analysis**: The conjoint analysis is a powerful technique for determining consumer preferences for product attributes.
- **Hypothesis Testing**: The null hypothesis in an experiment is the hypothesis that the independent variable has no effect on the dependent

variable. The null hypothesis is expressed as H0. This hypothesis is assumed to be true unless proven otherwise. The alternative to the null hypothesis is the hypothesis that the independent variable does have an effect on the dependent variable. This hypothesis is known as the alternative, research, or experimental hypothesis and is expressed as H1.

Once analysis is completed, make the marketing research conclusion. In order to analyze whether research results are statistically significant or simply by chance, a test of statistical significance can be run.

8. Prepare the research report

All the research findings have to be compiled in a report to be then presented to the organization. The format of the marketing research report varies with the needs of the organization. The report often contains the following sections:

- Authorization letter for the research
- Table of Contents
- List of illustrations
- Executive summary
- Research objectives
- Methodology
- Results
- Limitations
- Conclusions and recommendations
- Appendices containing copies of the questionnaires, etc.

The report has to be written with objectivity, coherence, clarity in the presentation of the ideas and use of charts and diagrams. Sometimes, the study might also throw up one or more areas where further investigation is required.

Summary:

Marketing Research reduces the uncertainty in the decision-making process and increase the probability and magnitude of success if conducted in a systematic, analytical, and objective manner.

The Marketing Research Process involves a number of inter-related activities which have bearing on each other. Each and every step plays an important role in the research process.

Questions:

1. List out the various steps involved in the Marketing Research Process.
2. It is very important to define the research problem, explain.
3. Classify research designs and explain the relevance of each.
4. What are the types of data sources?
5. Enumerate the methods available for data collection
6. Is it important to determine the sample size? Explain.
7. How will you analyze the data collected from the research?
8. How will you prepare a research report?
9. The various steps in the Marketing Research Process are inter-related. Explain.
10. Does the Market Research Process serve as a framework for finding solution to the research problem at hand? Evaluate critically.

Chapter 3

RESEARCH DESIGN

Objective:

- To understand the meaning of Research Design.
- To study about the various types of Research designs.
- To understand the type of research design to use for specific problems.

The Research Design

Research design provides the glue that holds the research project together. A design is used to structure the research, to show how all of the major parts of the research project -- the samples or groups, measures, treatments or programs, and methods of assignment -- work together to try to address the central research questions.

According to Green and Tull: A Research Design is the specification of methods and procedures for acquiring the information needed. It is the over-all operational pattern or framework of the project that stipulates what information is to be collected from which sources by what procedures.

Hence it is clear that Research design is the blueprint for research. It lays down the methodology involved in the collection of information and arriving at meaningful conclusions from the same.

There are many methods for studying and tackling a problem, but there are no perfect methods. Many times more than one method could be used in the research process.

There are many classifications accepted for a Research Design. One of the most accepted classification is grouping it under three types:

1. Exploratory
2. Descriptive and
3. Causal

This can be depicted as in figure 2 given below:

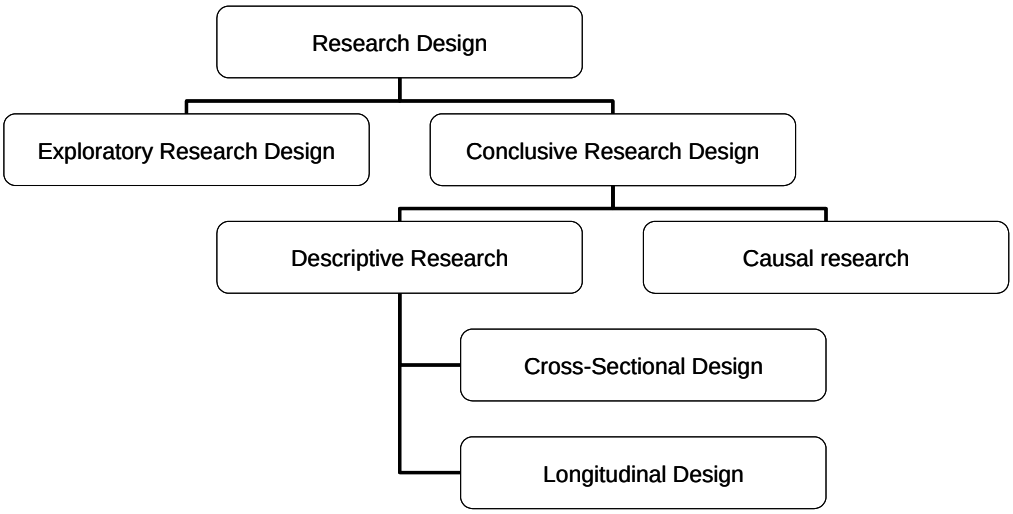


Figure 2: Classification of Research Designs

Exploratory Research Design

As the term suggests, exploratory research is often conducted because a problem has not been clearly defined as yet, or its real scope is as yet unclear. It is a process of discovery wherein you uncover as many ideas as possible. It allows the researcher to familiarize him/herself with the problem or concept to be studied, and perhaps generate hypothesis to be tested. It expands knowledge. It is the initial research, before more conclusive research is undertaken.

Exploratory research helps determine the best research design, data collection method and selection of subjects.

Another common reason for conducting exploratory research is to test concepts before they are launched in the marketplace, always a very costly endeavor. In concept testing, consumers are provided either with a written concept or a prototype for a new, revised or repositioned product, service or strategy.

Exploratory research relies more on secondary data. It does not have a rigid design as the researcher themselves are not very well versed with the subject and are trying to gain knowledge of the same. Hence it can be quite informal, relying on secondary research such as reviewing available literature and/or data, or qualitative approaches such as informal discussions with consumers, employees, management or competitors, and more formal approaches through in-depth interviews, focus groups, projective methods, case studies or pilot studies.

The results of exploratory research are not usually useful for decision-making by themselves, but they can provide significant insight into a given situation. The findings of this can be used to develop the research further. Points can be arrived at which requires to apply the other methodologies.

Conclusive Research Design

Descriptive research is also used to generate hypotheses but generally has more information available than in exploratory research. Descriptive research is

usually conducted to characterize one or more variables within a population, particularly in relation to person, place, and time.

As the name indicates, conclusive research is meant to provide information that is useful in reaching conclusions or decision-making. It is mostly quantitative in nature, in the form of numbers that can be quantified and summarized. It relies on both secondary data, particularly existing databases that are reanalyzed to shed light on a different problem than the original one for which they were constituted, and primary research, or data specifically gathered for the current study**.**

The purpose of conclusive research is to provide a reliable or representative picture of the population through the use of a valid research instrument. In the case of formal research, it will also test hypothesis.

Conclusive research can be sub-divided into two categories:

1. Descriptive or statistical research, and
2. Causal research

Descriptive Research

Descriptive research or statistical research provides data about the population or universe being studied. It describes the "who, what, when, where and how" of a situation and not what caused it. Therefore, descriptive research is used when the objective is to provide a systematic description that is as factual and accurate as possible. It provides the number of times something occurs, or frequency, lends itself to statistical calculations such as determining the average number of occurrences or central tendencies.

One of its major limitations is that it cannot help determine what causes a specific behaviour, motivation or occurrence. In other words, it cannot establish a causal research relationship between variables.

The two most common types of descriptive research designs are

1. Observation: Observation is a primary method of collecting data by human, mechanical, electrical or electronic means. The researcher may or may not have direct contact or communication with the people whose behaviour is being recorded. Observation techniques can be part of qualitative research as well as quantitative research techniques. The commonly used observation methods are:

- Participant and non participant observation: This depends on whether the researcher chooses to be part of the situation s/he is studying.
  (e.g. studying team dynamics by being a team member would be participant observation)
- Obtrusive and unobtrusive observation: Depends on whether the subjects being studied can detect the observation
  (e.g. hidden microphones or cameras observing behaviour)
- Observation in natural or contrived settings: Observing the behaviour in its natural setting and in a condition where the natural settings are created.
- Disguised and non-disguised observation: Depends on whether the subjects being observed are aware that they are being studied or not. In disguised observation, the researcher may not disclose his true identity and pretend to be someone else to keep away the bias in the findings.
- Structured and unstructured observation: This refers to guidelines or a checklist being used for the aspects of the behaviour that are to be recorded; for instance, noting who starts the introductory conversation between the group members and what specific words are used by way of introduction.
- Direct and indirect observation: This depends on whether the behaviour is being observed during the time it occurs or after the occurrence, as in

the case of TV viewing, for instance, where choice of program and channel flicking can all be recorded for later analysis.

One distinct advantage of the observation technique is that it records actual behaviour, not what people say they said/did or believe they will say/do. On the other hand, the observation technique does not provide us with any insights into what the person may be thinking or what might motivate a given behaviour/comment. This type of information can only be obtained by asking people directly or indirectly.

2. Surveys: The survey technique mainly involves the collection of primary data about subjects, usually by selecting a representative sample of the population or universe under study, through the use of a questionnaire. It is a very popular since many different types of information can be collected, including attitudinal, motivational, behavioral and perceptive aspects. It allows for standardization and uniformity in the questions asked and in the method of approaching subjects, making it easier to compare and contrast answers by respondent group. It also ensures higher reliability than some other techniques.

If properly designed and implemented, surveys can be an efficient and accurate means of determining information about a given population. Results can be provided relatively quickly, and depending on the sample size and methodology chosen, they are relatively inexpensive. However, surveys also have a number of disadvantages, which must be considered by the researcher in determining the appropriate data collection technique**.**

Since in any survey, the respondent knows that s/he is being studied, the information provided may not be valid insofar as the respondent may wish to impress (e.g. by attributing him/herself a higher income or education level) or please (e.g. researcher by providing the kind of response s/he believes the researcher is looking for) the researcher. This is known as response error or bias.

The willingness or ability to reply can also pose a problem. If the information sought is considered sensitive or intrusive the respondent may hesitate to reply, leading to a high rate of refusal. This can be overcome by framing such questions carefully.

There can be an interviewer error or bias as the interviewer can (inadvertently) influence the response elicited through comments made or by stressing certain words in the question itself. This is seen through facial expressions, body language or even the clothing that is worn.

Another consideration is response rate. Depending on the method chosen, the length of the questionnaire, the type and/or motivation of the respondent, the type of questions and/or subject matter, the time of day or place, and whether respondents were informed to expect the survey or offered an incentive can all influence the response rate obtained. Proper questionnaire design and question wording can help increase response rate.

Descriptive studies are also classified into:

1. Cross-sectional studies: It deals with a sample of elements from a given population. Number of characteristics from the sample elements are collected and analyzed. It is of two types: field studies and surveys.
2. Longitudinal studies. This is based on panel data and panel methods. A panel constitutes a group of respondents who are interviewed and reinterviewed from time to time. Hence the same variable is repeatedly measured. This helps in studying a particular behaviour over a period of time.

Causal Research

Causal research is undertaken to see if there is a cause and effect relationship between variables. In order to determine causality, it is important to hold the variable that is assumed to cause the change in the other variable(s) constant and then measure the changes in the other variable(s). This type of research is very

complex and the researcher can never be completely certain that there are not other factors influencing the causal relationship, especially when dealing with people's attitudes and motivations. There are often much deeper psychological considerations that even the respondent may not be aware of.

There are two research methods for exploring the cause and effect relationship between variables:

1. Experimentation or natural experimentation: This highly controlled method allows the researcher to manipulate a specific independent variable in order to determine what effect this manipulation would have on other dependent variables. Experimentation also calls for a control group as well as an experimentation group, and subjects would be assigned randomly to either group. The researcher can further decide whether the experiment should take place in a laboratory or in the field, i.e. the "natural" setting as opposed to an "artificial" one. Laboratory research allows the researcher to control and/or eliminate as many intervening variables as possible.

2. Simulation: Another way of establishing causality between variables is through the use of simulation.

A sophisticated set of mathematical formula are used to simulate or imitate a real life situation. By changing one variable in the equation, it is possible to determine the effect on the other variables in the equation.

For the natural experiments there are three classes of designs:

1. Time-series and trend designs
2. Cross-sectional designs and
3. A combination of the above two.

Time series and trend designs: In a time series design, data is collected from the sample or population at successive intervals. The trend data relate to matched samples drawn from the same population at successive intervals. It can be of many types.

A simple design can be represented as below:

X O

Where X indicates the exposure of a group to an experimental treatment and O indicates the observation or measurement taken on the subject or group after an experimental treatment. Another method also involves a control group. This can be represented as below:

O_1 O_2 O_3 X O_4 O_5 O_6

O'_1 O'_2 O'_3 O'_4 O'_5 O'_6

Where O's represent measurement of the control group. This is termed as multiple time-series design.

Cross-sectional designs: It studies the effect of different levels of treatments on several groups at the same time. It can be represented as below:

X_1 O_1

X_2 O_2

X_3 O_3

X_4 O_4

An example would be different kind of incentives given for the same product in various territories. This would help in understanding the effect of varying the incentive on the sales performance across territories.

Combinational Design: This design combines both the time-series and cross-sectional designs.

This design is generally seen while measuring advertising effectiveness in a panel. An advertisement is run and the respondents are asked if they have seen it earlier. Those who have seen it earlier constitute the test group and those who have not constitute the control group. The purchase made before and after the

advertisement by the test and the control group marks the advertising effectiveness.

So many research designs have been listed. The one that is ultimately selected should help in solving the problem. It should help in arriving at the desired conclusions.

Summary:

A Research Design is the specification of methods and procedures for acquiring the information needed. It is the blueprint for a research process.

There are many classifications accepted for a Research Design. One of the most accepted classification is grouping it under three types: Exploratory, Descriptive and Causal.

An exploratory research is often conducted because a problem has not been clearly defined as yet, or its real scope is as yet unclear. Conclusive research on the other hand is meant to provide information that is useful in reaching conclusions or decision-making. It is mostly quantitative in nature. A Causal research is undertaken to see if there is a cause and effect relationship between variables.

Causal research again can be: Time-series and trend designs, Cross-sectional designs and a combination of the above two.

Questions:

1. What is a Research Design? Explain.
2. Classify the Research Designs and define each of them.
3. 'Exploring helps in knowledge growth'. Explain with relevance to Exploratory Research design.
4. 'Conclusive research helps in drawing conclusions'. Explain.
5. What is descriptive Research? Classify and explain the same.
6. What are cross-sectional and longitudinal studies? Explain.
7. What is causal research? Explain the causal research methods.
8. What is Time series design? When is it used?
9. A good Research design is essential for solving a research problem. Explain.

10. Are more than one research designs used to find a solution to a problem? Critically evaluate the same.

Chapter 4

DATA SOURCES

Objective:

- To understand the meaning and importance of data sources.
- To read in detail about the sources of Primary data and secondary data sources.
- To understand the relevance of these data sources while solving a research problem.

Data sources

One of the most important components of Marketing Research is collection of data required to solve a defined research problem. The general tendency of the researcher is to organize a survey and collect the data from the field.

The most important point to be considered before this is to research the secondary sources and gather data already available. This gives a logical perspective to problem solving. Only then the actual data required to be collected from the primary survey can be well defined.

Hence it is imperative to know the advantages and drawbacks of Secondary and Primary data.

Secondary Data

Secondary data is defined as the data that has been collected by individuals or agencies for purposes other than those of the particular research study. For example, if a government department has conducted a survey of, say, school going children, then a uniform manufacturer might use this data for his research purpose.

As mentioned earlier, it is ideal to undertake a marketing research study after a prior search of secondary sources (also termed desk research). The reasons for this are summed up below.

- Many a time secondary data may itself be sufficient to draw conclusions and answer to solve the problem. Primary data collection may not be required.
- Secondary data is economical than collecting primary data. A thorough examination of secondary sources can yield a great deal more information than through a primary data collection exercise which needs to be critically evaluated.
- Searching secondary sources is much less time consuming than primary data collection.
- Secondary sources of information may at times consider a large sample and hence can yield more accurate data than that primary research. This is especially true for census information or syndicated reports by government departments which are generally large scale. This is likely to yield far more accurate results than custom designed surveys that are based on relatively small sample sizes.
- Secondary data can play a substantial role in the exploratory phase of the research when the main objective is to define the research problem and to generate hypotheses. The assembly and analysis of secondary data helps in better understanding of the marketing problem. This also gives an idea about the course of action and missing links which can be got from the primary research.
- Secondary sources are very useful to structure the sample and define the population.

Disadvantages of secondary data:

Even though the secondary data offers a lot of advantages; it also has its own shortcomings. This corresponds to both the source and the quality of the data. The main disadvantages may be listed as follows:

- The researcher has to be careful while using the units defined in the data. It is better to study the definitions used prior to accepting the same for research purpose.

For ex, the meaning of family might differ from urban and rural as it may consider the nuclear or the joint family system especially in India. Hence

while considering secondary data on the size of family, these definitions need to be kept in mind.

It should be noted that definitions may change over time and if this is not evaluated the conclusion derived may be wrong.

- The errors of measurements are not generally published in secondary sources and hence this should be considered while looking at data from secondary sources. The solution is to try to speak to the individuals involved in the collection of the data to obtain some guidance on the level of accuracy of the data. This is especially crucial if the stake is high in terms of commercial implications.
- The data has to be validated for source biases as it may have been prepared to appear exaggerated or otherwise. Hence it is better to go through details of the purpose for which the data had been collected.
- The reliability of published data may vary over time. Hence the data needs to be checked for time validity. For ex: New states have been formulated in India. Data pertaining to population studies conducted before the formation of the new state needs to be evaluated for its current application.
- Many a time the data collected may be outdated and hence it needs to be refreshed again. This may otherwise hinder the analysis.

All these drawbacks do exist, but still secondary data has it own merits. It is ideal to use multiple sources of secondary data. In this way, these different sources can be cross-checked and validated for the source of information. It is better to disregard the data whenever any controversy exists.

The below flowchart (figure 3) depicts the evaluation procedure for using secondary data. As can be seen, the flowchart divides into two phases. The early stages relate to the relevance of the data to the research objectives and the later stages of the flowchart are concerned with questions about the accuracy of secondary data.

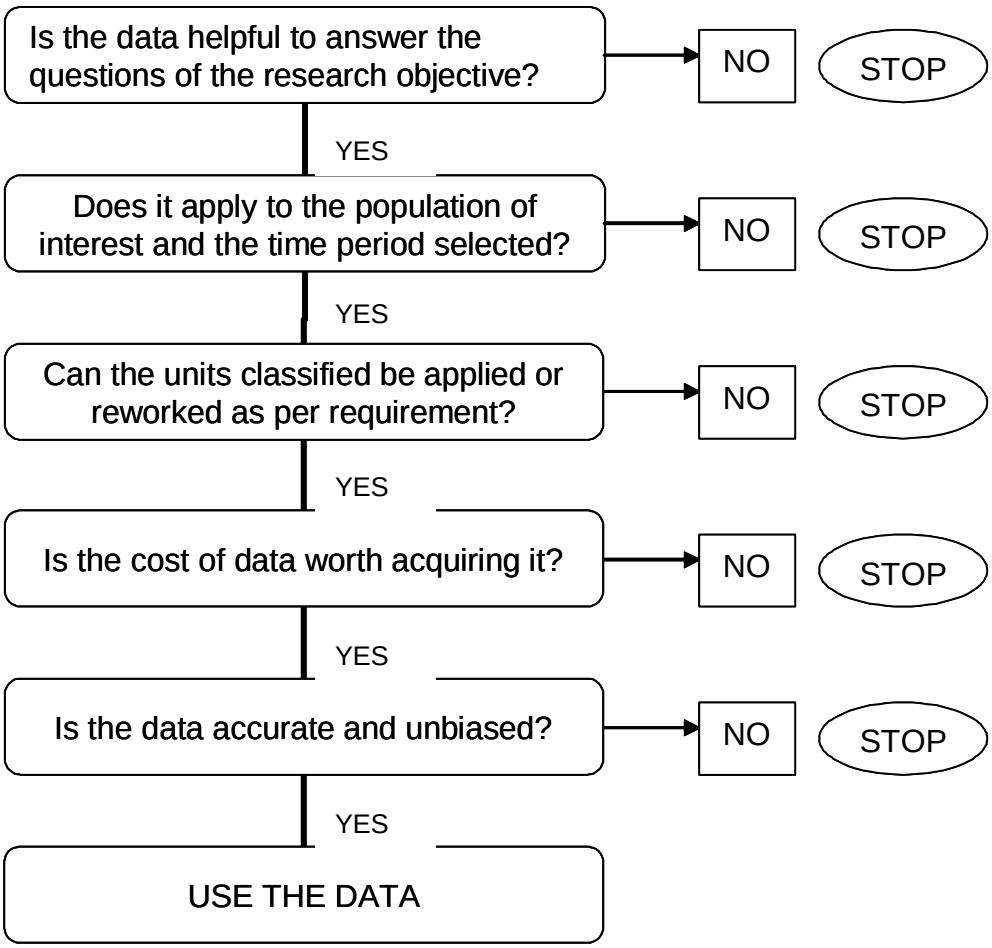


Figure 3: Evaluation of secondary data

Secondary sources of information:

Secondary sources of information can be collected from two sources: internal sources and external sources.

Internal sources of secondary information:

Lot of data is available within an organization regarding day to day operations. These data can be utilized wherever required. These include:

1. **Sales data:** Sales orders are received, invoiced and delivered. Cost of the goods supplied is also recorded. Sales across different territories are recorded via the reports received from the field. Most of these reports can be used for making marketing decisions. These resources are

generally overlooked while deciding on critical issues. Lot of information pertaining to sales by territory, sales by customer type, prices and discounts, average size of order by customer, customer type, geographical area, average sales by sales person and sales by pack size and pack type, etc.

This data can be used to identify the most profitable product and customers, tracking sales trends, analysis on discounts given, scattering pattern of sales orders, effect of seasonality on sales, etc.

2. **Financial data:** This relates to data on various costs involved in procurement of raw materials, production of goods, distribution of goods, conversion costs, labor costs, transportation cost, storage cost, etc. With such data the efficiency of operation can be determined. It helps in assessing the cost of production of a new product, analysis the cost of free capacity, etc.
3. **Transportation data:** A good record of the data relating to transport operations determine which route to use, which transporter to use, cost of effective routing patterns, etc. This helps in determining whether it would be sensible to have your own vehicle or hire a vehicle. This enables decision towards a trade off analysis towards a better profitability.
4. **Human Resource data:** Enormous information could be collected from an organizational perspective from the human resource department. Data on employee turnover, absenteeism, strength of employees could be obtained. This would help in man power planning for the present and future, succession planning, training and development for better productivity, etc.
5. **Storage data:** This may help in calculating the direct product profitability by calculating the rate of stock turn; stock handling costs, assessing the efficiency of certain marketing operations and the efficiency of the marketing system as a whole.

External sources of secondary information

Lot of secondary data is now available further to the discovery of World Wide Web and lot of institutions looking at such analysis.

Large number of organizations provides marketing information including national and local government agencies, quasi-government agencies, trade associations, universities, research institutes, financial institutions, specialist suppliers of secondary marketing data and professional marketing research enterprises. Dillon et al advise that searches of printed sources of secondary data

begin with referral texts such as directories, indexes, handbooks and guides. These sorts of publications rarely provide the data in which the researcher is interested but serve in helping him/her locate potentially useful data sources.

The main sources of external secondary sources are (1) government (federal, state and local) (2) trade associations (3) commercial services (4) national and international institutions.

1. Government statistics: These may include all or some of the following: Population census, Social surveys, family expenditure surveys, Import/export statistics, Production statistics, and Agricultural statistics.

   Some of the Indian government bodies are:

   - Population Statistics of Govt. of India – Provides statistics related to general population of India
   - Central Bureau of Health Intelligence – Provides health related statistics
   - Indian Council of medical Research – Provides information on research being conducted on major diseases
   - Policy Reform Options Database – Provides data on policy reforms
   - Ministry of Health and Family Welfare – Provides information on family welfare
   - Ministry of Statistics and Programme implementation – Gives information on various statistical indicators of Indian economy
   - India Brand Equity Foundation – Provides information on Indian economy and Industry
   - Insurance Regulatory and Development Authority – Data on Health Insurance in India.

2. Trade associations - They might produce a wide range of data. Normally it may produce a trade directory and, perhaps, a yearbook.
3. Syndicated reports – These are published market research reports from various organizations which charge for their information. These data relate to consumer information and media information. These are generally prepared to cater to all interested and not to any specific

organizational requirement. Hence the relevant data is extracted from this.

4. National and international institutions: Economic reviews, Research reports, journals and articles are all useful sources to contact. A lot of secondary data can be obtained from World Bank, WHO, International Monetary Fund, International Fund for Agricultural Development, United Nations Development Programme, Food and Agriculture Organization and International Labor Organization.

Advances in telecommunications technology have combined to allow people around the world to exchange information quickly and easily. Internet has made access to information relatively easier and quicker.

Primary Data:

Primary data collection begins when a researcher is not able to find the data required for his research purpose from the secondary sources. Market researchers are interested in a variety of primary data about demographic/socioeconomic characteristics, attitudes/opinions/interests, awareness/knowledge, intentions, motivation, and behavior.

There are two basic means of obtaining primary data depending on the nature of the problem and by the availability of time and money:

1. **Observation**

As the name implies, the researcher observes the situation of interest and records the relevant facts, actions, or behaviors. Observation provides accurate data about what consumers do in certain situations but do not provide details of why it happened.

There are several methods of observation:

- Structured – unstructured observation: In structured observation, the observer is given a set of behaviors to observe. In unstructured observation, the observer is allowed to observe anything that may be

relevant to the research objective. In the first instance there may be a bias and in the second the result may vary from observation to observation.

- Disguised – undisguised observation: In a disguised observation, the subjects do not know that they are being observed. This is a better way to observe as the subjects are not conscious that they are being observed and behave freely. In an undisguised observation the subjects are aware that they are being observed and tend to be cautious.
- Observation under normal setting – Laboratory setting: Normal setting would be a field survey; laboratory setting would be under a fixed roof or venue. The latter allows a prompt and economical way of collecting data and permit the use of more objective measurements.
- Direct – Indirect Observation – In the case of direct observation, the event or the behaviour of a person is observed as it occurs. In an indirect observation, a record of a past behaviour is observed.
- Human – Mechanical Observation – The observations are recorded manually in Human Observations. In Mechanical Observation, the research is carried out through hidden cameras and audiometers; hence there is no subjective bias.

2. Questionnaire

Questionnaires are data collecting instruments used to ask respondents questions to secure the desired information. Questionnaires may be administered by mail, over the telephone, by computer, or in person.

The design of a questionnaire depends on whether the researcher wishes to collect exploratory information (i.e. qualitative information for the purposes of better understanding or the generation of hypotheses on a subject) or quantitative information (to test specific hypotheses that have previously been generated).

The questionnaires can be classified into four types:

1. Structured – non-disguised
2. Structured disguised
3. Non-structured - non- disguised
4. Non-structured – disguised

Non-disguised are the direct questions and the object of enquiry is revealed to the respondent. Disguised are the indirect questions where the object of enquiry is not revealed to the respondent. In a structured questionnaire the questions are asked in a pre-determined order.

Formal standardised questionnaires: If the data is required to be analysed statistically, a formal standardised questionnaire is designed. The points to be remembered while designing such questionnaires are:

- The questionnaire has to be framed in such a manner that each respondent receives the same stimuli
- The questionnaire has to be well-defined so that the interviewer is able to answer the respondent's clarifications if necessary.
- The response format must be easy to complete during the interviewing process.
- A well-designed questionnaire should primarily meet the research objectives.
- A questionnaire should obtain the most complete and accurate information possible.
- The questionnaire should be brief and to the point and be so arranged that the respondent remains interested throughout the interview.

Development of a questionnaire:

The following steps are involved in the development of a questionnaire:

1. **Choose and decide on the information required**

The objective behind the survey should be kept in mind while designing a questionnaire. Though the researcher has an idea about the kind of information to be collected, additional help can be obtained from secondary data. In respect of secondary data, the researcher should be aware of what work has been done on the same or similar problems in the past, what factors have not yet been examined, and how the present survey questionnaire can build on what has already been discovered.

2. **Define the target respondents**

The researcher must define the population that he wishes to collect the data from. Firstly, in marketing research, researchers often have to decide whether they should cover only existing users of the generic product type or whether to also include non-users. Secondly, researchers have to draw up a sampling frame. Thirdly, in designing the questionnaire we must take into account factors such as the age, education, etc. of the target respondents.

3. **Selection of methodology to reach the target segment**

This will influence not only the questions the researcher is able to ask but the phrasing of those questions. The main methods available in survey research are: personal interviews, group or focus interviews, mailed questionnaires and telephone interviews.

Among these, the first two are used more extensively than the second pair. The general rule is that the more sensitive or personal the information, the more personal the form of data collection should be.

4. **Decide on question content**

There will be lot of temptation to use questions without critically evaluating their contribution towards the achievement of the research objectives. Hence researchers must proactively look at whether the question is really required and if it can be used in testing one or more of the hypotheses established during the research design.

5. **Decide on type of questions**

The questions can be classified into two forms, i.e. closed and open-ended. So

In a closed type of question, the respondent chooses between an alternative already stated. He does not get a chance to answer in a descriptive manner.

For ex.: Do you use Brand X? Yes _____ No_____.

The closed type of questioning has a number of important advantages:

- It is easy for the respondent to answer. He does not have to think and answer.
- It 'prompts' the respondent so that the respondent has to rely less on memory while answering a question.
- Analysis is easier as responses can be easily classified
- It permits categorization of the response to specify the answer categories.

It also has some disadvantages:

- It does not allow the respondent the opportunity to give a different response to those suggested.
- They 'suggest' answers that respondents may not have considered before.

In an open-ended question the respondent is asked to give a reply to a question in his/her own words. No answers are suggested. These responses are explanatory in nature and give some insights from the respondents end.

Example: "What do you like most about this Product?"

Open-ended questions have a lot of advantages when used in a questionnaire:

- They allow the respondent to answer in his own words, with no influence by any specific alternatives suggested by the interviewer.
- They often reveal the issues which are most important to the respondent, and this may reveal findings which were not originally anticipated when the survey was initiated.
- Respondents can 'qualify' their answers or emphasize their opinions.

The inherent problem of an open-ended question is that they must be treated with caution as:

- Respondents may find it difficult to 'articulate' their responses i.e. to properly and fully explain their attitudes or motivations.

- Respondents may not give a full answer simply because they may forget to mention important points. Some respondents need prompting of the types of answer they could give.
- Data collected is in the form of verbatim comments - it has to be coded and reduced to manageable categories. This can be time consuming for analysis and there are numerous opportunities for error in recording and interpreting the answers given on the part of interviewers.
- Respondents will tend to answer open questions in different 'dimensions'. Such responses need to be probed further for clarity on response.

There are a lot of advantages of this type of questions as

- The researcher can avoid the potential problems of poor memory or poor articulation by then subsequently being able to prompt the respondent into considering particular response options.
- Recording of the responses during the interview is relatively easy.

The one disadvantage of this form of question is that it requires the researcher to have a good prior knowledge of the subject in order to generate realistic/likely response options before printing the questionnaire.

In many situations a questionnaire will need to incorporate all these forms of question, because some forms are more appropriate for seeking particular forms of response.

6. **Putting questions into a meaningful order and format**

Opening questions: Opening questions generally should be easy to answer and not in any way threatening to the respondents. This is crucial because it is the respondent's first exposure to the interview and sets the tone for the nature of the task to be performed. If they find the first question difficult to understand, or beyond their knowledge and experience, or embarrassing in some way, or uninteresting they are likely to break off immediately. If, on the other hand, they

find the opening question easy and pleasant to answer, they are encouraged to continue.

Question flow: Questions should flow in some kind of psychological order, so that one leads easily and naturally to the next. There could be a continuity maintained on the flow of the questions where the response from on leads to another. This helps in creating a sequence and the respondent's interest is maintained. Questions on one subject, or one particular aspect of a subject, should be grouped together. Respondents may feel it disconcerting to keep shifting from one topic to another, or to be asked to return to some subject they thought they gave their opinions about earlier.

Question variety: Respondents become bored quickly and restless when asked similar questions for half an hour or so. Hence the questions need to be carefully keyed in to maintain the interest throughout the interview.

7. **Closing questions**

By the time the respondent comes to the end of the questionnaire it is quite natural for a respondent to become increasingly indifferent to the questionnaire. This is mainly due to impatience or fatigue. He might give careless answers to the later questions. Hence such questions should be included in the earlier part of the questionnaire. Potentially sensitive questions should be left to the end, to avoid respondents cutting off the interview before important information is collected.

8. **Physical appearance of the questionnaire**

The physical appearance of a questionnaire has a significant effect upon both the quantity and quality of marketing data obtained. Data quality can also be affected by the physical appearance of the questionnaire with unnecessarily confusing layouts making it more difficult for interviewers, or respondents in the case of self-completion questionnaires, to complete this task accurately.

In general it is best for a questionnaire to be as short as possible. A long questionnaire leads to a long interview and this may lead to decreasing interest on the part of the respondent.

- **Piloting the questionnaires**

Piloting is very mush essential to test whether the desired responses are being obtained for the questions. Many a time, the perception of the respondents varies from those of the researcher. Hence these issues can be corrected in the initial stage itself so that the research process is facilitated. The purpose of pretesting the questionnaire is to determine:

- If the wordings used help in achieving the desired results
- Are the questions in the right order?
- Are the questions easy to understand?
- If any questions needed to be added or deleted.
- Are the instructions to interviewers are adequate?

The few respondents selected for the pilot survey should be broadly representative of the type of respondent to be interviewed in the main survey.

If the questionnaire has been subjected to a thorough pilot test, the research objective is easy to achieve. This solves the difficulties that may arise later in terms of invalid and inadequate responses which might lead to wastage of time and resources.

Summary:

One of the most important components of Marketing Research is collection of data required to solve a defined research problem. There are two types of data sources: Secondary and Primary.

Secondary data is defined as the data that has been collected by individuals or agencies for purposes other than those of the particular research study.

Secondary data can be obtained from internal sources and external sources. It has to be validated before use to ensure that it suits the purpose of research.

Primary data is the data collected from the field for finding the answer to the research problem. It can be collected by two methods: Observation and Questionnaire. Designing a questionnaire is very important as it determines the quality of response sought and analyzed for finding solutions to a research problem at hand.

Questions:

1. Tabulate the difference between Primary and Secondary data.
2. Sum up the advantages of secondary data.
3. Does secondary data have some disadvantages? Explain.
4. Give a brief description about the internal sources of secondary information.
5. List out some of the sources of secondary data known to you.
6. What is primary data? How is it collected?
7. Explain the methods of observation used to collect primary data.
8. Describe the various steps involved in the development of a questionnaire.
9. Enumerate and tabulate the differences between open and closed questions.
10. Give the advantages and disadvantages of open and closed questions.

Chapter 5

MARKETING INFORMATION SYSTEM

Objective:

- To understand the concept of Marketing Information System or MIS.
- Learn about the components of an effective MIS.
- List out the types of decisions facilitated by MIS.

Marketing Information System

There are five important functions of managers: Planning, Organizing, Coordinating, Deciding and Controlling.

A marketing information system (MIS) is intended to bring together a lot of data into an understandable body of information. An MIS provides processed data which is suitable for decision making.

According to Kotler, an MIS is more than a system of data collection or a set of information technologies:

"A Marketing Information System is a continuing and interacting structure of people, equipment and procedures to gather, sort, analyse, evaluate, and distribute pertinent, timely and accurate information for use by marketing decision makers to improve their marketing planning, implementation, and control".

According to Smith, Brien and Stafford MIS is:

“A structured, interacting complex of persons, machines and procedures designed to generate an orderly flow of pertinent information, collected from both intra-and extra – firm sources, for use as the basis for decision-making in specified responsibility areas of marketing management.

