

Department of Distance and Continuing Education University of Delhi



Master of Business Administration (MBA)

Semester - II

Course Credit - 4.5

Core Course - MBAFT - 6207

MARKETING RESEARCH

Editorial Board

Prof. H. K. Dangi

Dr. Ameet Sao

Content Writers

Dr. Ameet Sao, Dr. Gurjeet Kaur, Dr. Supreet Kaur

Dr. Anupreet Kaur Mokha, Mr. Animesh Minz,

Ms. Shalu Garg, Mr. Jigmet Wangdus

Academic Coordinator

Mr. Deekshant Awasthi

© Department of Distance and Continuing Education

ISBN: 978-81-19169-13-9

1st edition: 2023

e-mail: ddceprinting@col.du.ac.in
management@col.du.ac.in

Published by:

Department of Distance and Continuing Education under
the aegis of Campus of Open Learning/School of Open Learning,
University of Delhi, Delhi-110 007

Printed by:

School of Open Learning, University of Delhi



DISCLAIMER

This book has been written for academic purposes only. Though every effort has been made to avoid errors yet any unintentional errors might have occurred . The authors ,the editors,the publisher and the distributor are not responsible for any action taken on the basis of this study module or its consequences thereof.



INDEX

LESSON – 1: INTRODUCTION TO MARKETING RESEARCH.....1

- 1.1 Learning Objectives
- 1.2 Introduction
- 1.3 Definition of Marketing Research
- 1.4 Need and Importance of Marketing research
- 1.5 Limitation of Marketing Research
- 1.6 Data Collection in Marketing Research
- 1.7 Scope of Marketing Research
- 1.8 Marketing Research Process
- 1.9 Formulation of Research Problem
- 1.10 Marketing Research Problems and Decisions
- 1.11 Research Design
- 1.12 Summary
- 1.13 Glossary
- 1.14 Answers to In-Text Questions
- 1.15 Self-Assessment Questions
- 1.16 References
- 1.17 Suggested Readings

LESSON-2: SCALING TECHNIQUES 24

- 2.1 Learning Objectives
- 2.2 Introduction
- 2.3 Measurement Scales
- 2.4 Scaling Techniques
- 2.5 Reliability and Validity
- 2.6 Summary
- 2.7 Glossary
- 2.8 Answers to In-text Questions
- 2.9 Self-Assessment Questions
- 2.10 References
- 2.11 Suggested Readings

**LESSON-3: DATA COLLECTION METHODS 48**

- 3.1 Learning Objectives
- 3.2 Introduction
- 3.3 Primary Data and Secondary Data
- 3.4 Sources of Secondary Data
- 3.5 Conducting Review of Literature
- 3.6 Quantitative Research and Qualitative Research
- 3.7 Summary
- 3.8 Glossary
- 3.9 Answers to In-text Questions
- 3.10 Self-Assessment Questions
- 3.11 References
- 3.12 Suggested Readings

LESSON-4: QUALITATIVE RESEARCH TOOLS.....72

- 4.1 Learning Objectives
- 4.1 Learning Objectives
- 4.2 Introduction
- 4.3 Qualitative Research Process
- 4.4 Advantages, Disadvantages and frequently asked questions
- 4.5 Frequently asked questions.
- 4.6 Summary
- 4.7 Glossary
- 4.8 Answers to In-text Questions
- 4.9 Self-Assessment Questions
- 4.10 References
- 4.11 Suggested Readings

LESSON-5: SAMPLE DESIGN.....96

- 5.1 Learning Objectives
- 5.2 Introduction
- 5.3 Steps in Sample Design
- 5.4 Characteristics of a good sample design
- 5.5 Selecting a Sampling Procedure



- 5.6 Sampling Design Method
- 5.7 Self - Assessment Questions
- 5.8 Self-Assessment Questions
- 5.9 Summary
- 5.10 Glossary
- 5.10 Answers to In-Text Questions
- 5.11 Answers to In-Text Questions
- 5.12 References
- 5.13 Suggested Readings

LESSON-6: SAMPLE SIZE DETERMINATION & DATA PREPARATION 115

- 6.1 Learning Objectives
- 6.2 Introduction
- 6.3 Sample Size Determination
- 6.4 Data Coding
- 6.5 Data Preparation
- 6.6 Tabulation and Graphical Analysis
- 6.7 Sampling Errors
- 6.8 Summary
- 6.9 Glossary
- 6.10 Answers to In-Text Questions
- 6.11 Self-Assessment Questions
- 6.12 References
- 6.13 Suggested Readings

LESSON-7: PARAMETRIC HYPOTHESIS TESTING AND ANOVA.....134

- 7.1 Learning Objectives
- 7.2 Introduction
- 7.3 Inferential Data Analysis
- 7.4 Characteristics and Types of Hypothesis
- 7.5 Parametric Hypothesis Testing
- 7.6 Testing hypothesis about a population mean using the z statistic (σ known)
- 7.7 Testing hypothesis about a population mean using the t statistic (σ unknown)



- 7.8 Testing hypothesis for the difference of means
- 7.9 Design of experiments
- 7.7 Glossary
- 7.8 Answers to In-text Questions
- 7.9 Self-Assessment Questions
- 7.10 References
- 7.11 Suggested Readings

LESSON-8: NON PARAMETRIC TEST.....162

- 8.1 Learning Objectives
- 8.2 Introduction
- 8.3 Chi-square test
- 8.4 Sign test
- 8.5 Mann-Whitney U test
- 8.6 Wilcoxon signed-rank test
- 8.7 Kruskal Wallis test
- 8.8 Summary
- 8.8 Answers to In-text Questions
- 8.9 Self-Assessment Questions
- 8.10 References
- 8.11 Suggested Readings

LESSON-9: REPORT WRITING.....200

- 9.1 Learning Objectives
- 9.2 Introduction
- 9.3 Importance of Report Writing
- 9.4 Types of Research Reports
- 9.5 Structure of Report
- 9.6 Guidelines for effective report writing
- 9.7 Summary
- 9.8 Glossary
- 9.9 In-Text Questions
- 9.10 Self-Assessment Questions
- 9.11 References
- 9.12 Suggested Readings



LESSON-10: MARKETING RESEARCH AND ITS APPLICATIONS.....219

- 10.1 Learning Objectives
- 10.2 Introduction
- 10.3 Applications of Marketing Research
- 10.4 Attitudes Measurement and Scaling Techniques
- 10.5 Challenges in Conducting Marketing Research in India
- 10.6 Ethical Issues in Marketing Research
- 10.7 Summary
- 10.8 Glossary
- 10.9 Answers to In-text Questions
- 10.10 Self-Assessment Questions
- 10.11 References
- 10.12 Suggested Readings



LESSON - 1

INTRODUCTION TO MARKETING RESEARCH

Dr. Ameet Sao

RICS SBE, Amity University
Email-Id.ameetsao@gmail.com

STRUCTURE

- 1.1 Learning Objectives
- 1.2 Introduction
- 1.3 Definition of Marketing Research
- 1.4 Need and Importance of Marketing research
- 1.5 Limitation of Marketing Research
- 1.6 Data Collection in Marketing Research
- 1.7 Scope of Marketing Research
- 1.8 Marketing Research Process
- 1.9 Formulation of Research Problem
- 1.10 Marketing Research Problems and Decisions
- 1.11 Research Design
- 1.12 Summary
- 1.13 Glossary
- 1.14 Answers to In-Text Questions
- 1.15 Self-Assessment Questions
- 1.16 References
- 1.17 Suggested Readings

1.1 LEARNING OBJECTIVES

- Understand the basic concepts and definitions of marketing research, including the research process, data collection methods, and types of research design.
- Develop the ability to identify and formulate research questions and hypotheses, and to design appropriate research methods to gather and analyze data.



- Learn how to analyse and interpret research data, and to communicate research findings effectively to different audiences.
- Understand the ethical considerations involved in conducting marketing research and how to apply them in real-world research projects.
- Understand the role of marketing research in the decision-making process, and how it can be used to inform strategic marketing decisions and improve overall business performance.

1.2 INTRODUCTION

It is a process used by companies to gather and analyse information about their target market, competitors, and industry in order to make informed decisions about their marketing strategies and tactics. This information can be used to gain a competitive advantage, improve customer satisfaction, and increase sales.

Marketing research is the process of gathering, analysing and interpreting data related to the market, consumers and competition to inform business decisions. It helps companies understand their target audience, identify trends and opportunities, and measure the effectiveness of their marketing efforts. The goal of marketing research is to provide businesses with the information they need to make informed decisions about product development, pricing, promotion, and distribution.

Marketing research can be divided into two categories: primary research and secondary research. Primary research is original research that is conducted specifically for a particular purpose, such as surveys, focus groups and interviews. Secondary research, on the other hand, is a research that has already been conducted and is available for use, such as industry reports, government statistics and online data.

One of the key benefits of marketing research is that it helps businesses gain a better understanding of their target audience. This can include information on demographics, purchasing habits, and preferences. By understanding their target audience, companies can develop more effective marketing strategies and create products and services that better meet the needs of their customers.

Marketing research can also be used to identify trends and opportunities in the market. For example, companies may use market research to identify new product or service opportunities, or to understand how changes in the economy or technology might affect their industry. Marketing research can also be used to measure the effectiveness of marketing



efforts. For example, companies may conduct research to evaluate the impact of a new advertising campaign or to understand how consumers responded to a new product launch. This information can then be used to make adjustments and improve future marketing efforts. In short, marketing research is a critical aspect for companies to gain insights about their target market, customers, and competitors. It enables the collection of information that can aid in creating strategies, developing products, and making informed decisions.

1.3 DEFINITION OF MARKETING RESEARCH

According to *American Marketing Association*, “Marketing Research is the function that links the consumer, customer and public to the marketer through information-information used to identify and define marketing opportunities and problems, generate, refine and evaluate marketing actions; monitor marketing performance; and improve understanding of marketing as a process.”

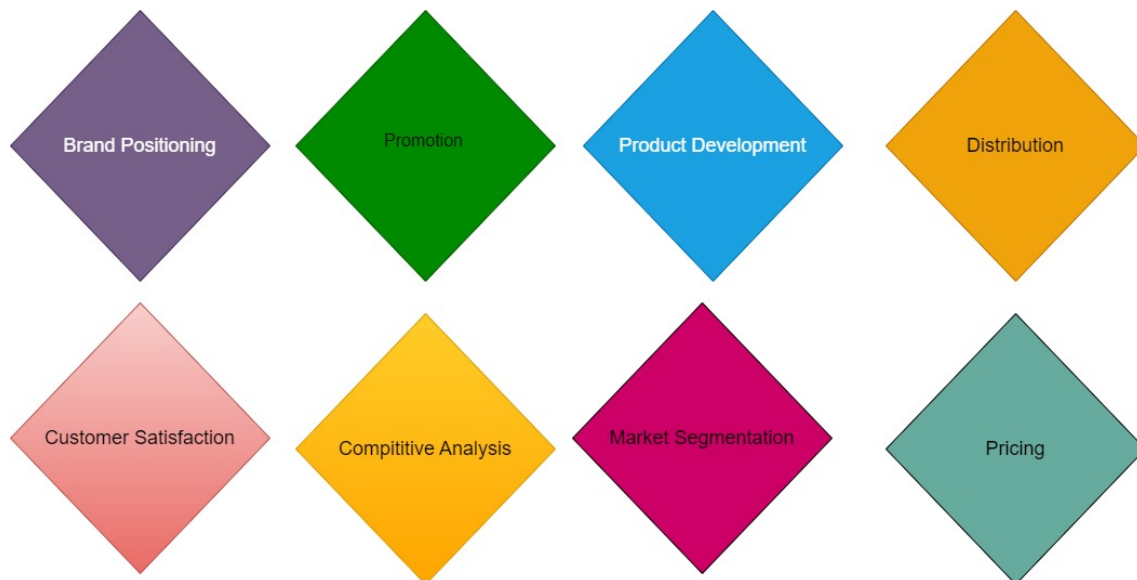
According to Philip Kotler defines marketing research as “Systematic problem analysis, model-building and fact-finding for the purpose of improved decision making and control in the marketing of goods and services.”

According to Green and Tull, “Marketing research is the systematic and objectives search for and analysis of information relevant to the identification and solution of any problem in the field of marketing.”

1.4 NEED AND IMPORTANCE OF MARKETING RESEARCH

Marketing research can be used for a variety of purposes, including:

1. **Product Development:** Companies can use marketing research to gather information about customer needs, preferences, and buying habits to inform the development of new products or services.
2. **Pricing:** Research can be used to understand consumer perceptions of pricing and to identify the price points that are most likely to be accepted by the target market.
3. **Promotion:** Research can be used to understand consumer attitudes and perceptions towards advertising, public relations, and other forms of promotion, and to evaluate the effectiveness of different marketing tactics.



4. **Distribution:** Research can be used to understand consumer preferences for different channels of distribution, such as online, brick-and-mortar, or direct-to-consumer, and to identify the most effective distribution strategies.
5. **Brand positioning:** Research will assist in understanding consumer perceptions of different brands and to identify key attributes that consumers associate with a brand.
6. **Market Segmentation:** Marketing research can be used to identify different segments of the market and to understand the needs and behaviors of each segment.
7. **Competitive Analysis:** Research will help to gather information about competitors and to understand their strengths, weaknesses, and market strategies.
8. **Customer satisfaction:** Research helps to measure customer satisfaction with products or services and to identify areas where improvements can be made.

Overall, marketing research provides valuable insights that can inform a wide range of business decisions, from product development and pricing to promotion and distribution. By using research to gain a deeper understanding of the market, customers, and competition, companies can make more informed decisions and achieve better business results.



IN-TEXT QUESTIONS

1. What is the primary purpose of marketing research?
 - a) To identify new market opportunities
 - b) To increase sales
 - c) To improve product design
 - d) To reduce costs
2. Why is marketing research important for businesses?
 - a) It helps businesses make informed decisions,
 - b) It helps businesses
 - c) It helps businesses understand their competitors
 - d) All of the aboveincrease their profits
3. Which of the following is NOT an example of a common marketing research method?
 - a) Surveys
 - b) Focus groups
 - c) Telenathv
 - d) Exneriments

1.5 LIMITATIONS OF MARKETING RESEARCH

Limited generalizability: The results of a marketing research study may not be generalizable to the entire population, as the sample size and selection methods can affect the study's representativeness.

Limited accuracy: Marketing research relies on self-reported data from participants, which may not always be accurate or reliable.

Limited control over external factors: External factors such as economic conditions, technological advances, and political and legal changes can impact the results of a marketing research study, limiting the ability of researchers to control for these factors.

Limited understanding of consumer behavior: Consumers are often not aware of their own behavior and motivations, making it difficult to fully understand their needs and preferences through research.

Time-consuming and expensive: Marketing research can be a time-consuming and expensive process, requiring significant resources to conduct and analyze data.

Limited ability to predict future trends: While research can provide insight into current market conditions and consumer behavior, it may not be able to accurately predict future trends.



Limited ability to measure attitudes and behaviors that are difficult to quantify: Some attitudes and behaviors, such as emotions and perceptions, are difficult to measure and quantify through research.

Limited ability to capture the complexity of consumer decision-making: The consumer decision-making process is complex and multifaceted, and research may not be able to fully capture all of the factors that influence it.

Limited ability to address ethical concerns: Marketing research may raise ethical concerns, such as participant privacy and confidentiality, and researchers must be aware of these concerns and take steps to address them.

1.6 DATA COLLECTION IN MARKETING RESEARCH

Data collection in marketing research is a thorough process in which the researcher conducts a planned search for all relevant information. There are two types of data: primary and secondary.

- (i) Primary data is original information that is collected specifically for the research being conducted. It is gathered to address the specific problem at hand.
- (ii) Secondary data, on the other hand, is information that has already been collected and is readily available from external sources. This type of data is typically less expensive and quicker to obtain than primary data, and may be the only option when primary data cannot be collected.

Data Collection Methods

- 1. Qualitative research** - This method is used to gain an initial understanding of a problem or topic. Unlike quantitative research, it is non-statistical and uses an inductive approach to gather data. Data is collected and grouped into meaningful categories, and the researcher develops explanations from the data itself. This type of research is often used in exploratory and descriptive studies. The data collected in qualitative research can come in various forms, such as interview transcripts, documents, diaries, and notes made during observation. There are two primary methods for collecting qualitative data:
 - a. Direct Collection Method-** This method involves collecting data directly from the source, often using disguised methods to ensure that the purpose of the research is not



known to the participants. This method can include techniques such as participant observation, case-study, in-depth interviews, and focus groups.

Focus groups involve bringing together a small group of individuals, typically six to twelve, and one or two moderators who lead the discussion. The moderators may adopt opposite positions to encourage diverse perspectives. The group discussion is guided and controlled by the moderators, and the group is often observed from an adjacent room. The discussion is recorded through various means. The objectives of focus group research include: obtaining primary information for a research project, developing survey questions, understanding the reasons behind a specific phenomenon, observing how people interpret certain phenomena, and testing initial ideas or plans.

The process of conducting a focus group research includes the following steps:

- (i) Clearly defining the research problem.
- (ii) Identifying and selecting a suitable sample of participants.
- (iii) Determining the number of groups needed for the research, usually a minimum of two groups.
- (iv) Planning and organizing the logistics of the study, including arranging a location for the focus group to meet.
- (v) Choosing and training moderators, who will lead the discussion.
- (vi) Preparing materials for the focus group, such as discussion guides or stimuli.
- (vii) Facilitating the focus group session.
- (viii) Analyzing the data collected and preparing a summary report of the findings.

Focus groups offer several advantages in research, including:

- The ability to gather primary information and conduct pilot studies.
- A relatively low cost compared to other research methods.
- The ability to be conducted in a relatively short amount of time.
- Flexibility in terms of the discussion topic and the ability to explore multiple angles.



- Moderators can observe and interpret nonverbal cues and body language of participants, which can provide insight into their thoughts and feelings.
- The ability to gather additional information through pre- or post-discussion questionnaires.

Focus group research also has some limitations, including:

- It is not suitable for collecting quantitative data.
- The presence of a dominant group member can influence the opinions of others in the group. Moderators may also inadvertently restrict the participation of certain group members.
- The success of the focus group relies heavily on the skills and abilities of the moderator.
- The participants in the focus group may not be representative of the larger population being studied.
- The use of recording equipment can make participants uncomfortable and limit their willingness to share their thoughts and opinions. The location of the equipment during the focus group also should be considered.

b. Indirect collection-Method -

Indirect data collection in research refers to obtaining information from sources that are not directly collected from the research subjects. This can include secondary data such as previously gathered documents, reports, and databases. These data were not collected for the current research project but can provide useful context and background information. This method is typically used in addition to direct data collection methods and can include techniques such as reviewing existing literature, analyzing content, and conducting historical research.

1.7 SCOPE OF MARKETING RESEARCH

Marketing research is the process of gathering, analyzing, and interpreting data related to a company's products, customers, and competition. The scope of marketing research includes:

Market analysis: This involves researching the size and growth of the market, as well as identifying trends and segmenting the market.



Product research: This involves researching the features and benefits of a product, as well as testing new products and getting feedback from customers.

Consumer research: This involves researching consumer needs, preferences, and purchasing behavior.

Competitive research: This involves researching the competition, including their products, pricing, and marketing strategies.

Distribution research: This involves researching the best ways to distribute and promote a product.

Advertising research: This involves researching the effectiveness of different advertising campaigns and media.

Sales research: This involves researching the effectiveness of sales promotions and strategies.

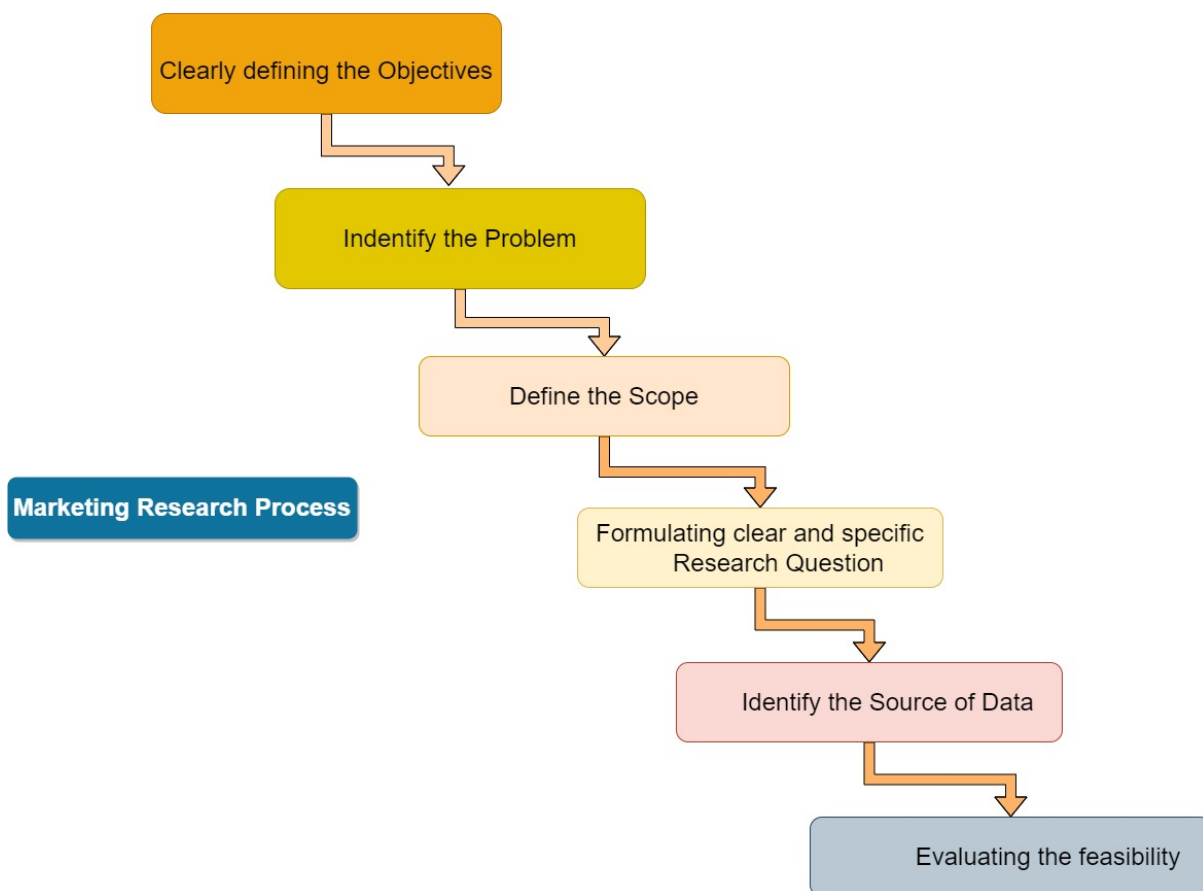
Social media research: This involves researching the impact of social media on consumer behavior and purchasing decisions.

Marketing research helps companies make informed decisions about product development, pricing, promotion, and distribution.

It helps companies to understand their target market and identify opportunities for growth, and it can also help them to identify and address problems with their current products and strategies.

1.8 MARKETING RESEARCH PROCESS

In the marketing research process, the formulation of a research problem is a crucial step. It entails identifying and defining the specific issue or question that the research will focus on. The research problem should be clear, specific and pertinent to the organization's decision-making needs. Some of the key steps in formulating a research problem in marketing research include:



Clearly defining the research objective, which should be a statement of the specific information that is needed to address the decision-making problem.

Identifying the decision-making problem, which should be directly related to a specific decision that the organization needs to make.

Defining the scope of the research, in terms of population, sample and time frame.

Formulating specific, clear and relevant research questions.

Developing a research hypothesis, a statement of expectations or predictions about the relationship between variables that can be tested through the research.

Identifying the data required and assessing its availability.

Evaluating the feasibility of the research, including the availability of resources, data, time, and ethical considerations.



Overall, formulating a research problem in marketing research should be a comprehensive and systematic process that is closely tied to the organization's decision-making needs. By clearly defining the research objective, problem, scope, questions, and hypothesis, researchers can ensure that the research is focused, relevant, and can provide actionable insights.