Hence, such a system facilitates decision-making in different areas of marketing management.

Marketing Research and Marketing Information System:

There are differences between Marketing Research and Marketing Information System.

Marketing Research is about seeking information from external sources. Its main purpose is to solve the research problem. It tends to focus on past information and is not computer based. It is one source of information for Market Information System.

Marketing Information System handles both data from internal sources like orders, sales, inventory levels, payables, etc and also data from external sources like developments in the macro environment.

Components of an effective Marketing Information System:

According to Kotler, an effective Marketing Information System has the following components:

1. Internal Accounting System: Maintains data pertaining to sales, receivables, costs, etc that are internal to the organisation.
2. Market Intelligence System: As the name itself implies, it speaks about the external changes occurring in the macro environment and prepares mangers to make effective strategies.
3. Marketing Research System: This undertakes studies on specific marketing problems and provides solution to the management.
4. Marketing Management Science System: It mainly helps in building models for better understanding of the marketing process.

The below diagram (figure 4) illustrates the major components of an MIS, the environmental factors monitored by the system and the types of marketing decision which the MIS seeks to help in.

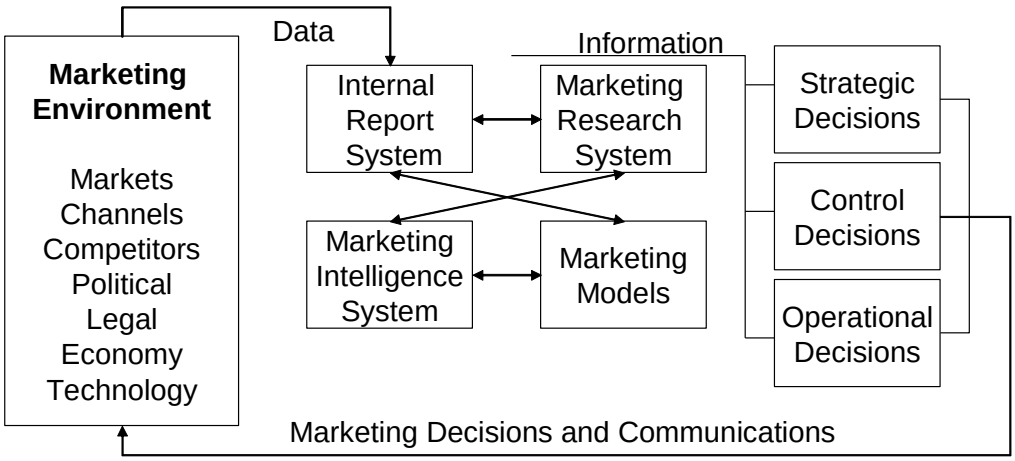


Figure 4: Components of an MIS

With the help of this model all the four important constituents can be explained. A fully fledged MIS should have the following components, the methods (and technologies) of collection, storing, retrieving and processing data.

Three levels of decision making can be observed here: strategic, control (or tactical) and operational. MIS has to support all these three decision making.

Strategic decisions are very important as they have implications on changing the structure of an organisation. Therefore the MIS must provide precise and accurate information.

Control decisions deal with broad policy issues.

Operational decisions concern the management of the organisation's marketing mix.

MIS should help the manager in his decision making process for problem identification, generation and evaluation of alternative courses of action, to acquire necessary feedback on implementing his decision and help him to take corrective action.

Summary:

There are five important functions of managers: Planning, Organizing, Coordinating, Deciding and Controlling.

A marketing information system (MIS) is intended to bring together a lot of data into an understandable body of information. An MIS provides processed data which is suitable for decision making.

Three levels of decision making can be observed here: strategic, control (or tactical) and operational. MIS has to support all these three decision making.

Questions:

1. How is a MIS beneficial to a Manager?
2. Define MIS according to various authors.
3. What is the difference between Marketing Research and Marketing Information System?
4. What are the components of an effective MIS? Explain.
5. What are the types of decision making facilitated by MIS?

References for this Unit:

1. Kotler, P., (1988) *Marketing Management: Analysis* Planning and Control, Prentice-Hall
2. Agnilar, F., (1967) *Scanning The Business Environment,* Macmillan, New York
3. Green, P.E. Tull, D.S. and Albaum G (1993) *Research methods for marketing decisions,* 5th edition, Prentice Hall
4. Joselyn, R. W. (1977) *Designing the marketing research,* Petrocellis/Charter, New York
5. Dillon, W.R., Madden, T. and Firtle, N. H., (1994) *Marketing Research in a Research Environment,* 3rd edition, Irwin.
6. Sudman, S. and Bradburn, N. M. (1973), *Asking Questions*
7. G. C. Beri, Marketing Research, second edition, Tata McGraw-Hill Publishing Company Limited.
8. Wikepedia – The free encyclopedia

Unit 2 Sampling

Objectives

In this section, we will introduce you to the concept of sampling, sampling methods, sample size and sampling error. After you go through this unit, you should be able to understand :

- the concept of sampling.
- the differences between census and sampling.
- various sampling terminologies
- how an appropriate sampling design can be determined.
- sampling plan and the steps in developing it.
- various probability and non-probability sampling methods.
- the concept of Random-Digit Dialing.
- the concept of sample size and the methods of determining it.
- the concept of sampling error and the types.

In this section, we have discussed the following :

- **2.1. Introduction to sampling**
  - *2.1.1. When a Census Is Appropriate*
  - *2.1.2. When a Sample Is Appropriate*
- **2.2. Sampling terminology**
- **2.3. Determining the Appropriate Sampling Design**
- **2.4. Sampling Plan**
  - *2.4.1. Steps in Developing a Sampling Plan*
- **2.5. Sampling methods**
  - *2.5.1. Types of probability sampling designs*
  - *2.5.2. Types of non-probability sampling designs*
- **2.6. Random-Digit Dialing (RDD)**
- **2.7. Sample size**
- **2.8. Sampling error**
- **2.9. Non-response Problems**

2.1. Introduction

Sampling is an important concept that we practice in our every day life. Sampling involves selecting a relatively small number of elements from a larger defined group of elements and expecting that the information gathered from the small group will allow judgments to be made about the larger group. If all the respondents in a population are asked to provide information, such survey is called a census. Information obtained from a subset of the population is known as the statistic (from sample). Researchers then attempt to make an inference about the population parameter with the knowledge of the relevant sample statistic. Sampling is often used when conducting a census is impossible or unreasonable. When using a census, the researcher is interested in collecting primary data about or from every member of the defined target population.

2.1.1. When a Census Is Appropriate

A census is appropriate if the population size itself is quite small. A census also is conducted if information is needed from every individual or object in the population. For example, if the researcher is interested in determining the number of foreign students enrolled in a university, it is necessary to get information from all the departments in the university because of possible variations within each department. If the cost of making an incorrect decision is high or if sampling errors are high, then a census may be more appropriate than a sample.

2.1.2. When a Sample Is Appropriate

Sampling may be useful if the population size is large and if both the cost and time associated with obtaining information from the population is high. The opportunity to make a quick decision may be lost if a large population must be surveyed. With sampling, in a given time period, more time can be spent on each personal interview, thereby increasing the response quality. It is easy to manage surveys of smaller samples and still exercise quality control in the interview process. Sampling may be sufficient in many instances. If the population being dealt with is homogeneous, then sampling is fine. If taking a census is not possible, then sampling is the only alternative.

2.2. Sampling terminology

2.2.1. *Population*

A population is an identifiable total group or aggregation of elements that are of interest to the researcher and pertinent to the specified problem. A defined target population consists of the complete group of elements that are specifically identified for investigation according to the objectives of the research project.

2.2.2. *Element*

An element is a person or object from which data and information are sought. In research, the element is a particular product or group of individuals. Elements must be unique, countable and when added together, make up the whole target population. Target population elements must include a particular consumer product, specific group of people or specific organisations.

2.2.3. *Sampling units*

Sampling units are the target population elements available for selection during the sampling process. In a simple, single-stage sample, thee sampling units and the population elements may be the same. But many studies involve complex problems that require the use of a multi-stage sampling process.

2.2.4. *Sampling frame*

It is the list of all eligible sampling units. Some common sources of sampling frames are list of registered voters and customer lists from magazine publishers, credit card companies and the like. There are specialized commercial companies that are in the business of developing databases that contain names,

addresses and telephone numbers of potential population elements. It is usually very difficult and expensive for a researcher to
gain access to truly accurate or representative , current sampling frames. In such situations, a researcher would have to employ an alternate method such as random-digit dialing or a location survey in order to generate a sample of prospective respondents.
The sampling frame contains the operational population from which the sample will be drawn. In an ideal situation, the operational population, the defined target population frame and the sampling frame are identical. In those situations where a sampling frame contains all of the eligible sampling units of the defined population plus additional ones,
Then it is said to have over-registration.

But if the eligible sampling units are accidentally left out of the sampling frame, hen the frame has an under-registration condition.

2.2.5. *Sampling gap*

Both over-registration and under-registration factors create sampling gaps. A sampling gap is the representation difference between the population elements and sampling units in the sampling frame. A sampling gap can also be viewed as sampling frame error and occurs when certain sample units are not excluded or complete segments of the defined target population are not accurately represented in the sampling frame. The larger the sampling frame error, the greater the chance of misleading and inaccurate data results.

2.2.6. *Sampling distribution*

One important assumption that underlies sampling theory is that the population elements are randomly distributed. If a researcher were able to do a census of the entire target population elements, then the probability distribution of the population, or the relative frequencies of a population's parameters would depict a normal bell-shaped distribution pattern.

Sampling distribution is the frequency distribution of a specific sample statistic (sample mean or sample proportion) from repeated random samples of the same size.

2.2.7. *Central Limit Theorem*

The Central Limit Theorem becomes the backbone for doing survey research and data collection through experimental designs. The theorem states that for almost all defined target populations, the sampling distribution of the mean or the percentage value derived from a simple random sample will be

approximately normally distributed, provided that the sample size is sufficiently large. When '*n*' is greater than or equal to 30, the sample is a large sample.

Activity 2.1.
Develop a population list or sampling frame for an attitude study, when the target population is :
i. the students of your college.
ii. High-income families in your residential area.
iii. Shops that sell tennis rackets in your city.

2.3. Determining the Appropriate Sampling Design

Selection of the most appropriate sampling design should incorporate the seven factors :

Research objectives

A full understanding of the overall information problem situation and the research objectives provides the initial guidelines for determining the appropriate sampling design. If the research objectives include the desire to generalise the sample data results to the defined target population, then the researcher must seriously consider using some type of probability sampling method rather than a non-probability sampling method. The stage of the research project and type of research (e.g., exploratory, descriptive, casual will influence the researcher's selection of sampling method.

Degree of accuracy

The degree of accuracy required or the researcher's tolerance for error may vary from project to project. If the researcher wants to make predictions or inferences about the "true" position of all members of the defined target population, then he or she must choose some type of probability sampling method. In contrast, if the researcher is solely trying to identify and obtain preliminary insights into the defined target population, non-probability methods might prove to be more appropriate.

Availability of resources

If the researcher's financial and human resources are restricted, these limitations will certainly eliminate some of the more time-consuming, complex probability sampling methods. Researchers who are influenced by the cost concerns versus the value of the information will often opt for a non-probability sampling method rather than conduct no research at all.

Critical factors in selecting the appropriate sampling design

Factor	*Questions*
Research objectives	Do the research objectives call for the use of qualitative or quantitative research designs ?
Degree of accuracy	Does the research call for making predictions about the defined target population or only preliminary insights ?
Availability of resources	Are there budget constraints with respect to the resources that can be allocated to the research project ?
Time frame	How quickly does the research project have to be completed ?
Advanced knowledge of Target population	Are there complete lists of the defined target population elements ?
Scope of the research	Is the research going to be international, national, regional or

Time frame

Researchers who need to meet a short deadline will be more likely to select a simple, less time-consuming sampling method rather than a more complex and accurate method. Researchers tend to opt for using some form of convenience sampling to gather data necessary to test the reliability of a newly developed construct or scale measurement.

Advanced knowledge of the target population

In many cases, a complete list of the population elements will not be available to the researcher. A lack of adequate lists may automatically rule out systematic random sampling, stratified random sampling, or any other type of probability sampling method. A preliminary study may be conducted to generate information to build a sampling frame for the study. The researcher must gain a strong understanding of the key descriptor factors that make up the true members of any target population.

Scope of the research

Whether the scope of the research project is to be international, national, regional, or local will influence the choice of the sampling method. The projected geographic proximity of the defined target population elements will influence the researcher's ability to compile needed lists of sampling units. When the target population elements are known or viewed to be unequally distributed geographically, a cluster sampling method may become much more attractive than other available methods. The broader the geographical scope of the research project, the more extensive and complex the sampling method becomes to ensure proper representation of the population.

Perceived statistical analysis needs

The need for statistical projections (i.e., estimates) based on the sample results is often a criterion. Only probability sampling techniques allow the researcher to appropriately use statistical analysis for estimates. While statistical analysis methods can be performed on data structures obtained from non-probability samples of people and objects, the researcher's ability to accurately

generalize the results and findings to the larger defined target population is very suspect and technically inappropriate.

2.4. Sampling Plan

A sampling plan is the blueprint or frame work needed to ensure that the raw data collected are representative of the defined target population. A good sampling plan will include, the following steps: (1) define the target population, (2) select the data collection method, (3) identify the sampling frames needed, (4) select the appropriate sampling method, (5) determine necessary sample sizes and overall contact rates, (6) create an operating plan for selecting sampling units, and (7) execute the operational plan.

2.4.1. *Steps in Developing a Sampling Plan*

Step 1: Define the target population

In any sampling pan, the first task of the researcher is to determine and identify the complete group of people or objects that should be investigated in the project. The target population should be given its identity by the use of descriptors that represent the characteristics of elements that make the target population's frame. These elements become the prosperity sampling units from which a sample will be drawn. Clear understanding of the target population will help the researcher successfully draw a representative sample. Devoting effort to identifying the target population usually will pay off. The following guidelines should be considered :

Look to the research objectives

If the research objectives are well thought out, the target population definition will be clear as well. The Research objectives include the research question, the research hypothesis and a statement of the research boundaries. Each of these elements contributes to refining the definition of the target population.

Consider alternatives

It is rare to find a study for which there are no alternative, reasonable, target population definitions. The task is to identify and evaluate several of the alternatives. The key point is to recognise that alternative definitions exist.

Know your market

If the research objective is to learn about the market response to some element of the marketing program, it is necessary to know something about the market. Without it, the population definition will have to be unnecessarily broad and, therefore, will lead to an unnecessary increase in research expenses.

Consider the appropriate sampling unit

The target population consists of sampling units. A sampling unit may contain people, households, or products. One task is to specify which sampling unit is appropriate.

Specify clearly what is excluded

The specification of target population should make clear what is excluded. For example, a study of voting intentions on certain candidates and issues might restrict the sampling population to those of voting age and even to those who intend to vote or to those who voted in the last election.

Don't over-define

The population, should be compatible with the study purpose and the research questions; but, the research should not over-define the population.

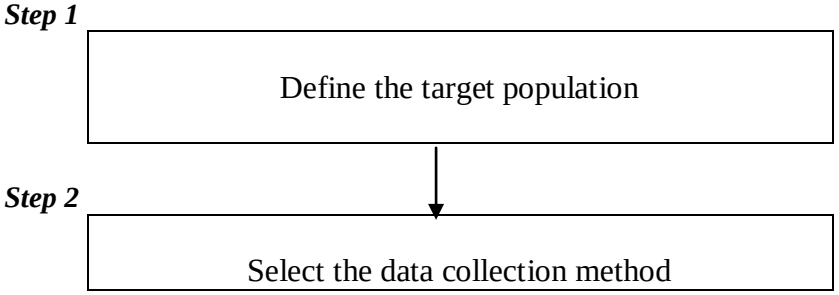
Consider convenience

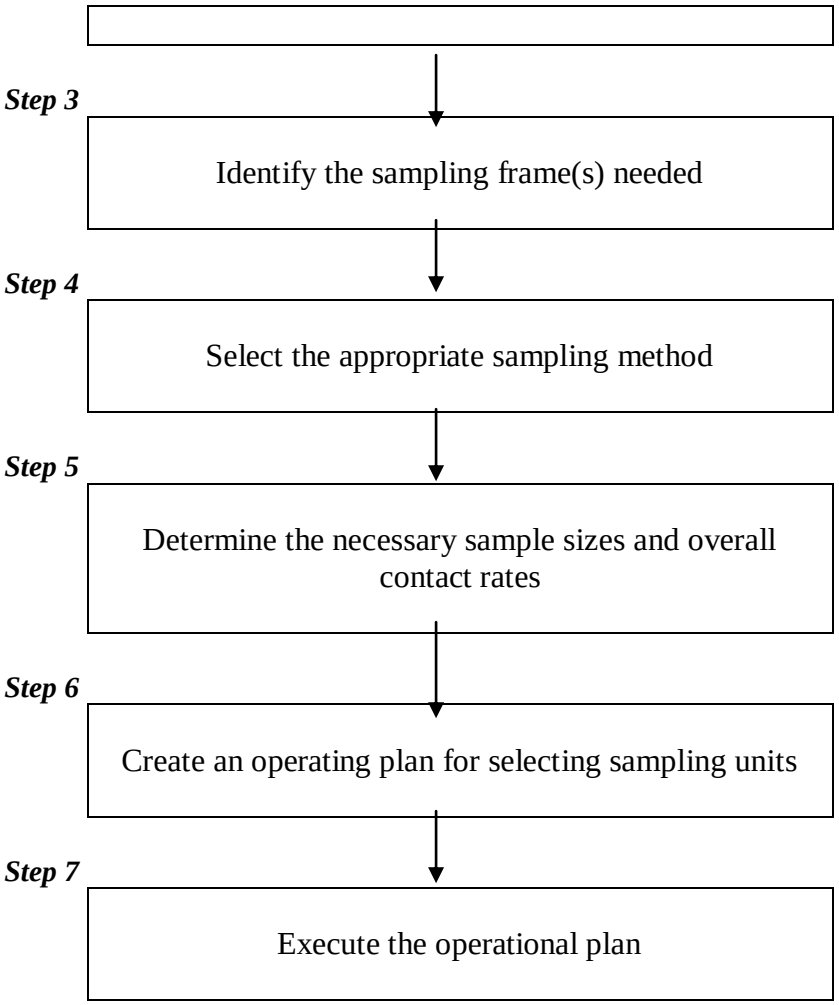
When there is a choice, preference should be given to populations that are convenient to sample.

Step 2: Select the data collection method

Using the information problem definition, the data requirements, and the established research objectives, the researcher must choose a method for collecting the required raw data from the target population elements. Choices include interviewing approach or a self-administered survey. The method of data collection guides the researcher in identifying and securing the necessary sampling frame(s) for conducting the research.

Steps in developing a Sampling plan





Step 3: Identifying the sampling frame(s) needed

After gaining an understanding of whom or what should be investigated, the researcher must assemble a list of eligible sampling units. This list needs to contain enough information about each prospective sampling unit so that the researcher can successfully contact them. An incomplete sampling frame decreases the likelihood of drawing a representative sample. Sampling frame lists can be created from a number of different sources. In creating the necessary sampling frames, the researcher must be aware of possible conditions of over-registration and under-registration of the prospective sampling units. These conditions will create sampling gaps or sampling frame errors that decrease the likelihood of being able to draw a representative sample.

Creating lists

The biggest problem in simple random sampling is obtaining appropriate lists. Lists do not exist for specialised populations. A solution for this problem is just to use a convenient list. When lists that do not match the population are used, biases are introduced. Sometimes several lists are combined in the hope of obtaining a more complete representation of the population. This approach introduces the problem of duplication. Those appearing on several lists will have an increased chance of being selected. Removing duplication can be expensive and must be balanced against the bias that is introduced. Another problem with lists is simply that of keeping them current. These lists can become outdated quickly as people move and change jobs within an organisation.

Creating lists for telephone interviewing

Telephone directories are used extensively as a basis for generating a sample. The concern with the use of directories is that population members may be omitted because they have changed residences, requested an unlisted number, or simply do not have a telephone. The incidence of unlisted numbers is extensive and varies dramatically from area to area. Another approach is to buy lists from magazines, credit-card firms, mail-order firms, or other such sources. One problem is that each such list has its own type of biases.

Dealing with population sampling frame differences

When a sampling frame does not coincide with a population definition, three types of problems arise : the subset problem, the superset problem and the intersection problem. A subset problem occurs when the sampling frame is smaller than the population. In other words, some of the elements in the population will not be present in the sampling frame. A superset problem occurs when the sampling frame is larger than the population but contains all the elements of the population. An intersection problem occurs when some elements of the population are omitted from the sampling frame, and when the sampling frame contains more elements than the population.

Step 4: Select the appropriate sampling method

The researcher must choose between two types of sampling orientations : probability and non-probability. Using a probability sampling method will always yield better and more accurate information about the target population's parameters than will any of the available non-probability sampling methods. Probability sampling has several advantages over non-probability sampling. First, it permits the researcher to demonstrate the sample's representativeness. Second, it allows an explicit statement as to how much variation is introduced, because a sample is used instead of a census of the populations. Finally, it makes possible the more explicit identification of possible biases.

Step 5: Determine necessary sample sizes and overall contact rates

In this step of a sampling plan, the researcher must consider how precise the sample estimates must be and how much time and money are available to collect the required raw data. To determine the appropriate sample size, decisions have to be made concerning (1) the variability of the population characteristic under investigation, (2) the level of confidence desired in the estimates, and (3) the degree of precision desired in estimating the population characteristic. The researcher must decide how many completed surveys will need to enter the data analysis activities of the overall research project.

Step 6 : Create an operating plan for selecting sampling units

In this step, the researcher wants to clearly lay out, in detail, the actual procedures to use in containing each of the prospective respondents who were drawn into the sample. All instructions should be clearly written so that interviewers know exactly what to do and how to handle any problems in the process of contacting prospective respondents.

Step 7 : Execute the operational plan

In some research projects, this step is similar to actually conducting the data collection activities. (e.g., actual calling of a prospective respondent to do a telephone interview). The important thing in this stage is to maintain consistency and control.

Activity 2.2.
A telephone survey is planned to determine the day-after recall of several test commercials to be run in Pondicherry. Design a sampling plan.

2.5. Sampling methods

There are two basic sampling designs : Probability and non-probability sampling methods. In probability sampling, each unit in the defined target population has a known, non-zero probability of being selected for the sample, The actual probability of selection for each sampling unit may or may not be equal depending on the type of probability sampling design used. It allows the researcher to judge the reliability and validity of raw data collected by calculating the probability to which the findings based on the sample would differ from the defined target population. The results obtained by the probability method can be generalized to the target population within a specified margin of error through the use of statistical methods. In non-probability sampling, the probability of selection of each sample unit is not known. Therefore, potential sampling error cannot be accurately known either. The selection of sampling units is based on some type of intuitive judgments, desire or knowledge of the researcher. The degree to which the sample nay or may not be representative of the defined target population depends on the sampling

approach and how well the researcher executes and controls the selection activities. There is always a temptation to generalize non-probability sample data results to the defined target population.

Comparative differences of probability and non-probability sampling methods

Factor sampling	Probability sampling	Non-probability
1.List of the population elements	Complete list necessary	Not necessary
2. Information about the-sampling units	Each unit identified	need detail on habits, activities, traits etc.
3. Sampling skill	skill required	little skill required
4. Time requirement	More time-consuming	Less time consuming
5. Cost per unit sampled	Moderate to high	Low
6. Estimate of population - parameters	Unbiased	Biased
7. Sample representativeness	Assured	Undeterminable
8. Accuracy and Relaiability	Computed with confidence intervals	Unknown
9. Measurement of sampling error	Statistical measures	No true measure available

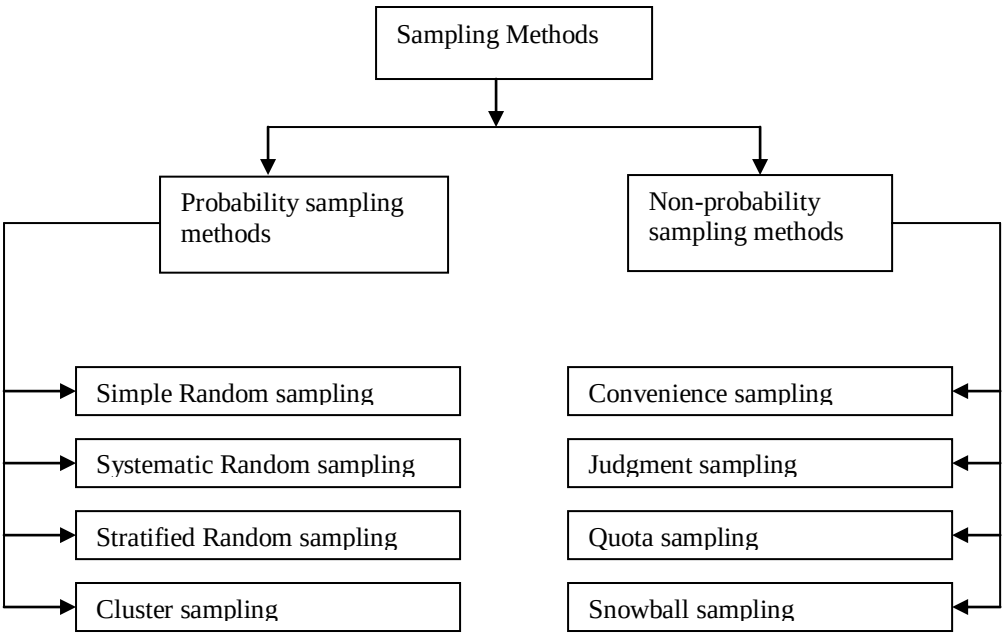
Simple Random Sampling is a probability sampling procedure which ensures that every sampling unit making up the defined target population has a known, equal, non-zero chance of being selected. For example, let's say an instructor decided to draw a sample of 10 students (n=10), from among all the students in a Marketing Research class that consisted of 30 students (N=30). The instructor could write each student's name on a separate, identical piece of paper and place all of the names in a jar. Each student would have an equal, known probability of selection for a sample of a given size that could be expressed by the formula :

$$\text{Probability of selection} = \frac{\text{Size of sample}}{\text{Population size}}$$

Size of population

Here, each student would have a 10 /30 (or 0.33) chance of being randomly selected in the drawn sample. When the defined target population consists of a larger

Types of sampling methods



number of sampling units, a more sophisticated method would be used to randomly draw the necessary sample. One of the procedures commonly used in marketing research is to incorporate a printed or computer generated table of random numbers to select the sampling units. A table of random numbers is a table that lists randomly generated numbers. Many computer programs have the ability to generate a table random numbers.

With the marketing research students defined above as the target population a random sample could be generated by assigning each students a unique two-digit code ranging from 01 to 30. Then we could go to the table of random numbers and select a starting point, which can be anywhere on the table. Using the partial table of random numbers given below, say we select the upper-left-hand corner of the table (31) as our starting point. We would then begin to read down the first column (or across the first row) and select those two-digit

numbers that matched the numbers within the acceptable range until 10 students had been selected. Reading down the first column, we would start with 31, then go to 14, 49,99, 54 and so on.

A partial table of random numbers

31 25	81 44	54 34	67 03
14 96	99 80	14 54	30 74
49 05	49 56	35 51	68 36
99 67	57 65	14 46	92 88
54 14	95 34	93 18	78 27
57 50	34 89	99 14	57 37
98 67	78 25	06 90	39 90
40 99	00 87	90 42	88 18
20 82	09 18	84 91	64 80
78 84	39 91	16 08	14 89

Source : M.G.Kendall and B. Babington Smith, "Table of Random Sampling Numbers", *Tracts for Computers*, 24 (Cambridge, England : Cambridge University Press, 1946), p.33.

Use only those random numbers that matched the numbers within the acceptable range of 01 to 30. Numbers that fall outside the acceptable range would be disregarded. Thus, we would select students with numbers 14, 20, 25, 05, 09, 18, 06, 16, 08, and 30. If the overall research objectives call for telephone interviews, drawing the necessary sample can be achieved using the random-digit dialing (RDD) technique.

Advantages and disadvantages

The simple random sampling technique has several advantages. The technique is easily understood and the survey's data results can be generalised to the defined target population with a pre-specified margin of error '*e*'.

Another advantage is that simple random samples allow the researcher to gain unbiased estimates of the population's characteristics. This method basically guarantees that every sampling unit of the population has a known and equal chance of being selected, no matter the actual size of the sample, resulting in a valid representation of the
Defined target population. The disadvantage of this method is the difficulty of obtaining a complete, current, and accurate listing of the population elements. Simple random sampling requires that all sampling units be identified. For this reason, simple random sampling often works best for small populations or those where computer-derived lists are available.

Activity 2.3.
By using a standard random number table, generate random numbers for selecting 50 samples from a population of 800 students who are doing a course on Marketing Research under ABC University.

System

Systematic random sampling (SYMRS) is similar to simple random sampling but requires that the defined target population be ordered in some way, usually in the form of a customer list, taxpayer roll, or membership roster. In research practices, SYMRS has become a very popular alternative probability method of drawing samples. Compared to simple random sampling, systematic random sampling is potentially less costly because it can be done relatively quickly. When executed properly, SYMRS can create a sample of objects or prospective respondents that is very similar in quality to a sample drawn using simple random sampling.

To employ SYMRS, the researcher must be able to secure a complete listing of the potential sampling units that make up the defined target population. Individual sampling units are selected according to their position using a skip interval. The skip interval is determined by dividing the number of potential sampling units in the defined target population by the number of units desired in the sample. The required skip interval is calculated using the formula :

$$\text{Skip interval} = \frac{\text{Defined target population list size}}{\text{Desired sample size}}$$

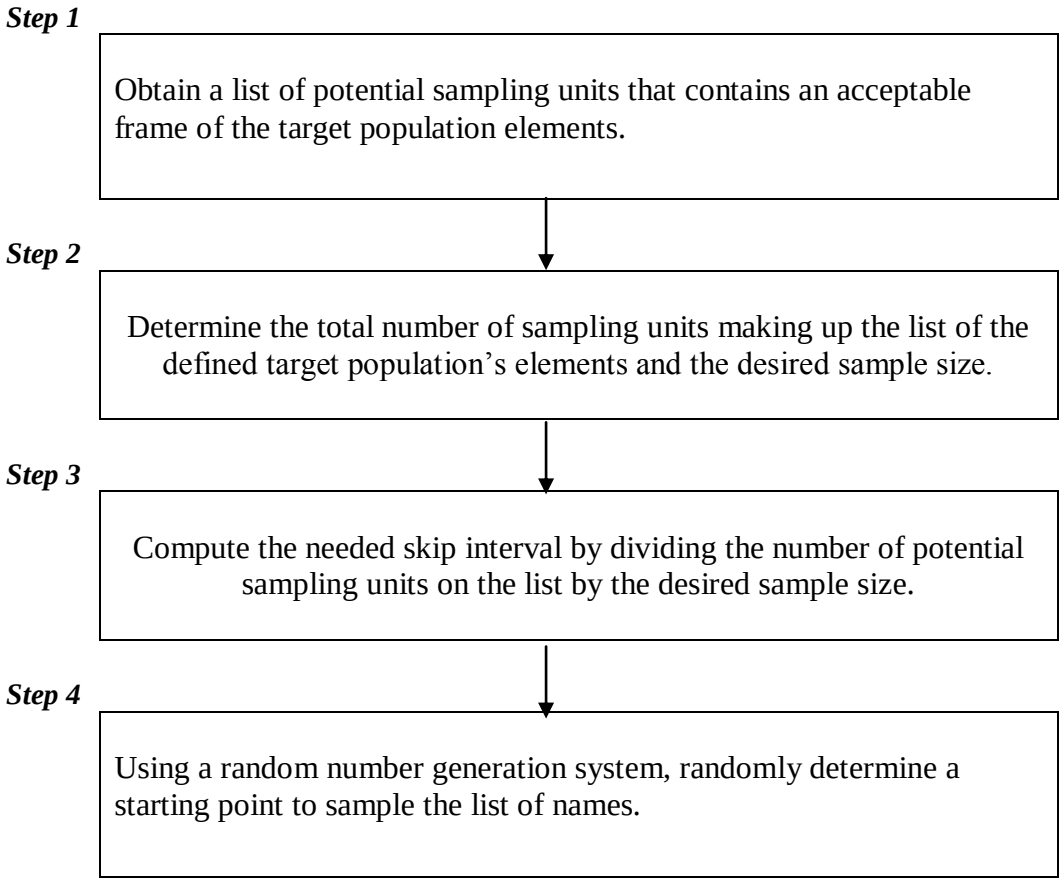
For instance, if the researcher wants a sample of 100 to be drawn from a defined target population of 1,000, the skip interval would be 10 (ie. 1,000/100). Once the skip interval is determined, the researcher would then randomly select a starting point and take every 10^{th} unit until he proceeded through the entire target population list.

There are two important considerations when using systematic random sampling. First, it is important that the natural order of the defined target population list be unrelated to the characteristic being studied. Second, the skip interval must not correspond to a systematic change in the target population.

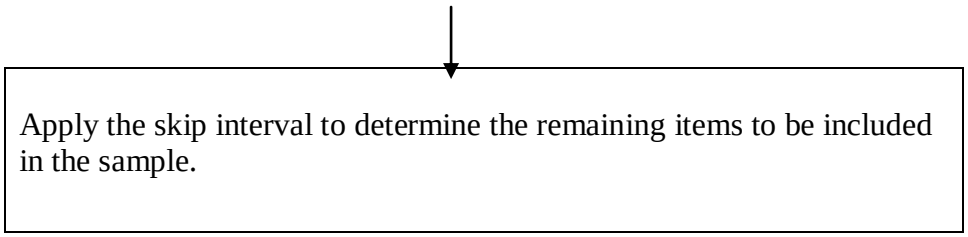
Activity 2.4.

Compare and contrast simple random sampling and systematic random sampling.
Take 10 samples from a population of all the students in your class.
Write a short report on how differently the samples came out of the two methods and which method generated a truly representative sample of your class.

Steps in drawing a Systematic Random Sample



Step 5



Advantages and disadvantages

Systematic sampling is frequently used because, if done correctly, it is a relatively easy way to draw a sample while ensuring randomness. The availability of lists and the shorter time required to draw a sample makes systematic sampling an attractive, economical methods for researchers. The greatest weakness of systematic random sampling is the potential for there to be hidden patterns in the data that are not found by the researcher. This could result in a sample that is not truly representative of the defined target population. Another difficulty is that the researcher must know exactly how many sampling units make up the defined target population. In research situations in which the size of the target population is extremely large or unknown, identifying the true number of units is difficult, and even estimates may not be accurate.

Stratified Random Sampling

Stratified random sampling (STRS) requires the separation of the defined target population into different groups, called strata, and the selecting of samples from each stratum. The goal in stratifying is to minimize the variability (or skewness) within each stratum and maximize the differences between strata. In some ways, STRS can be compared to segmentation of the defined target population into smaller, more homogeneous sets of elements.

To ensure that the sample maintains the required precision of the total population, representative samples must be drawn from each of the smaller population groups. Drawing a stratified random sample involves three basic steps:

i. Dividing the target population into homogeneous sub-groups or strata.
ii. Drawing random samples from each stratum.
iii. Combining the samples from each stratum into a single sample of the target population.

There are two common methods for deriving samples from the strata : proportionate and disproportionate. In proportionate stratified sampling, the sample size from each stratum is depended on the stratum's size relative to the defined target population. Therefore, the larger strata are sampled more heavily using proportionate stratified sampling because they make up a larger percentage of the target population. In disproportionate stratified sampling, the sample size selected from each stratum is independent of that stratum's proportion of the total defined target population. This approach is used when stratification of the target population produces sample sizes for sub-groups that contradict their relative importance to the study.