1.9 FORMULATION OF RESEARCH PROBLEM

Formulation of research Problem

The formulation of a research problem is an important step in the marketing research process. It involves identifying and defining the specific issue or question that the research will address. The research problem should be clear, specific, and relevant to the organization's decision-making needs. The following are some key steps in the formulation of a research problem in marketing research:

1. **Define the research objective:** The research objective should be a clear statement of the specific information that is needed to address the decision-making problem.
2. **Identify the decision-making problem:** The research problem should be directly related to a specific decision that the organization needs to make, such as launching a new product, determining a pricing strategy, or understanding consumer behavior.
3. **Determine the scope of the research:** The scope of the research should be defined in terms of the population, sample, and time frame of the study.
4. **Formulate research questions:** Research questions should be specific, clear, and relevant to the research objective and decision-making problem.
5. **Develop a research hypothesis:** A research hypothesis is a statement of expectations or predictions about the relationship between variables that can be tested through the research.
6. **Determine the data required:** The research problem should be formulated in a way that it can be answered by the data that is available or can be collected.
7. **Assess the feasibility of the research:** Feasibility assessment should include the availability of resources, data, and time, as well as ethical considerations.

Overall, the formulation of a research problem in marketing research should be a thorough and systematic process that is closely aligned with the organization's decision-making needs.



By clearly defining the research objective, problem, scope, questions, and hypothesis, researchers can ensure that the research is focused, relevant, and can provide actionable insights.

1.10 MARKETING RESEARCH PROBLEMS AND DECISION

Marketing research is the process of gathering information about the market, customers, competitors, and other factors that can impact a company's marketing strategy. Marketing research problems are the issues that need to be addressed in the research process to help a company make informed decisions. Some common marketing research problems include:

Defining the research problem: One of the biggest challenges in marketing research is defining the research problem. This involves understanding what information the company needs to make informed decisions and how to gather that information.

Identifying the target market: Identifying the target market is essential in marketing research. This involves understanding who the company's customers are, what they need and want, and how they make purchasing decisions.

Collecting and analyzing data: Collecting and analyzing data is a crucial part of marketing research. This involves choosing the right data collection method and analyzing the data to gain insights into the market and customers.

Validating research results: Validating research results is essential to ensure that the research is accurate and reliable. This involves testing research hypotheses and ensuring that the research findings are valid.

Marketing Management Decision Problem:

Marketing management decision problem is the issue that needs to be addressed to make a marketing decision. It is a problem or opportunity that requires a marketing decision or action. Some common marketing management decision problems include:

Product development: Product development is a common marketing management decision problem. It involves deciding what new products to develop, what features to include, and how to price and promote them.

Pricing strategy: Pricing strategy is another important marketing management decision problem. It involves deciding on the price of products and services based on market demand, competition, and production costs.



Promotion strategy: Promotion strategy is the process of communicating the value of a product or service to customers. It involves deciding on the best way to promote a product, such as through advertising, public relations, or sales promotions.

Distribution strategy: Distribution strategy involves deciding how to get products to customers. This includes choosing the best channels of distribution, such as wholesalers, retailers, or direct sales.

Example:

Suppose a company wants to enter a new market and wants to conduct marketing research to make informed decisions. The company faces the following marketing research problems:

Defining the research problem: The company needs to determine what information it needs to make informed decisions about entering the new market.

Identifying the target market: The company needs to identify the target market in the new market and understand their needs, wants, and purchasing behavior.

Collecting and analyzing data: The company needs to collect and analyze data on the new market, such as market size, competitors, and consumer preferences.

Validating research results: The company needs to ensure that the research results are accurate and reliable.

Once the company has gathered the necessary information, it can address the following marketing management decision problems:

Product development: The company needs to develop a product that meets the needs of the target market in the new market.

Pricing strategy: The company needs to determine the price of the product based on market demand, competition, and production costs.

Promotion strategy: The company needs to determine the best way to promote the product to the target market in the new market.

Distribution strategy: The company needs to determine the best distribution channels to get the product to the target market in the new market.

Let's take another example - A company wants to introduce a new line of organic skincare products in India. The marketing research problem is to understand the demand for organic skincare products in India, the target market's preferences, and the factors that influence purchase decisions.



Based on the research findings, the company needs to make the following decisions:

Product development: Develop a range of organic skincare products that meets the preferences and requirements of the target market in India.

Pricing strategy: Determine the price of the product that is competitive and attractive to the target market in India.

Promotion strategy: Determine the best channels and tactics to promote the organic skincare products to the target market in India.

Distribution strategy: Determine the best distribution channels to get the product to the target market in India.

Example 3:

A hotel chain wants to expand its business in India by opening a new branch in a new city. The marketing research problem is to understand the target market's preferences, demand, and competition in the new city.

Based on the research findings, the hotel chain needs to make the following decisions:

Product development: Develop a hotel experience that meets the preferences and requirements of the target market in the new city.

Pricing strategy: Determine the price of the hotel that is competitive and attractive to the target market in the new city.

Promotion strategy: Determine the best channels and tactics to promote the hotel to the target market in the new city.

Distribution strategy: Determine the best distribution channels to reach the target market in the new city, such as through online travel agencies, travel agents, or direct booking.

1.11 RESEARCH DESIGN

Research design is the overall plan or strategy for conducting research. It outlines the methods and procedures that will be used to collect and analyze data and addresses the key elements of the research, such as the research questions, population, sample, data collection methods, and data analysis techniques. A well-designed research plan will ensure that the research is conducted in a systematic and rigorous manner, and that the results are valid and reliable.

There are several different types of research designs, including:



- **Experimental design:** This type of design involves manipulating one or more variables to observe the effect on another variable. It is used to establish cause-and-effect relationships.
- **Quasi-experimental design:** This type of design is similar to experimental design but does not include random assignment of participants to groups.
- **Non-experimental design:** This type of design does not involve manipulation of variables and is used to explore relationships between variables or to describe a particular phenomenon.
- **Descriptive design:** This type of design is used to gather information about the characteristics of a particular group or population.
- **Correlational design:** This type of design is used to explore relationships between variables without manipulating any variables.
- **Longitudinal design:** This type of design involves collecting data over an extended period of time, typically to study changes or developments over time.

Each type of research design has its own strengths and weaknesses and the choice of design should depend on the research question, the availability of resources, time, and the ethical considerations.

The major 3 types of research are – Exploratory Research, Descriptive Research and Conclusive Research.

- (a) **Exploratory research** is a type of research that is conducted to gain a general understanding of a topic or problem, and to generate new ideas and insights. It is often used in situations where little is known about a topic, or when the research question is not well-defined. Exploratory research is typically conducted using qualitative methods, such as focus groups, interviews, and observation.

In marketing, exploratory research is often used to:

- Identify new opportunities and ideas for products or services
- Understand consumer behavior and attitudes
- Understand the competitive landscape
- Identify market segments and target audiences
- Understand the feasibility of new marketing strategies



The importance of exploratory research in marketing lies in its ability to generate new ideas and insights that can inform strategy development and decision making. It helps companies to identify gaps in their current understanding of the market and consumer behavior and to explore new possibilities.

Exploratory research allows companies to take a more flexible and open-minded approach to research, rather than being tied to a specific research question or hypothesis. It can also be less time-consuming and less costly than other forms of research, making it an efficient way to gather information in the early stages of a research project.

In summary, exploratory research is important in marketing research as it enables companies to gain a deeper understanding of the market, consumer behavior and competition, thus providing a solid foundation for informed decision making and strategy development.

Example: Exploring Consumer Attitudes Towards a New Product

A company is considering launching a new line of plant-based protein bars, but before doing so, they want to conduct exploratory research to gain a better understanding of consumer attitudes towards the product. The company's objective is to identify any potential issues or concerns that consumers might have, as well as uncover any opportunities for differentiation in the market.

The company decides to use focus groups as the method for conducting the exploratory research. The focus groups consist of 8-10 participants who are representative of the target market for the plant-based protein bars. The focus groups are moderated by a trained facilitator who encourages participants to share their thoughts, feelings, and experiences related to the product.

During the focus groups, participants are asked a series of open-ended questions designed to explore their attitudes towards the product. Some of the questions might include:

What do you think of the idea of a plant-based protein bar?

What do you think are the advantages and disadvantages of plant-based protein bars compared to other types of protein bars?

What are your concerns or hesitations about trying a plant-based protein bar?

How do you currently use protein bars in your diet, and how might a plant-based protein bar fit into your routine?

What features or qualities would you look for in a plant-based protein bar?



After conducting several focus groups, the company analyzes the data to identify themes and patterns in the responses. Some of the insights that emerge might include:

Consumers are generally positive about the idea of plant-based protein bars, but they are concerned about taste and texture.

Consumers are interested in plant-based protein bars as a healthy snack or meal replacement, but they want to see clear nutritional information and ingredient lists.

Consumers are interested in trying plant-based protein bars as a way to reduce their environmental impact, but they want to see evidence of sustainability in the product's production and packaging.

Based on these insights, the company might make decisions about the product's flavor profile, packaging design, marketing messages, and target audience. The exploratory research has helped the company gain a better understanding of consumer attitudes towards the product and identify opportunities for differentiation in the market.

- (b) Descriptive research** - is a type of research that is conducted to describe and understand the characteristics of a particular group or population. It is used to gather detailed information about the characteristics, behavior, or attitudes of a specific group, and to provide a detailed picture of the population being studied. Descriptive research is typically conducted using quantitative methods, such as surveys, questionnaires, and experiments. **Descriptive research is a type of research that is used to describe a phenomenon or a group of phenomena in a systematic way. The objective of descriptive research is to provide a detailed and accurate account of a particular phenomenon or group, without attempting to explain why it occurs.**

Descriptive research is often used in social sciences and business research to collect and analyze data on attitudes, behaviors, demographics, and other variables. The data collected in descriptive research is usually quantitative in nature, and it is often gathered through surveys, questionnaires, and other forms of standardized data collection methods.

Descriptive research can provide valuable insights into a particular phenomenon or group, including identifying patterns, trends, and associations between variables. This type of research is particularly useful when the researcher is trying to gain a comprehensive understanding of a particular phenomenon or group, without the need to establish cause-and-effect relationships.



Example 1: Retail Shopping Behavior Study

A retail company wants to conduct a descriptive research study to understand the shopping behavior of its customers. The research objective is to gather information about the frequency and types of purchases made by customers, the factors that influence their purchasing decisions, and their shopping habits.

The research method used is a survey. The company distributes a questionnaire to its customers asking them questions about their shopping behavior, such as:

How often do you shop at our store?

What products do you usually buy from our store?

What factors influence your decision to buy from our store?

Do you prefer to shop online or in-store?

How much time do you spend in our store during each visit?

The company collects the survey responses and analyzes the data using statistical methods to identify patterns and trends in the shopping behavior of its customers. The research findings can be used to improve the store's layout, product selection, marketing strategy, and customer experience to better meet the needs and preferences of its customers.

Example 2: Employee Satisfaction Survey

A company wants to conduct a descriptive research study to measure the job satisfaction levels of its employees. The research objective is to gather information about employees' perceptions of their job duties, work environment, salary, benefits, and management.

The research method used is an online survey. The company sends a questionnaire to all employees asking them to rate their level of satisfaction with various aspects of their job on a scale of 1-5. The questionnaire includes questions such as:

How satisfied are you with your current job duties?

How satisfied are you with your salary and benefits package?

How satisfied are you with the working conditions and work environment?

How satisfied are you with the communication and support provided by management?

The company collects the survey responses and analyzes the data using statistical methods to identify areas where employee satisfaction is high or low. The research findings can be used to develop strategies to improve employee satisfaction, such as implementing changes to job



duties, increasing salary and benefits, improving working conditions, and providing better communication and support from management.

Conclusive Research

Conclusive research, also known as conclusive or conclusive survey research, is a type of research that is conducted to provide specific and definitive answers to a research question. This type of research is designed to provide conclusive evidence about a particular phenomenon or to test a specific hypothesis.

Conclusive research typically involves the use of quantitative research methods, such as surveys, experiments, and observational studies. The data collected is analyzed using statistical methods, and the results are used to draw valid conclusions about the phenomenon being studied.

Conclusive research is used in marketing to:

- Test the effectiveness of marketing campaigns
- Measure customer satisfaction
- Establish cause-and-effect relationships between variables
- Understand consumer behavior and attitudes

Conclusive research is important in marketing research as it allows companies to test the effectiveness of marketing campaigns and measure customer satisfaction, providing specific and definitive answers to research questions. This type of research also allows companies to establish cause-and-effect relationships between variables, which can be used to inform strategy development and decision making.

Conclusive research is characterized by a high level of control over variables and a large sample size, which increases the generalizability of the research results. However, it can be costly and time-consuming, and may not be feasible for all research projects.

In summary, conclusive research is a type of research that is conducted to provide specific and definitive answers to a research question. This type of research is used to test the effectiveness of marketing campaigns, measure customer satisfaction, and establish cause-and-effect relationships between variables. It is characterized by a high level of control over variables and a large sample size, but it can be costly and time-consuming.

Case study on Exploratory Research

A small clothing retailer wants to expand its business by introducing a new line of clothing aimed at a younger demographic. The company currently has little information about this



target market and wants to conduct some exploratory research to gain a better understanding of their needs and preferences.

The company decides to conduct focus groups with a sample of young adults aged 18-25. The focus groups are moderated by a trained facilitator, and participants are asked a series of open-ended questions about their clothing preferences, shopping habits, and lifestyle.

The company also conducts in-depth interviews with a smaller sample of young adults to gather more detailed information about their clothing preferences and shopping habits.

The company also conduct ethnographic observation in the malls, shopping centers and other places where the target market hangs out, to understand their behavior and purchasing habits in a natural setting.

The data collected from the focus groups, interviews, and observations is analyzed qualitatively to identify common themes and patterns. The company discovers that the target market is highly influenced by social media and values sustainable and ethically-made clothing. They also prefer a mix of trendy and classic styles and are willing to pay more for quality and unique designs.

Based on these findings, the company decides to develop a line of sustainable and ethically-made clothing that incorporates trendy and classic styles, and to heavily promote the line on social media. They also set a slightly higher price point for the new line, as they believe the target market is willing to pay more for quality and unique designs.

The exploratory research conducted by the company provided valuable insights into the preferences and behavior of their target market, and helped inform the development of a new product line that is well-aligned with the needs and preferences of the target market.

Questions on Exploratory Research for exercise

1. What are the main pain points and unmet needs of consumers in the organic food industry?
2. How do consumers perceive and evaluate the sustainability of different e-commerce platforms?
3. What are the key factors that influence purchasing decisions of consumers in the luxury car market?



1.12 IN TEXT QUESTIONS

4. What is the first step in the marketing research process? a) Product development b) Market research c) Pricing strategy d) Promotion
5. What type of research is used to gather information about consumer demographics, behaviours, and attitudes? a) Exploratory research b) Descriptive research c) Conclusive research d) Experimental research
6. What type of research is used to test hypotheses and make causal inferences? a) Exploratory research b) Descriptive research c) Conclusive research d) Experimental research

1.13 SELF-ASSESSMENT QUESTIONS

1. What is the role of segmentation in the marketing process, and how can it be used to develop effective targeting and positioning strategies?
2. How do companies use consumer behavior insights in the marketing process to develop effective promotional campaigns and pricing strategies?
3. How can companies effectively measure the success of their marketing efforts in the different stages of the marketing process, and how can they use this information to make adjustments and improvements to their strategies?
4. How can descriptive research be used in the marketing process to gather data on consumer demographics, behaviours, and attitudes, and how can this information be used to develop effective marketing strategies?
5. Describe the steps involved in conducting a descriptive research study, including the design of the research, data collection methods, and data analysis techniques. Explain the potential limitations and challenges of using descriptive research and discuss how these can be overcome.



1.14 GLOSSARY

Marketing research: The process of gathering, analysing, and interpreting information about a market, product, or service.

Primary research: Research conducted directly by the researcher, such as surveys, interviews, and experiments.

Secondary research: Research that has already been conducted by others, such as published reports and statistics.

Research design: The plan or strategy for conducting research, including the methods and techniques that will be used.

Data collection: The process of gathering information for research, such as through surveys, interviews, and observations.

Data analysis: The process of interpreting and making sense of the data collected in research.

Conclusion: The summary of findings, results, and recommendations that are drawn from the research.

1.15 ANSWERS TO INTEXT QUESTIONS

1 – A, 2-D, 3-C, 4-B, 5-B, 6-C

1.16 REFERENCES

Saunders, M., Lewis, P., & Thornhill, A. (2015). Research methods for business students (7th ed.). Harlow, England: Pearson Education.

Malhotra, N. K. (2017). Marketing research: An applied orientation (7th ed.). Pearson.

Cooper, D. R., & Schindler, P. S. (2018). Business research methods (13th ed.). McGraw-Hill Education.

Hair, J. F., Money, A. H., Samouel, P., & Page, M. J. (2017). Research methods for business: A skill-building approach (7th ed.). John Wiley & Sons.

Zikmund, W. G. (2003). Business research methods (7th ed.). Thomson South-Western.



1.17 SUGGESTED READINGS

- Srinivasan, R. (2015). Marketing research: An Indian perspective (6th ed.). New Delhi, India: Tata McGraw-Hill Education.
- Malhotra, N. K. (2017). Marketing research: Text and cases (6th ed.). New Delhi, India: Tata McGraw-Hill Education.
- Green, P. E., & Tull, D. S. (2018). Research for marketing decisions (10th ed.). Upper Saddle River, NJ: Prentice Hall.
- Malhotra, N. (2018). Marketing research: An applied orientation (7th ed.). Harlow, England: Pearson Education.



LESSON-2

SCALING TECHNIQUES

Dr. Anupreet Kaur Mokha
Assistant Professor
SGTB Khalsa College
University of Delhi
anupreetmokha.6@gmail.com

STRUCTURE

- 2.1 Learning Objectives
- 2.2 Introduction
- 2.3 Measurement Scales
 - 2.3.1 Scale Characteristics and Levels of Measurement
 - 2.3.2 Types of Measurement Scales
 - 2.3.3 Comparison of Measurement Scales
- 2.4 Scaling Techniques
 - 2.4.1 Comparative Scaling
 - 2.4.2 Non-Comparative Scaling
 - 2.4.3 Selection of Measurement Scale
- 2.5 Reliability and Validity
- 2.6 Summary
- 2.7 Glossary
- 2.8 Answers to In-text Questions
- 2.9 Self-Assessment Questions
- 2.10 References
- 2.11 Suggested Readings

2.1 LEARNING OBJECTIVES

After reading this lesson, you should be able:

1. To define the concepts of measurement and scaling.



2. To understand the attributes of description, order, distance and origin and how they define the level of measurement in a scale;
3. To discuss and compare the types of measurement scales- nominal, ordinal, interval and ratio;
4. To classify and discuss various types of scaling techniques- comparative and non-comparative.
5. To understand how to assess the reliability and validity of scales.

2.2 INTRODUCTION

In the previous lesson, you understood the concept of marketing research and its process and also understood the significance and various types of research design- exploratory, descriptive and conclusive research. Once you have a clear knowledge of what you wish to ask your target participants, you should consider the concepts of measurement and scaling.

This lesson provides an overview of measurement and scaling. It discusses the four types of measurement scales - nominal, ordinal, interval and ratio and also provides a comparison between them in terms of attributes, mathematical operations and statistical techniques used. The next part of the lesson briefly explains the types of scaling techniques- comparative and non-comparative with the illustrated examples. The comparative techniques comprise paired comparison, constant sum, rank order and Q-sort scaling, whereas, non-comparative techniques comprise continuous and itemized rating scales (further classified into Likert, semantic differential and Stapel scales). Later on, this lesson also discusses the reliability and validity of a scale.

This lesson will help the learners to develop their questionnaire which is the most significant research tool and will help them to understand how to measure the data. The scales will help the learners to put their feelings, opinions and thoughts into a measurable form.

2.3 MEASUREMENT SCALES

Measurement is the process of assigning numbers or symbols to attributes of objects based on predefined rules. Objects include tangible items (such as consumers, employees, mobile phones, books, automobiles, etc) and intangible items (such as attitudes, social pressure, etc). Attributes of objects include physical attributes (height, weight, etc), social attributes (social status, leadership qualities, etc) and psychological attributes (attitudes, opinions, preferences, intelligence, etc). It is worth noting that we are not measuring the object here but rather some



attributes of the object. If we want to measure consumers, we can measure only their preferences, attitudes, perceptions or other relevant attributes. For example: “*How satisfied are you with the product?*”. This will help the marketers to measure the satisfaction level of the consumers towards their product i.e. whether the consumers are satisfied or not. In marketing research, numbers are typically assigned due to one of two reasons: (i) numbers allow for statistical analysis of the data collected, and (ii) numbers make it possible to communicate measurement rules and results in a universal and transparent manner.

Scaling is the process of establishing a continuum on which measured objects are placed. It is an extension of measurement. For example, consider a scale for categorizing consumers based on their satisfaction level towards any particular product, service or brand. Each participant is required to assign a number from 1 to 5 indicating highly dissatisfied (1), dissatisfied (2), neutral (3), satisfied (4) and highly satisfied (5). So, here, the measurement means the actual assignment of numbers (1 to 5) to each participant whereas scaling is the process of categorizing participants’ on a continuum concerning their satisfaction level towards any particular product, service or brand. In this example, scaling refers to the process by which the participants would be categorized as highly dissatisfied, dissatisfied, neutral, satisfied and highly satisfied.

2.3.1 Scale Attributes and Levels of Measurement:

The scales used in marketing research can be categorized in terms of four fundamental attributes i.e. **description, order, distance and origin**. These attributes, collectively, are termed as the scale levels of measurement. The **scale level of measurement** indicates which attributes of an object the scale is measuring or not measuring. To comprehend the types of measurement scales, one must first have a basic understanding of scale attributes.

Description

Numbers can be used as labels or identities to describe each value of scale. For example, to describe gender on a scale, we can label it as follows: 1. Female and 2. Male. Here, label 1(female) is not lesser than label 2 (male) or label 2 (male) is not greater than label 1 (female). Here, they are just representing the description. All the levels of measurement scales must possess this attribute of description i.e. all the scales must have a unique label or identity in order to define the scale item.

Order

Order represents the relative position or sizes of the descriptors such as “less than”, “equal to” and “greater than”. For example, to rank professors in a college on a scale, we can label as follows: 1. Professor, 2. Associate Professor and 3. Assistant Professor. Here, Rank 1



(Professor) is higher than Rank 2 (Associate Professor). Similarly, Rank 2 (Associate Professor) is higher than Rank 1 (Assistant Professor). All the levels of measurement scales do not possess this attribute of origin. Like in gender, we cannot say whether a male is greater or less than a female. Hence, the gender scale does not possess the attribute of order.

Distance

Distance indicates that absolute differences between the scale descriptors are known and may be represented in units. The distance between any pair of numbers is less than, equal to or greater than the distance between any other pair of numbers. For example, the wall clock moves between 12 o'clock to 12 o'clock without any absolute zero as time cannot be zero. The time between 5 p.m and 7 p.m is equal to the time between 8 p.m and 10 p.m. It is worth noting that distance possesses the attribute of order also. There is an order which is followed in the case of wall-clock i.e. 6 a.m is greater than 5 a.m and similarly, 5 a.m is greater than 4 a.m. Hence, distance indicates order but the reverse is not possible. In other words, the scale that has distance also has order and description.

Origin

Origin indicates that the scale has an absolute true zero point. For example, income can be zero which means that there is no income at all. Also, the origin attribute indicates that meaningful information can be revealed by dividing or taking the ratio of two numbers. So, we can say an income of Rs 40,000 is twice of income of Rs 20,000. Hence, income possesses the attribute of origin. The scale which has the origin attribute will also have distance, order and description.

You may have noticed that all the four fundamental attributes i.e. description, order, distance and origin signify successively higher-level attributes with the description being the lowest scale attribute and origin being the highest scale attribute. All the levels of measurement scales possess the attribute of description. A scale that has an order also has a description. A scale that has distance also has order and description. Similarly, a scale that has an origin also has distance, order and description. A scale, therefore, has all of the lower-level attributes if it possesses a higher-level attribute. However, the opposite may not always be true, i.e. a scale that has a lower-level attribute may or may not possess a higher-level attribute. Now that we have understood the scale attributes, we can talk about the types of measurement scales.

2.3.2 Types of Measurement Scales:

Measurement scales are classified into four types: **nominal, ordinal, interval and ratio** as depicted in Fig. 2.1. Let us discuss them one by one.