An alternative type of disproportionate stratified method is optimal allocation. In this method, consideration is given to the relative size of the stratum as well as variability within the stratum to determine the necessary sample size of each stratum. The basic logic underlying optimal allocation is that the greater the homogeneity of the prospective sampling units within a particular stratum, the fewer the units that would have to be selected to the estimate the true population parameter (*0* or *P*) accurately for that sub- group.

The different types of stratified sampling are :

Proportional Stratified Sampling

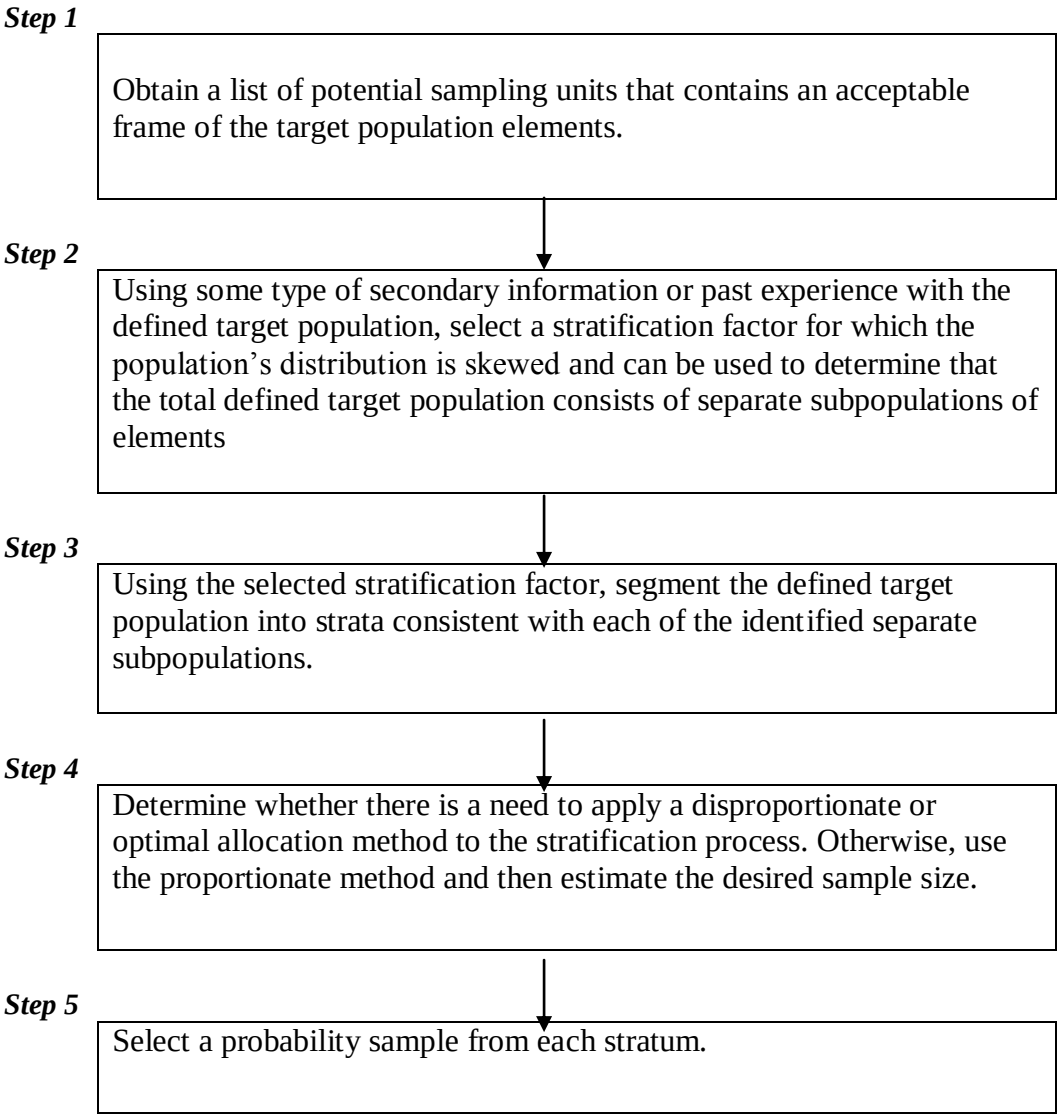
In this type of sampling procedure the number of objects or sampling units chosen from each group is proportional to the number in the population. Proportional stratified sampling can further be classified as *directly proportional* and *inversely proportional stratified sampling*.

Directly Proportional Stratified Sampling

Assume that a researcher is evaluating customer satisfaction for a beverage that is consumed by a total of 600 people. Among the 600 people, 400 are brand-loyal and 200 are variety-seeking. Past research indicates that the level of customer satisfaction is related to consumer characteristics, such as being either brand-loyal or variety-seeking. Therefore, it should be beneficial to divide the total population of 600 consumers into two groups 400 and 200 each and randomly sample from within each of the two groups. If a sample size of 60 is desired, then a 10 percent directly proportional stratified sampling is employed.

Consumer Type	**Group Size**	**10 Percent Directly Proportional Stratified Sample Size**
Brand-loyal	400	40
Variety-seeking	200	20
Total	600	60

Steps in drawing a Stratified Random Sample



Inversely Proportional Stratified Sampling

Assume that among the 600 consumers in the population, say 200 are very heavy drinkers and 400 are light drinkers. If a researcher values the opinion of the heavy drinkers more than that of the light drinkers, more people will have to be sampled from the heavy drinkers group. In such instances, one can use an

inversely proportional stratified sampling. If a sample size of 60 is desired, a 10 percent inversely proportional stratified sampling is employed.

Consumer Type	**Group Size**	**10 Percent Inversely Proportional Stratified Sample Size**
Heavy Drinkers	400	40
Light Drinkers	200	20
Total	600	60

In inversely proportional stratified sampling, the selection probabilities are computed as follows:

Denominator = 600/200 + 600/400 = 3 + 1.5 = 4.5

Heavy Drinkers proportional sample size = 3/ 4.5 = 0.667; 0.667 x 60 = 40

Light Drinkers proportional sample size = 1.5/4.5 = 0.333; 0.333 x 60 = 20

Disproportional Stratified sampling

In stratified sampling, when the sample size in each group is not proportional to the respective group sizes, it is known as disproportional stratified sampling. When multiple groups are compared and their respective group sizes are small, a proportional stratified sampling will not yield a sample size large enough for meaningful comparisons, and disproportional stratified sampling is used. One way of selecting sample sizes within each group is to have equal group sizes in the sample. In the example of heavy and light drinkers, a researcher could select 30 people from each of the two groups.

In general, stratified sampling is employed in many research projects, because it is easy to understand and execute.

Advantages and Disadvantages

Dividing the defined target population into homogeneous strata provides several advantages, including:

i. the assurance of representativeness in the sample;
ii. the opportunity to study each stratum and make relative comparison between strata; and
iii. the ability to take estimates for the target population with the expectation of greater precision or less error in the overall sample.

The primary difficulty encountered with stratified sampling is determining the basis for stratifying. It is imperative that the basis for stratifying be directly associated with the target population's characteristics of interest. Normally, the larger the number of relevant strata, the more precise the results. The inclusion

of excess or irrelevant strata will only waste time and money without providing meaningful results.

Cluster Sampling

While cluster sampling is similar to stratified random sampling, it is different in that the sampling units are divided into mutually exclusive and collectively exhaustive sub-populations, called clusters. Each cluster is assumed to be representative of the heterogeneity of the target population. Examples of possible divisions for cluster sampling include the customers who patronise a store on a given day, the audience for a movie shown at a particular time (e.g., the matinee), or the invoices processed during a specific week. Once the cluster has been identified, the prospective sampling units are drawn into the sample by either using a simple random sampling method or canvassing all the elements within the defined cluster.

Area sampling

A popular form of cluster sampling is **area sampling**. In area sampling, the clusters are formed by geographic designations. Examples include cities, sub divisions and blocks. Any geographical unit with identifiable boundaries can be used. When using area sampling, the researcher has to additional options : the one-step approach or the two -step approach. When deciding on using one-step approach, the researcher must have enough prior information about the various geographic clusters. By assuming that all the clusters are identical, the researcher can focus his attention on surveying the sampling units within one designated cluster group and then generalize the data results to the full target population. The probability aspect of this particular sampling method is executed by randomly selecting one geographic cluster and performing a census on all the sampling units located within that selected cluster.

Alternatively, the researcher may execute a two-step cluster sampling approach. First, the researcher would randomly sample a set of cluster and then would decide on the most appropriate probability method to sample individuals within each of the selected clusters. The two-step approach is preferable over the one-step approach, because there is a strong possibility that a single cluster will not be as representative of all other clusters as the researcher thinks.

Advantages and Disadvantages

The cluster sampling method is widely used in marketing research due to its overall cost-effectiveness and feasibility of implementation, especially in area sampling situations. In many cases, the only reliable sampling unit frame

available to researcher is one that describes and lists clusters. These lists of geographic regions, telephone exchanges, or blocks of residential dwellings can normally be easily compiled. Clustering method tends to be a cost-efficient way of sampling and collecting raw data from a defined target population.

One primary disadvantage related to cluster sampling is the tendency for clusters to be homogeneous. The more homogeneous the cluster, the less precise the derived sample estimate in representing the defined target population's parameters. The actual object or people within a cluster should be as heterogeneous as those in the target population itself. Another concern with cluster sampling methods is the appropriateness of the designated cluster factor used to identify the sampling units within clusters.

A comparison between the stratified sampling process and the cluster sampling process is given in the following table :

Stratified Sampling	**Cluster Sampling**
Homogeneity within group	Homogeneity between groups
Homogeneity between groups	Homogeneity within group
All groups are included	Random selection of groups
Sampling efficiency improved by increasing accuracy at a faster rate than cost	Sampling efficiency improved by decreasing cost at a faster rate than accuracy

Activity. 2.5.
Compare and contrast various probability sampling methods. Prepare a chart showing their relative merits and demerits.

2.5.2. *Types of non-probability sampling designs*

Non-probability sampling typically is used in situations such as :

i. the exploratory stages of a research project,
ii. pre-testing a questionnaire,
iii. dealing with a homogeneous population,
iv. when a researcher lacks statistical knowledge, and
i. when operational ease is required.

Convenience Sampling

Convenience sampling (or accidental sampling) is a method in which samples are drawn at the convenience of the researcher or interviewer. The assumptions are that the target population is homogeneous and the individuals interviewed are similar to the overall defined target population with regard to the characteristics being studied.

Advantages and Disadvantages

Convenience sampling allows a large number of respondents to be interviewed in a relatively short time. For this reason, it is commonly used in the early stages of research. The use of convenience samples in the development phases of constructs and scale measurements can have a seriously negative impact on the overall reliability and validity of those measures and instruments used to collect raw data. Another major disadvantage of convenience samples is that the raw data and results are not generalized to the defined target population with any measure of precision. It is not possible to measure the representativeness of the sample, because sampling error estimates cannot be accurately determined.

Judgment Sampling

In judgment sampling, (also referred to as purposive sampling), participants are selected according to an experienced individual's belief that they will meet the requirements of the study. Judgmental sampling is associated with a variety of biases. For example, shopping center intercept interviewing can over-sample those who shop frequently, who appear friendly, and who have uncertainty, because the sampling frame is unknown and the sampling procedure is not well specified.

There are situations where judgmental sampling is useful and even advisable.

First, there are times when probability sampling is either not feasible or expensive. For example, a list of sidewalk vendors might be impossible to obtain, and a judgmental sample might be appropriate in that case.

Second, if the sample size is to be very small - say, under 10 - a judgmental sample usually will be more reliable and representative than a probability sample. Suppose one or two cities of medium size are to be used to represent 200 such cities. Then it would be appropriate to pick judgmentally two cities that appeared to be most representative with respect to such external criteria as demographics, media habits, and shopping characteristics.

Third, sometimes it is useful to obtain a deliberately biased sample. If, for example, a product or service modification is to be evaluated, it might be possible to identify a group that, by its very nature, should be disposed toward the modification.

Advantages and Disadvantages

If the judgment of the researcher or expert is correct, then the sample generated from judgment sampling will be much better than one generated by convenience sampling. But, it is not possible to measure the representativeness of the sample. The raw data and information collected from sampling units generated though the judgment sampling method should be interpreted as nothing more preliminary insights.

Quota Sampling

The quota sampling method involves the selection of prospective participants according to pre-specified quota regarding either demographic characteristics (e.g., age, race, gender, income), specific attitudes (e.g., satisfied/dissatisfied, liking/disliking, great/marginal/no quality), or specific behaviours (e.g., regular/occasional/ rare customer, product user/non user). The underlying purpose of quota sampling is to provide an assurance that pre-specified sub-groups of the defined target population are represented on pertinent sampling factors that are determined by the researcher or client. Surveys frequently use quotas that have been determined by the specific nature of the research objectives.

In order to meet the quotas, researcher using quota sampling sometimes overlook the problems associated with adhering to the quotas. Assume that an oil company is interested in finding out if women assume responsibility for vehicle maintenance. The company is interested in interviewing women aged below 35 and with age equal to and above 35, as well as working women and nonworking women. Suppose the distribution of the population of women in a city (N=1,000) is as follows:

Population Characteristics

	<35 years	35 years and Above	*Total*	*Percentage*
Working women	300	200	500	50
Non-working women	200	300	500	50
Total	**500**	**500**	**1,000**	**100**
Percentage	50	50	100	

Assume that the researcher is interested in interviewing 100 women from this city and develops a quota system such that 50 percent of the sample should be working women and 50 percent of the sample should also be under 35 years old. A quota matrix can be developed for a sample size of 100.

Sample Characteristics

	<35 years	**35 years and Above**	***Total***	***Percentage***

Working women	50	0	50	50
Non-working women	0	50	50	50
Total	**50**	**50**	**100**	**100**
Percentage	50	50		

Advantages and Disadvantages

The greatest advantage of quota sampling is that the sample generated contains specific sub-groups in the proportions desired by researchers. In those research projects that require interviews, the use of quotas ensures that the appropriate sub-groups are identified and included in the survey. The quota sampling method may eliminate or reduce selection bias on the part of the field workers. An inherent limitation of quota sampling is that the success of the study will be dependent on subjective decisions made by the researchers. Also, it is incapable of measuring the true representativeness of the sample or accuracy of the estimate obtained. Hence, attempts to generalise data results beyond those respondents who were sampled and interviewed become very questionable and may misrepresent the defined target population.

Snowball Sampling

Snowball Sampling involves the practice of identifying and qualifying a set of initial prospective respondents who can, in turn, help the researcher identify additional people to be included in the study. This method of sampling is also called referral sampling, because one respondent refers other potential respondents. Snowball sampling is typically used in research situations where :

i. the defined target population is very small and unique, and

ii. compiling a complete list of sampling units is a nearly impossible task.

The snowball method would yield better results at a much lower cost. Here the researcher would identify and interview one qualified respondent, then solicit his or her help in identifying other people with similar characteristics. The main underlying logic of this method is that rare groups of people tend to form their own unique social circles.

Advantages and Disadvantages

Snowball sampling is a reasonable method of identifying and selecting prospective respondents who are members of small, hard-to-reach, uniquely defined target population. It is most useful in qualitative research practices, like focus group interviews. Reduced sample sizes and costs are primary advantages to this sampling method. Snowball sampling definitely allows bias to enter the overall research study.

Activity. 2.6.
Compare and contrast various non-probability sampling methods. Prepare a chart showing their relative merits and demerits.

2.6. Random-Digit Dialing (RDD)

If truly accurate or representative, current sampling frames are not available, a researcher would have to employ an alternate method such as random-digit dialing, if conducting telephone interviews. The only solution to the problem of unlisted telephone numbers is to generate phone numbers by some random process. This practice, referred to as Random-Digit Dialing involves generation of random numbers at random. But practical considerations complicate the problem greatly. The foremost is the relatively small proportion of the working numbers among all possible 10-digit telephone numbers. Only about 1 in 170 of all possible telephone numbers is actually in use. The proportion of the working residential numbers in RDD samples can be increased dramatically by selecting from only the 103 working area codes (first three digits). The approach yields approximately 1 working residential number for every 17 randomly generated. From a cost stand point, this rate is still low, entailing too many unproductive dialings while including a proportionate number of unlisted phone homes in the sample. There are three alternative approaches built around the use of a telephone book.

Four-Digit approach

Taking the four-digit approach, the researcher must, in addition to restricting the sample to the 103 working area codes, select numbers only from working central offices or exchanges. The lat four digits of the number are generated via some process that approaches randomness. Problem with this approach is that all exchanges have an equal probability of being selected, while some have a high proportion of all possible numbers in service and others have only a small proportion in service.

Three-Digit approach

The next logical progression in RDD is the three-digit dialing approach. This method increases the proportion of working numbers to better than one in three. Consulting the section of a criss-cross directory where phone numbers are listed numerically will show that within a particular exchange, certain sequences of 1000 numbers are totally unused while other groups of 1000 are in use. Generate the last three digits of each exchange by means of some random process. This method is more efficient in eliminating non working numbers, but increases bias due to missing new exchanges that have been activated.

The four digit method is safer from the standpoint of avoiding bias, but more expensive due to the greater number of calls that must be made. It is suggested that three-digit method is most appropriate when the directories for the area of interest are relatively current or when there has been little growth in the area since the publication of the re cent directory. In other cases, the four-digit method should be used.

Using telephone books

RDD samples can also be generated from the telephone book. This is accomplished by selecting numbers at random from the book and adding a random number as the sixth or seventh digit. Somewhere between one in two and one in three of the numbers generated will be working residential numbers. This is a viable approach because, all exchanges are proportionately represented in the book. The phone book is recommended as an RDD sample source only in those cases where the appropriate computer hardware and software are not available. There are two major reasons for making this recommendation. First, the construction of sample by this approach is time-consuming and expensive. Second, if the interviewers are given directions and left to generate numbers themselves, the researcher loses all control over the validity of the sample. Computer programmes can incorporate three- or four- digit approaches and generate RDD samples at a very low cost. The print-out can be set up to capture additional data and to help the researcher control field costs and proper execution of the sampling plan.

Activity 2.7.
Identify a situation where you would be in favour of using a non-probability sampling method over probability sampling method.

2.7. Sample size

Determining the appropriate sample size is not an easy task. The researcher must consider how precise the estimates must be and how much time and money are available to collect the required data, since the data collection is one of the most expensive components of a study. Three factors play an important role in determining appropriate sample sizes. They are :

i. *The variability of the population characteristic under consideration* : The greater the variability of the characteristic, the larger the size of the sample necessary.

ii. *The level of confidence desired in the estimate* : The higher the level of confidence desired, the larger the sample size needed.

iii. *The degree of precision desired in estimating the population characteristic* : The more precise the required sample results, the larger the necessary sample size.

2.7.1. ***Estimating the sample size by traditional methods***

There are four traditional approaches to determine the sample size. They are :

i. *Judgementally / arbitrarily* : The researcher can simply select a sample size arbitrarily or on the basis of some judgementally based criterion.. There may be instances where the sample size represents all that where available at a particular point of time.

ii. *Analysis considerations* : Analysis considerations may decide the sample size. Sample size may be determined from the minimum cell size needed.

iii. *The budget* : In certain cases, the budget may determine the sample size.

iv. *Applying standard error* : Sample size determination is based on specifying the desired precision in advance and then applying the appropriate standard error formula.

Two major classes of procedures are available for estimating the sample size

1. Confidence – interval approach : This is based on the idea of constructing confidence intervals around sample means or proportions.
2. Hypothesis-testing approach : This makes use of both type I error (rejecting a true null hypothesis) and Type II error (accepting a false null hypothesis).

Confidence – interval approach

In this method, a confidence interval is constructed around sample based mean or proportion. The standard error formulae are used for this purpose. This can be explained with an example. Consider a researcher my have taken a sample of 100 consumers and noted that their average per-capita consumption of orange juice was 2.6 litres per week. Pat studies indicate that the population standard deviation T can be assumed to be 0.3 litre. With this information, we can find a range around the sample mean level of 2.6 litres for which some pre-specified probability statement can be made about the process underlying the construction of such confidence intervals. Suppose that we want to set up a 95% confidence interval around the sample mean of 2.6 litres. The standard error of the mean can be computed as :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.3}{\sqrt{100}} = 0.03$$

From the table, we can find that the central 95% of the normal distribution lies within C 1.96 Z variates.
95% confidence interval ranges from 2.54 to 2.66 litres. Thus the pre- assigned chance of finding the true population mean to be within 2.54 and 2.66 litres is 95%.

The case of sample mean

Following are the steps involved :

1. *Specify the amount of error (E) that can be allowed*. This is the maximum allowable difference between the sample mean and the population mean. 8 C E defines the interval within which 0 will lie with some pre-specified level of confidence.
2. *Specify the desired level of confidence*. It can be 95%.
3. *Determine the number of standard errors (Z) associated with the confidence level*.
4. *Estimate the standard deviation of the population*. The standard deviation can be estimated by judgment, by reference to other studies or by the use of a pilot sample.
5. *Calculate the sample size using the formula for the standard error of the mean*.

$$\sigma_{\bar{x}} = \frac{E}{Z}$$

6. Neglacting the finite multiplier, we solve for *n* in the formula

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

7. In general we can find n from the formula

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

The case of sample proportion

The procedure for determining sample size for interval estimates of proportion are :

1. *Specify the amount of error that can be allowed*. Suppose that the desired reliability is such that an allowable interval of $p - A = \pm$ 0.05 is set, the allowable error *E* is 0.05.
2. *Specify the desired level of confidence*. Suppose that the level of confidence is 95%.
3. *Determine the number of standard errors (Z) associated with the confidence level*.
4. *Estimate the population proportion*. (A). The population proportion can be estimated by judgment, by reference to other studies or by the use of a pilot sample.
5. *Calculate the sample size using the formula for the standard error of the proportion*.

$$\sigma p = \frac{E}{Z}$$

6. Neglacting the finite multiplier, we solve for *n* in the formula

$$\sigma p = \frac{E}{Z} = \sqrt{A\,(1-A)/n}$$

7. We can find n from the formula

$$n = \frac{A\,(1-A)\,Z^2}{E^2}$$

Hypothesis testing approach

Sample size can also be determined by the hypothesis testing approach. For this, an assumed probability of making Type I error (called alpha risk) and the probability of making Type II error (called beta risk) are needed. These risks are based on the hypotheses :

H_0 : the null hypothesis

H_1 : the alternate hypothesis

In hypothesis testing, the sample results sometimes lead us to reject H_0 when its is true. This is a type I error. On other occasions, the sample findings may lead us to accept H_0 when it is false. This is a Type II error.

The case involving means

The steps are :

1. Specify the values for the null (H_0) and the alternate (H_1) hypotheses to be tested in terms of population means, μ_0 and μ_1 respectively.
2. Specify the allowable probabilities (α and β respectively) of Type I and Type II errors. The Type I error is the error of rejecting a true null hypothesis. The Type II error is made if the alternate hypothesis is rejected when it is true. α and β are the allowable probabilities of making those two types of errors respectively.
3. Determine the number of standard errors associated with each of the error probabilities α and β.
4. Estimate the population standard deviation σ.
5. Calculate the sample size that will meet the α and β error requirements. Since two sampling distributions are involved, a simultaneous solution of two equations is required to determine the sample size and critical value that will satisfy both equations. These equations are :

$$\text{Critical value} = \mu_0 + Z_\alpha \left[\frac{\sigma}{\sqrt{n}} \right]$$

$$\text{Critical value} = \mu_1 - Z_\beta \left[\frac{\sigma}{\sqrt{n}} \right]$$

6. Setting the right hand side of these two equations equal and solving for n gives

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

$$(\mu_1 - \mu_0)^2$$

The case involving proportions

The steps are :

1. Specify the values for the null (H_0) and the alternate (H_1) hypotheses to be tested in terms of population proportions, π_0 and π_1 respectively.
2. Specify the allowable probabilities (α and β respectively) of Type I and Type II errors.
3. Determine the number of standard errors associated with each of the error probabilities Z_α and Z_β.
4. Calculate the desired sample size n from the formula :

$$n = \left(\frac{Z_\alpha \sqrt{[\pi_0 (1- \pi_0)]} + Z_\beta \sqrt{[\pi_1 (1- \pi_1)]}}{\pi_1 - \pi_0} \right)^2$$

2.7.2. *Bayesian approach to sample size determination*

Bayesian procedures are based on the central principle that one should select the sample size that results in the largest positive difference between the expected payoff of sample information and the estimated cost of sampling. The difference between the expected payoff of the sample information and the estimated cost of sampling is frequently referred to as the expected net gain from sampling. An equivalent way of stating the principle is that one should select the sample size that leads to the largest expected net gain from sampling.

In a decisional situation in which one of the primary objectives is to maximize payoff, this rule is appropriate. The general approach to applying it requires the decision maker to :

- Determine the expected value of the sample information for a given sample size.

- Estimate the sampling cost for that specific option.
- Find the expected net gain from sampling under that option.
- Search through other sample sizes to find the one that leads to the highest expected net gain from sampling.

While logically sound concept, the Bayesian approach is difficult to implement. The primary problem comes in operationalising the first of the steps stated above. In order to determine the expected value of the sample information for a given sample size, one must relate the sample size being considered to the conditional probabilities of making errors, including the effects of non-sampling errors. In real life situations, this may become very difficult to do.

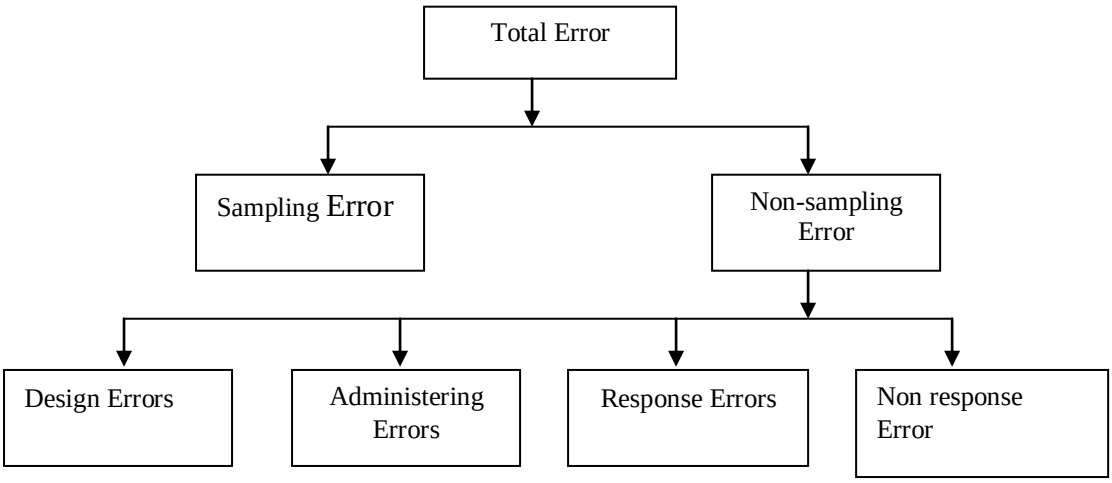
2.8. Sampling error

Several potential sources of error can affect the quality of a research process. The errors can influence the various stages of the research process and result in inaccurate or useless research findings. Researchers have numerous opportunities to make mistakes or errors in judgment, that result in creating some type of bias in any research study. All types of errors can be logically classified as either sampling or non-sampling errors. Random sampling errors can be detected by observing the difference between the sample results and the results of a census conducted using identical procedures. Two difficulties are associated with detection of the sampling error :

i. the fact that very seldom is a census conducted in survey research and

ii. sampling error can be determined only after the sample is drawn and data collection has been completed.

The Total Error in a research study is the difference between the true mean value (within the population) of the variable being studied and the observed mean value obtained through the research study.



Sampling error is any type of bias that is attributable to mistakes made in either the selection process of prospective sampling units or determining the sample size. Sampling error is the difference between a measure obtained from a sample representing the population and the true measure that can be obtained only from the entire population. This error occurs because no sample is a perfect representation of a given population, unless the sample size equals the population. Random sampling error tends to occur because of chance variations in the scientific selection of the needed sample units. Even if the sampling units were properly selected according to the guidelines of sampling theory, those units still might not be a perfect representation of the defined target population. When there is a discrepancy between the statistic calculated or estimated from the sample and the actual value from the population, a sampling error has occurred. Based on the principles of the central limit theorem, the size of the sampling error and its impact can be reduced by increasing the size of the sample.

Ways to minimize sampling error

1. Increase the sample size.
2. Use a statistically efficient sampling plan. That is making the sample as representative of the population as possible.

Non-sampling errors are those types of biases that occur in a research study regardless of whether a sample or census is used. There may be several reasons for these errors. Non-sampling errors can be broadly classified into four groups :

ii. Design errors
iii. Administering errors
iv. Response errors
v. Non response errors.

i. Design errors : Design errors also called researcher- induced errors are mainly due to flaws in the research design. These errors can take various forms.

Selection error : This occurs when a sample obtained through a non probability sampling method is not representative of the population.

Population specification error : This occurs when an inappropriate population is chosen from which to obtain data for the research study.

Sampling frame error : It occurs when the sample is drawn from an inaccurate sampling frame.

Surrogate information error : It is the difference between the information required for a research study and the information being sought by the researcher.

Measurement error : It is the difference between the information sought by the researcher for a study and the information generated by a particular measurement method employed by the researcher.

Experimental error : Any error caused by the improper design of the experiment induce an experimental error into the study.

Data analysis error : This can occur when the data from the questionnaires are coded, edited, analysed or interpreted.

ii. Administering errors : All errors that occur during the administration of a survey instrument to the respondents are classified as administering errors. They are caused by the person administering the questionnaire. They may be caused by three major factors :

Questioning error : This error arises while addressing questions to the respondents.

Recording error : This arises from improperly recording the respondent's answers.

Inference error : This error occurs when an interviewer interferes with or fails to follow the exact procedure while collecting data.

iii. Response errors : Also called data errors occur when the respondent intentionally or unintentionally provides inaccurate answers to the survey questions.

iv. Non-response errors : These occur if some members of a sample were not contacted or some of the members contacted provide no response to the survey.

2.9. Non-response Problems

The object of sampling is to obtain a body of data that are representative of the population. Unfortunately, some sample members become non-respondents because they :

i. refuse to respond,

ii. lack the ability to respond,

iii. are not at home, or

iv. are inaccessible.

Non-response can be a serious problem. If a sample size of 1,000 is needed and only a 50 percent response rate is expected, then 2,000 people will need to be identified as possible sample members. The seriousness of non-response bias depends on the extent of the non-response. If the percentage involved is small, the bias is small. The non-response problem depends on how the non-respondents differ from the respondents, particularly on the key questions of interest. To deal with non-response problem tendency will be to replace each non-respondent with a "matched" member of the sample. The difficulty is that the replacement cannot be matched easily on the characteristic that prompted the non-response, such as being employed or being a frequent traveller. Three approaches are :

i. to improve the research design to reduce the number of non-responses,
ii. to repeat the contact one or more times (call-backs) to try to reduce non-responses, and
iii. to attempt to estimate the non-response bias.

Improving the research Design

The challenge in personal and telephone interviewing is to gain initial interest and to generate rapport through interviewer skill and the design and placement of questions. In mail surveys, the task is to motivate the respondent to respond, through incentives and other devices. The number of not-at-homes can be reduced by scheduling calls with some knowledge of the respondents' likely activity patterns.

Call-Backs

Call-backs refer to new attempts to obtain responses. The use of call-backs is predicated on the assumption that they will generate a useful number of additional responses and that the additional responses will reduce meaningfully a non-response bias. If the non-response is due to refusals or the inability to respond, call-backs may not reduce significantly the number of non-respondents. They are most effective for the not-at-home non-respondent. The efficiency of the call-backs will be improved by scheduling them at different times of the day and week. In a mail survey, the call-back is particularly important, because the non-response level can be high. It is common practice to remind non-respondents at regular intervals.

Estimating the Effects of Non-response

One approach is to make an extra effort to interview a sub-sample of the non-respondents. In the case of a mail survey, the sub-sample might be interviewed by telephone. In a telephone or personal survey, an attractive incentive, such as a

worthwhile gift, might be employed to entice a sample of the non-respondents to co-operate.

Summary

Sampling is an important concept that we practice in our every day life. Sampling involves selecting a relatively small number of elements from a larger defined group of elements and expecting that the information gathered from the small group will allow judgments to be made about the larger group. If all the respondents in a population are asked to provide information, such survey is called a census. Information obtained from a subset of the population is known as the statistic (from sample).

Selection of the most appropriate sampling design should incorporate the seven factors. They are : Research objectives, Degree of accuracy, Availability of resources, Time frame, Advanced knowledge of the target population, Scope of the research and Perceived statistical analysis needs. A sampling plan is the blueprint or frame work needed to ensure that the raw data collected are representative of the defined target population.

There are two basic sampling designs : Probability and non-probability sampling methods. In probability sampling, each unit in the defined target population has a known, non-zero probability of being selected for the sample, The actual probability of selection for each sampling unit may or may not be equal depending on the type of probability sampling design used. It allows the researcher to judge the reliability and validity of raw data collected by calculating the probability to which the findings based on the sample would differ from the defined target population. The results obtained by the probability method can be generalized to the target population within a specified margin of error through the use of statistical methods. In non-probability sampling, the probability of selection of each sample unit is not known. Therefore, potential sampling error cannot be accurately known either. The selection of sampling units is based on some type of intuitive judgments, desire or knowledge of the researcher. The degree to which the sample may or may not be representative of the defined target population depends on the sampling approach and how well the researcher executes and controls the selection activities. There is always a temptation to generalize non-probability sample data results to the defined target population.

Various probability sampling methods are : a. Simple random sampling, b. Systematic Random Sampling, c. Stratified Random Sampling, d. Cluster Sampling , e. Area sampling. Non-probability sampling methods can be classified as : a. Convenience Sampling, b. Judgment Sampling, c. Quota Sampling, d. Snowball Sampling.

Random-Digit Dialing involves generation of random numbers at random. If truly accurate or representative, current sampling frames are not available, a researcher would have to employ an alternate method such as random-digit dialing, if conducting telephone interviews.

The traditional approaches in determining the sample size are :

i. Judgementally / arbitrarily, ii. Analysis considerations, iii. The budget, iv. Applying standard error. Bayesian procedures are based on the central principle that one should select the sample size that results in the largest positive difference between the expected payoff of sample information and the estimated cost of sampling.

Sampling error is the difference between a measure obtained from a sample representing the population and the true measure that can be obtained only from the entire population. The Total Error in a research study is the difference between the true mean value (within the population) of the variable being studied and the observed mean value obtained through the research study. Total error can be classified into sampling error and non-sampling error.

Some sample members become non-respondents because they : i. refuse to respond, ii. lack the ability to respond, iii. are not at home, or iv. are inaccessible.

Self Assessment Questions

1. Define sampling.
2. Give the differences between census and sampling.
3. Define the following :
    - Population
    - Element
    - Sampling units
    - Sampling frame
    - Sampling gap
    - Sampling distribution
4. What is Central Limit Theorem ?
5. What are the factors influencing the selection of a sampling design ?
6. Define sampling plan. What are the steps in developing a sampling plan ?
7. Differentiate between probability and non-probability sampling methods.
8. What are the types of probability sampling methods ?
9. Define the following :
    - Simple random sampling
    - Systematic Random Sampling
    - Stratified Random Sampling
    - Cluster Sampling
    - Area sampling
10. What are the types of non-probability sampling methods ?
11. Define the following :
    - Convenience Sampling
    - Judgment Sampling
    - Quota Sampling
    - Snowball Sampling
12. What is Random-Digit Dialing (RDD) ?
13. What are the factors in determining appropriate sample sizes ?
14. What are the traditional approaches in determining the sample size ?
15. What is Bayesian approach to sample size determination ?
16. What is sampling error ? What are the types of total error ?
17. What are the types of non-sampling errors ?
18. What are the reasons for non-response of respondents ?