Nominal Scales

A nominal scales only possess the attribute of description. It is used to describe or label any object. In this, each number is allocated to only one object, and also each object has only one number allocated to it, hence, one-to-one correspondence should be there between the numbers and objects. The variables measured on a nominal scale are usually categorical variables that are converted into numeric values for analyzing the data. For example, in the case of gender, we can assign 1 as Female and 2 as Male. These numbers have no quantitative values. Other examples include nationality, religion, season, marital status, etc. In marketing research, nominal scales are commonly used to identify individuals, attributes, brands, and other objects.

Ordinal Scales

An ordinal scales possess the attribute of 'order' in which numbers are assigned to objects representing the relative position or sizes of the descriptors. It does not depict the absolute differences between the objects. The ordinal scale only enables you to determine whether an object has a greater value, equal value or less value than some other object, but does not reveal the magnitude of differences. Hence, the ordinal scale represents the ranking order only. For example, preference ranking, quality rankings, occupational status, etc. In marketing research, ordinal scales are used to measure relative opinions, preferences, attitudes and perceptions.

Interval Scales

An interval scale possesses the attribute of both order and distance. In this, there is an equal difference between the scale values. The distance between any pair of numbers is less than, equal to or greater than the distance between any other pair of numbers, and this distance is meaningful. For example, the temperature difference between 50°C and 55°C is the same as the difference between 60°C and 65°C. The difference of 10°C in both intervals possesses the same meaning and interpretation. It is important to note that the interval scales do not provide any information regarding ratio. In other words, we cannot say that 40°C is twice as hot as of 20°C. In marketing research, attitudinal data, behavioral data, opinion data, etc. attained from rating scales are generally considered interval data.

Ratio Scales

Ratio scales possess all the attributes of scale level of measurement i.e. description, order, distance and origin. It has all the traits of nominal, ordinal and interval scales along with the absolute true zero point. With these, we can describe or identify objects, rank the objects and



also compare differences or intervals. In this, the values are distributed in equal intervals. Also, in this, scale values can be computed in the form of ratios. For example, the difference between 50 kgs and 60 kgs is the same as the difference between 60 kgs and 70 kgs. Also, 60 kgs are twice of 30 kgs in an absolute sense. Ratio scales are commonly used in measuring age, height, income, etc. In marketing research, a ratio scale is used in measuring a number of customers, costs, sales, market share, price, etc.

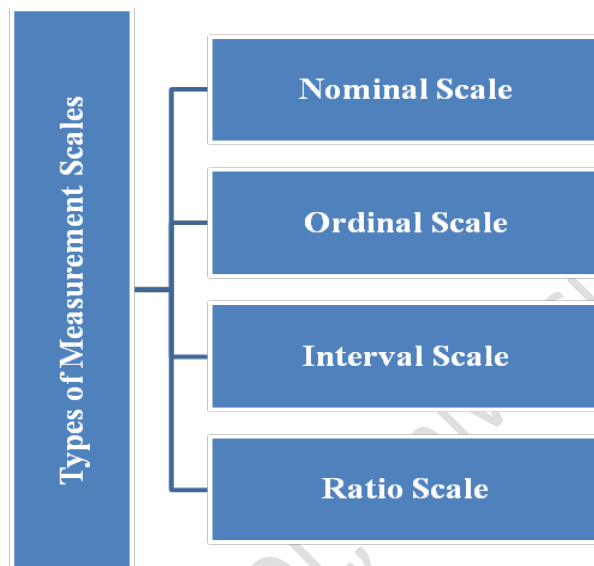


Fig 2.1: Types of Measurement Scales

ACTIVITY

Below are listed some objects of varying degrees of abstraction. Suggest properties of each of these objects that can be measured by each of the four basic types of scales.

- a) Length
- b) IQ Level
- c) Sales Revenue
- d) Season

2.3.3 Comparison of Measurement Scales

The four types of scales namely nominal, ordinal, interval and ratio are used for different purposes and also offer varying levels of information. In addition to scale attributes, the types



of scales also differ in terms of mathematical operations. The statistical techniques used to compute the data also differ according to the type of measurement scale used. A comparative analysis of each of the scales across attributes, mathematical operations and statistical techniques has been detailed in Table 2.1.

Table 2.1: Comparison of Measurement Scales

Scales		Nominal	Ordinal	Interval	Ratio
Attributes	Description	Yes	Yes	Yes	Yes
	Order	No	Yes	Yes	Yes
	Distance	No	No	Yes	Yes
	Origin	No	No	No	Yes
Mathematical Operations	= & $\neq$	Yes	Yes	Yes	Yes
	< & >	No	Yes	Yes	Yes
	+ & -	No	No	Yes	Yes
	X & $\div$	No	No	No	Yes
Statistical Techniques	Descriptives Statistics	Mode, Percentage	Median, Percentile	Mean, Range, Standard deviation.	Geometric mean, Harmonic Mean.
	Inferential Statistics	Binomial test, Chi-Square test	Correlation, ANOVA, Rank order	t-tests, ANOVA, Regression, product-moment correlation, factor analysis	Coefficient of Variation

IN-TEXT QUESTIONS

1. _____ attributes indicates that the scale has an absolute true zero point.
a) Description b) Order
c) Origin d) Distance
2. _____ scale only has the description attribute:
a) Nominal b) Ordinal
c) Interval d) Ratio
3. The ranking of teams in a Cricket match constitute a _____ scale.
a) Nominal b) Ordinal
c) Interval d) Ratio
4. In marketing research, attitudinal data, behavioral data, opinion data, etc. attained from rating scales are generally considered _____.
a) Nominal data b) Ordinal data
c) Interval data d) Ratio data
5. Chi-Square test is applicable on which type of data?
a) Nominal data b) Ordinal data
c) Interval data d) Ratio data

2.4 SCALING TECHNIQUES

The most common scaling techniques used in marketing research can be categorized into comparative scaling technique and non-comparative scaling technique (see Fig. 2.2).

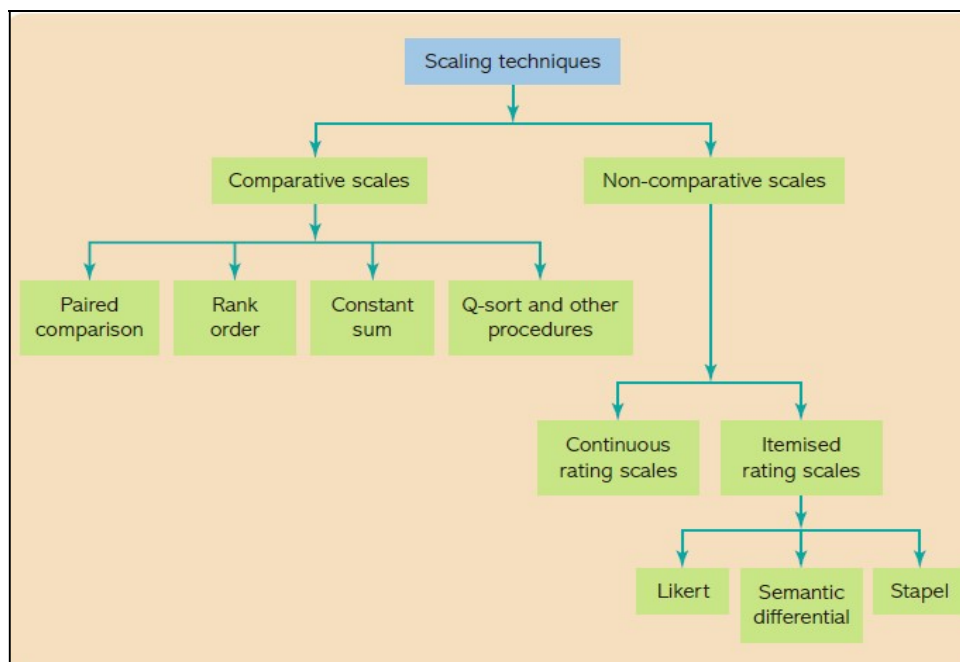


Fig 2.2: Types of Scaling Techniques (Source: Malhotra, Nunan and Birks (2017, p. 342))

2.4.1 Comparative Scaling

In a comparative scaling technique, there is a direct comparison of two or more objects. Data on a comparative scale must be evaluated in relative terms and have only ordinal or rank order attributes. Hence, for this reason, comparative scaling is also known as ‘*non-metric scaling*’. For example, respondents may be asked which laptop brand they prefer over the other. Different laptop brands would be scaled in relation to one another and the scale position of any one laptop brand would rely on the scale positions of the other laptop brands. Hence, in this, the respondents are forced to choose directly between the given objects. The commonly used comparative scaling techniques are:

Paired Comparison Scaling

In paired comparison scaling, the participants are given a choice between two objects and are asked to choose one object in the pair based on a set of criteria. The obtained data is ordinal in nature and the values are ranked in order of magnitude. In general, if n objects are taken, then $[n(n-1)/2]$, paired comparison are needed that includes all possible pairings of objects. For example, the respondents are asked to give their preference for three different mobile phone brands. For each pair, they are required to select one brand of mobile phone they prefer and to give a score of 1 to their preferred brand.

**Table 2.2:** Example of Paired Comparison Scaling

Mobile Phone Brands	One Plus	Samsung	Oppo	No. of times preferred
One Plus	-	1	1	2
Samsung	0	-	1	1
Oppo	0	0	-	0

In a particular box, 1 value indicates that the brand in that row was preferred over the brand in the corresponding column, whereas, a 0 value indicates that the column brand was preferred over the row brand. The response in Table 2.2 reveals that for this respondent, One Plus is the most preferred mobile phone brand followed by Samsung.

Constant Sum Scaling

In constant sum scaling, the participants are required to allocate a constant sum of units (such as 100 points) among a set of objects based on certain criteria. In this, the units allocated represent the importance the respondents attach to each attribute. The obtained data is interval in nature. For example, the respondents are asked to allocate 100 points to attributes of mobile phone brands as per their preference (see Table 2.3). The sum of all the points remains constant (i.e. 100) and therefore, it is called constant sum scaling.

Table 2.3: Example of Constant Sum Scaling

Mobile Phone Brands	Price	Screen Size	Battery Backup
One Plus	35	45	40
Samsung	30	25	35
Oppo	25	30	25
Total	100	100	100

Rank Order Scaling

In rank order scaling, respondents are provided with various objects simultaneously and are asked to rank or order them based on certain criteria. The obtained data is ordinal in nature. In this, the objects are ranked according to their preferences. The most preferred brand is



ranked as 1 and the second most preferred is ranked as 2 and so on. This procedure is continued till all the objects have been ranked as per the preferences. For example, the respondents are asked to compare different attributes of different mobile phone brands using rank order scaling. The ranks assigned are shown in Table 2.4.

Table 2.4: Example of Rank Order Scaling

Mobile Phone Brands	Price	Screen	Battery Backup
One Plus	1	3	2
Samsung	1	2	3
Oppo	3	2	1

Q-Sort Scaling

Q-Sort scaling technique uses a procedure of rank order in which the respondents are asked to sort a given number of objects into a predetermined number of sets based on certain criteria. This technique is used to distinguish among a quite large number of objects quickly. For better statistical reliability, it is advisable that the number of items to be sorted should be between 60 and 140, with a fair range of 60 to 90 objects. The Q-Sort procedure entails the creation of cards with single words, phrases, verbal statements, etc., related to the characteristics of objects being investigated. For example, participants are given 70 customer satisfaction statements on individual cards and asked to place them into 10 sets, ranging from “most highly agreed with” to “least highly agreed with”. The use of this ranking helps to capture what the participants think about the characteristics in relation to other characteristics.

2.4.2 Non-Comparative Scaling

In a non-comparative scaling technique, there is no comparison between the objects. In this, the objects are measured independently. These scales are also termed “metric” or “monadic scales” since only one object is evaluated at a time. The obtained data is generally assumed to be interval or ratio in nature. In marketing research, the non-comparative scaling technique is the most widely employed. The most commonly used non-comparative scaling techniques are the continuous rating scale and itemized rating scale.

Continuous Rating Scale

In a continuous rating scale, the participants are required to rate the objects by placing a mark on a line that extends from one extreme criterion variable at one end to the other. It is also



known as the graphic rating scale. The continuous rating scale can take many different forms namely lines (horizontal or vertical) and scale points (numbers or brief descriptions). Fig. 2.3 depicts three versions of the continuous rating scale.

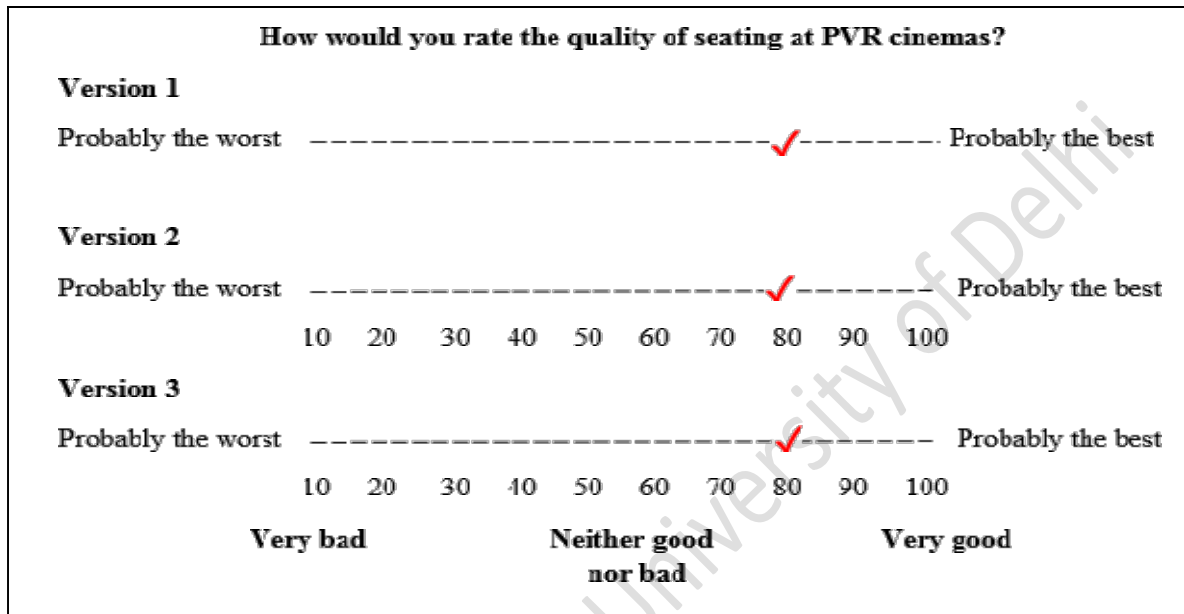


Fig 2.3: Example of Continuous Rating Scale

Itemized Rating Scale

In itemized rating scale, the participants are given a scale with a number or description for each category. The categories are arranged according to scale position and the participants are expected to choose one of the categories that best defines the objects being rated. In marketing research, itemized rating scales are often used and serve as the foundation for more complex scales like multi-item rating scales. The most commonly used itemized rating scales are Likert scale, Semantic differential scale and Stapel scale.

Likert Scale

The most popular itemized rating scale is the Likert scale which asks the participants to express their level of agreement or disagreement with a series of statements about the objects. Each scale item often contains five-point or seven-point responses ranging from 'strongly disagree' to 'strongly agree'. The scale points are assumed to have equal distance between them. An example of Likert scale is presented in Fig. 2.4.

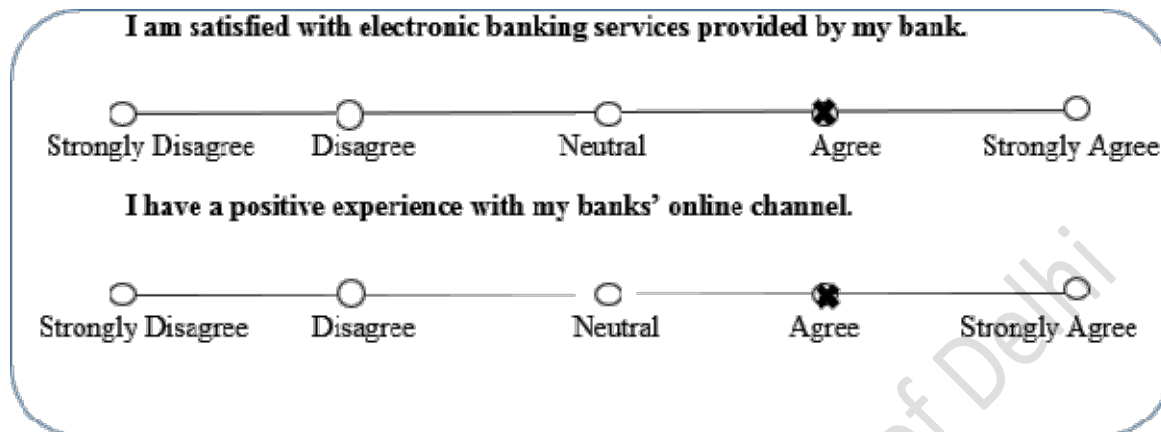


Fig 2.4. Example of Likert Scale

Semantic Differential Scale

In a semantic differential scale, the endpoints are associated with bipolar that have semantic meaning. To represent bipolar labels, sometimes negative phrases or adjectives are used labels (such as good-bad, strong-weak). Generally, a semantic differential scale uses a five-point or seven-point scale. An example of a semantic differential scale is presented in Fig. 2.5.

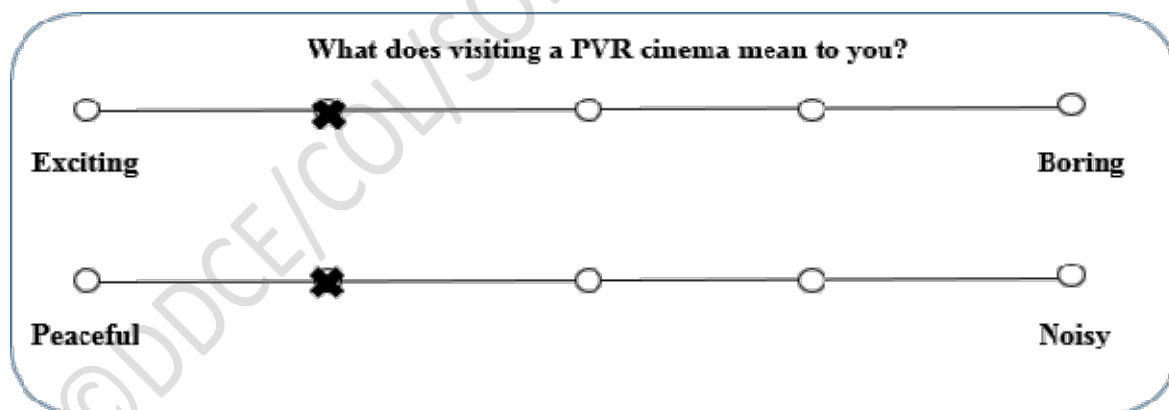


Fig 2.5. Example of Semantic Differential Scale

Stapel Scale

The Stapel scale is a uni-polar itemized rating scale in which only a phrase or single term is used. The scale is typically presented vertically and the participants are asked to rate how



accurately the phrase or single term defines the object by choosing an appropriate numerical response category. The highest (positive) number depicts that the description is accurate and a lower (negative) number depicts that the description is not accurate (see Fig. 2.6). It produces similar results to the semantic differential scale. Out of all the three itemised rating scales considered, the Stapel scale is used the least.

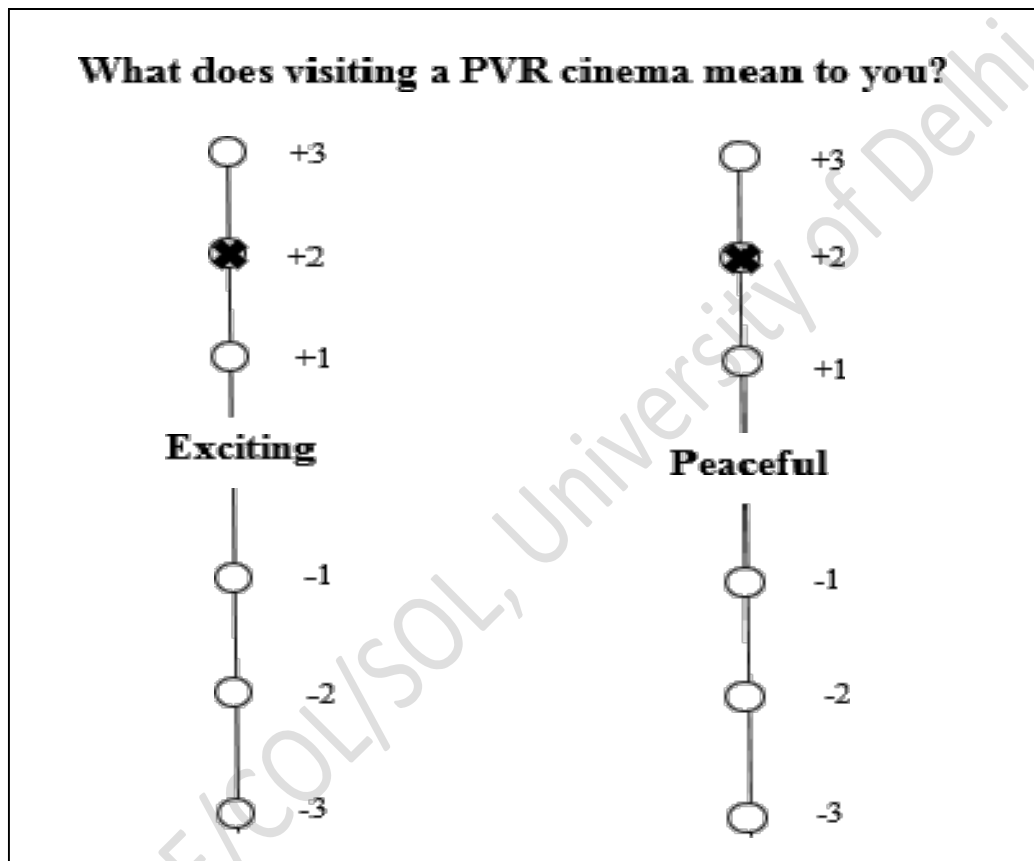


Fig 2.6. Example of Stapel Scale

2.4.3 Selection of Measurement Scale

Selection of measurement scale involves various keys factors:

Balanced and Unbalanced Scales

A scale is said to be balanced when it has an equal number of favorable and unfavorable categories above and below the midpoint whereas, an unbalanced scale has an unequal number of categories. It is always advisable to have a balanced scale if the study is not inclined towards one side and the aim is to obtain the most objective data.



Number of Scale Points

The number of scale points should suit the purpose of the research. It should not have too many points as it may become difficult to handle and also it should not have too fewer points as it will not be able to capture accurate results. Typically, five-point and seven-point scales are used in marketing research.

Forced and Unforced Choices

Unforced choice scale provides an option to express no opinion if the participant is unable to make a choice among the categories offered, whereas, it is not possible in the case of a forced choice scale where the participant is forced to make a choice. However, too many neutral options may distort the result, hence, researchers tend to avoid an unforced choice scale.

Physical Form of Scales

A scale can be presented in several options such as horizontal and verticals, whereas categories can be expressed in terms of graphics, numbers, boxes, discrete lines or units in a continuum. So, the researchers should try a number of options and then select the best one.

CASE STUDY

Assume you are Menu foods and you planned a major research study just prior to the largest pet food recall in our history. You plan to proceed with the study and feel you must add one or more questions to measure the consumer's confidence that your firm will be able to recover. Draft a scale for each of the following types that will measure that confidence:

- a) Likert scale
- b) Stapel scale
- c) Forced choice scale
- d) Constant sum scale

**IN-TEXT QUESTIONS**

6. Measurement scales are classified into ____ types.
- a) One
 - b) Two
 - c) Three
 - d) Four
7. In _____, the participants are given a choice between two objects and are asked to choose one object in the pair based on a set of criteria.
- a) Paired Comparison Scaling
 - b) Rank Order Scaling
 - c) Constant Sum Scaling
 - d) Q-Sort Scaling
8. When a scale has equal number of favourable and unfavourable categories above and below the midpoint, it is known as:
- a) Unbalanced Scaling
 - b) Balanced Scaling
 - c) Forced Choices Scaling
 - d) Q-Sort Scaling
9. The _____ is a uni-polar itemized rating scale in which only phrase or single term is used.
- a) Stapel Scale
 - b) Semantic Differential Scale
 - c) Likert Scale
 - d) Q-Sort Scaling
10. _____ scale provides an option to express no opinion if the participant is unable to make a choice among the categories offered.
- a) Balanced Scaling
 - b) Unbalanced Scaling
 - c) Forced Choices Scaling
 - d) Unforced Choices Scaling

2.5 RELIABILITY AND VALIDITY**2.5.1 Reliability**

Reliability of a measurement scale is the degree to which a scale yield consistent results when repeated measurements on a characteristic are made. There are two types of errors in this: systematic and random errors. Systematic errors do not have a negative influence on reliability because they remain constant across all measurements done using the scale in the same way. On the other hand, random error varies across each observation, hence, leading to lower reliability. Various approaches to estimate the reliability of a scale are:



Test-retest reliability

In test-retest reliability, the same measurement scale is provided twice to the same set of respondents at two different times, typically two to four weeks apart. Then, a correlation is made between the two measurements. If the scale is reliable and there is no change in respondents, then the result obtained will have a high correlation.

Alternate/Parallel forms of reliability

In this, for the same measurement, two or more scales are developed and the same participants are provided different scales at different times, typically two to four weeks apart. Then, the correlation is made between the scores of different scales to estimate reliability. However, this technique is difficult as it requires the development of two or more equivalent scales.

Internal consistency reliability

In internal consistency reliability, all the items on the scale relate to all other items. The simplest measure of this reliability is '*split-half reliability*'. In this, the scale items are divided into two halves and then the measurement scores for the two halves are correlated. The most common method to split scale items is the odd- and even-numbered or random strategy. High correlations between the two halves depict high internal consistency. However, this technique has a problem as the result will rely on how the scale items are split. To overcome this problem, Cronbach's coefficient alpha is used. Cronbach's alpha is the mean of all possible split-halves reliabilities that result from various ways of splitting the scale items. This coefficient value ranges from 0 to 1.

2.5.2 Validity

Validity is the degree to which the measurement scale accurately captures what it is designed to measure. There are two major forms of validity- Internal and External. **Internal validity** means the extent of certainty that the causal relationship being tested is reliable and unaffected by other factors or variables. **External validity** means the extent to which the results can be generalized to other groups, events or situations. Validity can be estimated by:

Content validity

Content validity is the degree to which the measurement scale captures all aspects of the concept being measured. It is subjective in nature and no empirical measurements are done.

Criterion validity

Criterion validity is the degree to which the outcome of the measurement is related to other valid measures of the same concept. It is of two types: (a) **concurrent validity** measures how



closely the outcomes of a specific measurement relate to those of a previously established measurement for the same construct, and (b) **predictive validity** measures how well scores on measurement can accurately predict future performance on another measure of the concept they represent.

Construct validity

Construct validity examines how effectively a set of measured items actually measures the variables that it is intended to measure. It is of three types: (a) **convergent validity** means the constructs that are theoretically supposed to be related to each other are in fact related to each other, (b) **discriminant validity** means the constructs that are theoretically not supposed to be related to each other are in fact unrelated to each other, and (c) **Nomological validity** means the validity that evaluates the relationship between theoretical constructs and aims to confirm substantial relationships between constructs suggested by theory.

IN-TEXT QUESTIONS

11. _____ is the degree to which a scale yield consistent results when repeated measurements on a characteristic are made.
 - a) Reliability
 - b) Validity
 - c) Scaling
 - d) Measurement
12. _____ validity is the degree to which the measurement scale captures all aspects of the concept being measured.
 - a) Concept
 - b) Predictive
 - c) Convergent
 - d) Construct
13. _____ means the constructs that are theoretically supposed to be related to each other are in fact related to each other.
 - a) Discriminant Validity
 - b) Construct Validity
 - c) Convergent Validity
 - d) Predictive Validity
14. Which of the following is not a type of construct validity?
 - a) Convergent Validity
 - b) Discriminant Validity
 - c) Nomological Validity
 - d) Predictive Validity
15. A validity that measures how closely the outcomes of a specific measurement relate to those of a previously established measurement for the same construct is called:
 - a) Concurrent Validity
 - b) Concept Validity
 - c) Nomological Validity
 - d) Predictive Validity



2.6 SUMMARY

- Measurement is the process of assigning numbers or symbols to attributes of objects based on predefined rules.
- Scaling is the process of establishing a continuum on which measured objects are placed.
- The scales used in marketing research can be categorized in terms of four fundamental attributes i.e. description, order, distance and origin.
- Measurement scales are classified into four types: nominal, ordinal, interval and ratio.
- A nominal scale has the only attribute of description which is used to describe or label any object.
- An Ordinal scale has the attribute of both description and order.
- An Interval scale has the attribute of description, order and distance.
- A Ratio scale has all the attributes of a description, order, distance and origin.
- The most common scaling techniques used in marketing research can be categorized into comparative scaling techniques and non-comparative scaling techniques.
- In a comparative scaling technique, there is a direct comparison of two or more objects, whereas in a non-comparative scaling technique, there is no comparison and the objects are measured independently.
- Comparative scaling techniques include paired comparison, constant sum, rank order and Q sort scaling.
- Non-comparative scaling techniques include continuous rating and itemized rating scales.
- Selection of measurement scale involves key factors namely number of scale categories, balanced or unbalanced scale, forced or unforced choice scale and physical form of the scale.
- Reliability of a measurement scale is the degree to which a scale yield consistent results when repeated measurements on a characteristic are made.
- Various approaches to estimate the reliability of a scale are test-retest, alternate/parallel form and internal consistency reliability.



- Validity is the degree to which the measurement scale accurately captures what it is designed to measure. Two major forms of validity- Internal and External validity.
- Validity can be estimated by content, criterion and construct validity.

2.7 GLOSSARY

Balanced scale: Equal number of favorable and unfavorable categories above and below the midpoint.

Coefficient alpha: Mean of all possible split-halves reliabilities that result from various ways of splitting the scale items

Comparative scaling technique: A type of technique in which a direct comparison of two or more objects is made.

Concurrent validity: A type of validity that measures how closely the outcomes of a specific measurement relate to those of a previously established measurement for the same construct.

Constant sum scaling: A type of technique in which the participants are required to allocate a constant sum of units among a set of objects based on certain criteria.

Construct validity: Examines how effectively a set of measured items actually measures the variables that it is intended to measure.

Content validity: Degree to which the measurement scale captures all aspects of the concept being measured.

Continuous rating scale: A type of scale in which the participants are required to rate the objects by placing a mark on a line that extends from one extreme criterion variable at one end to the other.

Convergent validity: A measure of construct validity which means constructs that are theoretically supposed to be related to each other are in fact related to each other.

Correlation: Measure of extent to which two or more variables are related to each other.

Criterion validity: Degree to which the outcome of the measurement is related to other valid measures of the same concept.

Criterion variable: the variable measured or predicted by the researcher; expected to be affected by a manipulation of the independent variable.

Description: Numbers that can be used as labels or identities to describe each value of scale.



Discriminant validity: A measure of construct validity which means constructs that are theoretically not supposed to be related to each other are in fact unrelated to each other.

Distance: Absolute differences between the scale descriptors are known and may be represented in units.

External validity: Extent to which the results can be generalized to other groups, events or situations.

Forced choice scale: A type of scale where the participant is forced to make a choice among the categories offered.

Internal consistency: attribute of an instrument in which items are homogeneous; a measure of reliability

Internal validity: Extent of certainty that the causal relationship being tested is reliable and unaffected by other factors or variables.

Interval scale: A type of scale in which there is an equal difference between the scale values.

Itemized rating scale: A type of scale in which the participants are given a scale with a number or description for each category.

Likert scale: A type of scale that asks the participants to express their level of agreement or disagreement with a series of statements about the objects.

Measurement: Process of assigning numbers or symbols to attributes of objects based on predefined rules.

Nominal scale: A type of scale used to describe or label any object.

Nomological validity: A measure of construct validity that evaluates the relationship between theoretical constructs and aims to confirm substantial relationships between constructs suggested by theory.

Non-comparative scaling technique: A type of technique in which there is no comparison and the objects are measured independently.

Order: Relative position or sizes of the descriptors such as “less than”, “equal to” and “greater than”.

Ordinal scale: A type of scale that determines whether an object has a greater value, equal value or less value than some other object, but does not reveal the magnitude of differences.

Origin: The scale has an absolute true zero point.



Paired comparison scaling: A type of technique in which the participants are given a choice between two objects and are asked to choose one object in the pair based on a set of criteria.

Predictive validity: A type of validity that measures how well scores on a measurement can accurately predict future performance on another measure of the concept they represent.

Q-Sort scaling technique: A type of technique that uses a procedure of rank order in which the respondents are asked to sort a given number of objects into a predetermined number of sets based on certain criteria.

Random error: A type of error that varies across each observation.

Rank order scaling: A type of technique in which the respondents are provided with various objects simultaneously and are asked to rank or order them based on certain criteria.

Ratio scale: A type of scale that has all the traits of nominal, ordinal and interval scales along with the absolute true zero point.

Reliability: Degree to which a scale yields consistent results when repeated measurements on a characteristic are made.

Scaling: Process of establishing a continuum on which measured objects are placed.

Semantic differential scale: A type of scale in which the endpoints are associated with bipolar that have semantic meaning.

Stapel scale: A uni-polar itemized rating scale in which only a phrase or single term is used.

Systematic error: A type of error that remains constant across all measurements done using the scale in the same way.

Unbalanced scale: Unequal number of favorable and unfavorable categories above and below the midpoint.

Unforced choice scale: A type of scale that provides an option to express no opinion if the participant is unable to make a choice among the categories offered.

Validity: Degree to which the measurement scale accurately captures what it is designed to measure.



2.8 ANSWERS TO IN-TEXT QUESTIONS

1. Distance	9. Stapel scale
2. Nominal	10. Unforced choices scaling
3. Ordinal	11. Reliability
4. Interval data	12. Concept
5. Nominal data	13. Convergent Validity
6. Four	14. Predictive Validity
7. Paired Comparison Scaling	15. Concurrent Validity
8. Balanced	

2.9 SELF-ASSESSMENT QUESTIONS

1. Discuss the difference between nominal, ordinal, interval and ratio scales.
2. Discuss various types of non-comparative scaling techniques.
3. *“It does not matter which scaling technique you use. As long as your measure is reliable you will get the right results”*. Discuss the statement.
4. What is validity? Discuss the various types of validity.
5. You have been asked by the head of marketing to design an instrument by which your college can evaluate the quality and value of its various curricula and courses. How might you try to ensure that your instrument has:
 - a) Internal consistency
 - b) Content validity
 - c) Construct validity
 - d) Predicted validity
6. Identify the appropriate scale and the associated statistics that can be used for the following:
 - a) Average monthly income
 - b) Occupation of respondents
 - c) Temperature
 - d) Speed



2.10 REFERENCES

- Malhotra, N. K., Nunan, D., & Birks, D. F. (2017). *Marketing Research: An Applied Orientation* (5th ed). New Jersey: Prentice Hall International Inc.
- Cooper, D. R., Schindler, P. S., & Sharma, J. K. (2018). *Business research methods* (12th ed). New Delhi: McGraw-Hill Education.
- Dangi, H. K., & Dewen, S. (2016). *Business research methods*. Delhi: Cengage India Limited.
- Surya, P. K. & Sharma, S. K. (2020). *Business Research Methods and Analytics*. New Delhi: Taxmann publication.

2.11 SUGGESTED READINGS

- Burns, A. C., Veeck, A. F. & Bush, R. F. (2017). *Marketing research* (8th ed). New Delhi: Pearson.
- Churchill, G., Iacobucci, D. & Israel, D. (2010). *Marketing research: A South-Asian Perspective*. Delhi: Cengage.
- Zikmund, W. G., & Babin, B. J. (2016). *Exploring Marketing Research* (11th ed.). USA: Cengage South-Western.



LESSON-3

DATA COLLECTION METHODS

Ms. Shalu Garg

Assistant Professor

Guru Gobind Singh Indraprastha University

Shalugarg11901@gmail.com

STRUCTURE

- 3.1 Learning Objectives
- 3.2 Introduction
- 3.3 Primary Data and Secondary Data
 - 3.3.1 Primary Data and Secondary Data
 - 3.3.2 Characteristics of Primary Data
 - 3.3.3 Advantages of Primary Data
 - 3.3.4 Disadvantages of Primary Data
 - 3.3.5 Primary Data Collection Sources
 - 3.3.6 Characteristics of Secondary Data
 - 3.3.7 Advantages of Secondary Data
 - 3.3.8 Disadvantages of Secondary Data
 - 3.3.9 Comparison Between Primary Data and Secondary Data
- 3.4 Sources of Secondary Data
 - 3.4.1 Internal Sources of Data
 - 3.4.2 External Sources of Data
- 3.5 Conducting Review of Literature
 - 3.5.1 Meaning of Literature Review
 - 3.5.2 Purpose of Literature Review
 - 3.5.3 Limitations of Literature Review
- 3.6 Quantitative Research and Qualitative Research
 - 3.6.1 Quantitative Research
 - 3.6.2 Qualitative Research
- 3.7 Summary
- 3.8 Glossary
- 3.9 Answers to In-text Questions
- 3.10 Self-Assessment Questions



3.11 References

3.12 Suggested Readings

3.1 LEARNING OBJECTIVES

In this lesson we are going to learn about the two important methods of data collection which are primary data collection and secondary data collections. We will learn the characteristics, advantages, and disadvantages of both the methods of data collection. We will also look at various sources of data collection under the two methods. We will also learn about conducting literature review.

After studying the lesson, students able to:

- Understand the meaning of Primary Data and Secondary Data.
- Explain the characteristics, advantages, limitations and sources of Primary Data and Secondary Data.
- Learn the Comparison Between Primary Data and Secondary Data.
- Get acquainted with internal and external sources of Secondary Data.
- Explain the conducting of review of literature.
- Understand the meaning of Quantitative Research and Qualitative Research.

3.2 INTRODUCTION

In all fields of business, it is prevalently viewed as that the result relies upon the info. The nature of an item relies upon the kind of unrefined components utilized. The idea is likewise prominently known as "GIGO" meaning garbage in garbage out. The nature of an item relies upon the kind of unrefined components utilized. Every one of these demonstrates the one basic truth that anything one sows are what one harvests. One cannot plant a wheat seed and anticipate that rice should develop. In research, information is the seed that will drive the outcomes. Information is a vital contribution to any exploration.

In this manner, an assortment of information becomes one of the main pieces of exploration. Previously examining the techniques for information assortment in business research, the analyst ought to comprehend two sorts of information, viz., primary, and secondary. Consequently, an assortment of information becomes one of the main pieces of examination. Assortment of information to such an extent that the information gives the expected data is vital. In this part, we will talk about the critical parts of data and data collection.



3.3 PRIMARY DATA AND SECONDARY DATA

3.3.1 Primary and Secondary Data:

Prior to examining the strategies for information assortment in business research, the analyst ought to grasp two sorts of information, viz., primary, and secondary. The primary data are those which are gathered afresh. The primary data is the one that is gathered for the first time. The primary data is unique in character. Secondary data is generally characterized as contrary to primary data. Secondary data are gathered afresh and include to the information gathered by somebody other than the analyst utilizing the information. The information was recently gathered for another reason. Such information might be exceptionally helpful for research purposes. Because of the initial collection of primary data, the analyst produces the information without any preparation instead of utilizing others' information.

The inventiveness of primary data brings many benefits. The information is gathered considering exploration targets. Accordingly, the information is quite certain for the examination. The analyst is likewise ready to control the information quality. Consequently, the dependability of Primary Data is likewise high. In any case, the major question with an assortment of Primary Data is that it is a time-consuming process. The information assortment costs are additionally higher. On the other hand, secondary data assortment can be faster and simpler. It for the most part requires lesser investment and assets to gather secondary data. The fundamental weakness of secondary data is that it was gathered by another person for another reason. Hence, the secondary data need not be adequately explicit for the exploration. The information may likewise be an old one which may not stand valid in the current date.

The strategies for gathering primary and secondary data are totally different. Primary data should be initially gathered. Then again, secondary data assortment includes an aggregation of information from existing sources of information.

For instance, let us say a cell phone organization needs to choose which programming stage to use for its next launch. To settle on this choice, they need to know the current fame of the stages. Significant data for this would be shared in the market of different cell phone platforms. To utilize secondary data, then they can peruse the industry data, existing company reports, etc., to figure out the portions of the market share of these platforms. Optional information alludes to the information gathered by somebody during these stages. On the other hand, they can do a Primary Data assortment by getting some information about their



most favoured mobile platforms. Consequently, for a given piece of data required, one can decide to utilize primary data or secondary data.

ACTIVITY

Which form of data collecting is best for conducting research regarding the family planning programme in India? Interpret its positive and bad aspects.

3.3.2 Characteristics of Primary Data:

Primary data comprise the premise data to tackle the showcasing issues. To throw the light on the importance of Primary Data, let us discuss the various features of Primary Data:

1. Primary Data are unique information.
2. Primary Data are costly.
3. Assortment of Primary Data takes significant time and efforts of the researcher.
4. They are gathered regarding issue close by.
5. They are gathered purposely from significant respondents.
6. Explicit strategies (like study techniques, perception strategies, exploratory strategy, and so on) and instruments (like printed structures, survey, camera, and so on) are utilized for gathering Primary Data.
7. They comprise an essential contribution to the examination project.
8. They are required be furnished, processed, or analysed before they are utilized.

3.3.3 Advantages of Primary Data:

Primary data is frequently dependable, legitimate, and objective in however much it was gathered and determined to address a specific exploration issue. It is imperative that primary data is not generally gathered because of the significant expense of execution. Other than these, there are various benefits of using Primary Data which are explained underneath:

- a) **Specific to the Needs of the Researcher**-Primary Data is intended for the requirements of the specialist right now of information assortment. The analyst can handle the sort of information that is being gathered.
- b) **Accurate in Nature**-It is exact contrasted with secondary data. The information is not exposed to individual bias and as such the authenticity can be relied upon.



- c) **Ownership of the Data**-The analyst display responsibility for information gathered through essential examination. The person in question might decide to make it accessible freely, patent it, or even sell it.
- d) **Up to Date**-Primary Data is as a rule modern since it gathers information progressively and does not gather information from old sources.
- e) **Control over the Data Collected**-The analyst has full command over the information gathered through essential exploration. He can conclude which plan, strategy, and information investigation methods to be utilized.

3.3.4 Disadvantages of Primary Data:

Despite being all the advantages, there are few shortcomings as well of collecting Primary Data which are given underneath:

- a) **Expensive**- Primary Data is expensive as compared to secondary data collection. information. Hence, gathering essential data may be troublesome.
- b) **Time-Consuming Process**- The procedure of collecting the data originally takes a lot of efforts in the research which takes a huge amount of time.
- c) **Complex**-It may not be practical to gather Primary Data at times because of its intricacy and required responsibility.

3.3.5 Primary Data Collection Sources:

Primary data can be gathered by leading experiments, noticing the respondents, or by direct correspondence with the respondents. The strategy directing the experiments is utilized particularly for exploratory research plans. For illustrative or causal examination plans, the information is for the most part gathered through overviews or registration. In an overview, the specialist looks at those peculiarities which exist in the universe autonomous of analyst's activities. The different techniques for information assortment for expressive and causal research are talked about below.

- a) **Observation Method**-In the observation method, the data is gathered by direct observation without getting clarification on pressing issues from the respondent. The primary benefit of this technique is that the respondent's inclination is diminished, assuming that perception is done precisely. Moreover, this technique is free of respondents' eagerness to answer. This technique is much of the time utilized in the investigation of customer conduct. The observation technique likewise has a few



burdens. It is a costly technique and the data given by this technique is extremely restricted.

The observation method is additionally delegated as participant and non-participant types of observation. If the specialist sees by making himself/herself an individual from the gathering being noticed, then, at that point, it is called participant observation. Here, the analyst can encounter what the individuals from the gathering experience. At the point when the specialist sees a different delegate without trying to turn into an individual from the gathering being noticed, then, at that point, it is named a non-participant observation. In a non-participant observation, if the researcher observes in such a way that his/her presence is unknown to the gathering being noticed, then it is named a disguised or hidden observation.

- b) **Interview Method**-The interview method includes coordinated addressing from the respondent. This strategy can be utilized through private meetings or through phone interviews. Individual meetings require the questioner to pose inquiries in face-to-face contact with the next individual or people. This technique permits researchers to do investigated examinations. Assortment of data through private interviews is done in an organized way. A bunch of predetermined questions is utilized to complete the interview.

The principal benefits of the individual interview are that data can be acquired in detail. The interviewer can assist the member with overcoming the reluctance to answer. The language of the interview can be embraced to the comfort of the respondent.

Being there so many benefits of the interview method, but there are also a few drawbacks of the interview method. It is additionally a costly strategy if the sample size is huge and spread geographically. The interviewer's predisposition and the respondents' inclination can sneak in. It is a very time-consuming process and very difficult to schedule and control these interviews. A portion of the samples might be unapproachable. The interview technique is generally additional tedious. We will be learning in detail about the interviews in detail in the next lesson.

- c) **Collection of Data Through Questionnaires**-This technique for information assortment includes posing inquiries from respondents utilizing a bunch of inquiries. This technique is especially utilized when the test size is enormous. The inquiries are carefully designed to guarantee that they cover all the data that is expected for research. Questionnaires are given to respondents, so they can answer themselves in



case of indirect administration. The questions are asked through direct interviews or telephone interviews utilizing questionnaires.

A questionnaire is an information assortment instrument that comprises a formalized arrangement of questions for getting data from respondents. The questionnaire ought to be built so that it can get the data required for the examination. In this manner, every one of the questions in a questionnaire is decided for its precision as well as importance. The questions should be framed in a manner so that the reaction gives the exact data expected for the exploration. The ambiguity in questions can effectively convert into ambiguity in reactions too. Additionally, the questions asked in the questionnaire ought to be significant for the examination. Inquiring unnecessary questions that do not try to give any valuable data for exploration ought to stay away. The questionnaire is made to guarantee that every one of the questions posed is pointed toward adding essential data for the research. We will be learning about questionnaires in detail in the next lesson.

- d) **Case Study Method**-The case study method includes a cautious and complete perception of a social unit. The social unit can be an individual, a family, a foundation, a social gathering, or even a whole local area. The social unit is frequently referred to as the "case." The case study method concentrates on technique, the attention is on gathering information inside and out rather than expansiveness. The case study analysis gives more weight to a full investigation of a predetermined number of occasions or conditions. The case study analysis is basically a concentrated examination of the case which is under consideration.

A common technique for qualitative analysis is the case study approach, which entails meticulous and thorough observation of a social unit, such as an individual, a family, an institution, a cultural group, or even the entire community. It is a method of study that emphasises depth over breadth. The case study emphasises more thoroughly examining a smaller set of circumstances and their connections. The case study discusses the processes that occur and how they relate to one another. So, a case study is essentially a thorough assessment of the specific unit being looked at. The goal of the case study approach is to identify the variables that together account for the behavior-patterns of the provided unit.

"The case study method of data collecting is a strategy through which individual factors, whether they be institutions, merely a single incident in a person's life, or a group, are evaluated in their relationship to any other in the group," says the report. A



life or case history is, thus, a rather thorough examination of a person or group (in terms of what they do and have done, think they do and have done, and anticipate to do and say they ought to do). Burgess referred to the case study approach as "the social microscope."

In a nutshell, we can say that the case study method is a type of qualitative analysis in which a person, situation, or institution is carefully and thoroughly observed. Efforts are then made to study each and every aspect of the relevant unit in great detail, and generalisations and inferences are drawn from the case data.

The following are some crucial attributes of the case study approach:

- With this approach, the researcher may choose one social unit or a number of such units for his or her investigation; he may even choose a circumstance to investigate thoroughly.
- Here, the chosen unit is thoroughly investigated, perhaps down to the last detail. In general, the investigation lasts a long time to determine the unit's natural history in order to gather enough data to make accurate inferences.
- With this methodology, we thoroughly examine the social unit from all angles. By using this approach, we want to comprehend the complex of variables that operate within a social group as a cohesive whole.
- This strategy uses a qualitative rather than a quantitative approach. There is no gathering of purely quantitative data. Every attempt is made to assemble data pertaining to all facets of life. As a result, case studies broaden our perspectives and provide us with a comprehensive understanding of life. While creating a case study on a man as a criminal, for instance, we will not just look at how many crimes he has committed but also the circumstances that led to those crimes. The study's goal could be to offer suggestions for strategies to help the offender change.
- An effort is made to understand the causal components' mutual interrelationships with regard to the case study methodology.
- The behaviour pattern of the relevant unit is directly researched using the case study approach rather than indirectly and abstractly.
- The case study method generates sound ideas and data that can be used to test them, resulting in a continuous expansion of generalised knowledge. Generalized social science might suffer without it.