Answer Key

1. Sampling involves selecting a relatively small number of elements from a larger defined group of elements and expecting that the information gathered from the small group will allow judgments to be made about the larger group.
2. If all the respondents in a population are asked to provide information, such survey is called a census. A subset of the population is known as the sample.

3. a. Population : A population is an identifiable total group or aggregation of elements that are of interest to the researcher and pertinent to the specified problem.

b. Element : An element is a person or object from which data and information are sought.

c. Sampling units : Sampling units are the target population elements available for selection during the sampling process.

d. Sampling frame : Sampling frame is the list of all eligible sampling units.

e. Sampling gap : A sampling gap is the representation difference between the population elements and sampling units in the sampling frame.

f. Sampling distribution : Sampling distribution is the frequency distribution of a specific sample statistic (sample mean or sample proportion) from repeated random samples of the same size.

4. The theorem states that for almost all defined target populations, the sampling distribution of the mean or the percentage value derived from a simple random sample will be approximately normally distributed, provided that the sample size is sufficiently large.

5. Research objectives, Degree of accuracy, Availability of resources, Time frame, Advanced knowledge of the target population, Scope of the research and Perceived statistical analysis needs.

6. A sampling plan is the blueprint or frame work needed to ensure that the raw data collected are representative of the defined target population. A good sampling plan will include, the following steps: (1) define the target population, (2) select the data collection method, (3) identify the sampling frames needed, (4) select the appropriate sampling method, (5) determine necessary sample sizes and overall contact rates, (6) create an operating plan for selecting sampling units, and (7) execute the operational plan.

7.

Factor sampling	Probability sampling	Non-probability
a. List of the population elements	Complete list necessary	Not necessary
b. Information about the-sampling units	Each unit identified	need detail on habits, activities, traits etc.
c. Sampling skill	skill required	little skill required
d. Time requirement	More time-consuming	Less time consuming
e. Cost per unit sampled	Moderate to high	Low
f. Estimate of population - parameters	Unbiased	Biased
g. Sample representativeness	Assured	Undeterminable
h. Accuracy and Reliability	Computed with confidence intervals	Unknown
i. Measurement of sampling error	Statistical measures	No true measure available

8. a. Simple random sampling
   b. Systematic Random Sampling
   c. Stratified Random Sampling
   d. Cluster Sampling
   e. Area sampling.

9.

a. Simple random sampling : Simple Random Sampling is a probability sampling procedure which ensures that every sampling unit making up the defined target population has a known, equal, non-zero chance of being selected.

b. Systematic Random Sampling : Systematic random sampling (SYMRS) is similar to simple random sampling but requires that the defined target population be ordered in some way, usually in the form of a customer list, taxpayer roll, or membership roster.

c. Stratified Random Sampling : Stratified random sampling (STRS) requires the separation of the defined target population into different groups, called strata, and the selecting of samples from each stratum.

d. Cluster Sampling : The sampling units are divided into mutually exclusive and collectively exhaustive sub-populations, called clusters. Each cluster is assumed to be representative of the heterogeneity of the target population.

e. Area sampling :The clusters are formed by geographic designations. Examples include cities, sub divisions and blocks.

10. a. Convenience Sampling
    b. Judgment Sampling
    c. Quota Sampling
    d. Snowball Sampling

11.

a. Convenience Sampling : Convenience sampling (or accidental sampling) is a method in which samples are drawn at the convenience of the researcher or interviewer.

b. Judgment Sampling : In judgment sampling, (also referred to as purposive sampling), participants are selected according to an experienced individual's belief that they will meet the requirements of the study.

c. Quota Sampling : The quota sampling method involves the selection of prospective participants according to pre-specified quota regarding either demographic characteristics, specific attitudes, or specific behaviours.

d. Snowball Sampling : Snowball Sampling involves the practice of identifying and qualifying a set of initial prospective respondents who can, in turn, help the researcher identify additional people to be included in the study.

12. Random-Digit Dialing involves generation of random numbers at random. If truly accurate or representative, current sampling frames are not available, a researcher would have to employ an alternate method such as random-digit dialing, if conducting telephone interviews.

13. i. The variability of the population characteristic under consideration.
    ii. The level of confidence desired in the estimate.
    iii. The degree of precision desired in estimating the population characteristic.

14. i. Judgementally / arbitrarily.
    ii. Analysis considerations.
    iii. The budget.
    iv. Applying standard error.

15. Bayesian procedures are based on the central principle that one should select the sample size that results in the largest positive difference between the expected payoff of sample information and the estimated cost of sampling.

16. Sampling error is the difference between a measure obtained from a sample representing the population and the true measure that can be obtained only from the entire population. The Total Error in a research study is the difference between the true mean value (within the population) of the variable being studied and the observed mean value obtained through the research study. Total error can be classified into sampling error and non-sampling error.

17.Non-sampling errors can be broadly classified into four groups :
    i. Design errors
    ii. Administering errors
    iii. Response errors
    iv. Non response errors

18. Some sample members become non-respondents because they :
    i. refuse to respond,
    ii. lack the ability to respond,
    iii. are not at home, or
    iv. are inaccessible.

Glossary

Area sampling :The clusters are formed by geographic designations. Examples include cities, sub divisions and blocks.

Cluster Sampling : The sampling units are divided into mutually exclusive and collectively exhaustive sub-populations, called clusters. Each cluster is assumed to be representative of the heterogeneity of the target population.

Convenience Sampling : Convenience sampling (or accidental sampling) is a method in which samples are drawn at the convenience of the researcher or interviewer.

Element : An element is a person or object from which data and information are sought.

Judgment Sampling : In judgment sampling, (also referred to as purposive sampling), participants are selected according to an experienced individual's belief that they will meet the requirements of the study.

Population : A population is an identifiable total group or aggregation of elements that are of interest to the researcher and pertinent to the specified problem.

Quota Sampling : The quota sampling method involves the selection of prospective participants according to pre-specified quota regarding either demographic characteristics, specific attitudes, or specific behaviours.

Random-Digit Dialing : Random-Digit Dialing involves generation of random numbers at random. If truly accurate or representative, current sampling frames are not available, a researcher would have to employ an alternate method such as random-digit dialing, if conducting telephone interviews.

Sampling distribution : Sampling distribution is the frequency distribution of a specific sample statistic from repeated random samples of the same size.

Sampling frame : Sampling frame is the list of all eligible sampling units.

Sampling gap : A sampling gap is the representation difference between the population elements and sampling units in the sampling frame.

Sampling units : Sampling units are the target population elements available for selection during the sampling process.

Simple random sampling : Simple Random Sampling is a probability sampling procedure which ensures that every sampling unit making up the defined target population has a known, equal, non-zero chance of being selected.

Snowball Sampling : Snowball Sampling involves the practice of identifying and qualifying a set of initial prospective respondents who can, in turn, help the researcher identify additional people to be included in the study.

Stratified Random Sampling : Stratified random sampling (STRS) requires the separation of the defined target population into different groups, called strata, and the selecting of samples from each stratum.

Systematic Random Sampling : Systematic random sampling (SYMRS) is similar to simple random sampling but requires that the defined target population be ordered in some way, usually in the form of a customer list, taxpayer roll, or membership roster.

Reference

1. Paul. E. Green, Donald. S. Tull, Gerald Albaum, *Research for Marketing Decisions*, Prentice Hall of India Pvt. Ltd. New Delhi.
2. Joseph. F. Hair, Robert. P. Bush, David. J. Ortinan, *Marketing Research*, Tata McGraw-Hill Publishing Company Ltd, New Delhi.
3. David. A. Aaker, V. Kumar, George.S.Day, *Marketing Research*, John Wiley & Sons Inc, Singapore.
4. William. M.K. Trochim, *Research Methods*, Biztantra, New Delhi.

Unit III

THE MEASUREMENT

3.1 Introduction
3.2 Definition of measurement
3.3 Functions of measurement
3.4 Advantages and disadvantages of measurement
3.5 Uses of measurement in research
3.6 Fundamentals of measurement
3.7 Types of measurement

Learning objectives

To understand the concept of measurement
To learn about the measurement process and how to develop a good measurement scale
To understand the four levels of scales and their typical usage
To explore the concepts of reliability and validity

The business research revolves around measurement, which refers to obtaining symbols to represent properties of objects, events or states. Measurement helps in identifying attitudes of individuals .When we measure an object, essentially all we are doing is counting the number of standard pieces it takes to be the same size as the object.

Measurement is defined as:

The determination of size in relation to some observed standard, e.g. metre, kilogram, second, ampere, degree Kelvin, candela, mole, or some unit derived from these seven basic units.

According to the famous British scientist, Lord Kelvin, "When you measure what you are speaking about and express it in numbers, you know something about it, but when you cannot (or do not) measure it, when you cannot (or do not) express it in numbers, then your knowledge is of a meager and unsatisfactory kind."

Measurement is the process observing and recording the observations. Measurement is the estimation or determination of extent, dimension or capacity, usually in relation to some standard or unit of measurement. The measurement is expressed as a number of units of the standard, such as distance being indicated by a number of kilometers.

The process of measuring involves estimating the ratio of the magnitude of a quantity to the magnitude of a unit of the same type length, time etc. A measurement is the result of such a process, expressed as the product of a real number and a unit, where the real number is the estimated ratio. Example is 100 meters, which is an expression of an object's length relative to a unit of length, the meter etc.,

The fundamental ideas involved in measuring are nominal, ordinal, interval and ratio. We consider four broad categories of measurements. Survey research includes the design and implementation of interviews and questionnaires. Scaling involves consideration of the major methods of developing and implementing a scale. Qualitative research provides an overview of the broad range of non-numerical measurement approaches.

Measurement is the assignment of numbers to objects in such a way that physical relationships and operations among the objects correspond to arithmetic relationships and operations among the numbers.

According to Dr D.D. Sharma, "Measurement is concerned with correspondence between empirical entities and a formal model of abstract elements or numbers ,the relationship among these elements and the operations which can be performed on them. Such a rule of correspondence determines a scale."

According to S.S.Stevens, "Measurement is the assignment of numerals to objects or events according to rules".

According to Campbell,'' Measurement is the assignment of numbers to represent properties."

CONCEPT OF MEASUREMENT

Measurement in quantitative research

Views regarding the role of measurement in quantitative research are somewhat divergent. Measurement is often regarded as being only a means by which observations are expressed numerically in order to investigate causal relations or associations. However, it has been argued that measurement often plays a more important role in quantitative research. For example, Thomas Kuhn (1961) argued that results that appear anomalous in the context of accepted theory potentially lead to the genesis of a search for a new, natural phenomenon. He believed that such anomalies are most striking when encountered during the process of obtaining measurements, as reflected in the following observations regarding the function of measurement in science:

When measurement departs from theory, it is likely to yield mere numbers, and their very neutrality makes them particularly sterile as a source of remedial suggestions. However, numbers register the departure from theory with an authority and finesse that no qualitative technique can duplicate, and that departure is often enough to start a search (Kuhn, 1961, p. 180).

In classical physics, the theory and definitions, which underpin measurement, are generally deterministic in nature. In contrast, probabilistic measurement models known as the Rasch model and Item response theory models are generally employed in the social sciences. Psychometrics is the field of study concerned with the theory and technique for measuring social and psychological attributes and phenomena. This field is central to much quantitative research that is undertaken within the social sciences.

Measurement may be defined as the assignment of numerals to characteristics of objects , persons, states, or events, according to rules. What is measured is not the objects, person, state, or event itself but some characteristic of it. When objects are counted, for example, we do not measure the object itself but only its characteristic of being present. We never measure people, only their age, height, weight, or some other characteristic.

LEVELS OF MEASUREMENT

There are four levels of measurement: nominal, ordinal, interval and ratio. We describe below the characteristics of each of these levels.

(1) Nominal measurement: Nominal measurement is the most elementary method of measurement which classifies persons, objects or events into a number of mutually exclusive categories on the basis of the simple presence or absence, applicability or inapplicability, possession or non-possession of certain property. This is 'have' us. 'have nots' measurement which assigns mutually exclusive labels to identify objects. Thus, the population of a town may be classified according to gender into 'males' and 'females' or according to religion into Hindus, Jains, Parsis, Muslims, Sikhs and Christians and each category of persons given certain labels either in the form of numerals (0,1,2,3) or in the form of letters (A,B,C,D).

These labels only tell us that the categories are qualitatively different from each other. They have no quantitative significance, i.e., they cannot be added, subtracted, multiplied or divided. Once can, if one likes, interchange the labels of various categories for they do not signify any ranking or ordering of categories . The numeral 1 given to a certain category does not imply its superior position to other category which is given numeral 0.

The only arithmetic operation possible in case of a nominal measurement is counting. Thus, mode is the only legitimate measure of central tendency. One can also calculate the percentage of objects failing within each category. But, it will not make sense to calculate the arithmetic mean of

gender in a sample consisting of 45 men and 55 women. All we can say is that there are more females than males in the sample or that 45% of the sample is male.

(2) Ordinal measurement. In ordinal measurement numerals, letters or other symbols are used to rank objects. This is essentially an advances form of categorization. Objects in this measurement are classified not only as to whether they share some characteristic with another object but also whether they have more or less of this characteristics than some other object on some characteristic. A significant amount of marketing research, particularly consumer-oriented research, relies on this type of data. Their most common use is in obtaining preferences measurements. For example, a consumer or a sample of experts may be asked to rank preference for several brands, flavors, or package designs. Attitude measures are also often ordinal in nature.

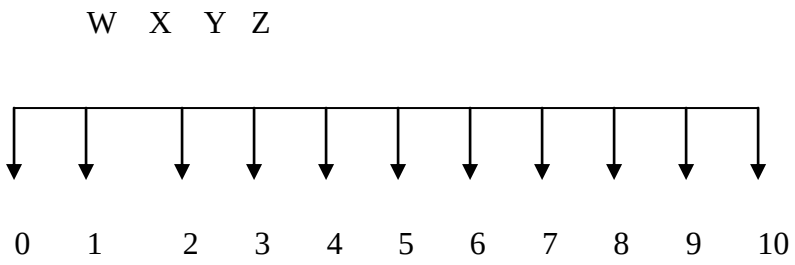
However, ordinal measurements do not provide information on how much more or less of the characteristic various objects possess. For example, if in respect of a certain characteristic two objects have the ranks 5 and 8 and two other objects the ranks 3 and 6, we cannot say that the differences between the two pairs are equal. There is also no way to know that any object has none of the characteristic being measured.

The king of descriptive statistics that can be calculated from these data are mode, median and percentages. For example, in the following table which shows the quality ratings given by 600 housewives to one brand of coffee one can usefully calculate the median and modal quality ratings. Both are '2' in this case. One can also calculate the percentages of the total appearing in each rank. But it is meaningless to calculate a mean because the differences between ordinal scales values are not necessarily, the same.

Quality Rating	No. of respondents giving rating
1	200
2	200
3	100
4	50
5	50

(2) Interval measurement. Interval measurements represent numerals to rank objects such that numerically equal distances on the scale represent equal distances in the property being measured. The distances between numerals are thus, very meaningful because, by comparing these distances we can know how far apart the objects are with respect to the property in question. For example, if we are measuring the achievements of 4 students W, X, Y, and Z on an interval scale and obtain the values 1, 4, 5 and 8 respectively (as Shown below) then by comparing the intervals, we can legitimately say that the difference between W and Y in their achievements is the same as the

difference between X and Z and That the difference between X and Z is four times the difference between W and Y.



One very important drawback of this measurement, however, is that We cannot compare objects on the basis of ratios of their absolute scores. Thus, in our example, we cannot say that the achievements of X is twice as great as that of Z. The reason is that in our measurement the zero point is arbitrary so that any change in this point will change the absolute scores and the ratios between scores. For example, if the arbitrary zero point of the above scale is shifted four scale-points downward, the position of X and Z on the scale will change to 8 and 12 and the previous ration of 1: 2 will now become 2:3

The most common examples of interval scales are the centigrade and Fahrenheit temperature scales which start with different points of origin. The point of origin (for the same natural phenomenon, the freezing point of water) is zero on the Centigrade scale and 32 on Fahrenheit. Because of this difference the ratio between any two readings on the Centigrade scale (e.g., between 10 to 30) is not the same as that on the Fahrenheit scale as shown below. One can only talk of a 20° C rise in temperature but not of 30° C as being three times as hot as 10°C.

Centigrade	0	10	30	100
Fahrenheit	32	53	86	212

The most frequent of interval measurement in social sciences is index numbers which are calculated on the basis of an arbitrary zero point. Another common form of interval measurement is a likert scale which is used in the measurement of attitudes and personality.

Virtually the entire range of statistical analysis can be applied to data measured on interval scales. One can use such descriptive measures as the mean, median, mode, range and standard deviation. Bivariate correlation analysis, t-test, analysis of variance test, and most multivariate techniques applied for purposes of drawing inferences can also be used.

(4) Ratio measurement. This measurement, besides possessing the property of the interval measurement, possesses one additional property , viz., it has a true, natural or absolute zero point, one for which there is universal agreement as to its location A true zero means that the object measuring zero possesses none of the property in question. Height and weight are obvious example. In general, wherever the objects, are being 'counted' for possessing or not possessing a certain characteristic ratio measurement is being done. In all such cases the number 0 has a true meaning – the object possesses none of the characteristic being measured. Other common examples of this type of measurement are sales, costs, number of purchasers, length, time, etc.

With a ratio measurement, the comparison between ratios of the absolute magnitude of the numbers becomes possible. Thus, a person weighing 100kg . is said to be twice as heavy as one weighing 50kg. and a person weighing 150kg is three times as heavy. Further, with a ratio scale we can compare intervals, rank objects according to magnitude, or use the

numbers to identify the objects. All descriptive measures and inferential techniques are applicable to ratio-measured data.

The following table summaries the above discussion about four types of measurement;

Types Of Measurement

Type	Basic empirical operation	Typical usage	Typical Statistics Descriptive Inferential
1. Nominal	Determination Of equality (0, 1,2,…….9)	Classification Male-female purchaser non-purchaser, Team A Team-B	Percentage chi-square, Binomial test
2. Ordinal	Determination Of greater or less (o<1<2……..<9)	Rankings: preference data, market position, attitude measures, many psychological measures	Median Rank-order correletion
3. Interval 4. Ratio	Determination of equality of intervals (2-1=7-6) Determination of equality of ratios (2/4 = 4/8)	Index numbers, attitude measures Sales, units produced, number of customers. Costs, age	Mean, Range, Product-moment Standard Deviation Geometric Coefficient of variation mean

COMPONENTS OF MEASUREMENT

Ideally speaking there should be only one component of a measurement and this component should be a direct reflection of the characteristic being measured. Unfortunately, such a situation seldom exists. Very often, a measurement consists of not one but two components- one representing the influence of the characteristic being measured and the other representing the influence of those characteristics which the researcher is not interested in measuring but which still creep in against his wishes. The characteristics are as follows:

1. Additional stable characteristics of the object or event, for example, the respondent's tendency to give only favorable responses independent of his true feelings.
2. Short-term characteristics of the object, for example, fatigue, health, hunger, and emotional state.
3. Situational characteristics, for example, the presence or absences or some other person or of location under which the measurement is taken
4. Characteristics of the measurement process, for example, Gender, age, and ethnic background. In addition, style of dress of the interviewer or the method of interviewing-telephone, mail, personal interview, etc.
5. Characteristics of the measuring instruments, for example, unclear instructions, ambiguous questions, confusing terms, omitted questions, etc.
6. Characteristics of the response process, for example, mistakes caused by checking a wrong responses

7. Characteristics of the analysis, for example, mistakes caused by wrong coding, tabulating, etc.

Six out of these seven characteristics (2 to 7) give rise to variable errors in a measurement, i.e., errors which occur randomly each time something is measured. First characteristics gives rise to a systematic error, i.e., error that occurs in a consistent manner each time something is measured. It is also caleed bias. Thus, the general situation is:

M=C+VE+SE where M stands for the measurement , C stands for the characteristic being measured, VE stands, for variable errors and SE stands for systematic errors.

ACCURACY MEASUREMENT

Accuracy of measurement depends upon the extent to which it is free from systematic and variable errors. Freedom from variable errors is known as the reliability of a measurement and freedom from systematic errors is known as the validity of a measurement. Reliability and Validity every researcher to know how they are measured.

The Measurement Process

The following are the steps involved in the measurement process

1. Identifying the concept of interest
2. Development of a construct
3. A constitutive definition
4. Operational definition
5. Measurement scales

6. Evaluating reliability and validity of the scale
7. utilization of scale
8. Research findings

We shall discuss the four numerical scales such as Nominal Scale, Ordinal scale, interval scale and ratio scale, functions of measurement and measurement error, benefits of measurement as described below.

Nominal Level of Measurement

a. Nominal Scales
    i. Partitions data into categories that is mutually exclusive and collectively exhaustive. Examples
    ii. Gender:

    (1) Male (2) Female

    iii. Geographic Area:

    (1) Urban (2) Rural (3) Suburban(4) Semi-urban

Ordinal Scales

b. Used strictly to indicate rank order.

Ranking

Please rank the following brands of Colour Television from 1-5 with 1 being the most and 5 the least preferred

Samsung_____ LG_____

Thomson _____ BPL _____

Videocon _____

- Interval Scales
  - Contains all the features of ordinal scales
  - Added dimension that the intervals between the data points on the scale are equal.
- Ratio Scales
  - All powers of those mentioned as well as a meaningful absolute zero or origin.

- Accurate Data Imply Accurate Measurement

M = A + E

Where M = measurement

A = accuracy

E = errors-random or systematic

Nominal Scale

Nominal Scale is defined as measurement of a variable which results in the classification of phenomena into a set of consistent and non-overlapping attributes (yes, no, male, female, etc.). Nonparametric statistics: Tests of significance that require few assumptions about the population. Use of these statistics should occur when samples are small (fewer than 30 subjects), when subjects were not randomly sampled, or data are not interval level. ...

A nominal scale allows for the classification and labeling of elements or objects into mutually exclusive and exhaustive categories based on defined features.

A scale of measurement with two or more categories that have no numerical (less than, greater than) properties. Proportion of map distance to ground distance at a point or line

Nominal Data

Classification data, e.g. male/ female

No ordering, e.g. it makes no sense to state that

Female >Male

Arbitrary labels, e.g., male/female, 0/1, etc

Nominal Scale. We can examine if a nominal scale data is equal to some particular value or to count the number of occurrences of each value. For example, gender is a nominal scale variable. We could examine if the gender of a person is Female or to count the number of males in a sample.

At the first level of measurement, numbers are used to classify data. In fact, words or letters would be equally appropriate. For example when we wanted to classify a cricket team into left footed and right footed players, we could put all the left footed players into a group classified as 1 and all the right footed players into a group as 2.

The only mathematical or statistical operation that can be performed on nominal scales is a frequency run or count. We cannot determine an average, except for the mode – that number which holds the most responses -, nor can we add and subtract numbers.

The purpose of this set of notes is to briefly summarize several aspects of scales of measurement including:

(a) the measurement principle involved for each scale,

(b) examples of the measurement scales,

(c) permissible arithmetic operations for each scale, and

(d) examples of statistics that are appropriate for each scale.

People or objects with the same scale value are the same on some attribute. The values of the scale have no 'numeric' meaning in the way that you usually think about numbers. People or objects with a higher scale value have more of some attribute.

The intervals between adjacent scale values are indeterminate. Scale assignment is by the property of "greater than," "equal to," or "less than." Intervals between adjacent scale values are equal with respect the the attribute being measured.

E.g., the difference between 8 and 9 is the same as the difference between 76 and 77.

There is a rationale zero point for the scale.

Ratios are equivalent, e.g., the ratio of 2 to 1 is the same as the ratio of 8 to 4.

Each "higher" level of measurement includes the measurement principle of the "lower" level of measurement. For example, the numbers 8 and 9 in an interval scale indicate that the object assigned a 9 has more of the attribute being measured than does the object assigned an 8 (ordinal property) and that all persons assigned a 9 have equivalent amounts of the attribute being measured (nominal property). This also implies that you can do lower level statistics on higher-level measurement scales.

Guttmann Scales

For thinking about numeric types, measurement, truth, social structures .Fiske argues that all social life is composed of patterns of interaction that are based on four types of scale: categorical, ordinal, interval, and ratio. The scales provide ways of perceiving, and thereby of organizing social interaction. The four scales are not mere manifestations of a single culture, but are different primary mathematical structures. They are different axiomatically. They are Trans cultural. In 1944, Louis Guttmann pointed out that all forms of measurement belong to one of four types of scale: categorical, ordinal, interval, and ratio.

Why is Level of Measurement Important?

First, knowing the level of measurement helps us to decide how to interpret the data from that variable. When we know that a measure is nominal then we know that the numerical values are just short codes for the longer names. Second,

knowing the level of measurement helps us to decide what statistical analysis is appropriate on the values that were assigned.

In nominal measurement, the numerical values just "name" the attribute uniquely. No ordering of the cases is implied. For example, the numerical numbers in football team l are measures at the nominal level. A player with number 7 is not more of anything than a player with number 5.

In ordinal measurement, the attributes can be rank-ordered. Here, distances between attributes do not have any meaning. For example, on a survey you might code Educational Qualification as 0=less than Higher .Secondary 1=Higher .Secondary. 2=Under Graduate degree; 3=Post Graduate degree; 4=professional degree; 5=doctoral degree . In this measure, higher numbers mean more education. But is distance from 0 to 1 same as 3 to 4? Of course not. The interval between values is not interpretable in an ordinal measure.

In interval measurement the distance between attributes does have meaning. For example, when we measure temperature (in Fahrenheit), the gap between 30 and 40 is same as gap between 70 and 80. The interval between values is interpretable. Because of this, it makes sense to compute an average of an interval variable, where it doesn't make sense to do so for ordinal scales. But we should note that in interval measurement ratios don't make any sense - 80 degrees is not twice as hot as 40 degrees (although the attribute value is twice as large).

Finally, in ratio measurement there is always an absolute zero that is meaningful. This means that we could construct a meaningful fraction (or ratio) with a ratio variable. Weight is a ratio variable. In applied business research, most "count" variables are ratio, for example, the number of clients in past six

months. Because we could have zero clients and because it is meaningful to say that "...we had twice as many clients in the past six months as we did in the previous six months."

The categories of the variable:

Level	are names	have an inherent order from more to less or higher to lower	are numbers with equal intervals between them	are numbers that have a theoretical zero point
Nominal level	X			
Ordinal level	X	X		
Interval level	X	X	X	
Ratio level	X	X	X	X

Source: Prof Candance Clark, social statistics

It's important to recognize that there is a hierarchy implied in the level of measurement idea. At lower levels of measurement, assumptions tend to be less restrictive and data analyses tend to be less sensitive. At each level up the hierarchy, the current level includes all of the qualities of the one below it and adds something new. In general, it is desirable to have a higher level of measurement (e.g., interval or ratio) rather than a lower one (nominal or ordinal).

Functions of Measurement

• Standardization

• Sharpen arguments and distinctions

• Enhance the precision of statements

- Allow application of mathematics

Causality

Ultimately, we seek to establish causal relationships among the phenomena we study. Necessary condition for causality: one that must be present for an event to occur, although its presence does not guarantee occurrence; its absence guarantees nonoccurrence.

Sufficient condition for causality: one that guarantees that the event will occur whenever it is present; but the event may still occur in its absence .

Ideal outcome of a scientific study - a set of conditions that are simultaneously necessary and sufficient for an event that would truly explain the event .The classic scientific technique for establishing causality is the controlled experiment. Social experiments rarely achieve the same kind of controlled conditions that are obtainable in a laboratory.

The cause must precede the effect in time.

The two variables must be empirically correlated. Otherwise stated as: association is a necessary condition for causality. The correlation between the two variables cannot be explained by the existence of a third variable. We could find out the cause and effect relationships. Correlation does not equal causality. Causal attributions always involve an interpretation that is relative to the categories and understandings of mechanisms available when the study is written. Causal inferences are based on assumptions. Human actions and interactions are rarely so simple that a single source of causality can be identified. Social science theories do help to provide possible explanations of tendencies or actions of groups with common characteristics. Yet there is always

a possibility of alternative inferences. Even assuming that some common set of explanations could be identified among selected subjects, are they representative of the whole population? The standard of acceptability for any experimental or observational evidence is: can someone else replicate the process such that the same results are produced.

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

Summary

In the early 1940's, the Harvard psychologist S.S. Stevens coined the terms nominal, ordinal, interval, and ratio to describe a hierarchy of measurement scales used in psychophysics, and classified statistical procedures according to the scales for which they were "permissible." Measurement may be defined as the assignment of numerals to characteristics of objects , persons, states, or events, according to rules. What is measured is not the objects, person, state, or event itself but some characteristic of it. When objects are counted, for example, we do not measure the object itself but only its characteristic of being present. We never measure people, only their age, height, weight, or some other characteristic. There are four levels of measurement: nominal, ordinal, interval and ratio.

KEY TERMS

Measurement

Nominal scale

Ordinal scale

Interval scale

Ratio scale

QUESTIONS FOR REVIEW

1. Define: Measurement
2. What are the steps involved in the Measurement Process.

Define the following: Nominal scale, Ordinal scale, Interval scale and Ratio scale

3. Indicate whether each of the following measures uses a nominal scale, ordinal scale, Interval scale and ratio scale.

a. Prices of commodities
b. Married and unmarried
c. Employed and unemployed
d. Professorial rank Professor, Associate Professor, Assistant professor.

4. What are the characteristics of Measurement?
5. What do you mean by the term Measurement error?

Further Reference

1. William G.Zigmand, 2003, Business Research Methods, Thompson Asia pvt ltd.,

2.D.D.Sharma , Marketing research,1999,SultanChand&Sons

3. Cooper, DonaldR, Schindler, Pamela.s 2006 Businees Research Methods

4.P.C.Tripathi,2005.A Text Book of Research Methodology, SultanChand&Sons.

5.O.R.Krishnaswami,M.Ranganatham,2005.Methodology of Research in Social Sciences,Himalaya Publishing House,Mumbai.

Unit-IV

DATA INSTRUMENTS

LEARNING OBJECTIVES

After reading this chapter , you should be able to understand

- **Features of a research instrument.**
- **Types of research instruments.**
- **Selection of research instrument for the study.**

INTRODUCTION

In today's unpredictable business environment, customer behavior and attitudes are changing, the economy is shifting, stakeholders are demanding, and the competition is getting fierce. Business firms can capitalize these changes by conducting marketing research. And good marketing research leads to great ideas that build business results. The foundation of good research is good data. Traditionally, the term Marketing Research evokes images of huge volumes of data collected by armies of interviewers. And the same images always involve data relating exclusively to responses to marketing initiatives, advertising, pricing, new product launches etc., Perhaps there might be some awareness of data on Customer Satisfaction and Retention.

There are very few hard and fast rules to define the task of data collection. Each research project uses a data collection technique appropriate to the particular research methodology. The two primary goals for both quantitative and qualitative studies are to boost response and maximize accuracy. To adequately address the research questions a variety of data collection methods and instruments will be used. These methods and instruments are not specific to any

one question but provide data that when used in combinations will address the research questions. In general researcher can opt for any instrument or combinations of various instruments for the study chosen.

i) TYPE OF INSTRUMENTS

Data instruments are the instruments employed by the researcher to collect the required data. Data collection instruments are designed to collect standard information from a large number of study participants. After deciding which type of information is needed, the methods of accessing data must be determined. There are different methods of collecting data. The actual design of the instrument, the data collection form that is used to ask and record the information is critical to the success of the research. Based on the objectives of the study researcher can design the instruments or forms to gather the data. The questionnaire is an example of research instrument, in an observation based study, the instrument can also be a mechanical or electronic device of observation like Tape recorder and camera. Survey methods may include mailed questionnaires, telephone interviews, or face-to-face interviews. Information gathered may be qualitative as well as quantitative. In experimental types of research some times an observation device is used, some times a questionnaire is used.

DATA COLLECTION INSTRUMENTS

a. **Questionnaires**
b. **Interview schedule**
c. **Observational / Scanning tools**
d. **Archival Data Collection Tools**
e. **Interviews / Focus Group Protocols**
f. **Experimentation**
g. **Multi-Modal Data Collection**

a. Questionnaires :

Questionnaire is a common instrument for collecting data beyond the physical reach of the researcher, that is, from a large size sample of people. It is an impersonal instrument for collecting information and must, therefore, contain clear questions, worded as simply as possible to avoid any confusion or ambiguity since the researcher probably will not be present to explain what was meant by any one particular question. 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.

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.

b. Interview schedule:

Interview schedule is a set of questions prepared by the researcher to conduct the survey through personal contact. The research worker or enumerator fills the schedule. With respect to primary research, the foremost tool is the personal interview. The face-to-face contact between researcher and respondent is not equal in terms of the potential quality of data that can be obtained. In the face-to-face interview it is possible to record more than the verbal responses of the

interviewee, which are often superficial. When human beings communicate directly with each other much more information is communicated between them. When two people face one another, the dialogue is conducted on several levels. It goes beyond verbal expression. The nature of words used, facial expressions and body language all communicate what the other party means.

c. Observational / Scanning tools:

Tools used to guide visual inspection and documentation of characteristics of people or settings of interest. The scanning process can include recording observations in writing or using checklists, videotaping, or other methods. Scanning processes may be supplemented by interviews, surveys, and other techniques.

d. Archival Data Collection Tools :

Forms or other procedures designed to collect or abstract data from existing records for research or evaluation purposes. Sometimes called 'indicator' data, examples of archival data sources include program files and records from student life programs, academic administration, social services agencies, police, courts, or other community-based agencies.

e. Interviews / Focus Group Protocols:

A set of questions and procedures designed to guide one-on-one or group interviews. Procedures fall on a continuum from unstructured, in which the questions are used as a general guide, to very structured, in which the questions are asked word-for-word in a set order.

f. Experimentation:

The popularity of experimentation in marketing research has much to do with the possibilities of establishing cause and effect. Experiments can be configured in such a way as to allow the variable causing a particular effect to be isolated. Other methods commonly used in marketing research, like surveys, provide

much more ambiguous findings. In fact, experimentation is the most scientific method employed in marketing research

g. Multi-Modal Data Collection

Many clients require more than one method of data collection to optimally execute their research. An example of this is a very complex survey that would be cost prohibitive on the telephone, but would be easy to complete on paper. Reaching the respondent by different roots certainly increases the response rate.