The case study approach is predicated on a number of assumptions. The following is a list of several key presumptions:

- The belief that human nature is constant at its core, despite the fact that behaviour can change depending on the circumstances.
- The presumption of researching the relevant unit's natural history.
- The presumption of thorough study of the relevant unit.

Principal stages involved Key stages of a case study include the following:

- Identifying and establishing the status of the phenomenon under study or the object of attention.
- Data gathering, analysis, and history of the specific phenomenon.
- Diagnosis and explanation of the underlying causes as a foundation for corrective or developmental therapy.
- Application of corrective methods, such as counselling and treatment (this phase is often characterised as case work).
- Follow-up programme to assess how well the treatment was received.

ACTIVITY

Their creative office recently gave print promotion to Saanchi, a showcasing supervisor. She needs to determine whether the model used in the creative is elevating the inventive or whether she is detracting and the focus from the object is shifting. What basic data collection techniques can be used to finish this review?

3.3.6 Characteristics of Secondary Data:

Secondary Data are distributed information. They have been gathered by others for their contemporary issues. Here and there, they are not valuable. They are advantageous for essential information. They support essential information. The degree to which secondary data are utilized for showcasing research relies on appropriateness, precision, and time. To throw the light on the importance of Secondary Data, let us discuss the various features of Secondary Data:

1. Secondary data are distributed information, not unique information, for the examination on hand.
2. They give the most recent data.



3. They can be handily gathered from different internal and external sources.
4. They are moderately less expensive; they need less effort, time, and cash.
5. They have been gathered by others for their own concerns and circumstance before.
6. They are utilized as a value to primary data. For the most part, they are utilized for characterizing and figuring out issues.
7. The utilization of secondary data is discretionary. Exploration can be directed even without the utilization of this sort of information.
8. They can be utilized without handling; do not bother dissecting them before they are utilized. They can be utilized straightforwardly
9. Significance, precision, and timing are the principal issues connected with secondary data.

ACTIVITY

"It is not safe to take distributed findings at face value without understanding their significance and obstacles." Explain this claim by outlining the several points that should be taken into account before adopting any published data.

3.3.7 Advantages of Secondary Data:

There are various benefits of secondary data which are explained below:

- a) **Accessible**-Secondary Data is effectively available contrasted with primary data.
- b) **Available on different platforms**-Secondary Data is accessible on various stages that can be gotten to by the specialist.
- c) **Affordable**- Secondary Data is truly reasonable. It expects practically zero expense to obtain them since they are at times given out free of charge.
- d) **Less Time-Consuming**-The time spent on gathering Secondary Data is generally very little contrasted with that of primary data.
- e) **Possible to carry out Longitudinal Studies**-Secondary Data makes it conceivable to do longitudinal examinations without waiting for long time to reach the conclusions.
- f) **Generate New Insights into Existing Primary Data**-It assists with producing new bits of knowledge into existing essential information.



3.3.8 Disadvantages of Secondary Data:

Despite so many advantages of using secondary data, there are certain drawbacks which we will discuss now:

- Data may not be Authentic and Reliable**-Secondary data may not be true and dependable. A scientist might have to additionally confirm the information gathered from accessible sources.
- Deal with Irrelevant Data**-Analysts might need to manage immaterial information before at long last tracking down the expected information.
- Exaggerated Data**-A portion of the information is overstated because of the individual predisposition to the information source.
- Outdated Source**-Secondary data sources are in some cases obsolete with no new information to replace the old ones.

3.3.9 Comparison between Primary Data and Secondary Data:

BASIS FOR COMPARISON	PRIMARY DATA	SECONDARY DATA
Meaning	Primary data refers to the first-hand data gathered by the researcher himself.	Secondary data means data collected by someone else earlier.
Data	Real time data	Past data
Process	Very involved	Quick and easy
Source	Surveys, observations, experiments, questionnaire, personal interview, etc.	Government publications, websites, books, journal articles, internal records etc.
Cost effectiveness	Expensive	Economical



BASIS FOR COMPARISON	PRIMARY DATA	SECONDARY DATA
Collection time	Long	Short
Specific	Always specific to the researcher's needs.	May or may not be specific to the researcher's need.
Available in	Crude form	Refined form
Accuracy and Reliability	More	Relatively less

Tabular Difference Source: S, S. (2020, July 13). *Difference Between Primary and Secondary Data (With Comparison Chart)*. Key Differences. <https://keydifferences.com/difference-between-primary-and-secondary-data.html>

IN-TEXT QUESTIONS

1. The data which is collected for the first time and is afresh known as _____.
2. The secondary data can be outdated. True/False.
3. The data which is readily available is known as _____.
4. The collection of primary data is an easy task. True/False
5. The primary data collection takes _____ period of time as compared to secondary data.



CASE STUDY

Case Study on Primary and Secondary Market Research at ABC Company

The Client-A prominent IT business with US headquarters required to interview senior management leaders in-depth to determine the value of evolving strategic theories in the wake of 9/11. They made the decision to contract ABC company to do this research.

The Requirement-A global online survey of more than 165 senior management professionals was part of the assignment. The survey's objective was to determine the applicability of strategic theories to evolving strategic needs and management practises for international organisations in the wake of 9/11.

The Solution-We chose one-on-one telephone interviews and online surveys because top management executives were the only source of the primary data. The main data to be gathered required advanced listening and comprehension abilities because the topic was the oretical strategy approaches. The team's members were all post-graduates with degrees in management and literature.

It was necessary to use a lot of secondary data in addition to the original investigation. Utilizing a number of paid global databases, business periodicals, and management reviews, extensive secondary research was conducted. The findings of the primary study were combined with all the data gathered from secondary research.

The Results-Since top management was required to provide all of the important facts, in order to develop a complete study outlining the strategy models that are applicable in the current post-9/11 climate, the findings from both primary and secondary sources were combined. The paper also identified the modern multinational firms' shifting strategic interests.

3.4 SOURCES OF SECONDARY DATA

Secondary data is information that has been gathered by people or agencies for purposes other than those of our specific examination study. Accordingly, the assortment of secondary data essentially includes distinguishing the right sources of information applicable for the examination. A scientist should never begin any examination expecting that no exploration



has at any point been finished around the subject of interest. On the off chance that a decent and thorough investigation is finished, significant data connected with examination can be gotten from optional sources. It is fundamental for a specialist to know about different sources from which optional information can be gotten.

There can be multiple sources of secondary data. Major sources of secondary data are internal records, externally published records, and syndicated data.

3.4.1 Internal Sources of Data:

These are the information sources that are created in the business for different purposes. The models can be bookkeeping records, exchange affiliation information, trade association data, sales performance record, customer database, etc. This information can be analyzed according to the point of view of the examination issue. For instance, a retailer sees a decrease in income and needs to research the causes. The exploratory examination can start by checking out the day-to-day exchange information. By examining inside accessible exchange information, one can distinguish without much of a stretch on the off chance that the fall in income is because of a fall in footfall, or a fall in average billing amount. Additionally, the item classes which have seen the most extreme falls can be examined. The accessible secondary data into the exploration issue for further examination.

Some examples of internal data sources are as follows:

- a) **Sales Data**-Organizations gather data generally through their day-to-day tasks. Records are made when orders are gotten and conveyed. Sales staff submit visit reports. The records are made as invoices, buy orders, sheets of returned products, and so on. The vast majority of this data can be of likely use in business research. Associations might disregard this important information asset. In this way, while doing business research, inner information sources, for example, sales invoices, orders, requests about items, getting back from clients and sales force client calling sheets, and so on can be of extraordinary use.

For instance, just from sales requests and invoices, one can determine data about deals by area, deals by client type, costs, limits, normal size of request by the client, client type, geographical region, normal sales by salesperson, and deals by pack size and pack type, and so forth. Such optional information can be utilized to recognize an association's most productive items and clients. It can likewise help in understanding client patterns and inclinations.



- b) **Financial Data**-Companies likewise have a ton of information connected with funds. Cost of creating, putting away, shipping and showcasing every one of its items, compensations paid to different workers, other functional costs, and so forth can likewise find many purposes in business research. For instance, such monetary data can be utilized to appraise the expenses connected to new items dispatches.
- c) **Transport Data**-Companies that keep great records connecting with their vehicle activities are well set to lay out which are the most productive courses, and loads, as well as the most practical directing examples. Great information on transport activities empowers the undertaking to perform compromise examination and, in this way, lay out whether it sounds good to possess or recruit vehicles, or the point at which an equilibrium of the two gives the best monetary result.
- d) **Capacity Data**-The pace of stock turn, stock taking care of expenses, surveying the productivity of certain showcasing tasks and the effectiveness of the promoting framework overall. More complex bookkeeping frameworks dole out expenses for the cubic space involved by individual items and the time period over which the item consumes the space. These frameworks can be additionally refined so that the productivity per unit, and pace of offer, are added. Along these lines, the immediate item productivity can be determined.

3.4.2 External Sources of Data:

Numerous associations give optional information. Public and nearby government organizations, semi-government organizations, exchange affiliations, colleges, research foundations, monetary establishments, and expert business research undertakings give a great deal of data. A portion of the vital outer sources of information is as per the following:

- a) **Externally Published Data**-This incorporate information distributed by different exchange affiliations, organizations, research firms, financier firms, government, and so forth. Auxiliary information can be tracked down in catalogues, guides, and files. Many counselling organizations likewise distribute market patterns information provides details regarding a standard premise. There are different literary works in the structure of diaries, magazines, books, and so on that structure the source of remotely distributed information. Government likewise gathers huge measure of information on customary spans.

Evaluation information, gathered by government, can be of extraordinary use for concentrating on the number of inhabitants in the nation and for acquiring a superior



down playing of the variety. Government additionally gathers different other measurable information in different fields like rainfall data, pollution data, income groups of population, farming production, industrial development, etc. An analyst can involve this promptly accessible information for the exploratory examination.

Government measurements in India is introduced on the site "www.indiastat.com" in a far-reaching way. It gives immense measure of information like population censuses, socioeconomics, accessibility of merchandise per capita, accessibility of assets like waterpower, and so forth, import/export measurements, production insights, farming measurements and market share information across different item types.

Public and global foundations give sources like college research reports, diaries furthermore, articles. Global offices like World Bank, International Labour Organization and Worldwide Monetary Fund likewise produce a great deal of optional information.

- b) **Syndicated Data-** There are numerous business research firms that gather enormous measure of information and disseminate it to organizations for an expense. This kind of information is called syndicated or partnered information. Distributed business research reports and different distributions are accessible from many associations which charge for their data. Such associations reserve an assortment of information, which is wide-ranging in its content, and they acquire income from selling this information. Syndicated information is generally utilized across various business areas.

For instance, by and large, TV-seeing information is gathered by an examination firm and disseminated to every one of the various telecasters for an expense. Also, customer disposition and purchase conduct information are likewise gathered by a business research firm for different item classifications and retail conditions. This firm then, at that point, gives the information to different item organizations.

- c) **The Internet Big Data-** Public and college libraries give a tremendous source of books, diaries, handbooks, research reports, periodicals, and reference books. Traditionally, specialist used to go through numerous hours looking through the racks of libraries to get the data they required. In any case, with development in the Internet, a ton of commendable material is typically accessible through the Internet. Essentially the references to such material can be effortlessly found. Essentially composing the watchwords can be sufficient to go over pertinent optional information through



normal web search tools. Advances in broadcast communications innovation empower individuals around the world to trade data rapidly and financially. There are specific web indexes and data sets which can give data intended for the examination needs.

For instance, Bloomberg is a supplier of nearby and worldwide monetary data, Econlit centres just around financial matters and the executives' distributions. A charge is by and large connected with admittance to most such information. Google Scholar is a public area web search tool which has become famous. It is not well versed in a specific subject. The web search tools can separate among scholastic and non-scholarly distributions. It likewise gives different choices like choosing field of exploration, composing in watchwords, phrases, and so forth to track down applicable data. Interpersonal organizations are likewise getting a significantly have an impact on how optional information is gathered. Utilizing informal communities, one can track down appropriate distributions and furthermore different researchers and experts.

One should comprehend that as how much data accessible on the net is unrealistic, one might wind up wasting a ton of life on looking through this enormous measure of information. In this way, it is prescribed to do altered look through unambiguous sources. Having a clear is likewise significant separation among real and produced information. One must just refer to trusted and reliable sources of information. Any information picked from a problematic source might bring blunders and make the research false.

ACTIVITY

Recently, a corporate retailer in Gujarat's territory launched its sixth location in one of the state's major cities. They had been concentrating solely on Gujarati area up to this moment. The organization's leadership was debating whether to expand the company into further Indian states. He had no idea how to go about making a decision. Offer the informational resources that can be used to learn a little bit about this company issue. Consider the direction that a secondary data analysis can provide.



CASE STUDY

Data Collection Method for a Small-Scale Enterprise

In the state of Maharashtra, a modest business had recently begun to prosper. This mobile phone company had offices in Maharashtra's five largest cities. The company's CEO, Sajeev, was debating whether to extend the company into further states. He was unsure about how to approach selecting a choice. Which states would be included in the enlargement, if it were necessary? He was most interested in understanding the market possibilities. He reasoned that it could be useful to survey the market to ascertain whether customers are content with their present handset providers. But it appeared to be a large effort. How on earth could he conduct such research for all of India? He wasn't sure if this was the best course of action. To assist him, he made contact with a management consulting company. The management consulting firm recognised the issue and advised him to begin the problem-solving process by doing secondary data-based research. From Sajeev, they gathered the internal data. They looked at information on sales, product categories, average revenue and profit, consumer trends, etc. Additionally, they examined secondary data on the sale of mobile phones in several Indian states at the same period. Secondary data was gathered on the pricing of mobiles in various states as well as the density of mobile usage by state. To narrow down the most suitable states with the greatest potential for the handsets made by Sajeev, all of the data was compiled and analysed. After narrowing the states down to two, Sajeev carried out extensive primary research in these two states to ascertain what the customer needs were for mobile phones. The outcomes of primary research were then used to determine the final cities to target and the handsets to launch. As a result, when a problem in business research becomes too big, secondary data study is recommenced. The research can be used to optimise the current issue. It is possible to cut down on the time and resource requirements. If primary research is necessary to gain additional insights, it can be done.



IN-TEXT QUESTIONS

11. The data which is available on the web is called the _____ data.
12. The data collected using secondary data can be availed with two sources- _____ and _____.
13. The companies which are made specifically to sell the secondary data provides _____ data.
14. Sales data comes under primary data collection. True/False.
15. Financial Reports come under secondary data collection. True/False.

3.5 CONDUCTING REVIEW OF LITERATURE

3.5.1 Meaning of Literature Review:

A Literature or narrative review is a far-reaching audit and examination of the distributed writing on a particular subject or exploration question. The literature that is explored contains books, articles, scholarly articles, meeting procedures, affiliation papers, and expositions. It contains the most appropriate investigations and focuses on significant past and flow exploration and practices. It gives a foundation and setting and shows how your research will add to the field.

A literature review ought to:

- Give a complete and refreshed review of the writing
- Make sense of why this review has occurred
- Articulate a position or theory
- Recognize and represent clashing and verifying perspectives

By basic definition, a literature review suggests a detailed summary of past inquiries that was led on a given point. This sort of work includes the reviewing of books, insightful articles, and whatever other sources are considered pertinent to a given area of study. This paper is intended to process, portray, sum up, and equitably do an assessment and explanation of the past investigation. The literature review ought to help us as a writer to figure out the idea of the research by offering a hypothetical base for the exploration obviously. The audit thinks about crafted by the past writers and in this manner guarantees the audience that basically everything has been simply perceived. It is, in this way, vital to choose the best essential and auxiliary sources of writing review preceding procedure to the genuine composition.



These sources of writing review in research show that we have obviously perused, analysed, and evaluated the first work and that you can create a significant paper. In this sort of work, we can likewise utilize hypothetical versus empirical research.

Theoretical vs Empirical Definition:

While managing empirical research, we concentrate around information assortment and summing up. The best summary is a model, and the best models are believed. The utilization of a particular model to direct our investigation is included as a hypothetical examination. We can likewise utilize the hypothetical structure in meta-examination. This is a sort of work that has the sole means to consolidate the experimental discoveries from existing inter-firm resourcefulness literature. It is additionally vital to see this multitude of ideas as well as different basics of writing to compose these papers totally and effectively.

A portion of the components of the literature review that we should simply comprehend considering our topic will incorporate; types of literary research, steps in the literature review, tertiary literature, primary and secondary sources of literature review, and how to best apply them in your paper, leading fitting survey of auxiliary writing, and hereafter. So, every one of the significant ideas and rules utilized will assist us with creating a great literature review that can help our audience obviously comprehend what different writers found on a point and how our argument is reasonable. Keeping in mind, if we can't see every one of the ideas, great organizations that help understudies, for instance, Peachy Essay, are dependably prepared to come to our help. The organization can offer incredible models for this sort of review including sources of literature review PPTX, sources of literature review PDF, methodology PDFs, and theoretical articles which can all assist us with better grasping this sort of work.

3.5.2 Purpose of Literature Review:

A literature review can be composed as an introduction to a review to:

- Exhibit how a review fills a gap in the research
- Contrast a review and other research that has been finished

Or on the other hand, it may be a different work (an exploration article all alone) which:

- Puts together or portrays a point
- Depicts factors inside a specific issue/issue



3.5.3 Limitations of Literature Review:

A portion of the restrictions of a literature review are:

- a) **Snapshot in Time**-It is a depiction of time. Dissimilar to different surveys, this one has starting, a centre, and an end. There might be future improvements that could make your work less important.
- b) **Too Focused**- It could be excessively engaged. Some specialty studies might miss the master plan.
- c) **Difficulty of Comprehensiveness**- Being comprehensive can be troublesome. It is basically impossible to ensure all the writing on a point was thought of.
- d) **Easy to be Biased**-It is not difficult to be one-sided in the event that you stick to top-level diaries. There might be different spots where individuals are publishing commendable research. Hope to open access publications and meetings to mirror a more comprehensive assortment. Additionally, make a point to incorporate contradicting views (and not simply supporting proof).

3.6 QUANTITATIVE RESEARCH AND QUALITATIVE RESEARCH

3.6.1 Quantitative Research:

Quantitative research depends on the amount estimation. It is an observational examination to create also, test speculations and hypotheses by estimating the quantitative properties. In this manner, it is appropriate to those peculiarities that can be communicated as far as the number of Statistical models utilized in quantitative research. Quantitative research utilizes organized inquiries with foreordained reaction choices and includes an enormous number of respondents. For instance, all out deals of the industry as far as worth in INR and volume in a huge number of pieces can be read up for a current year and contrasted and authentic information to estimate the deal in later years. The number of individuals who can review the brand with practically no guide can be examined and contrasted and the buyers who could do a helped review.

3.6.1 Qualitative Research:

Qualitative research is worried about peculiarities connecting with or including quality or kind. Qualitative research includes gathering, investigating, and interpreting data by observation. Qualitative research includes the implications, definitions, qualities, images, metaphors, and description of things. Qualitative research is substantially more abstract and



exploratory in nature. Attitude or assessment research to figure out how individuals feel about a specific subject or establishment is an illustration of qualitative research. Qualitative research is to find the hidden thought processes of the human way of behaving are significant in conducting sciences. Strategies, for example, word association tests, sentence completion tests, and story completion tests are frequently utilized in qualitative research.

Qualitative research can be additionally grouped into three kinds-phenomenology, ethnography, and case study.

- In phenomenology, the experience of at least one people is read up for a peculiarity of interest. For instance, research might be completed to concentrate on the experience of people who have done internet shopping through their cell phones.
- Ethnography concentrates around portraying the culture of a gathering by concentrating on their common credits, values, standards, rehearse, languages furthermore, and material things. For instance, an investigation of the way of life and the word related rehearses in Nagaland can be named as ethnography study.
- The Case Study is concentrated around doing a detailed examination of at least one recognized instance of interest. Doing an exploration to concentrate on a particular unit of disaster management the executives' force and its activities can be named as case study research.

IN-TEXT QUESTIONS

11. Conducting review of literature is the heart of research. True/False.
12. _____ research depends upon the estimations of the number.
13. Ethnography is the part of Quantitative research. True/False.
14. _____ research depends upon the features or qualities.
15. Case Study is a part of Qualitative research. True/False.

3.7 SUMMARY

In the lesson we have learnt the two important methods of collecting data. The choice of selection of method to collect will depend upon the different requirements of the researcher. Data is an essential commitment to any investigation or research. Thusly, a grouping of data becomes one of the fundamental bits of assessment. Variety of data so much that the data



gives the normal information is essential. In this part, we have discussed the basic pieces of information and information assortment. Also, we have learnt the meaning, importance and purpose for conducting review of literature which is fundamental and the heart of the research.

3.8 GLOSSARY

Snapshot of Time: It refers to the time taken.

Comprehensiveness: It means to perceive and understand anything.

3.9 ANSWERS TO IN-TEXT QUESTIONS

1. Primary Data	9. False
2. True	10. True
3. Secondary Data	11. True
4. False	12. Quantitative
5. Long	13. False
6. Published	14. Qualitative
7. Internal and External	15. True
8. Syndicated	

3.10 SELF-ASSESSMENT QUESTIONS

1. Differentiate between primary and secondary data. What are their advantages and disadvantages of each type of data?
2. What are the various sources of secondary data?
3. What are the main advantages and disadvantages of secondary data?
4. What are the different methods of collecting data?
5. Describe the merits and limitations of the observation method in data collection.



3.11 REFERENCES

S, S. (2020, July 13). *Difference Between Primary and Secondary Data (With Comparison Chart)*. Key Differences. <https://keydifferences.com/difference-between-primary-and-secondary-data.html>

LibGuides: Literature Reviews: What is a literature review? (2020). *. <https://guides.lib.utexas.edu/c.php?g=1060589&p=7710318>

Case Study on Primary & Secondary Market Research - O2I. (2019). *. <https://www.outsource2india.com/kpo/SuccessStories/PrimarySecondaryResearch.asp>

3.12 SUGGESTED READINGS

Dewen, S. D. H. K. |. (2022). *Business Research Methods* (1st ed.). CENGAGE INDIA.

Das, S. A. M. N. K. (2019). *Marketing Research An Applied Orientation Revised Edition*. Pearson India.

C. (2022a). *Business Research Methods* (11th ed.). Mc Graw Hill India.

Smith, S. M., & Albaum, G. S. (2004b). *Fundamentals of Marketing Research* (1st ed.). SAGE Publications, Inc.



LESSON-4

QUALITATIVE RESEARCH TOOLS

Dr. Ameet Sao
Assistant Professor
RICS SBE
Amity University
ameetsao@gmail.com

STRUCTURE

- 4.1 Learning Objectives
- 4.1 Learning Objectives
- 4.2 Introduction
- 4.3 Qualitative Research Process
 - 4.3.1 Qualitative Research Approaches
 - 4.3.2 Qualitative Research Methods-Depth interview, focus group, projective technique, observation method, survey and questionnaire, pilot testing.
 - 4.3.3 Analysing qualitative data
- 4.4 Advantages, Disadvantages and frequently asked questions
 - 4.4.1 Benefits of Qualitative Research
 - 4.4.2 Disadvantages of Qualitative Research
- 4.5 Frequently asked questions.
- 4.6 Summary
- 4.7 Glossary
- 4.8 Answers to In-text Questions
- 4.9 Self-Assessment Questions
- 4.10 References
- 4.11 Suggested Readings

4.1 LEARNING OBJECTIVES

After studying this lesson, you should be able to:

- Define qualitative research
- Explain the importance of qualitative research
- Analyze qualitative data



- Use qualitative tools in future research.

4.2 INTRODUCTION

Qualitative research is gathering and evaluating non-numerical data (such as text, video, or audio) in order to better comprehend concepts, views, or experiences. It can be utilised to gain in-depth insights into a topic or to develop fresh research ideas. The opposite of qualitative research is quantitative research, which includes gathering and interpreting numerical data for statistical analysis. In the humanities and social sciences, qualitative research is often employed in areas such as anthropology, sociology, education, health sciences, history, and so on.

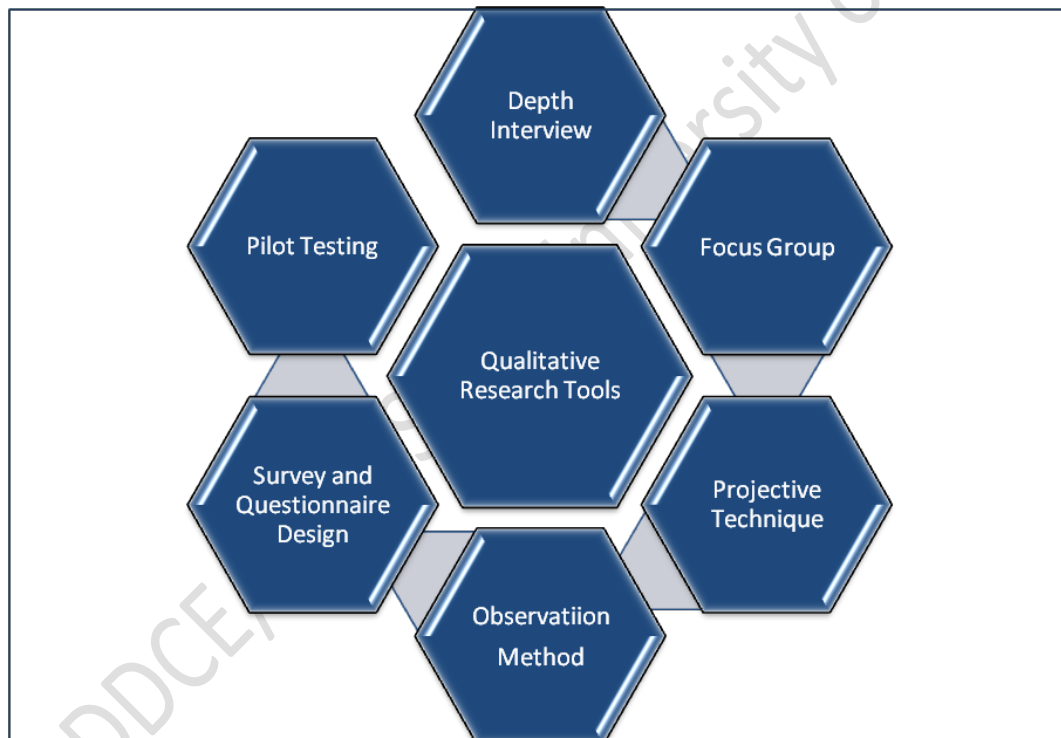


Fig 1.1: Qualitative Research Tools

Qualitative research can help in the following business situations:

- **Market Research:** Analyzing customer interest in a company's new idea in a certain demographic.
- Understanding the genuine demand of the end customer is essential for new product development.



- **Creative Development Research:** What should be stated and how should it be said in relation to branding.
- **Diagnostic Studies:** Determine how the company's category or brand is performing in comparison to the offers and image of competitors.

4.3 QUALITY RESEARCH PROCESS

Qualitative research provides critical insights into why people feel or behave the way they do. It aids in the identification and perception of underlying views, behavioural patterns, and motivations. Qualitative research assists in the formation of hypotheses for further investigation or quantification. In some ways, it enriches the knowledge acquired by quantitative research by comprehending rather than measuring. To summarise, the qualitative method aids in deciphering the consumer's less logical and more emotional decision-making nature.

To accomplish this, it follows a process covering

1. Quality Research Approaches
2. Quality Research Tools
3. Analyzing Quality Data.

4.3.1 Quality Research Approaches

Qualitative research is done to learn more about how individuals see the world. While there are several ways to qualitative research, they all have the flexibility of preserving rich meaning when evaluating data. Grounded theory, ethnography, action research, phenomenological research, and narrative research are all common methodologies. They have certain similarities, but emphasise distinct goals and points of view.

Inductive theory -development is used by researchers to obtain rich data on a topic of interest.

Ethnography- To understand the cultures of individuals or organisations, researchers immerse themselves in them.

Participatory action research- To generate social change, researchers and participants collaborate to connect theory and practise.

Phenomenological – A study that investigates a phenomena or event by reporting and analysing the lived experiences of individuals.



Narrative study- investigates how tales are delivered in order to better understand how participants interpret and make sense of their experiences.

4.3.2 Qualitative Research Tools

Each study methodology involves the use of one or more data gathering methods. Some of the most frequent qualitative approaches are as follows:

Depth Interview- one-on-one interactions in which you individually ask them questions.
Projective Technique- Respondents can project their real or subjective ideas and beliefs onto other persons or even inanimate things using projective methods.
Focus Group- a group of people who are asked questions and have a conversation.
Observation Method- make thorough field notes of what you see, hear, or encounter.
Content analysis: Content analysis involves analyzing and interpreting written or visual material, such as documents, social media posts, or video recordings, to identify patterns and themes.
Observation: In observational research, researchers observe and document behavior and interactions in natural settings, such as a workplace or community setting. Observation can be either participant (where the researcher is an active participant in the setting) or non-participant (where the researcher observes from a distance).
Narrative analysis: Narrative analysis involves analyzing and interpreting the stories that people tell about their experiences, with the goal of understanding the meanings and values that underpin those stories.
Survey and Questionnaire Design- sending out questionnaires with open-ended inquiries.
Pilot Testing- a trial run for your research study that enables you to evaluate your methodology with a small sample size before conducting your major study.

Another approach that is widely practiced is Case-based research, which can fall under both qualitative and quantitative research methods, depending on how the research is conducted and analyzed.



If case-based research involves a detailed examination of a specific case or cases in order to understand a particular phenomenon or to develop a theory, it is typically considered a qualitative research method. In this approach, the researcher collects data through observation, interviews, or other forms of data collection, and analyzes the data using methods such as coding, thematic analysis, or grounded theory. The goal of qualitative case-based research is to generate rich, detailed, and contextualized information about a particular case or cases, which can provide insights into broader patterns or trends.

However, if case-based research involves collecting data from a sample of cases and using statistical methods to analyze the data, it may be considered a quantitative research method. In this approach, the researcher typically collects data through surveys or other standardized instruments, and uses statistical methods such as regression analysis or factor analysis to identify patterns and relationships among the variables. The goal of quantitative case-based research is to generalize findings to a larger population of cases, and to test hypotheses or theories using statistical methods.

Overall, the classification of case-based research as qualitative or quantitative depends on the research design, data collection methods, and analysis techniques used. Some research may fall in between these two categories, or may involve a combination of both qualitative and quantitative methods.

Examples of qualitative case based approach and quantitative case based approach

Qualitative Case-Based Research:

A market research firm conducts a series of in-depth interviews with customers who have recently purchased a new smart phone, with the goal of understanding their decision-making process and the factors that influenced their choice of brand and model. The researchers use a qualitative approach to analyze the data, coding the interviews for themes and patterns in the customers' responses. The insights gained from this research can be used to inform the development of new marketing strategies and product features.

A company conducts a series of focus groups with customers who have recently switched from a competitor's product to their own, with the goal of understanding the reasons behind the switch. The researchers use a qualitative approach to analyze the data, identifying key themes and issues that emerged from the discussions. This information can be used to develop targeted marketing messages and product features that address the specific needs and concerns of this group of customers.

Quantitative Case-Based Research:



A company surveys a representative sample of customers who have recently purchased a new car, collecting data on their demographics, buying behavior, and brand preferences. The company uses a quantitative approach to analyze the data, using statistical techniques such as regression analysis to identify the factors that are most strongly associated with customer satisfaction and brand loyalty. The insights gained from this research can be used to guide future marketing campaigns and product development efforts.

A retailer collects data on sales and customer traffic at each of its store locations over a six-month period, with the goal of identifying patterns and trends in customer behavior. The company uses a quantitative approach to analyze the data, using statistical methods such as time series analysis and clustering to identify differences in customer behavior across different stores and regions. The insights gained from this research can be used to inform decisions about store layout, staffing, and product placement to optimize sales and customer satisfaction.

4.3.2.1 Depth Interview

An intensive series of one-on-one interviews with a small number of respondents is used as part of the qualitative research approach known as "in-depth interviewing" to learn about their opinions on a certain concept, initiative, set of emotions, behaviour, or circumstance. Typically, it involves direct, one-on-one interviews. For this, respondents are asked for a thorough history, and elaborate information about their beliefs, values, motivations, expressions, feelings, etc. is gathered. Their non-verbal cues are also noticed. They require more time to record lengthy observations. This makes it possible to record detailed descriptive information on people's actions, attitudes, and perceptions of a complex process as it develops.

These are carried out to cater to each person's replies. The questions will change depending on the type of responses provided. Even the environment of the interview affects the respondents. The relationship that interviewers build with their subjects is crucial to the interviews' success.

For instance, ask "Please describe the clinic's services" rather than "Do you know about the clinic's services?" • Fact-based inquiries should come before those that solicit opinions. For instance, before asking "What did you think of the events?" start by asking "What activities were conducted?"

The Benefits of a Depth Interview

1. Lots of information is given.



2. The information is relative more correct.
3. Since the interviewer and respondent have developed a personal rapport, private or intimate topics may also be discussed.

Negative aspects of depth interviews

1. Due to the non-standardization of the interviewers, generalisation is challenging
2. There is a potential of prejudice because the interviewer's performance is crucial to success.
3. Data analysis is time-consuming.

4.3.2.2 Projective Techniques

Projective Techniques are unstructured and indirect methods of investigation developed by psychologists that use projection of respondents to infer about hidden motives, urges, or intentions that cannot be secured through direct questioning because the respondent either refuses to reveal them or is unable to figure them out himself.

These strategies are beneficial in allowing respondents to express themselves without fear of personal shame. These strategies assist respondents in automatically projecting their own attitudes and feelings on the subject under examination. Thus, projective techniques are useful in motivational research and attitude surveys.

Important Projective Techniques

1. Word Association Test.
 2. Completion Test.
 3. Construction Techniques.
 4. Expression Techniques.
- **Word Association Test:** A clue or hint is presented, and the respondent is asked to state what first comes to mind. The association may appear as a word or a picture. The same thing can be interpreted in a variety of ways. You are given a list of terms, and it is unclear which word they are most drawn to. The interviewer writes down the respondents' comments that express their true feelings. For the researcher, it is crucial to know how frequently a word receives a response and how long it takes before receiving one. For instance, 20 out of 50 responders connected the words "Fair" with "Complexion."



- **Completion Test:** In this type of exercise, participants must finish a sentence or a story that has been left unfinished. Their attitude and mental state will be reflected in the finished product.
- **Construction Test:** This resembles a completion test in some ways. You may be requested to create a tale about a photograph that has been presented to you. Like the completion test, the original structure is constrained and lacking in specificity. A dialogue for two cartoons, for instance, needs to be written.
- **Expression Techniques:** In this exercise, participants are asked to describe how others feel or act.

4.3.2.3 Observation Method

The observation approach entails watching, either manually or mechanically, what people really do or what happens when they buy or consume something. Information is gathered by watching processes in action. A few examples are as follows: -

1. **Service Stations** – Visit one as a client and have a look around.
2. To determine how well Dunlop mattresses are shown. In a department shop, the observer records:
 - (a) How many people pass by;
 - (b) how many people pause to look at the display; and
 - (c) How many people decide to buy?
3. **Supermarket:** Which shelf position is best? They utilise hidden cameras.
4. To establish typical sales arrangements and discover the level of sales excitement displayed by distinct salesmen – Typically, an investigator does this task utilising a concealed camera.

Advantages of the Observation Method

1. The researcher does not have to rely on the desire or capacity of respondents to describe events properly if they observe and record occurrences.
2. The interviewer's prejudice is either eliminated or diminished. Thus, data gathered by observation are often more accurate and objective.

Problems/Disadvantages with the Observation Method

1. The biggest drawback of using the observation approach is the inability to examine things like attitudes, motivations, customers' or consumers' states of mind, their reasons for purchasing items, and their images.



2. It takes time for the investigator to wait for a certain event to occur.
3. Questionnaires make it easier to discuss private and intimate behaviours than it does to actually witness them, such as late-night television watching.

The observation method's fourth drawback is its cost. Observational data are typically more expensive to gather than other survey data. Between events to be watched, the observer must wait while doing nothing. Unproductive time results in higher costs.

4.3.2.4 FOCUS GROUPS

Group talks or group interviews are other names for focus groups. They are employed to comprehend the mindset or behaviour of the target audience. One or two moderators are chosen who guide the discussions—are chosen. 6-10 participants are allowed to be a part of the study. In the event that there are two moderators, they will take opposing stances. The moderator presents the subject then changes the subject to research. It aids in gathering first responses to marketing initiatives or understanding consumer perceptions of a novel product idea. These moderators oversee the conversation. The group is being seen from nearby rooms. These conversations are recorded using a variety of tools.

Goals of a Focus Group

1. To see how various phenomena are perceived by humans
2. To acquire primary data for a study project
3. To assist in the development of questionnaires for survey research
4. To test mainly plans or ideas.
5. To comprehend the causes of a certain phenomena.

Procedures for running focus groups

1. Identify the issue
2. Decide on a sample
3. Estimate the required number of groupings (minimum number should be two)
4. Prepare the mechanics of the research. Set up the responders in the area where the focus group will be gathered.
5. Choose and brief the moderators.
6. Get the focus group materials ready.
7. Lead the meeting.



8. Analyze the data and create a report summary.

Negative/Disadvantages aspects of focus groups

1. Quantitative data collection is not acceptable.
2. The self-designated group leader may push his or her viewpoint on the other participants. Users may be restricted by moderators.
3. It mainly depends on the moderator's abilities.
4. The focus group respondents could or might not be representative of the demographic from which they were selected.
5. Respondents may be constrained by recording devices. The location of the recording equipment is crucial.

4.3.2.5 Survey Method

The survey method is a way of acquiring data that involves posing questions to persons who are deemed to be knowledgeable. It is formalised to create a list of questions. Typically, a straightforward strategy is employed. Questions on the respondents' demographic interests and opinions are asked with them.

Benefits of the Survey Method

1. Surveys produce a wider range of information than other approaches (direct observation, testing). Surveys are useful for gathering data about socioeconomic traits, attitudes, views, and motivations as well as for determining product features, advertising media, sales promotion, distribution networks, and other marketing factors.
2. In general, questioning is quicker and less expensive than observation.
3. It's easy to administer questions.
4. Reliable data
5. The outcomes are less variable.
6. It is comparatively easy to interrelate, evaluate, and cite data gathered using the survey approach.

Disadvantages of the Survey Method

1. Respondents' unwillingness to divulge information—This necessitates some persuasive interviewing on the interviewer's behalf. The interviewer may use the promise that the information would be kept private or the tactic of giving out gifts.



2. The respondents' inability to offer information may be caused by one of the following:
a. ignorance; b. forgetfulness; c. failure to recognise their motivations and give "reasons why" for their actions.
3. The respondents' human biases are present, such as "Ego."
4. There are Symantec challenges; it is challenging, if not impossible, to phrase a question so that it would convey the same thing to every answer. The identical question will typically provide different results if it is phrased in two distinct ways.

How to get over Survey Method drawbacks

1. Careful question framing and wording.
2. Careful oversight of data collection by using investigators with specialised training who will watch closely and report on minor reactions of those interviewed.
3. Careful interpretations based on an understanding of the data's precise meaning and a clear understanding of its limits. This is particularly accurate when people answer questions like, "What price would you pay for this product?"

Techniques of Survey Method

1. Initial Telephonic Interview
2. Personal Interview
3. Mail Interview
4. Electronic Interview

Initial Telephone Interview-The greatest way for swiftly obtaining the information needed is interviewing. The researcher uses the telephone to obtain responses from the respondents.

1. Telephonic interviews provide a number of benefits, starting with their speedy data collecting.
2. It offers the interviewer the opportunity to speak with one or more people and explain his questions if necessary, which is a benefit over the "Mail Questionnaire."
3. Telephone interviews appear to have a somewhat higher response rate than postal questionnaires.
4. The information is of higher quality.
5. It is less expensive and there are less administrative issues.



Advantages

1. Telephonic interviews provide a number of benefits, starting with their speedy data collecting.
2. It offers the interviewer the opportunity to speak with one or more people and explain his questions if necessary, which is a benefit over the "Mail Questionnaire."
3. Telephone interviews appear to have a somewhat higher response rate than postal questionnaires.
4. The information is of higher quality.
5. It is less expensive and there are less administrative issues.

Disadvantages

1. It can't manage a free-form interview
2. It cannot be utilised for inquiries that need for lengthy, descriptive responses.
3. Respondents are unable to be seen
4. Callers are hesitant to give out personal information
5. You cannot contact those who do not have access to a telephone.

Personal Interviewing- It is the most adaptable of all approaches. They are employed when props are needed in addition to the vocal answer. Nonverbal reactions are also noted.

Personal Interview Benefits

1. The interviewee has the option of asking extra questions and supplementing the interview with personal observations.
2. They are more adaptable. The order of the questions can be modified.
3. It is conceivable to have knowledge of both the past and the future.
4. Extensive research is possible.
5. It is feasible to validate data from different sources.
6. The information acquired is quite accurate and dependable, and it aids in the early identification of cause and effect relationships.

Personal interviews have the following disadvantages:

1. They involve significantly more technical and administrative planning and monitoring;
2. They are more expensive;
3. They are time demanding.



4. The interviewer has an impact on data accuracy.
5. Several call banks may be necessary.

3. Mail Poll/Survey

Questionnaires are distributed to respondents, who complete them and return them.

Advantages of Mail Survey

1. It may reach individuals of different backgrounds.
2. By providing particular incentives, the response rate can be increased.

Mail Survey Disadvantages

1. It cannot be utilised for uncontrolled learning.
2. It is expensive.
3. It necessitates an existing email list.
4. It takes a long time.
5. There is an issue with difficult queries.

4. Online/e- Interview

Rather than asking for information, electronic interviewing involves identifying and noting persons, things, and occurrences. For example, when you go to the supermarket, you see the products people like. The Universal Product Code (UPC) is another way to track what consumers buy.

The Benefits of an Electronic Interview

1. There is no reliance on the respondent's desire or ability.
2. The information is more precise and objective.

Electronic Interview Disadvantages

1. It is impossible to monitor attitudes.
2. Events that last a long time cannot be witnessed.
3. Observer bias exists. It's not entirely objective.
4. If respondents are aware that they are being watched, their responses may be influenced.
5. It is an expensive procedure.



4.3.2.6 Questionnaire Design

The process of creating the structure and questions for a survey instrument that will be used to gather information about a certain occurrence is known as questionnaire design. All the different phases of survey design and implementation should be taken into account while creating a questionnaire.

These include the following nine elements: (1) determination of goals, objectives, and research questions; (2) definition of key concepts; (3) generation of hypotheses and proposed relationships; (4) choice of survey mode (mail, telephone, face-to-face, Internet); (5) question construction; (6) sampling; (7) questionnaire administration and data collection; (8) data summarization and analysis; (9) conclusions and communication of results.

Structured and unstructured questionnaires are the two main categories. The quasi-structured questionnaire, which is frequently used in social science research, is a combination of these two.

A questionnaire is a type of research tool used to collect data from respondents by asking them a series of questions. One way to think of questionnaires is as a type of written interview. They can be conducted in-person, over the phone, online, or by mail.

A set of questions used to gather information about someone or something is called a questionnaire. Finding trends and patterns or doing statistical analysis are not employed with it. An illustration would be when you go for a checkup or join a gym and are required to respond to a series of questions on your present physical health.

When conducting a survey, the researcher employs a questionnaire to get data from the participants in order to formulate research questions. A questionnaire is a very practical tool for quickly gathering data from a big number of individuals. Therefore, the questionnaire's design is crucial to ensuring that correct data is gathered and that the findings can be interpreted and extrapolated.

A survey can take many different forms, including postal surveys, phone interviews, in-person interviews, and online surveys. The design has to be slightly different for each kind of survey. A self-administered questionnaire should contain extremely clear instructions and questions, follow a logical order, and avoid complicated filters. An example of this would be a mail survey. Without urging from the interviewer, the respondents are more likely to provide honest responses. On the other hand, with a questionnaire conducted by an interviewer (such as a face-to-face interview or telephone interview), the questions might be more complicated since the interviewers can clarify them.



Features of a questionnaire

- The order of the questions. To enhance the rate of answer to the questions, a correct succession of questions should be used, Unstructured questionnaires, structured questionnaires, closed-ended questions or dichotomous questions, open-ended questions, uniformity, exploratory, easy to grasp, and so on.

What is a good Questionnaire and how to frame good Questionnaire?

A good questionnaire should be valid, reliable, clear, interesting and concise.

Valid

A legitimate questionnaire should ask the questions it is intended to ask, meaning the questions should be written so that the responder can comprehend their purpose. To do this, the "content expert" should examine the questionnaire during the pilot test (e.g. if the target respondent is a pneumonia patient, then a pneumonia patient should comment whether he understands the questionnaire). Until the question is fully understood, all ambiguities or questions should be addressed.

Reliable

If a responder is asked the same question repeatedly over the course of a short period of time, a valid questionnaire should produce the same response. This may be done by doing a "test-retest," which involves giving the responder the same questionnaire a second time and evaluating the answers for consistency. Any inconsistency in the replies may be the result of unclear questions, which should be evaluated and rephrased.

Concise

A concise questionnaire only includes questions that are intended to address the study goals. Any inquiries outside the purview of the study have to be disregarded. Researchers frequently "spread the net wider" in order to gather more information, whether or not that information is valuable. This typically occurs when the study objectives have not been carefully considered by the researcher. There is a chance that it will ask too many questions and the questionnaire will be lengthy.

Interesting

A more engaging questionnaire increases the likelihood that the responder will finish it, improving the response rate. In order to do this, the researcher must carefully consider which questions to pose in what order and if they are pertinent to the responder.



Constructing a theoretical framework

Building a conceptual framework is the first step in creating an effective questionnaire. The researcher must be extremely explicit about the "dependent" and "independent" aspects he plans to look into as well as his research questions. This framework is crucial in ensuring that all pertinent variables are included in the study and that any extraneous ones may be left out.

Asking 'Right' questions

It's time to create the questions such that they are legitimate and trustworthy now that you have built the conceptual framework and are clear on the questions you want to ask. The initial research questions must be generated by the researchers.

Open-ended vs. closed-ended questions

These queries may be "close-ended" or "open-ended," depending on the situation (Table 1a). "Closed-ended" inquiries provide respondents alternatives and demand that they select one or more things from a list. "Open-ended" inquiries provide the respondent the freedom to express their thoughts without being constrained by the available alternatives. The former is favoured if the alternatives are few and the range of responses is well known; the latter is chosen if the options are many and the range of answers is unknown. Prior to analysis, the responses to the open-ended questions need to be regrouped.

Examples of open and closed ended questions

Open-Ended- Here are two examples of open-ended questions that could be used in marketing research:

"What words or phrases come to mind when you think of our brand?" or What is/are the main reason(s) for you to purchase goods from shopping mall?

This open-ended question encourages respondents to provide their own thoughts and associations about a company's brand. The answers to this question can provide insights into how the brand is perceived by consumers, and what specific attributes or qualities are associated with it.

"Can you describe a time when you had a particularly positive experience with our product/service?"

This open-ended question encourages respondents to share a detailed account of their experience with a company's product or service. The answers to this question can provide insights into what specific aspects of the product/service are most valued by consumers, and how the company can continue to provide positive experiences for its customers.

**Closed Ended-**

Here are two examples of closed-ended questions that could be used in marketing research, along with their relevance:

"On a scale of 1-10, how likely are you to recommend our product to a friend?"

This closed-ended question asks respondents to rate their likelihood of recommending a product to others on a scale. The answer choices are limited to a specific range of values, which makes it easier to quantify the results and analyze them statistically. This type of question is useful in measuring customer satisfaction and loyalty, as well as identifying areas for improvement in the product or service.

"Which of the following best describes your current employment status: employed full-time, employed part-time, self-employed, unemployed, student, or retired?"

This closed-ended question asks respondents to choose one option from a list of pre-determined answer choices. The answer choices are mutually exclusive and collectively exhaustive, which means that they cover all possible options and do not overlap. This type of question is useful in gathering demographic information about a target audience, which can help inform marketing strategies and messaging.

The relevance of closed-ended questions in marketing research lies in their ability to provide structured and quantifiable data, which can be analyzed more easily and systematically than open-ended questions. Closed-ended questions also reduce the potential for response bias and ensure that all respondents are interpreting the question in the same way, which improves the reliability of the results. However, closed-ended questions can be limiting in terms of the depth and nuance of the responses they elicit, and they may not capture the full range of opinions and perspectives that respondents hold. Therefore, a combination of open-ended and closed-ended questions is often used in marketing research to gather both qualitative and quantitative data.