Construction of a good data instrument

In general researcher designs the instrument catering to the nature of the study and other environment. The following points may be considered for constructing a good instrument

- Reliability of the instrument .
- Relevancy
- Brevity
- Nonambiguity
- Specificity
- Test the validity of the data.
- unbiased reply.
- Cost effective.
- Possibility of pre-test .

ii) STEPS INVOLVED IN DESIGNING DATA INSTRUMENT

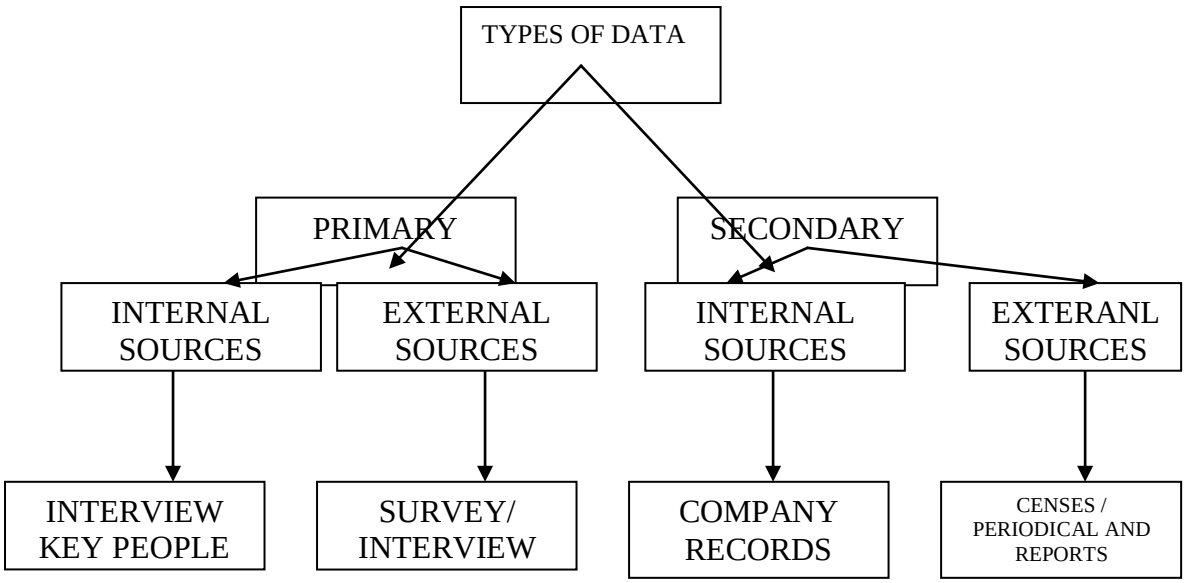
Data collection tools plays vital role in research. They serve two very important purposes. Firstly it helps the researcher to identify varies elements required to address for achieving the objectives of the study. Secondly it roots for how data

should be collected. While preparing the instrument due importance must be given to the environment. Companies desire and background of the respondent are to be considered for improving the effectives of the selected method. Many researchers elect to use instruments that were used previously by other researchers. An advantage of the strategy is that information about reliability and validity of the instruments may be established already. Instruments often have been revised many times and are designed to facilitate easy data collection. Disadvantages to using existing instruments are that the instruments may not have been used with the populations to be studied or may not address a particular issue .The following steps can be used to design a data instrument.

1. Decide what information is needed
2. Know the environment
3. Decide what method to use
4. Format the instrument
5. Pilot test
6. Get feedback on the instrument
7. Revise

IV.B. DATA COLLECTION METHODS

In marketing research collection of data plays a dominant role. Data is required to make decision in any business situation. the researcher is faced with one of the most difficult problems of obtaining suitable, accurate and adequate data. As soon as the research design has been decide, the next step is that of selecting the sources of data. Based on the nature and purpose of the study researcher can go for various modes of data collection .Data sources can be classified into two categories namely primary and secondary sources .Discussion about sources are as follows.



IV.B. a. <u>COLLECTION OF SECONDARY DATA</u>

<u>OR</u>

<u>DESK RESEARCH</u>

Secondary data is data which has been collected by individuals or agencies for purposes other than those of their particular research study. These secondary data source may also be termed as paper source. These secondary sources could include previous research reports, newspaper, magazine and journal content, government and NGO statistics. Sometimes secondary research is required in the preliminary stages of research to determine what is known already and what new data are required, or to inform research design. At other times, it may make be the only research technique used. For example, if a government department has conducted a survey of, say, family medical expenditures, and then medicine

manufacturer might use this data in the organization's evaluations of the total potential market for a new product. Similarly, statistics prepared by a ministry of health will prove useful to a whole host of people and organizations, including those marketing medicine supplies.

i).THE NATURE OF SECONDARY SOURCES OF INFORMATION

No marketing research study should be undertaken without a prior search of secondary sources (also termed desk research). There are several grounds for making such a bold statement.

1. Secondary data may be available which is entirely appropriate and wholly adequate to draw conclusions and answer the question or solve the problem.
2. It is far cheaper to collect secondary data than to obtain primary data. For the same level of research budget a thorough examination of secondary sources can yield a great deal more information than can be had through a primary data collection exercise.
3. The time involved in searching secondary sources is much less than that needed to complete primary data collection.
4. Secondary sources of information can yield more accurate data than that obtained through primary research. This is not always true but where a government or international agency has undertaken a large scale survey, or even a census, this is likely to yield far more accurate results than custom designed and executed surveys when these are based on relatively small sample sizes.
5. It should not be forgotten that secondary data can play a substantial role in the exploratory phase of the research when the task at hand is to define the research problem and to generate hypotheses.

6. Secondary sources help define the population. Secondary data can be extremely useful both in defining the population and in structuring the sample to be taken. For instance, government statistics on a country's agriculture will help decide how to stratify a sample and, once sample estimates have been calculated, these can be used to project those estimates to the population.

ii). CLASSIFICATION OF SECONDARY DATA

The sources of secondary data can be classified into four categories. They are:

1. **Published sources**: data available from various national and international institutions
2. **Unpublished sources:** in few cases data maintained by the organizations are not available to the public/researchers in published forms. Researchers can get this information through personal contacts.
3. **Internal source**: internal data are prepared by summaries of companies internal operations. Sales report, invoice, shipment records and consumer database maintained by the companies are the major sources.
4. **External data:** which means data collected from outside of the company. Collection of external data is more complex because the data have much greater variety and the sources are more.

Sources of secondary data
➢ Trade associations ➢ National and local press Industry magazines ➢ National/ international governments

- Informal contacts
- Trade directories
- Published company accounts
- Business libraries
- Professional institutes and organizations
- Omnibus surveys
- Previously gathered marketing research
- Census data
- Public records
- Web sites

Government publications: Official statistics are statistics collected by governments and their various agencies, bureaus, and departments. These statistics can be useful to researchers because they are an easily obtainable and comprehensive source of information that usually covers long periods of time. However, because official statistics are often "characterized by unreliability, data gaps, over-aggregation, inaccuracies, mutual inconsistencies, and lack of timely reporting", it is important to critically analyze official statistics for accuracy and validity. There are several reasons why these problems exist:The scale of official surveys generally requires large numbers of enumerators (interviewers) and, in order to reach those numbers enumerators contracted are often under-skilled. The size of the survey area and research team usually prohibits adequate supervision of enumerators and the research process; and resource limitations (human and technical) often prevent timely and accurate reporting of results.

Technical Reports: Technical reports are accounts of work done on research projects. They are written to provide research results to colleagues, research

institutions, governments, and other interested researchers. A report may come from completed research or on-going research projects.

Scholarly Journals: Scholarly journals generally contain reports of original research or experimentation written by experts in specific fields. Articles in scholarly journals usually undergo a peer review where other experts in the Same field review the content of the article for accuracy, originality, and relevance.

Literature Review Articles: Literature review articles assemble and review original research dealing with a specific topic. Reviews are usually written by experts in the field and may be the first written overview of a topic area. Review articles discuss and list all the relevant publications from which the information is derived.

Trade Journals: Trade journals contain articles that discuss practical information concerning various fields. These journals provide people in these fields with information pertaining to that field or trade.

Syndicate*:* The growing demand for marketing data has produced a number of companies that collect and sell standardized data designed to serve information needs shared by a number of organization; the most common are information needs associated with performance-monitoring research. Syndicated data sources can be classified as

1. Consumer data-collected from consumers regarding purchases.
2. Retail data-focus on products or services sold through retailers .
3. Wholesale data- report on movement of different brands .
4. Industrial data-data on companies to construct sales prospect
5. Advertising evaluation data- to measure the effectiveness of the advertisement
6. Media and audience data- to match market and media

Reference Books: Reference books provide secondary source material. In many cases, specific facts or a summary of a topic is all that is included. Handbooks, manuals, encyclopedias, and dictionaries are considered reference books .

International publications: The United Nations Organizations(UNO), International Labor Organization(ILO),International Monetary Fund(IMF),World Bank, etc., are publishing numerous items of data relating to the socio-economic conditions of different countries.

iii). **QUALITY OF INFORMATION SOURCES**

One of the advantages of secondary data review and analysis is that individuals with limited research training or technical expertise can be trained to conduct this type of analysis. Key to the process, however, is the ability to judge the quality of the data or information that has been gathered. The following tips will help you to assess the quality of the data.

Determine the Original Purpose of the Data Collection: Consider the purpose of the data or publication. Is it a government document or statistic, data collected for corporate and/or marketing purposes, or the output of a source whose business is to publish secondary data (e.g., research institutions).

Ascertain the Credentials of the Source(s) or Author(s) of the Information : What are the author's or source's credentials -- educational background, past works/writings, or experience -- in this area?

Are the methods sound?: Does the article have a section that discusses the methods used to conduct the study? If it does not, you can assume that it is a popular audience publication and should look for additional supporting information or data.

Date of Publication : When was the source published? Is the source current or out-of-date? Topic areas of continuing or rapid development, such as the sciences, demand more current information.

Who is the Intended Audience? Is the publication aimed at a specialized or a general audience? Is the source too elementary -- aimed at the general public?

What is the Coverage of the Report or Document? Does the work update other sources, substantiate other materials/reports that you have read, or add new information to the topic area?

After judging the competency of the data based on the above tests, researcher can chose the right one.

SECONDARY DATA	
ADVANTAGES	**DISADVANTAGES**
1. Low cost 2. speed 3. Often the only resource, for example historical documents 4. Only way to examine large-scale trends 5. Saves time 6. Wide range of availability 7. Serves as a source of comparative data	1. Lack of consistency of perspective 2. Biases and inaccuracies can not be checked 3. Published statistics often raise more questions than they answer (for example, what does church attendance tells us about religious beliefs?) 4. The concern over whether *any* data can be totally separated from the context of its collection

Marketing Research Agencies in India

- ✓ Operations Research Group (ORG)
- ✓ National Council of Applied Economic Research (NCAER)

- ✓ Indian Market Research Bureau (IMRB)
- ✓ Marketing and Research Group (MARG)
- ✓ Marketing Research and Advisory Services (MRAS)
- ✓ Marketing and Business Associate (MBA)

IV.B.b. COLLECTION OF PRIMARY DATA

Primary data is data collected for the first time. It is original and collected for a specific purpose, or to solve a specific problem. It is expensive, and time consuming, but is more focused than secondary research. There are many ways to conduct primary (data collection) research. The following are the prime methods:

1. **Observation**
2. **Questionnaire**
3. **Interview**
4. **Focus groups**
5. **Projective techniques**
6. **Product tests**
7. **Diaries**
8. **Omnibus Studies**
9. **Clinics**
10. **Surveys**
11. **Case study**

I-OBSERVATION METHODS

Observation involves the process of physically or mechanically recording some specific aspect of a consumer's activity or behavior .Observational techniques are methods by which an individual or individuals gather firsthand data on programs, processes, or behaviors being studied. They provide evaluators with an opportunity to collect data on a wide range of behaviors, to capture a great variety of interactions, and to openly explore the evaluation topic. By directly observing operations and activities, the evaluator can develop a holistic perspective, i.e., an understanding of the context within which the project operates. This may be especially important where it is not the event that is of interest, but rather how that event may fit into, or be impacted by, a sequence of events. Observational approaches also allow the evaluator to learn about things the participants or staff may be unaware of or that they are unwilling or unable to discuss in an interview or focus group.

When to use observations. Observations can be useful during both the formative and summative phases of evaluation. For example, during the formative phase, observations can be useful in determining whether or not the project is being delivered and operated as planned. In the hypothetical project, observations could be used to describe the faculty development sessions, examining the extent to which participants understand the concepts, ask the right questions, and are engaged in appropriate interactions. Some research designs call for this type of data. Some contented that observation is objective than communication .How ever , observation whether in a field setting or a laboratory setting ,is not very versatile.

Recording Observational Data

Observations are carried out using a carefully developed set of steps and instruments. The observer is more than just an onlooker, but rather comes to the

scene with a set of target concepts, definitions, and criteria for describing events. While in some studies observers may simply record and describe, in the majority of evaluations, their descriptions are, or eventually will be, judged against a continuum of expectations.

Observations usually are guided by a structured protocol. The protocol can take a variety of forms, ranging from the request for narrative describing events seen to a checklist or a rating scale of specific behaviors/activities that address the evaluation question of interest. The use of a protocol helps assure that all observers are gathering the pertinent information and, with appropriate training, applying the same criteria in the evaluation .The protocol goes beyond a recording of events, i.e., use of identified materials, and provides an overall context for the data. The protocol should prompt the observer to,

- ✓ **Describe the setting** of program delivery, i.e., where the observation took place and what the physical setting was like;
- ✓ **Identify the people** who participated in those activities, i.e., characteristics of those who were present;
- ✓ **Describe the content of the intervention,** i.e., actual activities and messages that were delivered;
- ✓ **Document the interactions** between implementation staff and project participants;
- ✓ **Describe and assess** the quality of the delivery of the intervention; and
- ✓ **Be alert to** unanticipated events that might require refocusing one or more evaluation questions.

METHODS OF OBSERVATION

There are several methods of observation of which any one or a combination of some of them can be used by the observer.

1. Structured-unstructured observation:

Structured observation is used when the research problem has been formulated precisely and the observers have been told specifically what is to be observed. They may be given a simple form to record their observations. Unstructured observation implies that observers are free to observe what ever they think is relevant and important .While structured observations are free from subjective bias, unstructured observations are subject to this limitation .The extent of the bias may vary to the extent an observation is unstructured.

2. Disguised-undisguised observation:

In the case of disguised observation, the subjects do not know that they are being observed. In some cases, disguised observation may be made by the observer by posing as one of the shoppers who are being observed. This type of observation is often preferred because it is feared that people may behave differently when they know they are being observed. It may be difficult to completely disguise an observation, though this apart, it poses an ethical question of its desirability when those who are being observed are kept in the dark.

3. Observation under natural setting-laboratory setting:

Observations can also be classified on the bases of their setting, i.e. natural or laboratory. Observations in field studies are in their natural setting and are, therefore undertaken in extremely realistic conditions. Sometimes, an experimental manipulation may be introduced in a field study. Observation in a laboratory setting, on the other hand, enables the observer to control extraneous variables which influence the behavior of people. Observational studies in laboratory settings have certain advantages over field studies. They enable the collection of data promptly and economically and in addition, permit the use of more objective measurements.

4. Direct-indirect observation:

In the case of direct observation, the event or the behavior of a person is to be observed as it occurs .In contrast; indirect observation implies that some record of past behavior is observed. In other words ,the behavior itself is not observed ,rather its effects are observed. An observer engaged in indirect observation generally looks for physical traces of behavior or occurrence of an event. Suppose he is interested in knowing about the liquor consumption of a household, he would like for empty liquor bottles in the garbage. Similarly, the observer may seek the permission of the housewife to see the pantry. He may carry a pantry audit to ascertain the consumption of certain types of products. It may be noted that the success of an indirect observation largely depends on how best the observer is able to identify physical traces of the problem under study. Direct observation is far more common than indirect observation.

5. Human-mechanical observation:

Another way of classifying observations is whether they are made manually or by machines. Most of the studies in marketing research based on human observation wherein trained observers are required to observe and faithfully record their observations. In some cases, mechanical devices such as eye cameras and audiometers are use for observation. One of the major advantages of electrical/mechanical devices is that their recordings are free from subjective bias. As against this advantage, such observations may be less valid than human observations. This is because the observer's power of integration can lead to a more valid evaluation of the observation.

OBSERVATIONS

ADVANTAGES	DISADVANTAGES

1. Provide direct information about behavior of individuals and groups 2. Permit evaluator to enter into and understand situation/context 3. Provide good opportunities for identifying unanticipated outcomes 4. Exist in natural, unstructured, and flexible setting	1. Expensive and time consuming 2. Need well-qualified, highly trained observers; may need to be content experts 3. May affect behavior of participants 4. Selective perception of observer may distort data 5. Investigator has little control over situation 6. Behavior or set of behaviors observed may be unusual

II-QUESTIONNAIRE METHOD OF DATA COLLECTION

This method of data collection is quite popular, particularly in case of big enquiries. It is being adopted by private individuals, research workers, private and public organizations and even by governments. In this method a questionnaire is sent (usually by post) to the persons concerned with a request to answer the questions and return the questionnaire. A **questionnaire** consists of a number of questions written or printed in a definite order on a form or set of forms. The questionnaire is send to respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself. The respondents have to answer the questions on their own. The method of collecting data by mailing the questionnaire to

respondents is most extensively employed in various economic and business surveys.

The **merits** claimed on behalf of this method are as follows:

1. There is low cost even when the universe is large and is widely spread geographically.
2. It is free from the bias of the interviewer; answers are in respondent's own words.
3. Respondents have adequate time to give well thought out answers.
4. Respondents, who are not easily approachable, can also be reached conveniently.
5. Large samples can be made use of and thus results can be made more dependable and reliable.

The major **demerits** of this system are listed here:

Low rate of return of the duly filled in questionnaire; bias due to no-response is often indeterminate.

1. It can be used only when respondents are educated and cooperating.
2. The control over questionnaire may be lost once it is sent.
3. There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaire have been dispatched.
4. There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.
5. It is difficult to know whether willing respondents are truly representative.
6. This method is likely to be the slowest of all.

Before using this method, it is always advisable to conduct 'pilot study' (Pilot Survey) for testing the questionnaires.

Main aspects of questionnaire:

Quite often questionnaire is considered as the heart of a survey operation. Hence it should be very carefully constructed. If it is not properly set up, then the survey is bound to fail. This fact requires us to study the main aspects of a questionnaire viz., the general form, question sequence and question formulation and wording. Researcher should note the following with regard to these three main aspects of a questionnaire:

1. **General form**:
    - **structured questionnaire.**
    - **unstructured questionnaire.**

So far as the general form of a questionnaire is co concerned it can either be structured or unstructured questionnaire. Structured questionnaires are those questionnaires in which there are definite, concrete and pre-determined questions. The questions are presented with exactly the same wording and in the same order to all respondents. Structured questions may also have fixed alternative questions in which responses of the informants are limited to the stated alternatives. Thus a highly structured questionnaire is one in which all questions and answers are specified and comments in the respondent's own words are held to the minimum. When these characteristics are not present in a questionnaire, it can be termed as unstructured or non-structured questionnaire.

2. Question sequence:

In order to make questionnaire effective and to ensure quality to the replies received, a researcher should pay attention to the question-sequence in preparing the questionnaire. A proper sequence of questions reduces considerably the chances of individual questions being misunderstood. The question-sequence must be clear and smoothly-moving, meaning thereby the relation of one question to another should be readily apparent to the respondent, with questions that are easiest to answer being put in the beginning. The first few questions are particularly important because they are likely to influence the attitude of the

respondent and in seeking his human interest. The following type of questions should be generally avoided as opening questions in the questionnaire:

- Questions that put too great a strain on the memory or intellect of the respondent.
- Questions of a personal character.
- Questions related to personal wealth etc,

Following the opening questions, we should have questions that are really vital to research problem and a connecting thread should run through successive questions. Ideally, the question-sequence should conform to the respondent's way of thinking.

2. Question formulation and wording:

With regard to this aspect of questionnaire, the researcher should note that each question must be very clear for any sort of misunderstanding can do irreparable harm to survey. Question should also be impartial in order not to give a biased picture of the true state of affairs. Questions should be constructed with a view to their forming a logical part of a well thought out tabulation plan. In general all questions should meet the following standards :

(a) Should be easily understood.

(b) Should be simple i.e., should convey only one thought at a time.

(c) Should be concrete and should conform as much as possible to the respondent's ways of thinking.

4. Kind of questions:

- **Open end question (format)**
- **Closed-end question**

In general, there are two types of questions one will ask, **open format or closed format.** Open format questions are those that ask for unprompted opinions. In other words, there are no predetermined set of responses, and the participant is free to answer however he chooses. Open format questions are good for

soliciting. An obvious advantage is that the variety of responses should be wider and more truly reflect the opinions of the respondents. This increases the likelihood of you receiving unexpected and insightful suggestions, for it is impossible to predict the full range of opinion. It is common for a questionnaire to end with and open format question asking the respondent for her unabashed ideas for changes or improvements.

Open format questions have several disadvantages. First, their very nature requires them to be read individually. There is no way to automatically tabulate or perform statistical analysis on them. This is obviously more costly in both time and money, and may not be practical for lower budget or time sensitive evaluations. They are also open to the influence of the reader, for no two people will interpret an answer in precisely the same way. This conflict can be eliminated by using a single reader, but a large number of responses can make this impossible

Finally, open format questions require more thought and time on the part of the respondent. Whenever more is asked of the respondent, the chance of tiring or boring the respondent increases.

Closed format questions usually take the form of a multiple-choice question. They are easy for the respondent, There is no clear consensus on the number of options that should be given in an closed format question. Obviously, there needs to be sufficient choices to fully cover the range of answers but not so many that the distinction between them becomes vague.

Qualities of a Good Question:

There are good and bad questions. The qualities of a good question are as follows:

1. Questions should evoke the truth.

2. Asks for an answer on only one dimension.
3. Can accommodate all possible answers.
4. Don't imply a desired answer.
5. Don't use emotionally loaded or vaguely defined words.
6. Don't ask the respondent to order or rank a series of more than few items.

III-INTERVIEW METHODS OF DATA COLLECTION

This is the technique most associated with marketing research. An interview is a specialized type of communication, usually verbal, between two or more people and is carried out for a specific purpose. It is different from an ordinary conversation in that its form and purpose is structured. Interviews can be telephone, face-to-face, or over the Internet. The use of interviews as a data collection method begins with the assumption that the participants' perspectives are meaningful, knowable, and able to be made explicit, and that their perspectives affect the success of the project. An interview, rather than a paper and pencil survey, is selected when interpersonal contact is important and when opportunities for follow-up of interesting comments are desired.

Generally two types of interviews are used in marketing research: structured interviews, in which a carefully worded questionnaire is administered; and in-depth interviews, in which the interviewer does not follow a rigid form. In the former, the emphasis is on obtaining answers to carefully phrased questions. Interviewers are trained to deviate only minimally from the question wording to ensure uniformity of interview administration. In the latter, however, the interviewers seek to encourage free and open responses, and there may be a

tradeoff between comprehensive coverage of topics and in-depth exploration of a more limited set of questions.

PREPARING FOR INTERVIEWS

If you are going to carry out an interview, you need to think about the steps involved. Good preparation must be done before the interview to make sure that you get what you need from it, and some thought given afterwards to the information gathered. The following steps can be followed for conducting the interview

1. Decide on the purpose of the interview:

Clear understanding of the purpose of the interview will definitely enhances the efficiency of the method. Before throwing questions the researchers themselves must sure of their motto. For example, Are you trying to find out someone's opinion about something? Are you trying to help someone with a problem? , are you interviewing someone for a job?

2. Decide what kind of information you need to get from the interview to achieve the purpose:

Every interview must be preplanned to ensure the desired outcome and to use time efficiently. Begin with an opening statement of purpose, goals, timing, confidentiality, and format. The researcher may have a list of specific goals. When the interview is conducted to study the level of customer satisfaction, the parameters must be specifically selected.

3. Decide what questions you are going to ask:

Researcher should draw up a list of questions so that the answers help to achieve the goals. Before selecting interviewing as a data collection method, the researcher must determine whether the research question can be answered appropriately by interviewing people who have experienced the phenomenon of interest. A hypothetical study will be used to illustrate one process the researcher could use to facilitate interviewing.

4. Carry out the interview:

Starting of an interview it self an art, the approaches used by the researcher heavily influences the respondent in the way of answering. Spend a little time setting the scene and getting acquainted with the interviewee (ice breaker) will certainly open the mind of the respondent. Giving assurance to the respondent about confidentiality and criticism positively increases the reliability.

5. Study the answers to the questions: The closing of the interview is very important. Researcher must ensure that the answers heard are the answers that were given. Restate the answers and get acceptance from the interviewee. Allow clarification of any misstatements or misunderstandings. Give reassurance of the purpose of the interview being data collection not oversight. Share any suggestions for improvement if considered advisable Before completing the interview the researcher make sure that , whether the goals have been achieved or not. If not, uncovered areas are to be identified and enough time must be asked from the respondent to continue the interview.

6. Make a decision about the purpose of the interview:

As soon as the interview is over if possible , the out come of the interview may be reported to the respondent. In marketing research for various reasons this may not be advisable.

The role of the interviewer:

The interviewer is really the "jack-of-all-trades" in marketing. The interviewer's role is complex and multifaceted. The researcher must posse’s capability of extracting the answer from the respondent .In most cases they are the backbone

of the functioning of the interview. To perform the task the following things are essential:

1. Locate and enlist cooperation of respondents
2. Motivate respondents to do good job
3. Clarify any confusion/concerns
4. Observe quality of responses
5. Conduct a good interview

TYPES OF INTERVIEW

a. Telephone interviews

Telephone ownership is very common in developed countries. It is ideal for collecting data from a geographically dispersed sample. The interviews tend to be very structured and tend to lack depth. Telephone interviews are cheaper to conduct than face-to-face interviews (on a per person basis).

Advantages of telephone interviews

1. Phone interviews can have a shorter data collection period than face-to-face interviews.
2. Phone interviews may have a better response rate .
3. Phone interview reduces travel costs while permitting interaction with remote participants.
4. Phone interviews can supplement site visits, face-to-face interviews, and other methods. For example, after a site visit or a survey you could gather additional data by conducting a phone interview
5. Cheaper than face-to-face interviews

Disadvantages of telephone interviews

1. Phone interviews can be quite tiring, so they are often shorter than face-to-face interviews.
2. Phone interviews can be difficult if the interviewer or interviewee has a strong accent.
3. Phone interviews are not as good as face-to-face interviews when you are dealing with complex issues.
4. If you have multiple interviewers, you have to worry about consistent approaches to the interviews.

5. Phone interviews are often conducted at times that are convenient to the participant, but not for the interviewer (evenings, early mornings, weekends)
6. Phone interviews at a person's office or home can involve many potential distractions like colleagues stopping by, calls on other lines, background noise, and the lure of using the computer to work during the phone interview.
7. Respondents can simply hang up
8. Interviews tend to be a lot shorter
9. Visual aids cannot be used
10. Researchers cannot see the behavior or body language of the rspondet.

b. Face-to-face interviews

When marketers want to discuss questions in great depth, they generally use personal interview, which are face-to-face interviews conducted with one person or a group of people at one time. personal interviews can be large and more detailed than other types of study. The research will probe and develop points of interest.

Advantages of face-to-face interviews

1. They allow more 'depth'
2. Physical prompts such as products and pictures can be used
3. Body language can emphasize responses
4. Respondents can be 'observed' at the same time

Disadvantages of face-to-face interviews

1. Interviews can be expensive
2. It can take a long period of time to arrange and conduct.
3. Some respondents will give biased responses when face-to-face with a researcher.

c. In-depth interviews

An in-depth interview is a dialogue between a skilled interviewer and an interviewee. Its goal is to detailed material that can be used in analysis. Such interviews are best conducted face to face, although in some situations telephone

interviewing can be successful. Extensive probing and open-ended questions characterize In-depth interviews. Typically, the project evaluator prepares an interview guide that includes a list of questions or issues that are to be explored and suggested probes for following up on key topics.

ADVANTAGES AND DISADVANTAGES OF IN-DEPTH INTERVIEW	
Advantages	**Disadvantages**
✓ Usually yield richest data, details, new insights ✓ Permit face-to-face contact with respondents ✓ Provide opportunity to explore topics in depth ✓ Afford ability to experience the affective as well as cognitive aspects of responses ✓ Allow interviewer to explain or help clarify questions, increasing the likelihood of useful responses ✓ Allow interviewer to be flexible in administering interview to particular individuals or circumstances	✓ Expensive and time-consuming ✓ Need well-qualified, highly trained interviewers ✓ Interviewee may distort information through recall error, selective perceptions, desire to please interviewer ✓ Flexibility can result in inconsistencies across interviews ✓ Volume of information too large; may be difficult to transcribe and reduce data

d. Computer Direct Interviews

These are interviews in which the Interviewees enter their own answers directly into a computer. They can be used at malls, trade shows, offices, and so on. The Survey System's optional Interviewing Module and Interview Stations can easily create computer-direct interviews. Some researchers set up a Web page survey for this purpose.

Advantages

1. The virtual elimination of data entry and editing costs.
2. You will get more accurate answers to sensitive questions.
3. The elimination of interviewer bias. Different interviewers can ask questions in different ways, leading to different results. The computer asks the questions the same way every time.
4. Ensuring skip patterns are accurately followed. The Survey System can ensure people are not asked questions they should skip based on their earlier answers. These automatic skips are more accurate than relying on an Interviewer reading a paper questionnaire.
5. Response rates are usually higher. Computer-aided interviewing is still novel enough that some people will answer a computer interview when they would not have completed another kind of interview.

Disadvantages

1. The Interviewees must have access to a computer or one must be provided for them.
2. As with mail surveys, computer directs interviews may have serious response rate problems in populations of lower educational and literacy levels. This method may grow in importance as computer use increases.

IV-FOCUS GROUPS

Focus groups are made up from a number of selected respondents based together in the same room. Highly experienced researchers work with the focus group to gather in depth qualitative feedback. Groups tend to be made up from 10 to 18 participants. Discussion, opinion, and beliefs are encouraged, and the research will probe into specific areas that are of interest to the company commissioning the research.

Typically focus groups are used as a means of testing concepts, new products, and messages. A focus group is qualitative research, which means that you do not obtain results with percentages, statistical testing, or tables. Instead, this methodology is less structured than surveys or other quantitative research and tends to be more exploratory as well. Rather than providing quantifiable

responses to a specific question obtained from a large sampling of the population, focus group participants provide a flow of input and interaction related to the topic or group of topics that the group is centered around. While they appear to be less formal than a survey, focus groups do provide an important source of information for making business decisions. It is important, however, to ensure that persons using the results of such a qualitative study understand how to correctly interpret the resulting information.

Quantitative research provides results that can be generalized to a specific population, because it is based on a statistical sampling of the target population. The results of qualitative research, such as focus groups, however, are not quantifiable. They reflect only a very small segment of the target market in question. Given the number of focus group participants, results are not necessarily representative of the general population from which participants are recruited and should not be considered as such.

When to Use and When to Avoid Focus Groups

Focus groups are best used when the concept or idea you wish to evaluate is new and when the best evaluation comes from letting the target customer (who can be either a consumer or a customer in the business-to-business sense) view the concept directly. A good example of this is a new advertising campaign. Typically, an advertising agency will want to test a new advertisement with the consumers it hopes to reach with the campaign. The agency needs to know whether the message is clear, whether consumers view the advertisement positively or negatively, whether the advertisement would prompt them to purchase the product, and so forth. The agency can explore these questions by showing the advertisement to consumers and discussing their reactions and their likes and dislikes. Their comments will allow the agency to fine-tune the advertisement or, if it is not well liked, to go back to the drawing board.

Another common use of focus group research is product concept testing. Participants in a group can discuss detailed diagrams and product descriptions, or they may test a product prototype in a hands-on fashion. This allows researchers to identify customer needs and, attitudes regarding a concept before investing in further product development.

A focus group study also is often used to design the questionnaire for a quantitative survey. The focus group covers general issues on a topic, and respondents' comments often help researchers identify pertinent issues that might otherwise be left out of a survey. Hypotheses generated by focus groups frequently lead to further testing using quantitative methods. Alternatively, focus groups can be used to further interpret quantitative research. For example, if a telephone survey produces a significant percentage of unexpected comments on a certain topic, a focus group study could investigate the issue with greater depth.

Another particularly useful application of focus group research is as a brainstorming mechanism. If you have a problem to solve, this type of methodology often provides fresh insights regarding the issue at hand. It can also provide an excellent forum for generating creative ideas or new-product ideas for new markets, as well as generating new ideas for your established markets.

Focus groups can be used for the following purposes

- ✓ To test new concepts
- ✓ To position a product/ service
- ✓ To assess product usability
- ✓ To evaluate advertising/ copy
- ✓ To evaluate promotions
- ✓ To develop questionnaires
- ✓ To generate ideas or support brainstorming

V. PROJECTIVE TECHNIQUES

The technique of inferring a subjects attitudes or values based on his or her description of vague objects requiring interpretation. Projective techniques are borrowed from the field of psychology. They will generate highly subjective qualitative data. There are many examples of such approaches including: Inkblot tests - look for images in a series of inkblots Cartoons - complete the 'bubbles' on a cartoon series Sentence or story completion Word association - depends on very quick (subconscious) responses to words .Psychodrama - Imagine that you are a product and describe what it is like to be operated, warn, or used.

Projective techniques consist of the following:

a. **Word association tests:** In word association test the subject is presented with a list of words, one at a time, and asked to respond with first that comes to the respondents mind. Both verbal and nonverbal responses are recorded. Word association frequently used to test potential brand names and company images among public .interpreting word association tests is difficult, and the marketing researcher should make sure to avoid subjective interpretations .when there is considerable agreement in the free-association process, the researcher assumes that the test has revealed the consumers inner feelings about the subject

b. **Sentence completion test:** A projective technique in which respondents are required to complete a number of partial sentences with the first word or phrase that comes to mind. The sentence completion method is also based on the principles of free association. For example:

   People who smoke are --------

   A man who smoke filter cigarette are------

   Imported cigarette is most liked by-------

The women in politics are------

When comparing to word association tests the answers given to sentence completion tests tends to more extensive .the intent of sentence completion questions are more apparent.

VI- PRODUCT TESTS

Product tests are often completed as part of the 'test' marketing process. Products are displayed in a mall of shopping center. Potential customers are asked to visit the store and their purchase behavior is observed. Observers will contemplate how the product is handled, how the packing is read, how much time the consumer spends with the product, and so on.

VII-DIARIES

Diaries are used by a number of specially recruited consumers. They are asked to complete a diary that lists and records their purchasing behavior of a period of time (weeks, months, or years). It demands a substantial commitment on the part of the respondent. However, by collecting a series of diaries with a number of entries, the researcher has a reasonable picture of purchasing behavior.

VIII-OMNIBUS STUDIES

An omnibus study is where an organization purchases a single or a few questions on a 'hybrid' interview (either face-to-face or by telephone). The organization will be one of many that simply want to a straightforward answer to a simple question. An omnibus survey could include questions from companies in sectors as diverse as heath care and tobacco. The researches are far cheaper, and commit less time and effort than conducting own research.

IX-CLINICS

The Clinic methodology constitutes a hybrid approach of both qualitative and quantitative, primarily for testing product or service concepts. A significant number of respondents are screened and recruited to a central facility (this may be a focus group facility or it may be a hotel conference room). Concepts are presented to the large audience of respondents; videos may be shown; product prototypes may be passed around; client technical people may make parts of the presentation. Each respondent is provided with a wireless remote answer device and, during the session, they are asked specific questions and reply in real-time through their wireless devices. Results can, as appropriate, be fed back instantaneously and are used at the end of the presentation phase to develop a list of those respondents who have answered in a certain way, and these are asked to stay for a focus group. Thus the focus group might have people who liked a new product concept but were only prepared to pay a little for it, or people who did not like some or all aspects of the product or offer. The result is a set of quantitative data from the broader audience, and qualitative insight from the highly focused focus group. The process can be undertaken quickly across multiple countries and the results made available within a few weeks.

X-SURVEYS

A survey is a research technique in which data are systematically collected directly from the people being studied by the questionnaire. Surveys are a form of questioning that is more rigid than interviews and that involve larger groups of people. Surveys will provide a limited amount of information from a large group of people and are useful when you want to learn what a larger population thinks.