Likert Scale- A scale with a range of replies is preferable than a yes/no answer for gauging attitudes or providing opinions. A 5-point or 7-point Likert scale is a regularly used approach.

It measures the strength of a given attitude or conviction. It is feasible to compute mean scores for every given set of statement replies (item scores)

We should always visit mall for purchasing goods.

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



Options/choices

Each question's alternatives should be as comprehensive as feasible. This ensures that the reply may pick an option that best fits his response. The researcher must brainstorm, analyse related published research, consult with experts, and, if required, conduct a focus group discussion with the target respondents to discover the viable possibilities. The researcher might include a "Other: please specify _____" category as one of the alternatives to allow for other conceivable options.

Filtering

Filtering is used in a questionnaire with multiple components, some of which the responder does not need to answer, to direct the respondent to answer just the important items. However, you should avoid applying too much filtering because it may confuse respondents and complicate the questionnaire.

Order of the questions

The questions should be asked in a sensible order. Begin with easy questions before progressing to more challenging ones. Some like to begin with the respondents' socio-demography, while others prefer to leave it till the end because it covers more personal questions such as household income, education level, and religion. However, this is dependent on how forward-thinking the target group is. To normalise the "sensitive" inquiry, explain to the respondent why you are asking the personal question or make a generic comment.

Avoid asking unclear questions.

When asking a question, be as detailed as possible. Terms like "often," "always," and "frequently" may signify various things to different individuals.

Create the questionnaire with an eye towards analysis.

It is critical to anticipate what approach will be used to analyse the data obtained while constructing a questionnaire. Consider the concept of age. If the goal of the age question is to determine the average age of the participants, then a precise age should be collected (e.g., "What is your age? (at your last birthday): _____ years"). If, on the other hand, you want to categorise them according to different age groups throughout the analysis, you may wish to frame the question accordingly (e.g., "What is your age? (at your latest birthday): 19, 19-29, 30-40, etc.)



Format

The final "touch-up" of the questionnaire is critical since the "appearance" of the questionnaire may determine whether or not the responder would fill it out. This is particularly important for consumer surveys. The title should be prominent and represent the main goal of the research. Divide the questionnaire into sections based on the content if feasible (e.g., boxes with bold titles), and it should flow smoothly from one area to the next with suitable filtering. Finally, a cover letter outlining the purpose of your research, your connections, and, if applicable, assuring confidentiality and how you intend to use the data you have gathered.

Pilot study

A pilot test is an important phase in the questionnaire design process before data gathering begins. It will aid in detecting faults in the questionnaire's content, grammar, and presentation. First, invite coworkers, relatives, or friends to provide feedback on the questionnaire. This will detect any errors in content, grammar, and format. The potential responders should next be asked to complete the questionnaire and offer comments. During the pilot test, you should particularly ask respondents to remark on any questions that you believe are unclear or sensitive.

What exactly is pilot testing?

Pilot testing encompasses a variety of actions that allow you to examine various areas of your project ahead of time. This may also be referred to as a test run or rehearsal to assist you identify any flaws or gaps in your strategy before implementing it.

Three essential areas are involved in pilot experiments: project modelling, observation, and analysis. To begin, you must reproduce real-life circumstances that convey the essence of your endeavour. If you're conducting research, for example, gather a sample size of your population and distribute your survey or questionnaire, depending on your data gathering method.

Throughout the procedure, you should keep a close eye on how things go. Do any of your survey questions cause participants difficulty? Is it possible to export survey findings to your project management software or spreadsheet? Consider speaking with important stakeholders to gain direct input from them, and document what you learn.

The last stage involves organising and evaluating the data you've collected and using the results to improve your project. For example, if you're creating a new product, this



knowledge will assist you enhance important aspects to better suit the expectations of your consumers.

Advantages of Pilot Testing in Research

1. Pilot testing assists you in determining the viability of your study approach.
2. It reveals how to best allocate various resources during your methodical inquiry.
3. Pilot experiment data can help you establish your major research topic.
4. Pilot testing can be used as a type of baseline survey.
5. Pilot trials allow researchers to customise their systematic study to the demands of their research population.
6. You can learn new research approaches.

Questionnaires for Pilot Testing

1. How would you grade this procedure on a scale of 1 to 10?.....
2. What was your favourite part of this process?.....
3. What didn't you like?.....
4. How simple were the questions to understand for you?
Very Easy • Somewhat Easy • Neither Easy Nor Difficult • Difficult • Extremely Difficult
5. Do you comprehend the research's purpose? Yes \ No
6. How long did it take you to complete the research task?
10 minutes • 20 minutes • 30 minutes • More than 30 minutes
7. What were the three most important findings from the research?
8. Were there any parts of the procedure that you found rude or judgmental? Yes \ No
9. What features of this product or project would you want to see?.....
10. Do you want to know how we utilise this information?
• **Certainly** • **Not at all**

Because all observations, interpretations, and analyses are done via the qualitative researcher's personal perspective, they typically see themselves as "instruments" in the research process.

This is why, while writing up your methodology for qualitative research, it is vital to analyse your technique and completely clarify the judgments you made when acquiring and interpreting the data.



4.3.3 Qualitative data analysis

Texts, images, videos, and audio may all be used to collect qualitative data. You may be working with interview transcripts, survey results, field notes, or recordings from natural settings, for example.

The following five steps are common to most forms of qualitative data analysis:

1. Gather and arrange your data. This might include interview transcription or putting up field notes.
2. Examine and investigate your data. Examine the data for patterns or recurring concepts.
3. Create a data coding scheme. Create a collection of codes that you may use to categorise your data based on your first thoughts.
4. Label the data using codes. In qualitative survey analysis, for example, this may entail going over each participant's replies and marking them with codes in a spreadsheet. You may develop additional codes to add to your system as you browse through your data.
5. Look for repeating motifs. Connect codes to form cohesive, overarching themes.

Examples with Methods

Method	Qualitative data analysis	
Content Analysis	In qualitative data, characterise and categorise prevalent words, phrases, and concepts.	A market researcher might do content analysis to determine the terminology used in software app descriptions.
Thematic	Identify and understand qualitative data patterns and themes.	A psychologist may use thematic analysis on trip blogs to investigate how tourism influences self-identity.
Textual Analysis	To investigate the text's substance, structure, and design.	A textual analysis might be used by a media researcher to assess how celebrity news coverage has changed over the last decade.



IN-TEXT QUESTIONS

1. study investigate to understand how respondents make sense of their experiences.
2. The goal of ----- research is to fully understand the essence of some phenomenon.
3. ----- type of research can be done for studying cultural anthropology.
4. In ----- techniques, participants are allowed to express their feelings

4.4 ADVANTAGES, DISADVANTAGES AND FAQ

4.4.1 Qualitative research has several advantages.

Qualitative research frequently attempts to maintain the voice and perspective of participants and is adaptable when new research topics emerge. Qualitative research is useful for the following purposes:

- **Flexibility**-As new ideas or trends arise, the data collecting and analysis method can be modified. They are not predetermined in advance.
- **Natural setting**-Data is collected in naturalistic or real-world settings.
- **Valuable insights**-In creating, testing, or enhancing systems or goods, detailed descriptions of people's experiences, feelings, and perceptions can be employed.
- **New idea generation**-Researchers might discover unexpected challenges or possibilities as a result of open-ended replies that they would not have considered otherwise.

4.4.2 Qualitative research's disadvantages

When evaluating and interpreting data, researchers must consider both practical and theoretical constraints. Qualitative research is plagued by:

- **Unreliability**-Because of uncontrollable factors that alter the results in the actual world, qualitative research is frequently unreliable.
- **Subjectivity**-Qualitative research cannot be reproduced since the researcher's major involvement is in data analysis and interpretation. In data analysis, the researcher



selects what is significant and what is unimportant, therefore interpretations of the same data might differ substantially.

- **Limited applicability**-Small samples are frequently employed to get extensive information on certain circumstances. Despite thorough research processes, drawing generalizable findings is challenging since the data may be skewed and unrepresentative of the larger population.
- **Labor-intensive**- Although software may be used to organise and record enormous quantities of text, data analysis is frequently done manually.

What types of questionnaire designs are there?

In research technique, what is a questionnaire?

4.4.3 Frequently asked questions about qualitative research

What is the distinction between quantitative and qualitative research methods?
What are the most common qualitative research methods?
What exactly is data collection?
How do you go about analysing qualitative data?

4.5 MAIN BODY SECTION-3

IN-TEXT QUESTIONS

1. What are the advantages and disadvantages of quality research?
2. State the relevance of Pilot study in a research?
3. Give an example where you can use focus group as qualitative research tool in research?

4.6 SUMMARY

Qualitative research is conducted to understand more about how people see the world. While there are several approaches to qualitative research, they all offer the flexibility of retaining rich meaning while assessing data. The key tools that are used to derive the output of qualitative research are-

Depth interview, focus group, survey and questionnaire, pilot testing, projective techniques.



4.7 GLOSSARY

Depth Interview- one-on-one interactions in which you individually ask them questions

Projective Technique- Respondents can project their real or subjective ideas and beliefs onto other persons or even inanimate things using projective methods.

Focus Group- a group of people who are asked questions and have a conversation.

Observation Method- make thorough field notes of what you see, hear, or encounter.

4.8 ANSWERS TO IN-TEXT QUESTIONS

- | | |
|---|--|
| <ol style="list-style-type: none">1. Inductive2. Phenomenological3. Observatory4. Expression | |
|---|--|

4.9 SELF-ASSESSMENT QUESTIONS

1. What types of questionnaire designs are there?
2. In research technique, what is a questionnaire?
3. Difference between Open ended and closed ended questions?
4. Discuss with suitable examples the relevance of Observation method in research?
5. What kind of time investment is needed to do a qualitative research study?

4.10 REFERENCES

Strauss, A., & Corbin, J. M. (1990). Basics of qualitative research: Grounded theory procedures and techniques. Sage Publications, Inc.

4.11 SUGGESTED READINGS

Kothari, C.R.. (2004). Research methodology : Methods and techniques (2nd revised edition). New Delhi: New Age International (P) Limited, Publishers.



LESSON-5

SAMPLE DESIGN

Dr. Ameet Sao
Assistant Professor
RICS SBE, Amity University
Email-ID : ameetsao@gmail.com

STRUCTURE

- 5.1 Learning Objectives
- 5.2 Introduction
- 5.3 Steps in Sample Design
- 5.4 Characteristics of a good sample design
- 5.5 Selecting a Sampling Procedure
- 5.6 Sampling Design Method
- 5.7 Self - Assessment Questions
- 5.8 Self-Assessment Questions
- 5.9 Summary
- 5.10 Glossary
- 5.10 Answers to In-Text Questions
- 5.11 Answers to In-Text Questions
- 5.12 References
- 5.13 Suggested Readings

5.1 LEARNING OBJECTIVES

- Understanding the differences between probability and non-probability sampling techniques.
- Recognizing the advantages and disadvantages of each sampling method.
- Identifying the appropriate sampling technique for a given research question.
- Knowledge of common probability sampling techniques (e.g. simple random sampling, stratified sampling, cluster sampling).



- Knowledge of common non-probability sampling techniques (e.g. convenience sampling, quota sampling, snowball sampling).
- Understanding the concept of sampling frame and its impact on the representativeness of a sample.

5.2 INTRODUCTION

Sample design refers to the process of selecting a subset of individuals from a larger population for the purpose of collecting data. The goal of sample design is to select a sample that is representative of the population and provides reliable information about the population parameters. Field work refers to the process of collecting data from the sample through various methods such as surveys, observations, or experiments. Both sample design and field work are critical components of any research study and the quality of the data collected can greatly impact the validity and reliability of the findings. Effective sample design and field work require careful planning and consideration of various factors such as research objectives, budget, time constraints, and ethical considerations.

Sample design is an important aspect of research as it can greatly impact the accuracy and generalizability of the results. There are two main types of sample design: probability and non-probability sampling. Probability sampling methods, such as simple random sampling, stratified sampling, and cluster sampling, use random selection to ensure each member of the population has an equal chance of being included in the sample. This type of sampling is commonly used when the population size is known and the goal is to make inferences about the population based on the sample data.

On the other hand, non-probability sampling methods, such as convenience sampling, quota sampling, and snowball sampling, do not rely on random selection. Instead, individuals are selected based on criteria such as availability, ease of access, or specific characteristics of interest. This type of sampling is commonly used when the population size is unknown or the goal is to provide a qualitative understanding of a specific phenomenon.

Field work involves the actual collection of data from the sample. This can be accomplished through various methods such as surveys, observations, or experiments. Effective field work requires careful planning and attention to detail, as the quality of the data collected will greatly impact the validity and reliability of the results. For example, in survey research, effective field work might involve creating clear and concise survey questions, pre-testing the survey, and properly training interviewers. In observational studies, effective field work might involve carefully defining observation categories and training observers to ensure



consistent coding. In experimental studies, effective field work might involve controlling extraneous variables and properly randomizing participants to treatments.

To conclude we can say that sample design refers to the plan and techniques used to select a representative sample from a population for research purposes. It outlines the method for selecting items for the sample, as well as the sample size. The sample design must be determined before collecting data and researchers should choose a reliable and suitable design for their study.

5.3 STEPS IN SAMPLE DESIGN

In developing a sampling design, the researcher should consider the following:

Define the population: Determine the group of individuals or items to be studied.

Select the sample design: Choose the appropriate method for selecting a sample from the population (e.g. random sampling, stratified sampling, etc.).

Determine the sample size: Decide on the number of items or individuals to include in the sample.

Implement the sample design: Apply the chosen method to select the sample from the population.

Review and adjust: Evaluate the results of the sample design and make any necessary adjustments.

5.4 CHARACTERISTICS OF A GOOD SAMPLE DESIGN

Characteristics of a Good Sample Design:

Representativeness: The sample should accurately reflect the characteristics of the population.

Ease of implementation: The sample design should be simple and practical to execute.

Adequate sample size: The sample size should be large enough to provide reliable results.

Avoid bias: The sample design should minimize the potential for bias in the selection process.

Cost-effectiveness: The sample design should be cost-effective and efficient.

Flexibility: The sample design should allow for adjustments and changes if necessary.



IN-TEXT QUESTIONS

1. Which of the following is NOT a factor to consider when selecting a sampling procedure?

- a) Purpose of the sample
- b) Size of the sample
- c) Weather conditions
- d) Sampling method

Answer C) Weather Conditions

b) What is the first step in the sampling design process?

- a) Determining the sample size
- b) Selecting the sampling method
- c) Defining the population
- d) Collecting data from the sample

Answer: C) Defining the population.

5.5 SELECTING A SAMPLING PROCEDURE

When selecting a sampling procedure, it is important to choose the appropriate sampling procedure to ensure the sample accurately represents the population and to minimize bias by considering the following factors:

Purpose: What is the purpose of the sample (e.g., estimation, hypothesis testing)?

Population: What is the target population and is it accessible?

Sample size: How large is the sample needed to achieve the desired level of precision?

Sampling method: What type of sampling method (e.g., random, stratified, cluster) is appropriate for the study?

Feasibility: Is the sampling procedure practical and feasible given the resources and time available?

Sampling error refers to the deviation between the sample and the population that the sample represents. Here are some examples of sampling error from a marketing research perspective:

Sampling bias: Sampling bias occurs when the sample is not representative of the population. For example, if a company conducts a survey only among its current customers, the results may not reflect the opinions of potential customers.

Small sample size: A small sample size may not accurately represent the population. For example, if a company wants to know the opinions of a city's residents on a particular issue,



but only surveys 50 people, the results may not be representative of the city's entire population.

Non-response bias: Non-response bias occurs when the people who choose to participate in a survey or study are different from those who do not participate. For example, if a company conducts an online survey and only a certain demographic group responds, the results may not be representative of the entire population.

Random sampling error: **Even with a large and representative sample, random sampling error can** still occur. This is due to chance variations in the sample. For example, if a company conducts a survey on a topic, there is always a chance that a random sample of participants may not provide a true representation of the population.

Self-selection bias: Self-selection bias occurs when individuals self-select into a study or survey. For example, if a company offers a reward for completing a survey, only those who are motivated by the reward may participate, leading to biased results. **Cluster sampling error:** occurs when the sample is drawn from clusters or groups of individuals instead of individual members of the population, leading to potential over- or under-

Stratified sampling error: occurs when the sample is not evenly distributed among the different strata or subgroups in the population. In order to comprehend stratified sampling, let us first understand stratified sampling – It is a sampling method where the population is divided into subgroups, or strata, based on certain characteristics (e.g., age, gender, income, education). Then, a random sample is taken from each stratum in proportion to its size in the population. Stratified sampling aims to reduce the sampling error and increase the precision of the estimates, especially when the population is heterogeneous and the variable of interest is related to the strata.

An example of stratified sampling in marketing research could be a company conducting a customer satisfaction survey. The company could stratify their customer base by the type of product or service they purchase, as different types of customers may have different opinions about the company's performance. The company could then randomly sample from each stratum to ensure that the sample accurately represents the different types of customers in proportion to their size in the population.

By using stratified sampling, the company can increase the precision of the estimates and reduce the sampling error, as they are ensuring that the sample is representative of the population with respect to the stratification variable. This can help the company to make more informed decisions about how to improve their products or services, and tailor their marketing strategies to specific customer segments.



Stratified sampling error is the difference between the sample estimates (e.g., mean, proportion) and the true population values that is due to using a stratified sampling method instead of a census or a more accurate sampling method. Stratified sampling aims to reduce the sampling error by ensuring that the sample is representative of the population with respect to the stratification variable. However, even with a stratified sampling method, some degree of sampling error may still occur due to chance variations in the sample.

The size of the stratified sampling error depends on various factors such as the degree of variability within strata, the size of the strata, the sample size, and the level of precision desired. Generally, larger strata and larger sample sizes can help reduce the stratified sampling error.

In marketing research, stratified sampling is often used when the population is heterogeneous with respect to some variables, and the variable of interest is related to the stratification variable. By using stratified sampling, marketers can obtain more accurate estimates of the variable of interest within each stratum, and make more informed decisions based on the specific needs and preferences of different customer segments.

Confounding: occurs when the relationship between variables is obscured by the presence of other variables that have a significant impact on the outcome.

Measurement error: occurs when the measurement instrument used to collect data is not accurate or precise.

Example: Suppose you are a market researcher who wants to estimate the proportion of people in Delhi, India, who own a smartwatch.

Purpose: Estimation of population proportion

Population: All people in Delhi

Sample size: A sample of at least 1,000 individuals to achieve a desired level of precision

Sampling method: Simple random sampling to ensure each person in the population has an equal chance of being selected

Bias: Ensure the sampling method is random to avoid selection bias

Feasibility: Ensure the sample size is feasible given the resources and time available for the study.

Based on these considerations, the appropriate sampling procedure would be to randomly select a sample of at least 1,000 individuals from Delhi's population and survey them to estimate the proportion of smart watch owners.



5.6 SAMPLING DESIGN - METHODS

There are different types of sample design based on two factors 1. Representation Basis and 2. Element selection Technique. On the representation basis, the sample methods can be probability sampling or non-probability sampling. On element selection basis, it can be unrestricted sampling or restricted sampling. Fig1 shows the different types of sampling techniques that can be used:

	Representation basis	
Element Selection Technique	Probability Sampling	Non- Probability Sampling
Unrestricted Sampling	Simple Random Sampling	Convenience Sampling
Restricted Sampling	Complex Random Sampling	Purposive Sampling Such as Quota Sampling, Judgemental Sampling

- (a) **Probability sampling** – In Probability sampling, each unit of the population has a probability of being selected as an unit of the sample. It is based on the concept of random selection. But this probability varies from one method to another method. This type of sampling is more rigorous and free from biases. The different methods of probability sampling are:

1. Simple Random Sampling

Simple Random Sampling is a method of selecting a sample from a larger population where each individual has an equal chance of being selected.

Example: Suppose you want to estimate the average daily household income in a specific city in India. The population consists of all households in the city. To conduct a simple random sample, you would generate a list of all households in the city and then randomly select a specified number of households to survey.

This method helps to reduce selection bias and increase the representativeness of the sample, but it may not always produce the most precise estimates. However, it is often used as a baseline method for comparison with other sampling methods.

2. Systematic Sampling

Systematic sampling is a method of selecting a sample from a population by selecting every kth member from a randomly selected starting point. The selection of k depends on the sample size and the size of the population.

**Example:**

Suppose you want to estimate the proportion of households in India that have access to the internet. The population consists of all households in the country. **To conduct a** systematic sample, you would first randomly select a starting point and then select every k th household from a list of all households in the country.

This method can be useful in large populations where it is difficult to list all members, but it can be subject to sampling bias if the list of households is not random. For example, if households are listed by city and the starting point is selected from a city with higher internet access, the sample will not be representative of the entire population.

Advantages of Systematic Sampling:

Ease of implementation: Systematic sampling is a relatively straightforward and simple method to implement, making it accessible for researchers with limited resources or time.

Reduced sampling error: By selecting every k th member from a randomly selected starting point, systematic sampling can reduce sampling error and increase the representativeness of the sample.

Increased precision: Compared to other simple random sampling methods, systematic sampling can increase precision by reducing the variability in sample size and selection method.

Disadvantages of Systematic Sampling:

Sampling Bias: Systematic sampling is susceptible to sampling bias, which can occur when the selection of sample units is not truly random.

Inadequate Representation: If the sampling interval is not properly chosen, some groups or sub-populations in the population may be inadequately represented in the sample.

Inflexibility: Systematic sampling does not allow for changes in the sample size or structure, making it less flexible than other sampling methods.

3. Stratified Sampling- Stratified sampling is a method of sampling where the population is divided into subgroups (strata) based on some common characteristic, and then a random sample is taken from each subgroup. This ensures that the sample represents the characteristics of the subgroups in the population, rather than just the overall population.

Examples:



- **Surveying a population by age:** dividing the population into age groups (e.g. 18-25, 26-35, 36-45) and taking a random sample from each group.
- Sampling patient data in a hospital: dividing patients into different illness categories (e.g. cancer, heart disease, diabetes) and taking a random sample from each category.
- **Polling voters:** dividing voters into different political affiliations (e.g. Republican, Democrat, Independent) and taking a random sample from each affiliation.

Advantages of Stratified Sampling:

- **Increased precision:** By taking a random sample from each subgroup, stratified sampling can lead to a more accurate representation of the population.
- **Better control of population characteristics:** Stratified sampling allows the researcher to ensure that the sample includes a sufficient number of individuals from each subgroup, which can be particularly important when studying a population with important subgroup differences.
- **Improved efficiency:** In some cases, stratified sampling can be more efficient than other sampling methods because it reduces the variance of the sample.

Disadvantages of Stratified Sampling:

- **Complexity:** Stratified sampling can be more complex than other sampling methods, as it requires dividing the population into subgroups and selecting a sample from each subgroup.
 - **Bias:** Stratified sampling can be biased if the subgroups are not selected or defined correctly, leading to an unrepresentative sample.
 - **Cost:** Stratified sampling can be more expensive than other sampling methods because it requires more resources to identify and select the subgroups and to collect data from each subgroup.
4. **Cluster Sampling** - Cluster sampling is a method of sampling where the population is divided into groups (clusters) and then a random sample of these clusters is selected. The entire population within the selected clusters is then included in the sample. This method is used when it is difficult or expensive to sample individuals directly from the entire population.



Examples:

- **Surveying households in a city:** dividing the city into neighbourhoods (clusters) and selecting a random sample of neighbourhoods, then surveying every household in the selected neighbourhoods.
- **Studying school children:** dividing schools in a district into clusters and selecting a random sample of schools, then including all the children in the selected schools in the sample.
- **Measuring the quality of products in a factory:** dividing the production line into stations (clusters) and selecting a random sample of stations, then testing every product produced at the selected stations.

Advantages of Cluster Sampling:

Simplicity: Cluster sampling is a simple and straightforward method that is easy to implement and requires fewer resources compared to other methods.

Cost-effectiveness: Cluster sampling can be less expensive than other methods because it reduces the number of individuals or elements that need to be directly sampled.

Feasibility: Cluster sampling is useful in cases where it is difficult or impossible to sample individuals directly from the entire population.

Disadvantages of Cluster Sampling:

Reduced precision: Cluster sampling can lead to a higher degree of sampling error compared to other methods because the sample is less representative of the population.