The Steps in a Survey

1. Establish the goals of the survey - What you want to learn
2. Determine your sample - Whom you will interview
3. Choose interviewing methodology - How you will interview
4. Create your questionnaire - What you will ask
5. Pre-test the questionnaire, if practical - Test the questions
6. Conduct interviews and enter data - Ask the questions
7. Analyze the data - Produce the reports

The first step in any survey is deciding what you want to learn. The goals of the project determine whom you will survey and what you will ask them. If your goals are unclear, the results will probably be unclear. Some typical goals include learning more about:

- ✓ The potential market for a new product or service
- ✓ Ratings of current products or services
- ✓ Employee attitudes
- ✓ Customer/patient satisfaction levels
- ✓ Reader/viewer/listener opinions
- ✓ Association member opinions
- ✓ Opinions about political candidates or issues
- ✓ Corporate images

Selecting the Sample

There are two main components in determining whom you will interview. The first is deciding what kind of people to interview. Researchers often call this group the **target population**. If you conduct an employee attitude survey or an association membership survey, the population is obvious. If you are trying to determine the likely success of a product, the target population may be less obvious. Correctly determining the target population is critical. If you do not interview the right kinds of people, you will not successfully meet your goals. The next thing to decide is how many people you need to interview. Statisticians know that a small, representative **sample** will reflect the group from which it is

drawn. The larger the sample, the more precisely it reflects the target group. However, the rate of improvement in the precision decreases as your sample size increases. For example, to increase a sample from 250 to 1,000 only doubles the precision. You must make a decision about your sample size based on factors such as: time available, budget and necessary degree of precision.

Avoiding a Biased Sample

A biased sample will produce biased results. Totally excluding all bias is almost impossible; however, if you recognize bias exists you can intuitively discount some of the answers. The consequences of a source of bias depend on the nature of the survey. For example, a survey for a product aimed at retirees will not be as biased by daytime interviews as will a general public opinion survey. A survey about Internet products can safely ignore people who do not use the Internet.

Quotas

A **Quota** is a sample size for a sub-group. It is sometimes useful to establish quotas to ensure that your sample accurately reflects relevant sub-groups in your target population. For example, men and women have somewhat different opinions in many areas. If you want your survey to accurately reflect the general population's opinions, you will want to ensure that the percentage of men and women in your sample reflect their percentages of the general population.

If you are interviewing users of a particular type of product, you probably want to ensure that users of the different current brands are represented in proportions that approximate the current market share. Alternatively, you may want to ensure that you have enough users of each brand to be able to analyze the users of each brand as a separate group.

TELEPHONE SURVEYS

Surveying by telephone is the most popular interviewing method in advanced countries. Where the time is short and distance is too far the research questions will be asked through telephone .

Advantages

1. People can usually be contacted faster over the telephone than with other methods. If the Interviewers are using CATI (computer-assisted telephone interviewing), the results can be available minutes after completing the last interview.
2. You can dial random telephone numbers when you do not have the actual telephone numbers of potential respondents.

Disadvantages

1. Many people are reluctant to answer phone interviews and use their answering machines to screen calls.
2. You cannot show your sample products by phone.

MAIL SURVEYS

Mail and telephone surveys are a method of collecting information by sending surveys via email or postal mail. Participants return completed forms to the researcher. Surveys may ask respondents to rate items on a scale .Some surveys also allow respondents to write their feelings or attitudes about a particular event or to elaborate in more detail on an item, or to express suggestions, etc.

Advantages

1. Mail surveys are among the least expensive.
2. This is the only kind of survey you can do if you have the names and addresses of the target population, but not their telephone numbers.
3. The questionnaire can include pictures - something that is not possible over the phone.
4. Mail surveys allow the respondent to answer at their leisure, rather than at the often inconvenient moment they are contacted for a phone or personal interview. For this reason, they are not considered as intrusive as other kinds of interviews.

Disadvantages

1. Time! Mail surveys take longer than other kinds. You will need to wait several weeks after mailing out questionnaires before you can be sure that you have gotten most of the responses.

2. In populations of lower educational and literacy levels, response rates to mail surveys are often too small to be useful.

EMAIL SURVEYS

Email surveys are both very economical and very fast. More people have email than have full Internet access. This makes email a better choice than a Web page survey for some populations. On the other hand, email surveys are limited to simple questionnaires, whereas Web page surveys can include complex logic.

Advantages

1. Speed. An email questionnaire can gather several thousand responses within a day or two.
2. There is practically no cost involved once the set up has been completed.
3. You can attach pictures and sound files.
4. The novelty element of an email survey often stimulates higher response levels than ordinary "snail" mail surveys.

Disadvantages

1. You must possess (or purchase) a list of email addresses.
2. Some people will respond several times or pass questionnaires along to friends to answer. Many programs have no check to eliminate people responding multiple times to bias the results. The Survey System's Email Module will only accept one reply from each address sent the questionnaire. It eliminates duplicate and pass along questionnaires and checks to ensure that respondents have not ignored instructions (e.g., giving 2 answers to a question requesting only one).
3. Many people dislike unsolicited email even more than unsolicited regular mail. You may want to send email questionnaires only to people who expect to get email from you.
4. You cannot use email surveys to generalize findings to the whole populations. People who have email are different from those who do not, even when matched on demographic characteristics, such as age and gender.
5. Email surveys cannot automatically skip questions or randomize question or answer choice order or use other automatic techniques that can enhance surveys the way Web page surveys can.

INTERNET/INTRANET (WEB PAGE) SURVEYS

Web surveys are rapidly gaining popularity. They have major speed, cost, and flexibility advantages, but also significant sampling limitations. These limitations make software selection especially important and restrict the groups you can study using this technique.

Advantages

1. Web page surveys are extremely fast. A questionnaire posted on a popular Web site can gather several thousand responses within a few hours. Many people who will respond to an email invitation to take a Web survey will do so the first day, and most will do so within a few days.
2. There is practically no cost involved once the set up has been completed. Large samples do not cost more than smaller ones (except for any cost to acquire the sample).
3. You can show pictures. Some Web survey software can also show video and play sound.
4. Web page questionnaires can use complex question skipping logic, randomizations and other features not possible with paper questionnaires or most email surveys. These features can assure better data.
5. Web page questionnaires can use colors, fonts and other formatting options not possible in most email surveys.
6. A significant number of people will give more honest answers to questions about sensitive topics, such as drug use or sex, when giving their answers to a computer, instead of to a person or on paper.
7. On average, people give longer answers to open-ended questions on Web page questionnaires than they do on other kinds of self-administered surveys.

Disadvantages

1. Current use of the Internet is far from universal. Internet surveys do not reflect the population as a whole. This is true even if a sample of Internet users is selected to match the general population in terms of age, gender and other demographics.
2. People can easily quit in the middle of a questionnaire. They are not as likely to complete a long questionnaire on the Web, as they would be if talking with a good interviewer.

3. If your survey pops up on a web page, you often have no control over who replies - anyone from Antarctica to Zanzibar, cruising that web page may answer.
4. Depending on your software, there is often no control over people responding multiple times to bias the results.

Things to consider when conducting surveys:

Who are you planning on surveying? Decide what group you are going to focus on surveying based on who you have access to and what your research is focused on.

How many people are you going to survey? You want to choose a target number of surveys to conduct. You don't want too few surveys because you won't have enough answers to support any generalizations or findings you may make. At the same time, you do not want too many surveys because you will be overwhelmed with analyzing your data.

How are you going to survey people? You can choose to conduct your survey in person (i.e. walk up to people and ask them questions); on paper (i.e. hand out surveys and ask people to return them); or even via the Internet. The survey method should be chosen based on the length of your survey and types of questions.

How long is your survey going to be? The answer to this question depends on what information you are attempting to discover and how much you want to find out. Longer surveys sometimes involve the same question asked in multiple ways to see if people are consistent in their answering strategies. For your first survey, however, it is better to keep things simple. Short questions are usually more effective than longer ones.

What type of questions are you going to ask? Do you want open-ended questions or closed questions? Open-ended questions are questions that allow the participant any type of response. An example of an open-ended question is: How are you feeling today? A closed question is one with a set of possible

responses or yes/no responses. An example is: Did you feel that the new campus regulation about parking is fair? While closed questions are much easier to analyze they do not provide the rich responses you may get with open-ended questions. Ultimately, what type of question you ask depends on what you want to discover.

SURVEY CHARACTERISTICS

Survey characteristics	Mail survey	Telephone survey	Personal interview survey
Cost	Lowest	Moderate	High
Geographic distribution	Wide	Wide	Moderate
Flexibility and questioning	Low	Moderate	High
Interviewer bias	No	Moderate	High
Speed and data collection	Slowest	Fastest	Moderate
Control and data collection	Lowest	Moderate	Highest
Response rates	Poor	Moderate	Highest

XI-CASE STUDY

Case study is a vaery popular method of qualitative research.It involves a careful and complete observation rather than using large samples and following

a rigid protocol to examine a limited number of variables, case study methods involve an in-depth, longitudinal examination of a single instance or event: a case. They provide a systematic way of looking at events, collecting data, analyzing information, and reporting the results. As a result the researcher may gain a sharpened understanding of why the instance happened as it did, and what might become important to look at more extensively in future research. Case studies lend themselves especially to generating (rather than testing) hypothesis
The design involves the intense investigation of situations which are relevant to the problem situation.The concept is to select several target cases where an intensive analysis will help to

- Know the environment
- Identify relevant variables
- Indicate nature of relationship among the variable
- Identify the naure of roblems and opportunies in the orginal case.

The case history method is especially useful in situations in which a complicated series of variables interact to produce the problem or opportunity .

PRIMARY vs. SECONDARY SOURCE OF DATA

Primary sources are the raw material of the research process, they represent the records of research or events as first described. Secondary sources are based on primary sources. These sources analyze, describe, and synthesize the primary or original source. Usage of primary source and secondary source should be complementary. Savvy entrepreneurs will do secondary research first and then conduct primary research. For example, the owner of a video-rental shop would want to know all about a neighborhood before opening a new store there. Using information gleaned from secondary sources, the owner can find all kinds of demographic data, including detailed income data and spending patterns. They can then send out a questionnaire to a sampling of households to find out what

kinds of movies people like to rent. That primary-research technique will help when it comes time to stock the store with the latest Hollywood releases. Secondary research lays the groundwork and primary research helps fill in the gaps. By using both types of market research, business owners get a well-rounded view of their market and have the information they need to make important business decisions. The selection of methods depends upon various factors like

1. Scope of the study.
2. Objectives of the study.
3. Time availability for the study.
4. Degree of accuracy needed.
5. Study environment, respondents.
6. Budget estimated.

IV.C. FIELD OPERATIONS

Learning objectives

After reading this part, you may be able familiar with the following topics

- **Objectives of field operations in marketing research.**
- **Functions involved in field operation.**

The actual data collection process is rarely done by the person who designs the research. An irony of marketing research is that highly educated and trained individuals will design the research, but the people who gather the data typically have little research training and experience. Field operation is that phase of the project during which researchers make contact with the respondents, administrator the data collection instruments, record the data, and return the data to a central location for processing. The wisdom behind the research design and skill involved in developing the data collection instrument will be wasted in the

field operation is poorly administrated. The planning of the field operation is highly influenced by data collection method employed.

Marketing managers preparing for the fieldwork operation should concentrate on the following areas.

1. Briefing sessions for field workers.
2. Supervision of the fieldwork
3. Verification

Briefing for field workers:

A proper research design will eliminate numerous sources of error, but careful execution of the fieldwork is necessary to produce results without substantial error. For the best execution the fieldworkers must be properly trained. Whether the fieldworkers/interviewers have just completed their training in basics or are already experienced, there is a need to inform them about the current research. Both experienced and inexperienced fieldworkers must be briefed on the background of the project, study method sampling technique. Instruction s for handling certain key situations are always important. To train the fieldworkers about the questionnaire, a field supervisor can elaborate the questions and its purpose. The general cautions to be taken by the fieldworkers can also be outlined

The briefing session will also cover the sampling procedure. A number of research projects allow the interviewer to be at least partially responsible for selecting sample. When this is the case, the potential for selection bias exists. For example in quota sample the field worker may select a wrong sample due to poor awareness about the importance of the quota representation. In addition to technical procedure few general guidelines can also be given to these fieldworkers, like the ethics in conducting field work.

Supervision of Fieldwork:

Irrespective of the training given to the field workers, they may commit few errors in execution. Direct supervision of personal interviewers, telephone

interviewers and other fieldworkers is necessary to ensure that the technique communicated in the training sessions are implemented in the field. The supervision of fieldworkers, like other forms of supervision, refers to controlling the efforts of workers. Field supervision of fieldworkers requires checking to see that field procedures are being properly followed. The supervisor checks field operations to ensure that the interviewing schedule is being met. In addition to quality control, continual training may be provided. For example, a telephone supervisor may notice that interviewers are allowing the phone to ring continuously before considering the call a no answer. The supervisor can instruct interviewers that if a telephone is allowed to ring too long and then answered, the respondent may be annoyed

Verification: It is the quality control procedures in fieldwork to ensure that interviewers are following the field plan. At the field the most important job of the supervisor is to verify that the interviews are being conducted according to the schedule and the plan. An interviewer might be tempted to skip the process for various reasons. Careful recording of the number of completed interviews will help ensure that the sampling procedure is being properly conducted. It is the supervisors function to motivate interviewers to carefully follow the plan. Where ever the deficiency is found the supervisor should take remedial actions to solve the complication.

IV.D. ERRORS AND DIFFICULTIES

Learning objectives

After reading this part, you may be able familiar with the following topics

- **Sources error in marketing research**
- **Impact of errors on research**
- **Types of errors**

i) ERRORS :

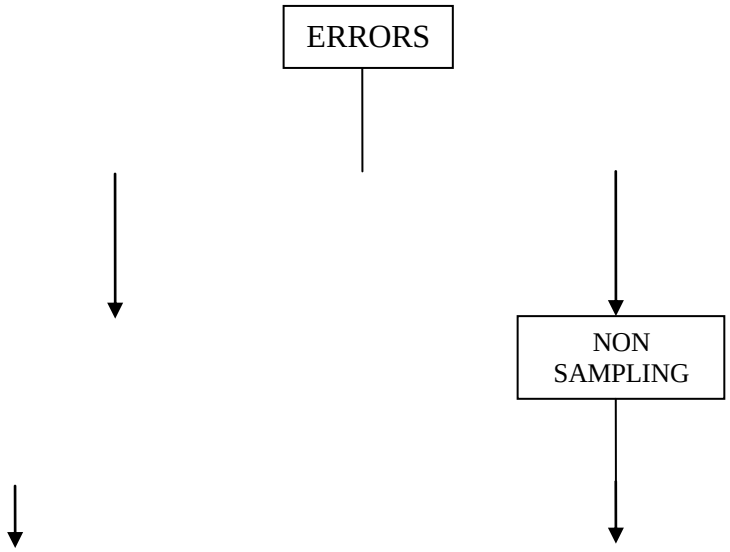
Since the usability of any market research depends upon the accuracy of the results, error control plays a critical role in the research process. Every step in the marketing research progression can produce serious errors. The control of these errors is of critical concern in marketing research. Avoiding many of the simpler marketing research errors takes only common sense, but avoiding many of the more complex mistakes requires a much deeper level of awareness. The common marketing research errors are highlighted below.

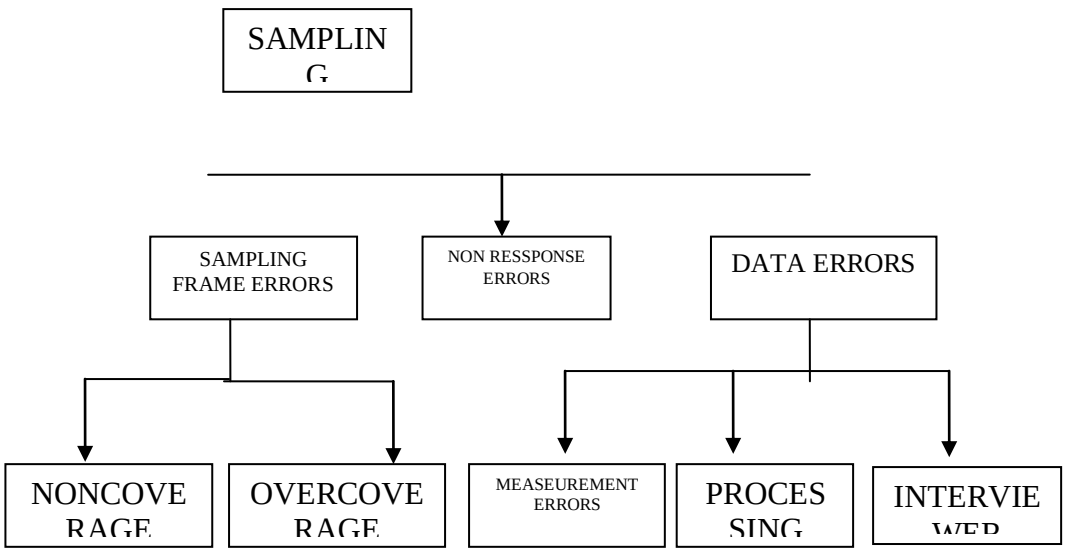
Types of errors

- ✓ Sampling errors
- ✓ Non sampling errors

Sampling errors

Most marketing research studies utilize samples of people, product or stores. Based upon these sample results, the researcher and the manager make conclusion of the whole population from which the sample was selected. For example the attitudes of all Maruti car owners could be inferred from a sample of a 1000 owners because the





sample is used to estimate the population. The difference between the sample value and the corresponding population value is called as sampling error .

Non sampling errors

Non sampling errors are all the errors that may occur in the marketing research process except the sampling error. This concept simply includes all the aspects of the research process where mistakes and deliberate deceptions can occur.. Unfortunately these mistakes and deceptions occur with great frequency in the marketing research process. An important point to note is that sampling error is measurable while it is not easy to measure as non-sampling error.

Hence, the researcher should take care of the following points to deal with non-sampling error.

- ✓ What kind of non- sampling errors may occur.
- ✓ What effects these errors may have on our results, and
- ✓ What steps we can take to reduce these errors.

The effect of non sampling errors:

Sampling error has two properties that make it useful to the researcher.

- ✓ measurable

- ✓ relevance with sample size

Unfortunately non sampling errors are not easily measurable and they do not decrease with sample size. Infact in all likelihood non sampling errors increase as sample size increases. What non sampling errors do is put a bias in results of unknown direction and magnitude

TYPES OF NON SAMPLING ERRORS:

Defective problem definition:

Problem of the study should be clearly stated so that they can be linked directly to the research results. Research objectives should always be clear so that research results can be presented in relation to specific objectives. A product manager requests a study to test a media mix. If the true problem is pricing strategy then any research that is conducted no matter how technically correct will not be helpful to the manager.

Defective population definition:

The study population must be defined to fit the study objectives. A universe which is relevant to the problem being studied, and a sample which adequately represents that universe, are vital requirements of high quality research. Consider the case of the manager of one of the restaurants in a major metropolitan airport would like to know what sort of image the restaurant has among those who have some likelihood of eating in the airport. The population is defined as people over 18 years old, getting off planes in the third week of October. If the sample is selected from this population one might get misleading results. It does not include significant numbers of potential customers that is people who are just taking off. Also the sample included people who has no chance of eating in the restaurants. That is people who change planes without going into the main terminal where the restaurants are located. Conclusions from this study are questionable.

Frame non representative of the population:

There are many acceptable methods of sampling, each with advantages and disadvantages in specific situations. The following guidelines are not meant to specify how sampling must be designed or managed, but rather are aimed at insuring that sample design and management are disclosed in sufficient detail to allow clear judgments of a sample's adequacy for the stated research purpose. The sampling frame must match the defined population. Consider the case of the investment company that uses the telephone book (the frame)to select a sample of "potential stock buyers" this frame would not cover the defined population well as a

significant number of high income people have unlisted phone numbers. These high income people are the prime potential stock buyers. Again conclusions are suspects.

Non responsive errors:

Errors occur because people in selected sample either refuse to be a part of the sample or they are not at home during the sample periods. a sample is a representative sample as selected. Some of the selected elements do not form part of the realized sample, it is not a truly representative sample. The resulting error is called non–response error. As an example of this problem consider the case of a resort developer who attempts to interview people during the day. The study yields some refusals and a lot of "not at homes". One must wonder whether the refusals as a group hold different attitudes of the development from those who respond.

Questionnaire structure error:

The error made when the structure and layout of the survey instrument leads to inaccurate responses. For example asking probing questions regarding viewpoints on potentially negative experiences before asking an overall

satisfaction question, where overall satisfaction would be incorrectly affected by the recent recall of potentially bad experiences.

Measurement error :

This is caused when information gathered is different from the information sought. For example, respondents are asked to indicate whether they own a car or not. Some of them may respond in the positive just to boost their image before an interviewer, even though they may not be owning a car. Such responses will result in measurement error.

Data analysis error:

The error that occurs when analysis is incorrectly executed. Simple mathematical errors are common, which is why data analysis should be checked over by more than one qualified person for quality. A more significant data analysis error is when simple frequency reporting (straight number percentage reporting) is executed when far greater information can be mined from the results (often inexpensively) through additional analysis such as cross-tabulation analysis, multiple regression (driver analysis), cluster analysis, factor analysis, perceptual mapping (multidimensional scaling), structural equation modeling tests, etc.

Reporting error:

The best approach and program design combined with the best analysis is only as good as the researcher's capability to synthesize and report on the results. The most common reporting error by far is the improper representation of the significant findings in a format conducive to creating management understanding and buy-in of survey results. It could be something as simple as poor language syntax to as complex as choosing the wrong results to report or not choosing the best way to graphically represent the results. More common in the current environment is not selecting the best delivery vehicle. For example, a

quality online reporting system is much preferred when distributing results across a company that is geographically spread out.

Common sources of error in field work:

Five common sources of error in field work are identified in the following discussion:

1. errors in selecting respondents
2. non-response errors (i.e. failure to get data from selected respondents)
3. errors created by the method of seeking data
4. errors resulting from interviewer's misinterpreting or misrecording answers; and
5. interviewer cheating

Respondent selection errors:

Telephone errors:

In telephone surveys the interviewer is typically given a list of numbers to call, or numbers are dialed on the basis of random-digit dialing. If the interviewer in the former case is also given the name of the individual at each number to whom she is to speak, there is no problem, as the interviewer simply asks for that individual by name. Unfortunately, names are seldom available, and the interviewer must select the individual to be interviewed at the number called-usually a household.

Mail intercept interviews:

In the case of shopping center interviews, respondents are often selected by convenience. It cannot be said that errors are made in respondent selection when this is the case, but the procedure is often biased because interviewers are likely to select those individuals who look friendly and appear easy to interview. To introduce more objectivity into the process, specific times and locations within the shopping center can be selected randomly, and it can be specified that the

first person passing a given point after each interview or attempted interview will be sought as the next respondent.

Door-door surveys:

When quota samples are used in door-door surveys, interviewers select the individuals to be interviewed subject only to quotas for various population groups such as sex, age and income. This interviewer control of the selection of the respondents is unlikely to result in the equivalent of a random sample. Interviewers tend to follow the paths of least resistance and greatest convenience.

Not -at -homes:

The percentage of not-at-homes varies by city size, day of the week, time of day, season of the year, age and the sex of the respondent, as well as with the provisions made to control not-at-homes in the individual studies. It is, however, almost always surprisingly large. Failure to obtain data from not-at-homes may bias survey results because population groups vary in the probability of being at home.

Refusals:

Refusal rates vary from project to project and may range up to 20%. Since refusals are often the result of personality and mood, it can be argued that they will occur randomly and will not bias results. Moreover, refusals are a matter of degree and of circumstances such as convenience at time of call. Repeated efforts to obtain compliance can reduce the refusal rate, but only to a degree. In addition to general refusals, refusals may occur on specific questions, particularly those relating to income and other personal questions.

ii) MINIMIZING OF ERRORS:

In research it is not possible to eliminate all the errors .However, to the extent that can be minimized. The following precautions can be followed to reduce the errors :

- ✓ Selection of suitable study method
- ✓ Selection of appropriate instrument
- ✓ Adequate sample size
- ✓ Using of trained and experiences researchers
- ✓ Planned data processing

iii) DIFFICULTIES IN DATA COLLECTION

Marketing research basically a problem solving tool. Starting from the definition of the problem till the end of the presentation of the report researchers are facing numerous problems. The major problems faced by the researchers are mentioned below:

1. Volatile changes in the business environment make marketing research more complicated. Continuous changes in market make the results of the study unsuitable.
2. Lack of scientific training and application in marketing research methodology is a major problem in our country .
3. The Research and Development Department has become a common feature in many organizations. But decisions makers do not appear to be very eager on implement the findings of the study.
4. Many of the organizations are not reach conscious and feel that investment in research is wastage of resources and does not encourage research.
5. Many people largely depend on customs, traditions and routine practices in their decision making, as they feel research does not have any useful purpose to serve in the management of their business.
6. The secrecy of business information is sacrosanct to business organizations. Most of the business organizations in our country do not part with information to researchers.
7. There is insufficient support between the business organizations and research institutions, which essential for the development of good and meaningful research.
8. In India companies are not in a position to allocate huge funds for the research.

9. Lack of availability secondary data makes marketing research as baseless start and time consuming one.
10. Poor awareness among the consumers about research makes the study more burning.
11. Researchers in India are not familiar with the new research instruments available for conducting market research,
12. poor library facilities at any places, because of which researchers have to spend much of time and energy in tracing out relevant material and information.
13. there is a difficulty of timely availability of upto date data from published sources
14. Lack of code of conduct among the researchers brings bad image on research. There is a need for developing code of conduct for researchers to educate them about ethical aspects of research.

IV.E. DATA PROCESSING

Learning objectives

In this chapter you will learn about

- **Objective of data processing**
- **Functions involved in data processing**
- **Process of editing and coding.**

Once the collection of data is over, the next step is to organize data so that meaningful conclusions may be drawn. The information content of the observations has to be reduced to a relatively few concepts and aggregates. The data collected from the field has to be processed as laid down in the research plan. This is possible only through systematic processing of data. Data processing involves editing, coding, classification and tabulation of the data collected so that they are amenable to analysis. This is an intermediary stage between the collection of data and their analysis and interpretation.

i) EDITING OF DATA

Editing is the first stage in data processing. Editing may be broadly defined to be a procedure, which uses available information and assumptions to substitute inconsistent values in a data set. In other words, editing is the process of examining the data collected through various methods to detect errors and omissions and correct them for further analysis. While editing, care has to be taken to see that the data are as accurate and complete as possible, units of observations and number of decimal places are the same for the same variable.

The following practical guidelines may be handy while editing the data:

1. The editor should have a copy of the instructions given to the interviewers.
2. The editor should not destroy or erase the original entry. Original entry should be crossed out in such a manner that they are still legible.
3. All answers, which are modified or filled in afresh by the editor, have to be indicated.
4. All completed schedules should have the signature of the editor and the date.

For checking the quality of data collected, it is advisable to take a small sample of the questionnaire and examine them thoroughly. This helps in understanding the following types of problems:

1. Whether all the questions are answered,
2. Whether the answers are properly recorded,
3. Whether there is any bias,
4. Whether there is any interviewer dishonesty,
5. Whether there are inconsistencies.

At times, it may be worthwhile to group the same set of questionnaires according to the investigators (whether any particular investigator has specific problems) or according to geographical regions (whether any particular region

has specific problems) or according to the sex or background of the investigators, and corrective actions may be taken if any problem is observed.

Mechanics of editing:

For editing purpose the researcher must draw a proper mechanism which will simplify the process as well as reduces the duplication work. Data frequently are written in with a colored pen or pencil .When space on the questionnaire permits, the original data usually are left in to permit a subsequent edit and to identify the originals concepts. Before tabulation of data it may be good to prepare an operation manual to decide the process for identifying inconsistencies and errors and also the methods to edit and correct them.

Types of editing

1.Field editing: Preliminary editing by a field supervisor on the interview to catch technical omissions, check legibility of hand writing and clarify responses that are logically inconsistent

2.In-house editing: A rigorous editing job performed by a centralized office staff. The researcher normally has centralized office staff to perform editing and coding. The researcher must setup a centralized office with all facilities for editing and coding purposes by which coordination can be accomplished

ii) CODING OF DATA

Coding refers to the process by which data are categorized into groups and numerals or other symbols or both are assigned to each item depending on the class it falls in. Hence, coding involves:

- ✓ Deciding the categories to be used, and
- ✓ Assigning individual codes to them.

For example, for the open-ended question "Do you enjoy milkshakes, if so, how much would you say you enjoy milkshakes." The researcher in the coding process will then have to observe different answers and give them a numeric value. For example, If we use a five-point scale with 1 being low ('don't enjoy') and 5 being high ('favorite treat'), and the response is "I like milkshakes" – the researcher would code the response as a "3," if the response was "I absolutely love milkshakes!" the researcher would code the response as a "5".

In general, coding reduces the huge amount of information collected into a form that is amenable to analysis. A careful study of the answers is the starting point of coding. Next, a coding frame is to be developed by listing the answers and by assigning the codes to them. A coding manual is to be prepared with the details of variable names, codes and instructions. Normally, the coding manual should be prepared before collection of data, but for open-ended and partially coded questions. These two categories are to be taken care of after the data collection. The following are the broad general rules for coding:

1. Each respondent should be given a code number (an identification number).
2. Each qualitative question should have codes. Quantitative variables mayor may not be coded depending on the purpose. Monthly income should not be coded if one of the objectives is to compute average

monthly income- But if it is used as a classificatory variable it may be coded to indicate poor, middle or upper Income group.

3. All responses including "don't know", "no opinion" "no response" etc., are to be coded.
4. Sometimes it is not possible to anticipate all the responses and some questions are not coded before collection of data. Responses of all the questions are to be studied carefully and codes are to be decided by examining the essence of the answers. In partially coded questions, usually there is an option " Any other (specify)". Depending on the purpose, responses to this question may be examined and additional codes may be assigned.

Production coding:

The actual process of transferring the data from the questionnaire or data collection form after the data have been collected is called production coding. Based on the nature of the data collection tool, codes may be written directly on the instrument or special coding sheet. Coding should be done in a central location so that a supervisor can help to solve interpretation problems.

Computerized coding:

Studies having large sample size uses a computer system for coding as well as data processing. The process of transferring data from a research project, such as answers to survey questionnaire, to computer is referred to as data entry. Several alternative means of entering data in to the computer are also available. A

researcher using computer–assisted telephone interviewing or with on-line direct data entry equipment automatically stores and tabulates responses as they are collected. Optical Scanning Systems may also be used to directly read material from marked sheets.

CHAPTER SUMMARY

In marketing research the process of data collections plays a vital role .In this chapter you have studied about data instruments, features of different data collection methods and data processing. Based on the study selected the researcher either can go for secondary source or primary data collection and for both .The process of editing and coding shapes the data to get the maximum results.

KEY WORDS

Data: Quantitative or/ and qualitative information, collected for study and analysis.

Case study: the exploratory research technique that intensively investigates one or a few situations similar to the problem situation.

Coding: The process of identifying and assigning score or other character symbol to previously edited data.

Editing: the process of checking completeness, consistency, and legibility of data and making the data ready for coding and transfer to storage.

Interview: A method of collecting primary data by meeting the informants and asking the questions.

Observation: The process of observing individuals in controlled situations.

Questionnaire: is a device for collection of primary data containing a list of questions pertaining to enquiry, sent to the informants, and the informant himself writes the answers.

Primary Data: Data that is collected originally for the first time.

Secondary Data: Data which were collected and processed by someone else but are being used in the present study.

Published Sources: Sources which consist of published statistical information.

Schedule: is a device for collection of primary data containing a list of questions to be filled in by the enumerators who are specially appointed for that purpose.

Survey : a method of primary data collection in which information is gathered by communicating with a representative sample of people.

Questions for practice

1. What are the qualities of a good research instrument?
2. Enumerate the different methods of data collection.
3. Explain the merits and limitations of using secondary data.
4. What precautions would you take while using the data from secondary sources?
5. Explain the various sources of primary data.
6. What is observation? Explain the role of observer in the process of observation.
7. Examine the merits and limitations of the observation method.
8. Explain different methods of conducting interviews.
9. How does the case study method differ from the survey method? analyse the merits and limitations of case study method in marketing research
10. What are the guidelines in the construction of questionnaire? Explain.
11. What is field operations? Explain the cautions to be taken by the researcher in field operations.
12. What is data processing? Explain the various functions of data processing.
13. Explain the importance of data editing in marketing research.
14. Explain the main sources of error in field operations.
15. "its is never safe to take published statistics at their face value without knowing their meaning and limitations". Elucidate this statement by enumerating and explaining the various points which you would consider before using any published data. Illustrate your answer by examples wherever possible.

G.Thamizh chelvan

S.G.Lecturer in commerce and International Business

Dr.G.R.D.College of science
Avinashi road
Coimbatore

UNIT V - MARKETING RESEARCH

CHAPTER I

Data Analysis – Univariate Analysis

Objective:

The objective of this chapter is to understand:

- Methods of analysing Market Research data
- Methods of Univariate Analysis

Introduction:

After collection of data from marketing research, the data has to be analyzed. This can be carried out by various statistical methods.

Data analysis begins with univariate analysis. Univariate analysis also is the foundation for the bivariate and multivariate analysis.

Univariate analysis is the assessment of the distributional properties of a variable. It serves two broad purposes: description and preparation for multivariate analysis. These functions correspond to the two primary forms of univariate analysis, the assessment of *central tendency* and of *dispersion*, or convergence and divergence. This mainly deals with the meaning of a typical value and to what extent do values differ from this typical value. Descriptive research focuses on identifying what is most characteristic of a set of

observations. Any variation from this typical value usually is the most important concern with regard to subsequent multivariate analysis.

There is a possibility that many times univariate analysis itself is the research goal. For example, we might calculate the percentage of women going for work. Descriptive research emphasizes what are most typical using estimates of central tendency.

When univariate analysis is preliminary to multivariate analysis, dispersion takes center stage. This analysis often uncovers at least some technical problems that need to be resolved before other forms of analysis can proceed.

Measures of Central Tendency

Measures of central tendency summarize the entire distribution of values as one single quantity or quality that can be thought of as the average value. Measures of central tendency are measures of the location of the middle or the center of a distribution. The mean is the most commonly used measure of central tendency.

The three most commonly-used measures of central tendency are:

1. **Mean**
    a. **Arithmetic Mean**

    The arithmetic mean is commonly called the average. The sum of the values divided by the number of values--often called the "average."

    The mean is the sum of all the scores divided by the number of scores.

This is denoted as

$\mu = \Sigma X/N$

where μ is the population mean and N is the number of scores. If the scores are from a sample, then the symbol M refers to the mean and N refers to the sample size. The formula for M is the same as the formula for μ.

Example: The mean of 5, 10, 22, 25, 17 is (5 + 10 + 22 + 25 + 17) / 5 = 15.8.

The mean is a good measure of central tendency for roughly symmetric distributions. This can be misleading in skewed distributions since it can be greatly influenced by extreme scores. Therefore, other statistics such as the median may be more informative for distributions such as reaction time or family income that are frequently much skewed.

b. Geometric Mean

The geometric mean is the nth root of the product of the scores. Thus, the geometric mean of the scores: 1, 2, 3, and 10 is the fourth root of 1 x 2 x 3 x 10 which is the fourth root of 60 which equals 2.78. The formula can be written as:

Geometric mean = $\left(\prod X\right)^{\frac{1}{N}}$

where ΠX means to take the product of all the values of X.

The geometric mean can also be computed by:

1. taking the logarithm of each number
2. computing the arithmetic mean of the logarithms
3. raising the base used to take the logarithms to the arithmetic mean.

c. **Harmonic Mean**

The harmonic mean is used to take the mean of sample sizes. If there are k samples each of size n, then the harmonic mean is defined as:

$$n_h = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k}}$$

For the numbers 1, 2, 3, and 10, the harmonic mean is:

$$\frac{4}{\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{10}}$$

= 2.069. This is less than the geometric mean of 2.78 and the arithmetic mean of 4.

The arithmetic mean in the case of a frequency distribution is calculated using the following steps:

- The middle point of each class interval is found and is multiplied by the number of observations (frequencies) in that class.
- The resultant values are summed up.
- The total is divided by the number of observations.

$$\overline{X} = \frac{\sum_{i=1}^{h} f_i x_i}{n}$$

where,

$\overline{X}$ = the sample mean

f_i = the frequency of the i th class

x_i= the mid-point of the i th class

h = the number of classes

n= the total number of observations in the sample

2. Median

The median is the middle score of a distribution: half the scores are above the median and half are below the median. The median is less sensitive to extreme scores than the mean and this makes it a better measure than the mean for highly skewed distributions. The median income is usually more informative than the mean income, for example.

Example: The median of the same five numbers (5, 10, 22, 25, and 17) is 17.

The sum of the absolute deviations of each number from the median is lower than is the sum of absolute deviations from any other number. The mean, median, and mode are equal in symmetric distributions. The mean is higher than the median in positively skewed distributions and lower than the median in negatively skewed distributions

For a grouped series, the median is calculated as:

$$M = l_1 + \frac{l_2 - l_1}{f} (m-c)$$

where,

M = Median

l_1 = the lower limit of the class in which the median lies

l_2 = the upper limit of the class in which the median lies

f = the frequency of the class in which the median lies

m = the middle item or n/2

c = the cumulative frequency of the class preceding the one in which the median lies.

3. Mode

The third common measure of central tendency is the mode. The advantage of the mode as a measure of central tendency is that its meaning is obvious. Further, it is the only measure of central tendency that can be used with nominal data.

Some of the important characteristics of the mode are:

- It can be applied to both qualitative and quantitative distribution
- It is not affected by the extreme values in the distribution
- It can be ascertained in an open-ended distribution

Mode denotes the most frequently-occurring value (or values). The mode is the value (or values) with the highest frequency.

Example: For men having the following ages -- 19, 18, 21, 23, 23, 23, 24 and 21, the mode is 23.

In case of a grouped data, the mode is calculated using the following formula:

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) - (f_1 - f_2)} \times i$$

where,

l_1 = the lower value of the class in which the mode lies

f_1 = the frequency of the class in which the mode lies

f_0 = the frequency of the class preceding the modal class

f_2 = the frequency of the class succeeding the modal class

i = the class interval of the modal class

The mode is greatly subject to sample fluctuations and is therefore not recommended to be used as the only measure of central tendency. A further disadvantage of the mode is that many distributions have more than one mode. These distributions are called "multimodal." In contrast to the mode, the median and the mean both pertain exclusively to ordered data.

The choice of a measure of central tendency depends upon the level of measurement (nominal, ordinal, interval, or ratio) of the variable and the shape of its distribution. The mode is the only indicator of central tendency for a nominal variable. It may be computed for other types of variables as well, but is not especially useful unless there is a distinct peak, that is, when one value clearly predominates. At least an ordinal level of measurement is required for the median; the mean additionally requires an interval level of measurement. In

general, the mean is preferred over the median and the median over the mode because the mean utilizes the most information about the distribution whereas the mode preserves only one piece of information. This also depends on the manner in which the variable is distributed.

It is necessary to examine the entire set of values to determine which value best typifies the set as a whole. It is inadvisable to calculate a measure of central tendency for a variable without first examining its distribution.

Dispersion

As seen earlier, the measures of central tendency are used to estimate "normal" values of a dataset. Measures of dispersion are important for describing the spread of the data, or its variation around a central value. Two distinct samples may have the same mean or median, but completely different levels of variability, or vice versa. A proper description of a set of data should include both of these characteristics. There are various methods that can be used to measure the dispersion of a dataset, each with its own set of advantages and disadvantages.

1. **Range**

It's calculation is one of the simplest. It is defined as the difference between the largest and smallest sample values. The range depends only on extreme values and provides no information about how the remaining data is distributed.

2. **Variance and Standard Deviation**

The standard deviation is the square root of the sample variance. These measures of dispersion are very important.

$$\text{variance} = \sigma^2 = \frac{\sum (x_r - \mu)^2}{n}$$

$$\text{standard deviation} \quad \sigma = \sqrt{\frac{\sum (x_r - \mu)^2}{n}}$$

μ = mean

What the formula means:

(1) $x_r - \mu$ means take each value in turn and subtract the mean from each value.
(2) $(x_r - \mu)^2$ means square each of the results obtained from step (1). This is to get rid of any minus signs.

(3) $\Sigma(x_r - \mu)^2$ means add up all of the results obtained from step (2).
(4) Divide step (3) by n, which is the sum of the numbers
(5) For the standard deviation, square root the answer to step (4).

Grouped Data

The formula for the standard deviation when the data is grouped is:

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$\text{Variance, } \sigma^2 = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$$

Example:

Find out the variance for the following. The table shows scores (out of 10) obtained by 20 people in a test

Scores (x)	**Frequency (f)**
1	0
2	1
3	1
4	3
5	2
6	5
7	5
8	2
9	0
10	1

Solution:

Scores (x)	**Frequency (f)**	**fx**	**fx^2**
1	0	0	0
2	1	2	4
3	1	3	9
4	3	12	48
5	2	10	50
6	5	30	180
7	5	35	245
8	2	16	128
9	0	0	0
10	1	10	100
	20	**118**	**764**

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$\text{Variance, } \sigma^2 = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$$

Variance = $\frac{764}{20} - \left(\frac{118}{20}\right)^2$

= 38.2 - 34.81 = 3.39

Summary:

The chapter dealt with the measures of central tendency which are the arithmetic mean, the median and the mode. This was followed by the measures of dispersion; the standard deviation and the coefficient of variation.

The measures of central tendency are used to estimate "normal" values of a dataset. Measures of dispersion are important for describing the spread of the data, or its variation around a central value. Two distinct samples may have the same mean or median, but completely different levels of variability, or vice versa.

Questions:

1. What are the measures of central tendency?
2. What is a mean? Explain the various forms of mean.
3. What are the characteristics of the median?
4. What are the characteristics of the mode?
5. What are the measures of the central tendency?
6. Explain standard deviation.
7. How you measure variance?
8. How are the measures of central tendency different from the measures of dispersion?

CHAPTER II

HYPOTHESIS TESTING

Objective:

The objective of this chapter is to understand:

- Meaning of hypothesis
- Hypothesis testing
- Types of hypothesis
- Two types of error in hypothesis testing
- One-tailed and two-tailed tests

Introduction:

A **statistical hypothesis** is an assumption about a population parameter. This assumption may or may not be true.

The best way to determine whether a statistical hypothesis is true is to examine the entire population. Since this is often impractical, researchers typically examine a random sample from the population. If the sample data are consistent with the statistical hypothesis, the hypothesis is accepted; if not, the hypothesis is rejected.

There are two types of statistical hypothesis.

- **Null hypothesis**. The null hypothesis is usually the hypothesis that sample observations result purely from chance effects.
- **Alternative hypothesis**. The alternative hypothesis is the hypothesis that sample observations are influenced by some non-random cause.

For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in Heads and half, in Tails. The alternative hypothesis might be that the number of Heads and Tails would be very different. Suppose we flipped the coin 50 times, resulting in 40 Heads and 10 Tails. Given this result, we would be inclined to reject the null hypothesis and accept the alternative hypothesis.

The term “null” means nothing or invalid. It may be written as:

$H_0: \mu = \mu_0$

Where, H_0 is the null hypothesis and μ_0 is the mean of the population.

The alternative hypothesis is

H_A: $\mu \neq \mu_0$

The rejection of the null hypothesis will show that the mean of the population is not μ_0. If this happens, then it implies that the alternative hypothesis is accepted. There can be more than two or more alternative hypothesis though only one can be tested at a time against the null hypothesis.

Hypothesis Testing

Statisticians follow a formal process to determine whether to accept or reject a null hypothesis, based on sample data. This process, called **hypothesis testing**, consists of four steps.

- **Formulating the hypothesis**: The first step in hypothesis testing is to formulate the hypothesis to be tested. This means stating the null hypothesis and the alternative hypothesis.

- **Identifying the test statistic**: The test statistic is a statistic that will be used by the researcher to determine whether the null hypothesis should be accepted or rejected. Typically, the test statistic is the sample estimate of the population parameter in the null hypothesis. Therefore, since we are testing a hypothesis about a population mean, the test statistic will be the sample mean. When the hypothesis pertains to a large sample (30 or more), the z-test implying normal distribution is used. When a sample is small (less than 30), the use of the z-test will be inappropriate. Instead the t-test will be more suitable. The test criteria frequently used in hypothesis testing are Z, t. F and χ^2.

- **Formulating a decision rule**: The decision rule consists of two parts: (1) a test statistic and (2) a range of values, called the region of

acceptance. The decision rule determines whether a null hypothesis is accepted or rejected. If the test statistic falls within the region of acceptance, the null hypothesis is accepted; otherwise, it is rejected.

- **Accepting or rejecting the null hypothesis**: Once the region of acceptance is defined, the null hypothesis can be tested against sample data. The test statistic is computed. For example consider that the test statistic is the sample mean. If the sample mean falls within the region of acceptance, the null hypothesis is accepted; if not, it is rejected.

Decision Rules

There are four possibilities that can arise when a hypothesis is tested:

1. The hypothesis is true but our test leads to its rejection.
2. The hypothesis is false but our test leads to its acceptance.
3. The hypothesis is true and our test leads to its acceptance.
4. The hypothesis is false and our test leads to its rejection.

Out of these four, the first two lead to an error in decision. The first possibility leads to a Type I error and the second possibility leads to a Type II error. This can be shown as follows:

		STATE OF NATURE	
		H_o True	H_o False
DECISION	Retain H_o	Correct Retention	Type II Error
	Reject H_o	Type I Error	Correct Rejection

The decision rule is a procedure that a researcher uses to decide whether to accept or reject the null hypothesis. There are two types of errors that can result from a decision rule.

- **Type I error**. A Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the **significance level**. This probability is also called **alpha**, and is often denoted by α.

- **Type II error**. A Type II error occurs when the researcher accepts a null hypothesis that is false. The probability of committing a Type II error is called **Beta**, and is often denoted by β. The probability of *not* committing a Type II error is called the **Power** of the test.

In practice, the decision rule has two parts: (1) a test statistic and (2) a range of values. The range of values is called the **region of acceptance**. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level. If the test statistic falls within the region of acceptance, the null hypothesis is accepted.

Note: The set of values outside the region of acceptance is called the **region of rejection**. If the test statistic falls within the region of rejection, the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the α level of significance.

One-Tailed and Two-Tailed Tests

To understand the difference between these look at the table below. It shows three sets of hypothesis. Each makes a statement about how the population mean μ is related to a specified value *M*.

Set	Null hypothesis	Alternative hypothesis	Number of tails
1	$\mu = M$	$\mu \neq M$	2
2	$\mu \geq M$	$\mu < M$	1
3	$\mu \leq M$	$\mu > M$	1

The first set of hypotheses (Set 1) is an example of a two-tailed test, since an extreme value on either side of the sampling distribution would cause a researcher to reject the null hypothesis. The other two sets of hypotheses (Sets 2 and 3) are one-tailed tests, since an extreme value on only one side of the sampling distribution would cause a researcher to reject the null hypothesis.

A test of a statistical hypothesis, where the region of rejection is on only one side of the sampling distribution, is called a **one-tailed test**. For example, suppose the null hypothesis states that the mean is less than or equal to 10. The alternative hypothesis would be that the mean is greater than 10. The region of rejection would consist of a range of numbers located on the right side of sampling distribution; that is, a set of numbers greater than 10.

A test of a statistical hypothesis, where the region of rejection is on both sides of the sampling distribution, is called a **two-tailed test**. For example, suppose the null hypothesis states that the mean is equal to 10. The alternative hypothesis would be that the mean is less than 10 or greater than 10. The region of rejection would consist of a range of numbers located on both sides of sampling distribution; that is, the region of rejection would consist partly of numbers that were less than 10 and partly of numbers that were greater than 10.

Other Considerations

There may be other considerations to be made while testing the null hypothesis. These are:

- Assumptions need to be made about the sampling distribution of the mean score. If the sample is relatively large (i.e., greater than or equal to 30), you can assume, based on the central limit theorem, that the sampling distribution will be roughly normal.

On the other hand, if the sample size is small (less than 30) and if the population random variable is approximately normally distributed (i.e., has a bell-shaped curve), you can transform the mean score into a t score. The t score will have a t distribution.

- Assume that the mean of the sampling distribution is equal to the test value M specified in the null hypothesis.
- In some situations, the standard deviation needs to be computed from the sampling distribution s_x. If the standard deviation of the population σ is known, then

  $s_x = \sigma * sqrt[(1/n) - (1/N)]$

  where n is the sample size and N is the population size. On the other hand, if the standard deviation of the population σ is unknown, then

  $s_x = s * sqrt[(1/n) - (1/N)]$

  where s is the sample standard deviation.

Hypothesis testing in respect of interval data:

1. **Test of sample mean (Single population):**

On the basis of the sample drawn from the population one needs to infer about the population parameter which could be the population mean. We need to carry out an appropriate statistical test of significance for testing the hypothesis concerning the population mean. We can consider two cases, sample size being large ($n>30$) and sample size small ($n\leq30$).

If n is large, the sample distribution of mean follows normal distribution as per central limit theorem and we can go for Z test. However, when sample size is small, we may have to choose between Z and t test depending upon whether standard deviation (σ) is known or not. In case σ is known, we should go for Z test or t test.

In case of a normal distribution of sample means (x bar) with mean μ and standard deviation σ is represented by

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

The standard normal distribution is calculated as:

$$Z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$$

This is used for testing the hypothesis.

2. **Test of Proportion:**

In order to test the hypothesis that population proportion (P) takes a specified value P_o against a two sided alternative, we test the hypothesis as follows:

H_o (Null Hypothesis): $\pi = \pi_o$

H_1 (Alternative Hypothesis): $\pi \neq \pi_o$

Here it can be shown that:

E (P) = π and

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}}$$

Using the Z-test, we find that

$$Z = \frac{P - \pi}{\sigma_P}$$

Summary:

A statistical hypothesis is an assumption about a population parameter. This assumption may or may not be true.

There are two types of statistical hypothesis.

- Null hypothesis. The null hypothesis is usually the hypothesis that sample observations result purely from chance effects.

- Alternative hypothesis. The alternative hypothesis is the hypothesis that sample observations are influenced by some non-random cause.

This process, called hypothesis testing, consists of four steps: Formulating the hypothesis, identifying the test statistic, formulating a decision rule and accepting or rejecting the null hypothesis

There are two types of errors that can result from a decision rule: Type I error which occurs when the researcher rejects a null hypothesis when it is true and Type II error which occurs when the researcher accepts a null hypothesis that is false.

A test of a statistical hypothesis, where the region of rejection is on only one side of the sampling distribution, is called a one-tailed test. A test of a statistical hypothesis, where the region of rejection is on both sides of the sampling distribution, is called a two-tailed test.

Questions:

1. What is a hypothesis? Explain.
2. Differentiate between Null hypothesis and alternative hypothesis.
3. What are the steps involved in testing a hypothesis?
4. Is hypothesis testing useful in marketing decision making?
5. What are Type I error and Type II errors?
6. Explain the relationship between Type I and Type II errors.
7. What are one-tailed and two-tailed tests?
8. What are the other considerations to be made while testing the null hypothesis?
9. How will you test the sample mean from a single population?

10. How will you test the hypothesis of a population proportion?

CHAPTER III

Bivariate Analysis

Objective:

The objective of this chapter is to understand:

- Association between Dependent Variable (DV) and Independent Variable (IV).
- Importance and methodology of correlation analysis.
- Importance and methodology of regression analysis.

Introduction:

In bivariate analysis, the hypothesis of "association" and causality are tested. In its simplest form, association simply refers to the extent to which it becomes easier to know/predict a value for the Dependent Variable (DV) if we know a case's value on the Independent Variable (IV).

This association could be understood by a measure of association. A measure of association often ranges between –1 and 1. Where the sign of the integer represents the "direction" of correlation (negative or positive relationships) and the distance away from 0 represents the degree or extent of correlation – the farther the number away from 0, the higher or "more perfect" the relationship is between the IV and DV.

Measures of association and statistical significance that are used vary by the level of measurement of the variables analyzed.

For Nominal Variables:

Measure of association is Lambda

The formula for Lambda is

$$\frac{\text{(Reduction in Error from guessing to predicting based on IV)}}{\text{Number of Original Error}}$$

This gives a ratio of how much improvement your prediction has by knowing values on the IV. Lambda ranges from 0 to 1. The higher the number, the stronger the relationship between the two variables.

The measure of statistical significance for nominal variables (and limited scale ordinal variables) is Chi Square. In fact, Chi Square can measure the statistical significance of any crosstab. It tells us how different the values in the cells of a crosstab are from expected values (or values predicted if no real relationship between the two variables existed – uses marginals to calculate these expected values).

The Chi Square is based on these factors:

1. The distribution of cases among the cells (can show the extent to which differences are observed;
2. The number of cells (degrees of freedom), and
3. The size of the sample (n).

The Chi Square requires the following steps:

1. State the null hypothesis and calculate the number in each category assuming that the null hypothesis is correct.
2. Determine the level of significance.
3. Calculate the Chi Square (x^2) as follows

$$X^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Where, O_i is the observed frequency in the i^{th} category

E_i is the expected frequency in the i^{th} category

k is the number of categories

4. Determine the number of degrees of freedom. For the specified level of significance and the degrees of freedom, find the critical or theoretical value of x^2.
5. The calculated value of the x^2 is then calculated with the theoretical value of x^2 and the region of rejection is determined.

For ordinal variables:

The appropriate measures of association all attempt to measure how values of ordered variables relate for the sample of cases. For instance, how many times high values are associated with high values, how many times are they associated with low values. They each use discordant and concordant pairs to create a value between –1 and 1. O indicates no relationship between how the values for the cases pair up. The closer to –1 means the stronger the negative (inverse) relationship, and the closer to 1 the more "perfect" the positive relationship.

There are several measures of association which measure ordinal variables' relationships. Somer's d, tau b, tau c, and gamma are the most usual ones. All are slight variations of the formula related in layman's terms above. While Somer's d, tau b, and tau c will all have very close to the same value, gamma usually will appear to have a slightly higher (stronger) relationship.

For interval variables

Regression and correlation analysis are used for this. The two terms correlation and regression are distinct from each other and many times are used interchangeably. Correlation is a statistical technique used for measuring the

relationship or interdependence of two or more variables. Correlation does not necessarily indicate a causal relationship between two or more variables. Regression analysis refers to the technique for deriving an equation that relates the dependent variable to one or more variables. It is used to predict one variable on the basis of another and hence helps in bringing out the causal relationship between two or more variables.

Correlation Analysis:

Correlation is a statistical technique used for measuring the relationship or interdependence of two or more variables. The measurement scales used should be at least interval scales, but other correlation coefficients are available to handle other types of data. Correlation coefficients can range from -1.00 to +1.00. The value of -1.00 represents a perfect negative correlation while a value of +1.00 represents a perfect positive correlation. A value of 0.00 represents a lack of correlation.

In statistics, the **Correlation coefficient** (r) is a measure of how well a linear equation describes the relation between two variables X and Y measured on the same object or organism. It is defined as the sum of the products of the standard scores of the two measures divided by the degrees of freedom:

$$r = \frac{\sum z_x z_y}{N - 1}$$

The result obtained is equivalent to dividing the covariance between the two variables by the product of their standard deviations. In general the correlation coefficient is one of the two square roots (either positive or negative) of the

coefficient of determination (r^2), which is the ratio of explained variation to total variation:

$$r^2 = \frac{\sum(Y' - \overline{Y})^2}{\sum(Y - \overline{Y})^2}$$

where:

Y = a score on a random variable Y

Y' = corresponding predicted value of Y, given the correlation of X and Y and the value of X

$\overline{Y}$ = sample mean of Y (i.e., the mean of a finite number of independent observed realizations of Y, not to be confused with the expected value of Y)

The correlation coefficient adds a sign to show the direction of the relationship. This applies when the relationship is linear.

The coefficient ranges from −1 to 1. A value of 1 shows that a linear equation describes the relationship perfectly and positively, with all data points lying on the same line and with Y increasing with X. A score of −1 shows that all data points lie on a single line but that Y increases as X decreases. A value of 0 shows that a linear model is inappropriate – that there is no linear relationship between the variables.

If we have a series of n measurements of X and Y written as x_i and y_i where i = 1, 2, ..., n, then the coefficient can be used to estimate the correlation of X and Y . This is also known as the "sample correlation coefficient". It is especially important if X and Y are both normally distributed. The Pearson correlation

coefficient is then the best estimate of the correlation of X and Y . The Pearson correlation coefficient is written:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

where and are the sample means of x_i and y_i , sx and sy are the sample standard deviations of x_i and y_i and the sum is from i = 1 to n.

This can be written as:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\ \sqrt{n\sum y_i^2 - (\sum y_i)^2}}.$$

The absolute value of the sample correlation must be less than or equal to 1

Regression Analysis:

In statistics, regression analysis is used to model relationships between random variables, determine the magnitude of the relationships between variables, and can be used to make predictions based on the models.

Regression analysis models the relationship between one or more response variables (also called dependent variables) usually named Y, and the predictors (also called independent variables) usually named X_1,...,Xp.

A line in a two dimensional or two-variable space is defined by the equation Y=a+b*X; in full text: the Y variable can be expressed in terms of a constant (a)

and a slope (b) times the X variable. The constant is also referred to as the intercept, and the slope as the regression coefficient or B coefficient.

Regression analysis can predict the outcome of a given key business indicator (dependent variable) based on the interactions of other related business drivers (independent variable). For example, it allows predicting sales volume, using the amount spent on advertising and number of sales people that a company employs.

The coefficient of determination, r^2, is useful because it gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable. It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph. The coefficient of determination is the ratio of the explained variation to the total variation. The coefficient of determination is such that $0 < r^2 < 1$, and denotes the strength of the linear association between x and y. The coefficient of determination represents the percent of the data that is the closest to the line of best fit.

Every sample has some variation in it. The total variation is made up of two parts, the part that can be explained by the regression equation and the part that can't be explained by the regression equation.

$$\sum (y - \bar{y})^2 = \sum (y' - \bar{y})^2 + \sum (y - y')^2$$

$$total = explained + unexplained$$

In regression analysis, the coefficient of determination is a measure of goodness-of-fit (i.e. how well or tightly the data fit the estimated model). The coefficient is defined as the ratio of two sums of squares:

$r^2 =$ SSR / SST

where SSR is the sum of squares due to regression, SST is the total sum of squares. By "sum of squares" we mean the sum of squared deviations between actual values and the mean (SST), or between predicted values and the mean (SSR). The coefficient of determination takes on values between 0 and 1, with values closer to 1 implying a better fit.

The objectives for which market researchers use regression analysis are:

1. Pattern connecting the dependent variable and the independent variable by establishing a functional relationship between the two.
2. For solving problems involving prediction and forecasting.
3. To study the quantum of variation in the dependent variable using the set of independent variables.

The Scatter Diagrams:

The Scatter Diagram is a tool for determining the potential correlation between two different sets of variables, i.e., how one variable changes with the other variable. This diagram simply plots pairs of corresponding data from two variables. The scatter diagram does not determine the exact relationship between the two variables, but it does indicate whether they are correlated or not. It also does not predict cause and effect relationships between these variables.

The scatter diagram is used to:

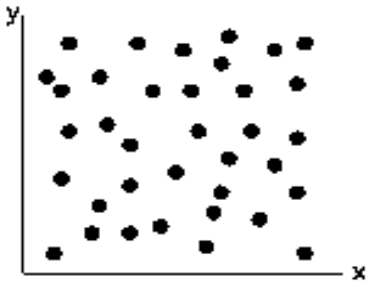
1) Quickly confirm a hypothesis that two variables are correlated;

2) Provide a graphical representation of the strength of the relationship between two variables; and
3) Serve as a follow-up step to a cause-effect analysis to establish whether a change in an identified cause can indeed produce a change in its identified effect.

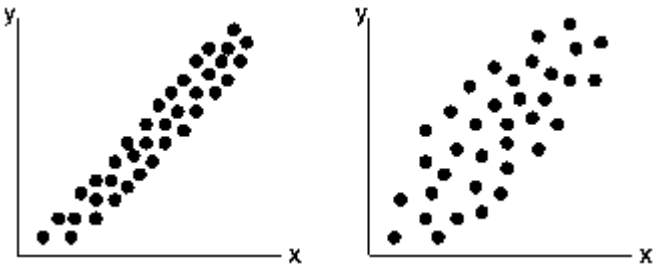
To make a scatter diagram for two variables requiring confirmation of correlation, the following simple steps are usually followed:

1) 50-100 pairs of data are collected for this and tabulated;
2) The x- and y-axes of the diagram, along with the scales that increase to the right for the x-axis and upward for the y-axis are drawn;
3) The data for one variable is assigned to the x-axis (the independent variable) and the data for the other variable is assigned to the y-axis (the independent variable);
4) Plot the data pairs on the scatter diagram, encircling (as many times as necessary) all data points that are repeated.

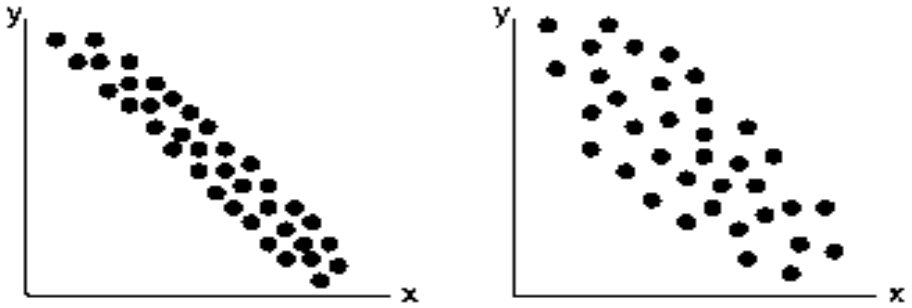
Interpretation of the resulting scatter diagram is very simple and can be analyzed looking at the pattern formed by the points. If the data points plotted on the scatter diagram are all over the place with no discernible pattern whatsoever, then there is no correlation at all between the two variables of the scatter diagram. An example of a scatter diagram that shows no correlation is shown below



The below scatter diagrams show positive correlation (Strongly positive and weak positive correlation):



The below scatter diagrams show negative correlation (Strongly negative and weak negative correlation):



If the variables are correlated, the points will fall along a line or curve. The better the correlation, the tighter the points will hug the line.

Summary:

In bivariate analysis, the hypothesis of "association" and causality are tested. In its simplest form, association simply refers to the extent to which it becomes easier to know/predict a value for the Dependent Variable (DV) if we know a case's value on the Independent Variable (IV).

The two terms correlation and regression are distinct from each other and many times are used interchangeably. Correlation is a statistical technique used for measuring the relationship or interdependence of two or more variables. Regression analysis refers to the technique for deriving an equation that relates the dependent variable to one or more variables. It is used to predict one variable on the basis of another and hence helps in bringing out the causal relationship between two or more variables.

The Scatter Diagram is a tool for determining the potential correlation between two different sets of variables, i.e., how one variable changes with the other variable. This diagram simply plots pairs of corresponding data from two variables.

Questions:

1) What is a chi-square test? What are the uses of the same?
2) What is correlation analysis?
3) What is regression analysis?
4) What is the difference between correlation and regression analysis?
5) What are scatter diagrams? What do they depict?

CHAPTER IV

Multivariate Analysis

Objective:

The objective of this chapter is to understand:

- The process of Multivariate analysis
- Study about analysis of variation, Multiple Regression Analysis, Discriminant Analysis, Conjoint Analysis, Factor Analysis, Cluster Analysis and Multi-dimensional Scaling

Introduction:

Multivariate analysis is the analysis of the simultaneous relationships among three or more phenomena. In Univariate analysis the focus is on the level (average) and distribution (variance) of the phenomenon, in a bivariate analysis the focus shifts to the degree of relationships (correlations or co variances) between the phenomena. In a multivariate analysis, the focus shifts from paired relationships to more complex simultaneous relationships among phenomena.

The various methods of Multivariate analysis are:

1. Analysis of Variance
2. Dependence Methods
    - Multiple Regression Analysis
    - Discriminant Analysis
    - Conjoint Analysis
3. Inter Dependence Methods

- Factor Analysis
- Cluster Analysis
- Multi-dimensional Scaling

Analysis of Variance

This can be studied under ANOVA, ANCOVA, MANOVA and MANCOVA.

ANOVA: An ANOVA (Analysis of Variance), sometimes called an F test, is closely related to the t test. The t test measures the difference between the means of two groups whereas an ANOVA tests the difference between the means of two or more groups.

ANOVA can be a one-way ANOVA or a factorial ANOVA.

One-way or single factor ANOVA, tests differences between groups that are only classified on one independent variable. A factorial ANOVA can show whether there are significant main effects of the independent variables and whether there are significant interaction effects between independent variables in a set of data. Interaction effects occur when the impact of one independent variable depends on the level of the second independent variable.

The advantage of using ANOVA rather than multiple t-tests is that it reduces the probability of a type-I error. Making multiple comparisons increases the likelihood of finding something by chance—making a type-I error.

An F indicates that there is a significant difference between groups, not which groups are significantly different from each other. This is one potential drawback to an ANOVA, which is loss of specificity.

A factorial ANOVA can examine data that are classified on multiple independent variables. More than two independent variables can be compared in an ANOVA (e.g., three-way, four-way).

ANCOVA: In ANCOVA, we can analyze both qualitative (class) and quantitative (continuous) independent variables. The mixed procedure allows the user to model both class and continuous variables. In ANOVA-type models, hypotheses about class or interactions among class variables are tests of means or differences among means. In regression-type models, in which all "factors" are continuous variables (rather than categories), hypothesis tests are tests about regression coefficients.

As one might expect, the assumptions of ANCOVA combines both the assumptions of regression and ANOVA. In addition, the tests of adjusted means are based on the assumption that the class variable by covariate interaction is negligible, that is the regression lines are parallel.

MANOVA: Multivariate analysis of variance (MANOVA) is simply an ANOVA with several dependent variables. For example, we may conduct a study where we try two different textbooks, and we are interested in the students' improvements in Physics and Chemistry. In that case, improvements in physics and chemistry are the two dependent variables, and our hypothesis is that both together are affected by the difference in textbooks. A multivariate analysis of variance (MANOVA) could be used to test this hypothesis. Instead of a

univariate F value, we would obtain a multivariate F value (Wilks' lambda) based on a comparison of the error variance/covariance matrix and the effect variance/covariance matrix. The "covariance" here is included because the two measures are probably correlated and we must take this correlation into account when performing the significance test.

MANOVA is useful in experimental situations where at least some of the independent variables are manipulated. It has several advantages over ANOVA.

1. By measuring several dependent variables in a single experiment, there is a better chance of discovering which factor is truly important.
2. It can protect against Type I errors that might occur if multiple ANOVA's were conducted independently. Additionally, it can reveal differences not discovered by ANOVA tests.

However, there are several cautions as well. It is a substantially more complicated design than ANOVA, and therefore there can be some ambiguity as to which independent variable affects each dependent variable. Moreover, one degree of freedom is lost for each dependent variable that is added. Finally, the dependent variables should be largely uncorrelated. If the dependent variables are highly correlated, there is little advantage in including more than one in the test given the resultant loss in degrees of freedom.

Some of the assumptions made here are:

- Normal Distribution: The dependent variable should be normally distributed within groups. Overall, the F test is robust to non-normality if it is caused by skewness rather than outliers. Tests for outliers should

be run before performing a MANOVA, and outliers should be transformed or removed.

- Homogeneity of Variances: Homogeneity of variances assumes that the dependent variables exhibit equal levels of variance across the range of predictor variables.
- Homogeneity of Variances and Covariance's: In multivariate designs, with multiple dependent measures, the homogeneity of variances assumption described earlier also applies. However, since there are multiple dependent variables, it is also required that their covariances are homogeneous across the cells of the design. There are various specific tests of this assumption.

Two Special Cases arise in MANOVA:

- Unequal sample sizes: As in ANOVA, when cells in a factorial MANOVA have different sample sizes, the sum of squares for effect plus error does not equal the total sum of squares. This causes tests of main effects and interactions to be correlated.
- Within-subjects design: Problems arise if the researcher measures several different dependent variables on different occasions.

MANCOVA: MANCOVA is an extension of ANCOVA. It is simply a MANOVA where the artificial direct variables are initially adjusted for differences in one or more covariates. This can reduce error "noise" when error associated with the covariate is removed.

MULTIPLE REGRESSION ANALYSIS

Multiple regression is used to account for (predict) the variance in an interval dependent, based on linear combinations of interval, dichotomous, or dummy independent variables. Multiple regression can establish that a set of independent variables explains a proportion of the variance in a dependent variable at a significant level (through a significance test of R^2), and can establish the relative predictive importance of the independent variables.

The multiple regression equation takes the form

$$y = a + b_1x_1 + b_2x_2 + ... + b_nx_n$$

where y is the dependent variable which is to be predicted, x_1, x_2 and x_n are the n known variables on which the predictions are to be based and a, b_1, b_2, ….b_n are parameters, the values of which are to be determined by the methods of least squares.

Associated with multiple regression is r^2 (multiple correlation), which is the percent of variance in the dependent variable explained collectively by all of the independent variables.

$$r^2 = \frac{\Sigma (Y_i - \overline{Y})^2 - (Y_i - Y_c)^2}{(Y_i - \overline{Y})^2}$$

where r^2 is the co-efficient of determination, Y_i is the value of i^{th} item in Y series, Y (bar) is the mean of the Y series and Y_c is the computed value of the i^{th} item in Y series on the basis of the regression.

Multiple regression shares all the assumptions of correlation: linearity of relationships, the same level of relationship throughout the range of the independent variable, interval or near-interval data, absence of outliers, and data whose range is not truncated. In addition, it is important that the model being tested is correctly specified. The exclusion of important causal variables or the inclusion of extraneous variables can change the interpretation of the importance of the independent variables.

Multiple regressions with dummy variables yield the same inferences as multiple analysis of variance (MANOVA), to which it is statistically equivalent. When the dependent variable is a dichotomy the assumptions of multiple regressions cannot be met, discriminant analysis or logistic regression is used instead. Partial least squares regression is sometimes used to predict one set of response variables from a set of independent variables

DISCRIMINANT ANALYSIS

Discriminant analysis is used to classify the sample into two or more categories. Example: Consumers may be classified as heavy and light users; Sales people can be classified as successful and unsuccessful and so on.

Discriminant function analysis is used to determine which variables discriminate between two or more naturally occurring groups.

For example, a researcher may want to investigate which variables discriminate between engineers who decide

(1) To seek employment in private companies,

(2) To take up government services, or

(3) To seek opportunities abroad.

For that purpose the researcher could collect data on numerous variables after the graduation of the engineers. *Discriminant Analysis* could then be used to determine which variable(s) are the best predictors of the engineers' choice of employment.

The objectives of two group discriminant analysis are to find a linear composite of the predictor variable to help the analyst to separate the groups, establishing procedures for assigning new individuals, testing for significant differences between the mean predictor variables and determining the variable which accounts for the most intergroup differences.

This is commonly carried out with the help of a computer program.

CONJOINT ANALYSIS

Conjoint analysis deals with the measurement of the combined effect of two or more attributes that are important from the view of the consumer. The use of the conjoint analysis will be more appropriate in a situation where a company would like to know the most desirable attribute for a new product or service.

For example, a hotel would like to know whether choice of menu or prompt service would attract a customer to visit them frequently.

For this, it will seek data from the consumer in the firm of response to identify product attributes. The various options available for this are direct interview with the customer or focus group interviews. All the attributes are weighed and compared.

The main steps involved in the application of conjoint analysis are

1. **Determination of salient attributes**: the attributes have to be selected based on the marketers experience or through interviews. Only valuable attributes need to be considered.
2. **Assigning levels to the selected attributes**: this can vary from most preferred to the least preferred.
3. **Fractional factorial design of experiments**: during the comparison of the profile of different products, it is essential to have a minimum number of designs which provide us all the information required. This will ensure easy management of the design.
4. **Physical design of the stimuli**: a prototype or a picture of the concept may be given to the consumer or customer to get a realistic picture.
5. **Data collection**: The customers are asked to rank all the alternatives using a rating scale. This will ensure ease of data collection and analysis.
6. **Determination of part-worth utilities**: Regression methods, mathematical programming methods, econometric methods may be used for the part-worth utility values.

The applications of conjoint analysis are

- It can be used for optimum product design based on the attributes considered. Simulations can be created to represent competitors' action or a fresh scenario.

- Consumers can be segmented based on their sensitivity to product attributes.
- It may help a manager to conduct SWOT analysis of the brand as the part-worth utility speaks about the relative brand strength.

There are certain limitations of conjoint analysis. It may not be perfect and convincing and may fail to capture utility functions and decision roles.

FACTOR ANALYSIS

Factor analysis is a name given to a class of techniques whose purpose is data reduction and summarization. The data from market research are vast and factor analysis helps in reducing the number of variables. Factor analysis is an explorative technique.

Factor analysis was invented nearly 100 years ago by psychologist Charles Spearman, who hypothesized that the enormous variety of tests of mental ability--measures of mathematical skill, vocabulary, other verbal skills, artistic skills, logical reasoning ability, etc.--could all be explained by one underlying "factor" of general intelligence that he called g. He hypothesized that if g could be measured and you could select a subpopulation of people with the same score on g, in that subpopulation you would find no correlations among any tests of mental ability. In other words, he hypothesized that g was the only factor common to all those measures.

The objectives of factor analysis are simplifying the data by reducing a large number of variables to a set of a small number of variables and analyzing the interdependence of relationship among a total set of variables.

Factor analysis can be used in several ways as given below:

1. It brings out the hidden dimensions relevant to a researcher among product preferences.
2. Helps to find out relationships among observed values.
3. Used when the data is large and has to be simplified and condensed.

The limitations of factor analysis are:

- It is a complicated tool and should be used if the researcher has a good understanding of the technique.
- The reliability of the results is some times questionable.
- It suitability depends on the judgment of the researcher.

Factor analysis is used in the case of exploratory research and has to be used were the concepts are well formulated and tested.

CLUSTER ANALYSIS

Cluster analysis is a technique used to segment a market. It is used to classify a person or object into a small number or mutually exclusive and exhaustive groups. Its object is to sort cases (people, things, events, etc) into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. Each cluster thus describes, in terms of the data collected, the class to which its members belong;

and this description may be abstracted through use from the particular to the general class or type.

CA lacks an underlying body of statistical theory and is heuristic in nature. Cluster analysis requires decisions to be made by the user relating to the calculation of clusters, decisions which have a strong influence on the results of the classification. CA is useful to classify groups or objects and is more objective than subjective.

Cluster analysis, like factor analysis and multi dimensional scaling, is an interdependence technique: it makes no distinction between dependent and independent variables. The entire set of interdependent relationships is examined. It is similar to multi dimensional scaling in that both examine inter-object similarity by examining the complete set of interdependent relationships. The difference is that multi dimensional scaling identifies underlying dimensions, while cluster analysis identifies clusters. Cluster analysis is the obverse of factor analysis. Whereas factor analysis reduces the number of variables by grouping them into a smaller set of factors, cluster analysis reduces the number of observations or cases by grouping them into a smaller set of clusters.

In marketing, cluster analysis is used for:

- Segmenting the market and determining target markets
- Product positioning and New Product Development
- Selecting test markets

Example: A supermarket might gather data on all of their existing customers and survey them regarding their buying criteria relative to the product line. They could then use cluster analysis to group customers with similar buying patterns

together. This type of cluster analysis, also known as market segmentation, is performed at increasing rates, due to the advent of high-speed computers and the ready availability of demographic data. Based on the broader descriptions of individuals within each cluster, the retail managers could make decisions that would be appropriate for the individuals within.

Clusters for this example might include:

- Price-sensitive shoppers
- Indifferent shoppers
- Quality-focused shoppers
- High-end status shoppers
- Monthly shopper

Clustering methods may be top down and employ logical division, or bottom up and undertake aggregation. Aggregation procedures which are based upon combining cases through assessment of similarities are the most common and popular will be the focus of this section.

Care should be taken that groups (classes) are meaningful in some fashion and are not arbitrary or artificial. To do so the clustering techniques attempt to have minimal internal variation as compared to maximal variation between groups. Homogeneous and distinct groups are delineated based upon assessment of distances or an F-test.

Steps in Cluster Analysis:

The two key steps within cluster analysis are the measurement of distances between objects and to group the objects based upon the resultant distances (linkages).

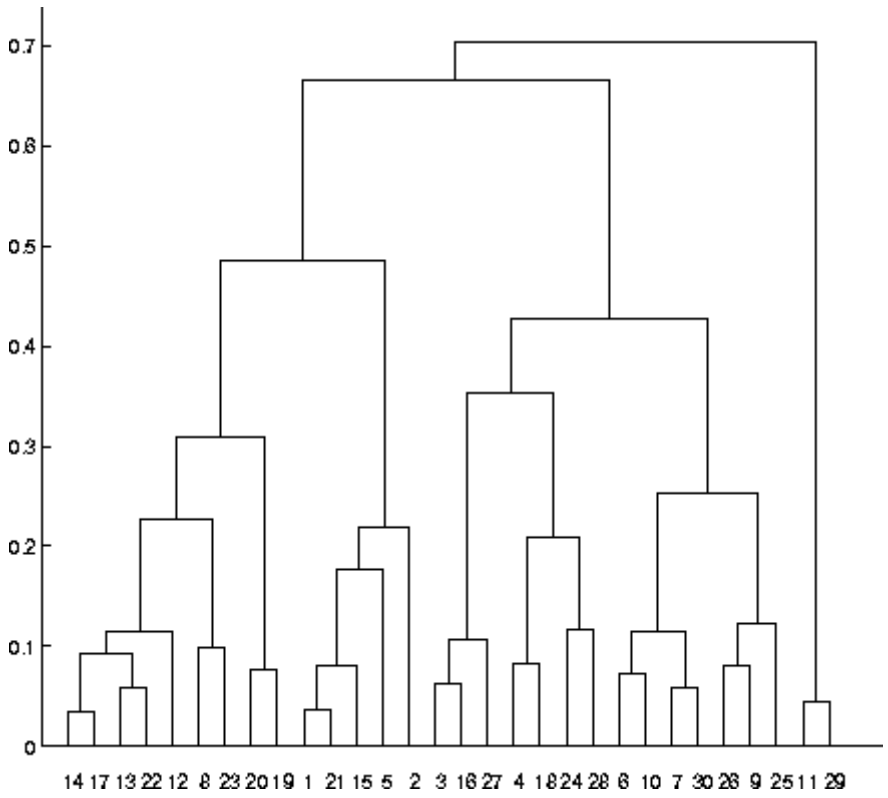
The distances provide for a measure of similarity between objects and may be measured in a variety of ways, such as Euclidean and Manhattan metric distance. The criteria used to then link (group) the variables may also be undertaken in a variety of manners, as a result significant variation in results may be seen.

Linkages are based upon how the association between groups is measured. For example, simple linkage or nearest neighbor distance, measures the distance to the nearest object in a group while furthest neighbor linkage or complete linkage, measures the distance between furthest objects. These linkages are both based upon single data values within groups, whereas average between group linkages is based upon the distance from all objects in a group. Centroid linkage has a new value, representing the group centroid, which is compared to the ungrouped point to weigh inclusion.

Ward's method is variance based with the groups variance assessed to enable clustering. The group which sees the smallest increase in variance with the iterative inclusion of a case will receive the case. Ward's is a popular default linkage which produces compact groups of well distributed size. Standardization of variables is undertaken to enable the comparison of variables to minimize the bias in weighting which may result from differing measurement scales and ranges. Z score format accounts for differences between mean values and reduces the standard deviation when variables have multivariate normality

Choosing number of groups:

The ideal number of groups to establish may be assessed graphically or numerically. Graphically the number of groups may be assessed with an icicle plot or dendrogram. The dendrogram bisected at a point which will divide the cases into a cluster based upon groupings up to the point where the bisection occurred. Numerically the number of cases may be assessed on the agglomeration schedule, by counting up from the bottom to where a significant break in slope (numbers) occurs.



The optimal number of groups may be assessed based upon knowledge of the data set. Discriminant analysis may also be employed to assess optimality and

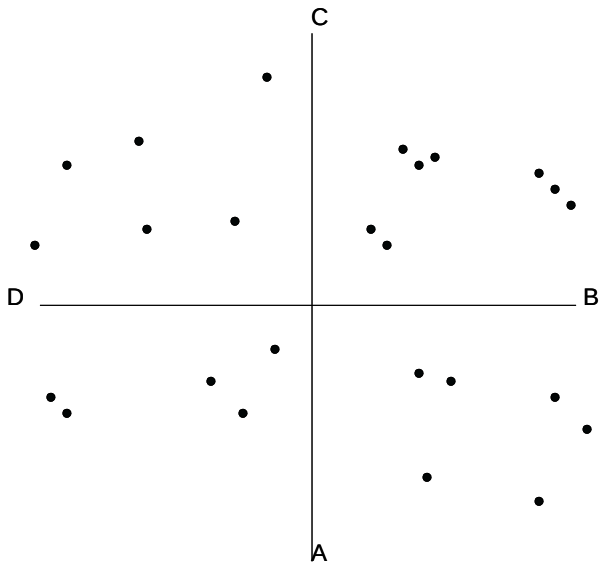
efficiency of computed groups, by imputing the cluster analysis derived classes for analysis with the original data.

Like the other techniques, cluster analysis presents the problem of how many factors, or dimensions or clusters to keep. One rule followed here is to choose a place where the cluster structure remains stable for a long distance. Some other possibilities are to look for cluster groupings that agree with existing or expected structures, or to replicate the analysis on subsets of the data to see if the structures emerge consistently.

MULTI-DIMENSIONAL SCALING

Multi-dimensional scaling (MDS) or perceptual map or positioning map is used for measuring human perception and preferences. It is spatial representation of relationships. It helps in the identification of attributes and the positioning of different products or brands on the basis of these attributes.
MDS is of two types, metric MDS and non metric MDS.

Given below is an example of a perceptual map. AC represents the x-axis and BD represents the y-axis. The value of the variable may be low to high from one end of the axis to the other.



Two approaches can be used for analyzing multi dimensional data. It can be done by measuring the attributes or distance between objects.

For MDS, a set of number called proximities and a computer based algorithms must be available.

The applications of MDS in marketing are in market segmentation, vendor evaluation, attitude scaling, advertisement evaluation, product repositioning, new product development and test marketing. Thus MDS measures the psychological distance or the dis-similarities to evaluate the external environment.

The limitations of MDS are

- Concepts of similarity and preferences may differ
- The selection of attributes are subjective
- It is at time difficult to interpret the results
- Different computer programmes may produce different results

Summary

This chapter deals with the different methods used in multivariate analysis and briefly describes their application. The methods are chosen based on the data and requirements of analysis.

Questions:

1. What is multivariate analysis? Explain.
2. Briefly explain the methods of multivariate analysis.
3. Explain variance of analysis. Write a note on the different types.
4. What is multiple regression analysis?
5. What is discriminant analysis? How is it useful in Marketing?
6. What are the steps involved in the application of conjoint analysis?
7. What are the limitations and uses of factor analysis?
8. What is cluster analysis? What are the steps involved in cluster analysis?
9. What is Multi-dimensional scaling? Where is it used?
10. How do you choose a method for analysis?

CHAPTER V

REPORT WRITNG

Objective:

The objective of this chapter is to understand:

- Importance of report writing
- Types of reports
- Considerations for oral and written reports

- Steps in the preparation of the report
- Presentation of data and feedback

Report Writing

It is not sufficient if the market researcher has collected information on the research problem, he also has to interpret the data and draw specific conclusions from it. The results of marketing research must be effectively communicated to management. This report has to be clear and concise.

Types of Report:

Reports can be broadly classified into two types: **oral and written**.

Presenting the results of a marketing research study to management generally involves a formal written report as well as an oral presentation. The report and presentation are extremely important. This is because:

1. The results of marketing research are often intangible (there is very little physical evidence of the resources, such as time and effort, that went into the project); the written report is usually the only documentation of the project.
2. The written report and the oral presentation are the only aspect of the study that marketing executives are exposed to, and consequently the overall evaluation of the research project rests on how well this information is communicated. They might not have been part of the marketing research study and hence will not know what information was collected.

3. Since the written research report and oral presentation are typically the responsibility of the marketing research supplier, the communication effectiveness and usefulness of the information provided plays a crucial role in determining the choice of the particular marketing research supplier for the future.

Differences between oral and written reports:

An oral report is any presentation that is verbally done to the management. Written reports are documentation of the research findings.

There are three major differences between oral and written reports.

1. Oral reports are difficult to interpret as they lack visual advantage. Charts, diagrams or pictures cannot be used to stress on important points. They have to rely only on pauses and volume emphasis.
2. The pace of the presentation cannot be controlled and regulated by the audience who is being presented with the oral report. In a written report, the reader can clarify a certain point by reading it two or more times, if needed slowly and carefully.
3. A researcher will write the report very precisely and with more accuracy since he is aware that a written report is bound to receive considerable attention and scrutiny from the readers. In contrast, an oral report will not be so precise nor will the researcher give as much time in its presentation since it cannot be subjected to the same degree of scrutiny as a written report.

Considerations for Oral Reporting

A researcher has to consider the following when he has been asked to make an oral presentation:

- Consider the audience to whom the report has to be presented and prepare the same so as to incorporate the technical requirements.
- It should be properly planned. The researcher has to rehearse what he is going to say or recommend. He has to collect and organize the data in a logical manner.
- Suitability of the language is another point worth considering. The reporting has to be simple, concise and clear.
- Visual aids like charts, graphs and handouts can be used judiciously if required to create a better impact. However this also needs to be used sparingly as it might disturb the proceedings.

Consideration for Written Reporting

Every person has a different style of writing. There is not really one right style for a report, but there are some basic principles for writing a research report clearly.

Types of Report:

There are many classifications available for reports. Some of them are listed below:

Sl. No.	Classification	Types
1	Time interval	Daily, Weekly, Fortnightly, Monthly or Annual
2	Functional Basis	Informational, Examinational and Analytical
3	Subject Matter	Economics, Finance, Industry, and other subjects of interest
4	Physical-form	Short-form and Long-form
5	Relationship between the reader and the writer	Administrative, Professional and Independent
6	Employment status of authors	Private and Public reports
7	Formality	Formal and Informal

None of these are mutually exclusive. A report may be a combination of some of the above.

Preparation of the Report

Preparing a research report is just not writing alone, but involves other activities besides writing. In fact, writing is actually the last step in the preparation process. Before writing can take place, the results of the research project must be fully understood and thought must be given to what the report will say. The objectives of the research should be matched with the results and the report has to be written.

Thus, preparing a research report involves three steps:

1. Understanding: Understand the objective of the research and then provide the solution.
2. Organizing: Organize your thoughts and findings and build a logical flow.
3. Writing: Now, draft the outline of your report and then write the same.

The general guidelines that should be followed for any report or research paper are as follows:

- **Consider the audience**: The information resulting from the study is going to be used by the marketing managers, who will use the results to make decisions. Thus, the report has to be understood by them; the report should not be too technical and not too much jargon should be used.

  This is a particular difficulty when reporting the results of statistical analysis where there is a high probability that few of the target audience have a grasp of statistical concepts. Hence, this needs to be translated into simple language.

- **Be concise, but precise**: Many a time, the researcher in order to convey his effort, tends to overcrowd the report with data. This leads to loss of focus. On the one hand, a written report should be complete in the sense that it stands by itself and that no additional clarification is needed. On the other hand, the report must be concise and must focus on the critical elements of the project and must exclude unimportant issues. Hence, a research report has to be concise and precise.

- **Understand the results and draw conclusions**: The managers who read the report are expecting to see interpretive conclusions in the report. The researcher must therefore understand the results and be able to interpret these. He should analyze the results and present them to the managers.

Research Report Format

The following outline is the suggested format for writing the research report:

1. Title page
2. Letter of Authorization
3. Summary of findings
4. Table of contents
    - List of tables
    - List of figures
5. Introduction
    - Background to the research problem
    - Objectives
    - Hypotheses
6. Methodology
    - Data collection
    - Sample and sampling method
    - Statistical or qualitative methods used for data analysis
    - Sample description
7. Findings
8. Limitations
9. Results, interpretation and conclusions.
10. Recommendations
11. Appendices
12. Bibliography

The title page indicates the topic on which the report has been prepared, the date of submission, Prepared by and Prepared for details.

The letter of authorization is provided to facilitate the research process.

The summary of findings is perhaps the most important component of the written report, since many of the management team who are to receive a copy of the report will only read this section. The summary of findings is usually put right after the title page, or is bound separately and presented together with the report.

The table of contents guides the reader as to what it contains and helps him to navigate the document. A separate table has to be provided for the charts and graphs if they are part of the research report.

The introduction should describe the background of the study and the details of the research problem. Following that, automatically the broad aim of the research can be specified, which is then translated into a number of specific objectives. Furthermore, the hypotheses that are to be tested in the research are stated in this section.

In the methodology chapter the sampling methods and procedures are described, as well as the different statistical methods that are used for data analysis. Finally, the sample is described, giving the overall statistics, usually consisting of frequency counts for the various sample characteristics.

Once the sample has been described, the main findings are to be presented in such a way that all objectives of the study are achieved and the hypotheses are tested. As mentioned before, it is essential that the main findings are well interpreted and conclusions are drawn wherever possible.

The limitations or the caveats of the study should also be mentioned so that the manager understands the pitfalls in the research process and he can make suitable assumptions for his decision making.

Recommendations are required if the manager has specifically asked the agency to suggest certain measures at the end of the research.

Appendix contains supporting materials for the report which cannot be given in the body of the text. The bibliography details provide the details of the sources which were referenced for the preparation of the report. The index shows the various topics and the relevant page numbers in the report in an alphabetical order.

Data presentation

Easy-to-understand tables and graphics will greatly enhance the readability of the written research report. As a general rule, all tables and figures should contain:

1. Identification number corresponding to the list of tables and the list of figures
2. A title that conveys the content of the table or figure, also corresponding to the list of tables and the list of figures, and
3. Appropriate column labels and row labels for tables, and figure legends defining specific elements in the figure.

There are a number of ways to produce tables and figures. When typing a report on a typewriter or word-processor, it is sometimes easiest to type a table out by

hand. However, when complicated tables have to be produced, it is advisable to use spreadsheet software like Lotus 123 or Excel.

Feedback on the Report

A feedback should be sought from the manger to whom the report has been submitted. This will help the researcher to understand the drawbacks of his report and improve on the same. Feedback is generally not given and it has to be sought. This will help the researcher to improve his services during the future projects.

Summary:

It is not sufficient if the market researcher has collected information on the research problem, he also has to interpret the data and draw specific conclusions from it.

Presenting the results of a marketing research study to management generally involves a formal written report as well as an oral presentation. The report and presentation are extremely important.

Preparing a research report involves three steps:

1. Understanding: Understand the objective of the research and then provide the solution.
2. Organizing: Organize your thoughts and findings and build a logical flow.
3. Writing: Now, draft the outline of your report and then write the same.

The following outline is the suggested format for writing the research report:

1. Title page
2. Letter of Authorization
3. Summary of findings
4. Table of contents
    - List of tables
    - List of figures
5. Introduction
    - Background to the research problem
    - Objectives
    - Hypotheses
6. Methodology
    - Data collection
    - Sample and sampling method
    - Statistical or qualitative methods used for data analysis
    - Sample description
7. Findings
8. Limitations
9. Results, interpretation and conclusions.
10. Recommendations
11. Appendices
12. Bibliography

Questions:

1. List out the differences between an oral and a written report.
2. When oral reports preferred and what are the considerations for it?

3. What are the points to be remembered which making a written report?
4. How are reports classified? Explain.
5. What is an outline? How is it useful for the preparation of a report?
6. Describe the various steps involved in writing a report.
7. Is feedback important for a report? Why?
8. How important do you think are writing skills and language important in presenting the research findings?
9. Imagine that you are presenting a report to a head of an education institution. Write a report on the current educational system and how your institute can make a difference.
10. "Writing a report is different from conducting a research". Explain.

CHAPTER VI

PARETO ANALYSIS

Objective:

The objective of this chapter is to understand:

- The meaning of Pareto analysis
- Defining the Pareto effect
- Steps involved in creating a Pareto chart

- Importance of Pareto analysis

Introduction:

Pareto analysis is a very simple technique that helps a manager to choose the most effective changes to make. It is represented as a bar graph used to arrange information in such a way that priorities for process improvement can be established. Pareto charts are constructed to provide a before-and-after comparison of the effect of control or quality improvement measures.

The 80-20 theory was first developed in 1906, by an Italian economist, Vilfredo Pareto, who observed an unequal distribution of wealth and power in a relatively small proportion of the total population. Joseph M. Juran adapted Pareto's economic observations to business applications.

Pareto Analysis is also used in inventory management through an approach called "ABC Classification". The ABC classification system works by grouping items by annual sales volume. This helps identify the small number of items that will account for most of the sales volume and that are the most important ones to control for effective inventory management. Thus it helps in effective inventory management.

The Pareto effect

In practically every industrial country a small proportion of all the factories employ a disproportionate number of factory operatives. In some countries 15 percent of the firms employ 70 percent of the people. This same state of affairs

is repeated time after time. In retailing for example, one usually finds that up to 80 percent of the turnover is accounted for by 20 percent of the lines.

This effect, known as the 80 : 20 rule, can be observed in action so often that it seems to be almost a universal truth. As several economists have pointed out, at the turn of the century the bulk of the country's wealth was in the hands of a small number of people.

This fact gave rise to the Pareto effect or Pareto's law: a small proportion of causes produce a large proportion of results. Thus frequently a vital few causes may need special attention wile the trivial many may warrant very little. It is this phrase that is most commonly used in talking about the Pareto effect – 'the vital few and the trivial many'. A vital few customers may account for a very large percentage of total sales. A vital few taxes produce the bulk of total revenue. A vital few improvements can produce the bulk of the results.

The Pareto effect is named after Vilfredo Pareto, an economist and sociologist who lived from 1848 to 1923. Originally trained as an engineer he was a one time managing director of a group of coalmines. Later he took the chair of economics at Lausanne University, ultimately becoming a recluse. Mussolini made him a senator in 1922 but by his death in 1923 he was already at odds with the regime. Pareto was an elitist believing that the concept of the vital few and the trivial many extended to human beings.

This method stems in the first place from Pareto's suggestion of a curve of the distribution of wealth in a book of 1896. Whatever the source, the phrase of 'the vital few and the trivial many' deserves a place in every manager's thinking. It

is itself one of the most vital concepts in modern management. The results of thinking along Pareto lines are immense.

For example, there may be a lot of customer complaints, a lot of shop floor accidents, a high percentage of rejects, and a sudden increase in costs etc. The first stage is to carry out a Pareto analysis. This is nothing more than a list of causes in descending order of their frequency or occurrence. This list automatically reveals the vital few at the top of the list, gradually tailing off into the trivial many at the bottom of the list. Management's task is now clear and unavoidable: effort must be expended on those vital few at the head of the list first. This is because nothing of importance can take place unless it affects the vital few. Thus management's attention is unavoidably focused where it will do most good.

Another example is stock control. You frequently find an elaborate procedure for stock control with considerable paperwork flow. This is usually because the systems and procedures are geared to the most costly or fast-moving items. As a result trivial parts may cost a firm more in paperwork than they cost to purchase or to produce. An answer is to split the stock into three types, usually called A, B and C. Grade A items are the top 10 percent or so in money terms while grade C are the bottom 50-75 percent. Grade B are the items in between. It is often well worthwhile treating these three types of stock in a different way leading to considerable savings in money tied up in stock.

Production control can use the same principle by identifying these vital few processes, which control the manufacture, and then building the planning around these key processes. In quality control concentrating in particular on the most

troublesome causes follows the principle. In management control, the principle is used by top management looking continually at certain key figures.

Thus it is clear that the Pareto concept – 'the vital few and the trivial many' – is of utmost importance to management.

Pareto Charts:

Pareto charts provide a tool for visualizing the Pareto principle, which states that a small set of problems (the "vital few") affecting a common outcome tend to occur much more frequently than the remainder (the "useful many"). A Pareto chart can be used to decide which subset of problems should be solved first, or which problems deserve the most attention.

The Pareto Chart is used to illustrate occurrences of problems or defects in a descending order. It is used for making decisions at critical points in different processes. This means that it can be used both during the development process as well as when products are in use, e.g. customer complaints.

How to create a Pareto Chart:

- First step is to list all elements of interest
- Use the same unit of measurement and measure each of the elements.
- Order the elements according to their measure
- Calculate the percentage for each element out of the total measurement
- Accumulate the percentage from top to bottom to equal 100 %

- Create a bar and line graph, line representing cumulative percentage
- Work on the most important element first

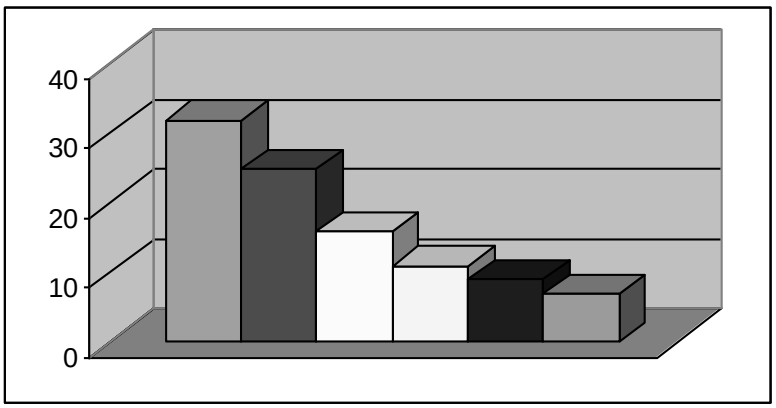
Advantages of Pareto Charts:

Pareto charts are a key improvement tool because they help us identify patterns and potential causes of a problem. Several Pareto charts can be created out of the same set of data. This will help a manager to quickly scan a number of factors that might contribute to a problem and focus on those with the greatest potential payback for his efforts

It is difficult to choose which issue to work on first when faced with a range of issues. To resolve this dilemma, the most useful thing to do is to apply Pareto's rule. This rule says - "eighty percent of your troubles will come from 20 per cent of your problems". In other words, problems will rarely have equal impact, so it is best to first concentrate on the most important.

This does not provide a scientifically accurate estimation of the weightage with respect to the range of alternatives. This reminds a manager to always look for 'the vital few' issues, and to separate them from 'the trivial many', before attempting to solve problems. The next step is to identify which particular problems are the most important. This is done by collecting appropriate data and displaying it in the form of a histogram with each measured characteristic shown in descending order of magnitude. Such a histogram is known as a Pareto chart.

Following is an example of a Pareto chart.



The high value items to the left hand side of the chart are the ones that are needed to be concentrated on first.

Some problems and difficulties associated with Pareto Analysis:

- Misrepresentation of the data.
- Inappropriate measurements depicted.
- Lack of understanding of how it should be applied to particular problems.
- Knowing when and how to use Pareto Analysis.
- Inaccurate plotting of cumulative percent data.
- Overcoming the difficulties.

In conclusion

Even in circumstances which do not strictly conform to the 80: 20 rule the method is an extremely useful way to identify the most critical aspects on which to concentrate. When used correctly Pareto Analysis is a powerful and effective

tool in continuous improvement and problem solving to separate the ‘vital few’ from the ‘many other’ causes in terms of cost and/or frequency of occurrence.

It is the discipline of organizing the data that is central to the success of using Pareto Analysis. Once calculated and displayed graphically, it becomes a selling tool to the improvement team and management, raising the question why the team is focusing its energies on certain aspects of the problem.

Summary:

Pareto analysis is a very simple technique that helps a manager to choose the most effective changes to make. . This effect, known as the 80: 20 rules, can be observed in action so often that it seems to be almost a universal truth.

It is represented as a bar graph used to arrange information in such a way that priorities for process improvement can be established. Pareto charts are constructed to provide a before-and-after comparison of the effect of control or quality improvement measures

Pareto charts provide a tool for visualizing the Pareto principle, which states that a small set of problems (the "vital few") affecting a common outcome tend to occur much more frequently than the remainder (the "useful many"). A Pareto chart can be used to decide which subset of problems should be solved first, or which problems deserve the most attention.

Once calculated and displayed graphically, it becomes a selling tool to the improvement team and management, raising the question why the team is focusing its energies on certain aspects of the problem.

Questions:

1. What is Pareto principle?
2. What is the importance of this in Marketing?
3. What is Pareto effect? Explain.
4. How is Pareto charts created?
5. What are the advantages of creating Pareto charts?
6. What are the difficulties associated with the creation of Pareto charts?

CHAPTER VI

ISHIKAWA DIAGRAMS

Objective:

The objective of this chapter is to understand:

- Origin of Ishikawa Diagrams
- Concept of Cause and Effect
- Steps involved in creating an Ishikawa Diagram

Introduction:

Ishikawa Diagrams are a graphical method for finding the most likely causes for an undesired effect. The method was first used by Kaoru Ishikawa in the 1960s.

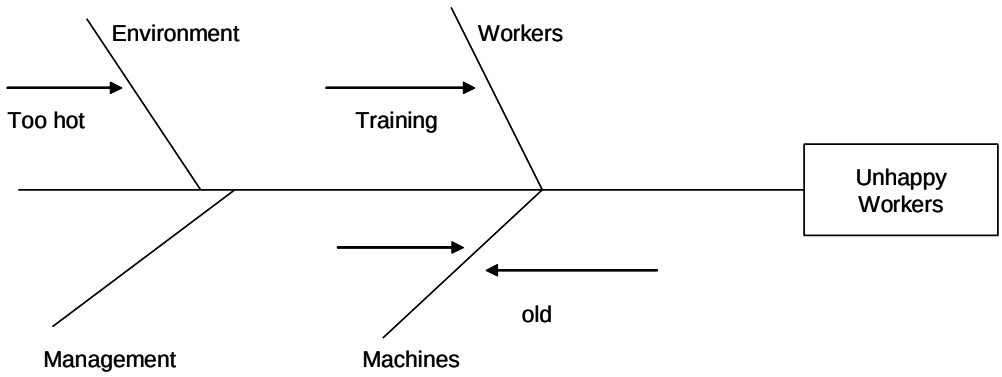
The Cause and Effect diagram also known as the "fishbone" or "Ishikawa" diagram after its creator Kaoru Ishikawa is used to systematically list all the different causes that can be attributed to a specific problem (or effect). A cause-and-effect diagram can help identify the reasons why a process goes out of control.

Because of its shape, it is also known as the fishbone diagram. Another name for this technique is: the cause-and-effect diagram. The fishbone diagram is a method/tool used in a root cause analysis.

The Ishikawa diagram is one of the seven basic tools of quality control, which include the histogram, Pareto chart, check sheet, control chart, cause-and-effect diagram, flowchart, and scatter diagram.

The purpose of this diagram is to arrive at a few key sources that contribute most significantly to the problem being examined. These sources are then targeted for improvement. The diagram also illustrates the relationships among the wide variety of possible contributors to the effect.

The figure below shows a simple Ishikawa diagram



The basic concept in the Cause-and-Effect diagram is that the name of a basic problem of interest is entered at the right of the diagram at the end of the main “bone". The main possible causes of the problem (the effect) are drawn as bones off of the main backbone. The "Four-M" categories are typically used as a starting point: "Materials", "Machines", "Manpower", and "Methods". Different names can be chosen to suit the problem at hand, or these general categories can be revised. The key is to have three to six main categories that encompass all possible influences. Brainstorming is typically done to add possible causes to the main "bones" and more specific causes to the "bones" on

the main "bones". This subdivision into ever increasing specificity continues as long as the problem areas can be further subdivided. The practical maximum depth of this tree is usually about four or five levels. When the fishbone is complete, one has a rather complete picture of all the possibilities about what could be the root cause for the designated problem.

The Cause-and-Effect diagram can be used by individuals or teams; probably most effectively by a group. A typical utilization is the drawing of a diagram on a blackboard by a team leader who first presents the main problem and asks for assistance from the group to determine the main causes which are subsequently drawn on the board as the main bones of the diagram. The team assists by making suggestions and, eventually, the entire cause and effect diagram is filled out. Once the entire fishbone is complete, team discussion takes place to decide on the most likely root causes of the problem. These causes are circled to indicate items that should be acted upon, and the use of the tool is complete.

How to Construct an Ishikawa Diagram:

- Place the main problem under investigation in a box on the right.
- Generate and clarify all the potential sources of variation.
- Use an affinity diagram to sort the process variables into naturally related groups. The labels of these groups are the names for the major bones on the Ishikawa diagram.
- The process variables are then placed on the appropriate bones of the Ishikawa diagram.
- Combine each bone in turn, insuring that the process variables are specific, measurable, and controllable. If they are not, branch or

"explode" the process variables until the ends of the branches are specific, measurable, and controllable.

The Ishikawa diagram, like most quality tools, is a visualization and knowledge organization tool. Simply collecting the ideas of a group in a systematic way facilitates the understanding and ultimate diagnosis of the problem. Several computer tools have been created for assisting in creating Ishikawa diagrams.

Summary:

Ishikawa Diagrams are a graphical method for finding the most likely causes for an undesired effect.

The Cause and Effect diagram also known as the "fishbone" or "Ishikawa" diagram after its creator Kaoru Ishikawa is used to systematically list all the different causes that can be attributed to a specific problem (or effect). A cause-and-effect diagram can help identify the reasons why a process goes out of control.

The Ishikawa diagram, like most quality tools, is a visualization and knowledge organization tool. Simply collecting the ideas of a group in a systematic way facilitates the understanding and ultimate diagnosis of the problem.

Questions:

1. What are Ishikawa diagrams?
2. Ishikawa diagrams are useful tools in Marketing. Explain.
3. What are the considerations for creating an Ishikawa diagram?
4. List out the steps involved in the creation of Ishikawa diagram.

References for this Unit:

1. Naresh K. Malhotra: MARKETING RESEARCH: AN APPLIED ORIENTATION, Pearson Education, Asia.
2. Aaker, Kumar & Day: MARKETING RESEARCH, John Wiley & Sons, 1998.
3. Paul E. Green & Donald S. Tull: RESEARCH FOR MARKETING DECISIONS.
4. Carol S. Aneshensel; University of California, Los Angeles; **UNIVARIATE ANALYSIS: CENTRAL TENDENCY, SPREAD, AND ASSOCIATIONS**
5. Cooley, W.W. and P. R. Lohnes. 1971. Multivariate Data Analysis. John Wiley & Sons, Inc.
6. George H. Dunteman (1984). Introduction to multivariate analysis. Thousand Oaks, CA: Sage Publications. Chapter 5 covers classification procedures and discriminant analysis.
7. Morrison, D.F. 1967. Multivariate Statistical Methods. McGraw-Hill: New York.
8. Overall, J.E. and C.J. Klett. 1972. Applied Multivariate Analysis. McGraw-Hill: New York.
9. Tabachnick, B.G. and L.S. Fidell. 1996. Using Multivariate Statistics. Harper Collins College Publishers: New York.