Clustering bias: The sample may be biased if the clusters are not selected or defined correctly, leading to an unrepresentative sample.

Difficulty in estimating population parameters: It can be difficult to estimate population parameters using cluster sampling, as the sample includes only a portion of the population and the relationship between the clusters and the population is not well understood.

5. **Multi-stage Sampling-** Multi-stage sampling is a method of sampling where multiple stages of sampling are used to select a sample from a population. The population is divided into smaller and smaller subgroups until a final sample is selected. This method is used when the population is too large or complex to be sampled in a single stage.

**Examples:**

- **Surveying a country's population:** dividing the country into regions, then cities within regions, then neighborhoods within cities, and finally households within neighborhoods to select a final sample.
- **Measuring the quality of products in a factory:** dividing the factory into production lines, then stations within production lines, and finally products within stations to select a final sample.
- **Studying the health of a population:** dividing the population into age groups, then gender, then geographic regions, and finally individuals to select a final sample.

Case Study :

Maruti Suzuki India Ltd is looking to grab the Electric Vehicle market in Mid-income segment. For this the company wants to understand consumer attitudes and preferences towards electric vehicles (EVs). To comprehend the customer attitude, the company decided to use simple random sampling by selecting a sample of 1000 customers from a list of all customers who have purchased a car from the company in the past two years. A survey was sent to these 1000 customers asking about their attitudes and preferences towards EVs.

1. How does simple random sampling ensure that the sample is representative of the population of customers who have purchased a car from the company in the past two years?
2. How could the results of this study be improved by using a different type of probability sampling method?
3. What are the potential sources of sampling error in this study and how could they be minimized?



5.7 SELF-ASSESSMENT QUESTIONS

3) What is the main advantage of stratified sampling?

- a. Increased precision
- b. Better control of population characteristics
- c. Improved efficiency
- d. All of the above

Answer: d. All of the above

4) What is the main disadvantage of simple random sampling?

- a. Increased precision
- b. Better control of population characteristics
- c. Reduced representativeness of the sample
- d. None of the above

Answer: c. Reduced representativeness of the sample

5) What is the main purpose of multi-stage sampling?

- a. To reduce the size of the population to be sampled
- b. To increase the efficiency of the sampling process
- c. To sample a large and complex population
- d. To reduce the cost of the sampling process

Answer: c. To sample a large and complex population

6) What is the main advantage of cluster sampling?

- a. Simplicity
- b. Cost-effectiveness
- c. Feasibility
- d. All of the above

Answer: d. All of the above

Non-Probability sampling:

Non-probability sampling is a type of sampling method in which individuals or elements are selected for a study based on non-random criteria, meaning that not all elements in the population have equal chance of being selected. The types of non-probability sampling include:



- **Convenience sampling:** selecting participants based on accessibility or proximity.
 - **Quota sampling:** selecting participants based on specific attributes to match the proportions of a population.
 - **Snowball sampling:** selecting participants based on referrals from previously selected participants.
 - **Purposive sampling:** selecting participants based on specific characteristics or qualities.
 - **Expert sampling:** selecting participants based on their expertise or knowledge in a certain area.
- (i) **Convenience sampling:** Convenience sampling is a type of non-probability sampling where participants are selected based on their ease of accessibility or proximity to the researcher. The main advantage of convenience sampling is its cost-effectiveness and simplicity, but it can result in biased samples that are not representative of the population. Here are two examples of where convenience sampling can be used in research: For Example:

Market research: Convenience sampling can be used to quickly gather data from a sample of customers at a shopping mall or store to understand their opinions and preferences.

Online surveys: Researchers can use convenience sampling by posting online surveys to social media platforms or email lists to gather data from a large number of participants quickly and easily.

- (ii) **Quota sampling:** Quota sampling is a type of non-probability sampling where participants are selected based on predefined proportions of certain characteristics, such as age, gender, or ethnicity. The goal of quota sampling is to ensure that the sample is representative of the population in terms of specific characteristics. Here are two examples of where quota sampling can be used in research:

Political surveys: In political surveys, quota sampling can be used to ensure that the sample includes an equal number of participants from different political affiliations.

Health studies: In health studies, quota sampling can be used to ensure that the sample includes an equal number of participants from different ethnic or racial groups, as certain health conditions are more common in certain populations.



(iii) **Snowball sampling** is a type of non-probability sampling where participants are selected based on referrals from previously selected participants. The main advantage of snowball sampling is that it can be used to reach hard-to-find populations or those who are difficult to access through traditional sampling methods. Here are three examples of where snowball sampling can be used in research:

- **Substance abuse studies:** In substance abuse studies, snowball sampling can be used to reach participants who use drugs or alcohol and are difficult to find through traditional methods. Participants are asked to refer others who use drugs or alcohol to the study.
- **Studies on marginalized communities:** Snowball sampling can be used to reach marginalized communities who are often difficult to access through traditional methods. Participants are asked to refer others from their community to the study.
- **Studies on rare diseases:** Snowball sampling can be used to reach patients with rare diseases who are difficult to find through traditional methods. Participants are asked to refer others with the same condition to the study.

(iv) **Purposive sampling** is a type of non-probability sampling where participants are selected based on specific criteria or characteristics that are deemed important for the research study. The main advantage of purposive sampling is that it allows researchers to select participants who have specific knowledge, experiences, or characteristics that are relevant to the research questions. Here are three examples of where purposive sampling can be used in research:

- **Qualitative studies:** In qualitative studies, purposive sampling can be used to select participants who have specific knowledge or experiences related to the research topic.
- **Studies on specific populations:** Purposive sampling can be used to select participants who belong to a specific population of interest, such as women or elderly people.
- **Expert studies:** In expert studies, purposive sampling can be used to select participants who are recognized experts in a specific field, such as doctors or lawyers. This allows researchers to gather insights from individuals who have specialized knowledge in the research area.

(v) **Expert Sampling** - Expert sampling is a type of non-probability sampling where participants are selected based on their expertise or knowledge in a specific area. The



main advantage of expert sampling is that it allows researchers to gather insights and information from individuals who have specialized knowledge and experience in the research area. Here are three examples of where expert sampling can be used in research:

- **Studies on specific industries:** In studies on specific industries, expert sampling can be used to select individuals who have in-depth knowledge and experience in that industry.
- **Studies on technical or scientific fields:** In studies on technical or scientific fields, expert sampling can be used to select individuals who have specialized knowledge and experience in the area of study.
- **Studies on public policy:** In studies on public policy, expert sampling can be used to select individuals who have specialized knowledge and experience in policy making, such as government officials or policy analysts. This allows researchers to gather insights from individuals who have a deep understanding of the policy-making process.

Case for Practice:

Consider a case of a fast-moving consumer goods (FMCG) company say HUL can be an excellent example of the use of non-probability sampling in research. The company wants to understand consumer behavior and preferences in order to improve its product offerings.

For this research, the company decided to use convenience sampling by conducting a survey among customers at a shopping mall. A total of 500 customers were approached and asked to fill out a survey regarding their purchasing behavior and preferences. Based on the information, answer the following questions:

1. What are the benefits and limitations of using convenience sampling in research on FMCG companies?
2. How could the results of this study be improved by using a different type of non-probability sampling?
3. What are the potential biases that could arise from using convenience sampling in this study?



5.8 SELF-ASSESSMENT QUESTIONS

7. What type of non-probability sampling involves selecting participants who are easily accessible or available?
 - a) Convenience Sampling
 - b) Quota Sampling
 - c) Snowball Sampling
 - d) Expert Sampling
8. What type of non-probability sampling involves dividing the population into subgroups and then selecting a specified number of participants from each subgroup?
 - a) Convenience Sampling
 - b) Quota Sampling
 - c) Snowball Sampling
 - d) Expert Sampling
9. What type of non-probability sampling involves asking initial participants to refer others to the study?
 - a) Convenience Sampling
 - b) Quota Sampling
 - c) Snowball Sampling
 - d) Expert Sampling
10. What type of non-probability sampling involves selecting participants based on their knowledge or expertise in a specific area?
 - a) Convenience Sampling
 - b) Quota Sampling

5.9 SUMMARY

Sampling design refers to the plan for selecting a subset of individuals from a population for data collection. There are two main types of sampling: probability and non-probability. Probability sampling involves methods such as random, systematic, and stratified sampling,



where each member of the population has a known chance of being selected. Non-probability sampling includes methods such as convenience, quota, and snowball sampling, where selection is not based on chance.

Field work is the practical implementation of the sampling design, including data collection and processing. This involves selecting the sample, obtaining information from the sample members, and documenting the data.

In conclusion, sampling design is a crucial aspect of research as it determines the representativeness of the sample and the validity of the results. Field work is the actual execution of the design, and it is important to ensure that it is carried out accurately to obtain reliable results.

5.10 GLOSSARY

Sampling design - the plan for selecting a subset of individuals from a population for data collection.

Probability sampling - a method of sampling in which each member of the population has a known chance of being selected, including methods such as random, systematic, and stratified sampling.

Non-probability sampling - a method of sampling in which selection is not based on chance, including methods such as convenience, quota, and snowball sampling.

Field work - the practical implementation of the sampling design, including data collection and processing.

Random sampling - a method of probability sampling in which each member of the population has an equal chance of being selected.

Systematic sampling - a method of probability sampling in which members of the population are selected at regular intervals.

Stratified sampling - a method of probability sampling in which the population is divided into subgroups (strata) and a sample is selected from each stratum.

Convenience sampling - a method of non-probability sampling in which the sample is selected based on accessibility and convenience.

Quota sampling - a method of non-probability sampling in which the sample is selected based on predefined proportions of certain subgroups in the population.



Snowball sampling - a method of non-probability sampling in which the sample is selected through referrals from initial sample members.

5.11 ANSWERS TO INTEXT-QUESTIONS

- | | |
|--|----------------------------|
| 1. (c) Weather Condition | 6 (d) All of the above |
| 2. (c) Defining the Population | 7 (a) Convenience Sampling |
| 3. (d) All of the above | 8 (b) Quota Sampling |
| 4 (c) Reduced representativeness of the sample | 9 (c) Snowball Sampling |
| 5 (c) To sample a large and complex population | 10 (d) Expert Sampling |

5.12 SELF-ASSESSMENT QUESTIONS

1. Discuss the main types of probability sampling and their advantages and disadvantages.
6. Describe the different methods of non-probability sampling and their suitability for different research purposes.
7. Discuss the trade-offs between using probability and non-probability sampling techniques in terms of representativeness, generalizability, and practicality.
8. Discuss the main types of probability sampling and their advantages and disadvantages.
9. Describe the different methods of non-probability sampling and their suitability for different research purposes.
10. Discuss the trade-offs between using probability and non-probability sampling techniques in terms of representativeness, generalizability, and practicality.

5.13 REFERENCES

Kumar, R. (2014). Research Methodology: A Step-by-Step Guide for Beginners. Sage Publications.

Cochran, W. G. (1977). Sampling Techniques for Researchers. John Wiley & Sons.



Kish, L. (1965). Survey Sampling. John Wiley & Sons.

Lumley, T. (2002). Non-Probability Sampling. Sage Publications.

Lohr, S. L. (1999). Sampling: Design and Analysis. Duxbury Press.

Research Methodology: A Step-by-Step Guide for Beginners" by C.R. Kothari (New Age International Publishers).

5.14 SUGGESTED READINGS

"Research Methodology: Methods and Techniques" by C.R. Kothari (New Age International Publishers)

"Quantitative Research Methods for Social Sciences" by S.P. Gangrade (Sage Publications)

"Research Design: Qualitative, Quantitative, and Mixed Methods Approaches" by J. Creswell (Sage Publications)

"Sampling Techniques" by A.K. Singh (Vikas Publishing House)

"Design and Analysis of Experiments" by Douglas C. Montgomery (John Wiley & Sons).



LESSON-6

SAMPLE SIZE DETERMINATION & DATA PREPARATION

Dr. Ameet Sao

RICS, SBE Amity University

Email-Id.ameetsao@gmail.com

STRUCTURE

- 6.1 Learning Objectives
- 6.2 Introduction
- 6.3 Sample Size Determination
- 6.4 Data Coding
- 6.5 Data Preparation
- 6.6 Tabulation and Graphical Analysis
- 6.7 Sampling Errors
- 6.8 Summary
- 6.9 Glossary
- 6.10 Answers to In-Text Questions
- 6.11 Self-Assessment Questions
- 6.12 References
- 6.13 Suggested Readings

6.1 LEARNING OBJECTIVES

- Understanding the importance and purpose of sample size determination and how to calculate it.
- Knowing the different methods and techniques of data coding and preparation, and being able to apply them in practice.
- Being able to organize and present data using tabulation and graphical presentation techniques.
- Understanding the concepts of sampling and non-sampling errors, and how to identify and mitigate them in research.



6.2 INTRODUCTION

Sample and sampling are fundamental concepts in research. Sampling is the process of selecting a subset of individuals or observations from a larger population for the purpose of studying and understanding the population as a whole. The selected subset is known as a sample. The sample is then used to make inferences about the population from which it was selected. The process of sampling is essential in research, as it allows researchers to study a manageable number of observations while still being able to draw valid conclusions about the population. There are various types of sampling methods, each with their own advantages and disadvantages, and the choice of which method to use depends on the research question and the resources available. In this paper, we will explore the concept of sample and sampling in depth, including the different types of sampling methods, the advantages and disadvantages of each method, and the importance of sample size determination.

Sampling is an essential part of research, which allows investigators to study a manageable number of observations while still being able to draw valid conclusions about the population. There are a variety of sampling methods, each with its own advantages and disadvantages. The selection of a specific method depends on the research question and the characteristics of the population. Some examples of the different types of sampling methods include:

Simple random sampling, which involves randomly selecting individuals from the population without any bias.

Stratified random sampling, which involves dividing the population into subgroups and then randomly selecting individuals from each group.

Cluster sampling, which involves randomly selecting groups of individuals from the population and then selecting individuals from within those groups.

Systematic sampling, which involves selecting individuals from the population at regular intervals.

Convenience sampling, which involves selecting individuals from the population who are easily accessible or convenient to study.

It is crucial to use a representative sample, which should be similar to the population in terms of key characteristics, in order to draw valid conclusions.



6.3 SAMPLE SIZE DETERMINATION

Determining the appropriate sample size is an important aspect of statistical design and analysis. Sample size determination is the process of identifying the number of observations or replicates that should be included in a statistical sample. The chosen sample size can have an impact on the precision of the estimate, the power of the test, and the chance of making a type I or type II error. There are various methods to determine the sample size, some of which include:

Power analysis, which is used to determine the sample size required to detect a statistically significant difference of a specified magnitude with a given level of power and level of significance.

Precision analysis, which is used to determine the sample size required to achieve a specified level of precision or margin of error in the estimate of a population parameter.

Effect size analysis, which is used to determine the sample size required to detect a specified effect size with a given level of power and level of significance.

Bayesian analysis, which uses prior knowledge and information to determine the sample size required.

Minimum detectable difference, which is used to determine the sample size required to detect a specified minimum detectable difference in the population with a given level of precision.

Let's say a real estate company wants to conduct a marketing survey research to understand consumer behavior towards the purchase of an affordable house. They want to estimate the proportion of people who are willing to buy an affordable house in a specific location, with a margin of error of 5% and a 95% confidence level.

To determine the appropriate sample size for this research, the following steps can be taken:

Estimate the population size: The company needs to estimate the number of people who fit their target population, i.e., those who are interested in buying an affordable house in the specific location. Let's say the estimated population size is 50,000.

Determine the expected proportion: The company needs to estimate the expected proportion of people in the target population who are willing to buy an affordable house. Let's say the expected proportion is 50%.

Determine the margin of error and confidence level: The company has decided on a margin of error of 5% and a confidence level of 95%.



Use a sample size calculator: The company can use a sample size calculator, which takes into account the estimated population size, the expected proportion, the margin of error, and the confidence level, to determine the appropriate sample size. Based on these inputs, the calculator suggests a sample size of 385.

Adjust for potential non-response: The company should also consider that not all individuals in the sample may respond to the survey. They may want to increase the sample size by a certain percentage to account for non-response. For example, if they assume a non-response rate of 20%, they may want to increase the sample size by 25%, resulting in a final sample size of 481.

Here's how the calculation for the sample size can be done using the formula:

$$n = [Z^2 * p * q * N] / [(N - 1) * e^2 + Z^2 * p * q]$$

where:

n = sample size

Z = z-score for the desired confidence level (95% = 1.96)

p = expected proportion of people willing to buy an affordable house (0.5)

q = 1 - p (0.5)

N = estimated population size (50,000)

e = margin of error (0.05)

Plugging in the values, we get:

$$n = [1.96^2 * 0.5 * 0.5 * 50,000] / [(50,000 - 1) * 0.05^2 + 1.96^2 * 0.5 * 0.5]$$

$$n = 384.16$$

Since we cannot have a fraction of a person, we round up to the nearest whole number, giving a sample size of 385.

If we want to adjust for non-response, we can multiply the sample size by the inverse of the response rate. For example, if we expect a non-response rate of 20%, we can adjust the sample size as follows:

$$\text{Adjusted sample size} = 385 / (1 - 0.2) = 481.25$$

Again, we round up to the nearest whole number, giving a final sample size of 481.



Once the appropriate sample size has been determined, the real estate company can randomly select participants from the target population to participate in the survey. By using this sample size, they can obtain estimates of the proportion of people who are willing to buy an affordable house with a margin of error of 5% and a 95% confidence level, which can help them make more informed decisions about their marketing strategies.

When determining sample size, it is important to consider the trade-offs between sample size and cost. Larger sample sizes may lead to more precise estimates, but they also increase the cost of collecting data. It is recommended to consult with statisticians or use sample size calculator software to determine the sample size needed for the study.

6.4 DATA CODING

Data coding is a process that assigns numerical or categorical codes to data to make it more manageable and easier to analyze. These codes can represent different categories or attributes of the data. There are various types of data coding, including:

Nominal coding, which is used for categorical data that has no inherent order, such as gender, race, and religion. For example, in a survey asking participants their gender, "1" could represent Male, "2" could represent Female, and "3" could represent Other.

Ordinal coding, which is used for categorical data that has an inherent order, such as levels of education, levels of income, and levels of satisfaction. For example, in a survey asking participants to rate their satisfaction on a scale of 1 to 5, "1" could represent "very dissatisfied" and "5" could represent "very satisfied".

Interval or ratio coding, which is used for numerical data that has an inherent order and an absolute zero point, such as temperature, weight, and time. For example, in a survey asking participants their age, "20" could represent 20 years old.

It is important to ensure that the data coding is done consistently and accurately to guarantee that the data is reliable and valid. To achieve this, a clear and detailed codebook that explains the coding scheme and the meaning of each code should be created. The choice of coding method depends on the type of data and the purpose of the analysis, it's recommended to consult with statisticians or refer to the literature on the specific field to choose the appropriate coding method.

Data coding is a technique that has several benefits, including:

Improved data management: Data coding can make it easier to organize, store, and retrieve data, making the data analysis process more efficient and less prone to errors.



Enhanced data analysis: By assigning codes to data, it can be analyzed using various statistical techniques such as frequency distributions, cross-tabulations, and regression analysis. This can uncover patterns and relationships in the data that may not be obvious when the data is in its raw form.

Increased data reliability and validity: By ensuring that the data is entered and recorded consistently, data coding can help to reduce errors and inconsistencies in the data. This can increase the reliability and validity of the data and the conclusions that are drawn from it.

Better data visualization: Coded data can be used to create visualizations such as charts and graphs, which can help to communicate the findings and insights more effectively.

Easier data sharing and collaboration: Coded data can be shared and collaborated on more easily with other researchers, as the data is in a standardized and easily understood format.

Cost-effective: Data coding can save time and money by reducing the need for manual data entry and data cleaning. This can also reduce the risk of errors and inconsistencies in the data.

Data coding has several limitations, including:

Limited flexibility: Data coding can be quite rigid and inflexible as the codes are assigned to specific categories or attributes of the data. This can make it difficult to handle missing data or to analyze the data in different ways.

Loss of information: Data coding can result in the loss of information, as the data is condensed into a smaller set of codes. This can make it difficult to capture the nuances and complexity of the data.

Risk of human error: Data coding is often done manually, which can increase the risk of errors and inconsistencies in the data, leading to unreliable or invalid conclusions.

Time-consuming and labor-intensive: Data coding can be a time-consuming and labor-intensive process, particularly for large datasets.

Limited ability to capture context: Data coding can reduce the ability to capture the context of the data, as the codes are often assigned to the data without considering the context in which the data was collected.

Bias: Data coding can introduce bias in the data if the coder has some sort of bias or if the codes are not clearly defined.

Case study on Sampling - Product

A case study of sampling in action in product development could be a study on the durability of a new smartphone model. The population in this case would be all the smartphones of that specific model that are sold to customers.



The research team decides to use stratified random sampling to select their sample. They first divide the population into strata based on the manufacturing date of the smartphones because they suspect that the durability might vary based on when the phone was manufactured. They then randomly select a certain number of smartphones from each stratum to form their sample.

Next, they conduct a series of durability tests on the selected smartphones, such as drop test, water resistance test and screen scratch test. They then analysed the data and compare the results of the sample smartphones with the industry standards for smartphones durability.

The team finds that the durability of the smartphones manufactured in the first quarter of the year is higher than the industry standards, while the durability of the smartphones manufactured in the second quarter of the year is lower than the industry standards. They use these findings to identify the issues in the production process and make necessary changes to improve the durability of the smartphones manufactured in the second quarter.

This study illustrates how sampling can be used to identify issues with a product and make necessary changes to improve it. It also shows how stratified random sampling can be used to identify specific subgroups within a population that may be of particular interest.

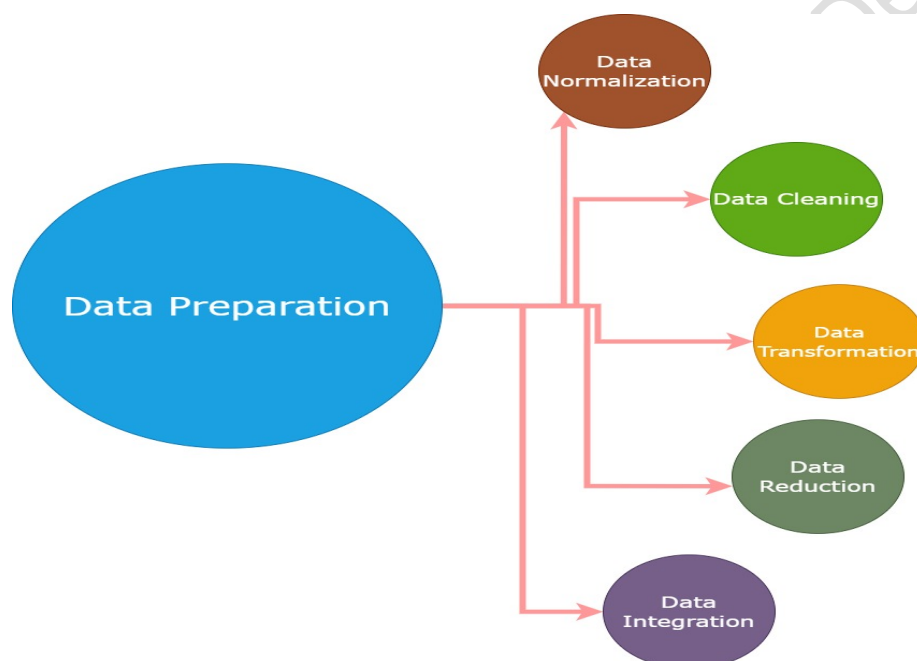
IN-TEXT QUESTIONS

1. Which of the following is not a type of data coding?
 - a) Nominal coding
 - b) Interval coding
 - c) Ordinal coding
 - d) None of the above
2. Which of the following is used for numerical data that has an inherent order and an absolute zero point?
 - a) Nominal coding
 - b) Interval or ratio coding
 - c) Ordinal coding
 - d) None of the above
3. Which of the following sampling technique involves dividing the population into subgroups and then randomly selecting individuals from each group?
 - a) Simple random sampling
 - b) Stratified random sampling
 - c) Cluster sampling
 - d) Systematic sampling
4. Which of the following sampling technique involves selecting individuals from the population who are easily accessible or convenient to study?
 - a) Simple random sampling
 - b) Stratified random sampling
 - c) Cluster sampling
 - d) Convenience sampling



6.5 DATA PREPARATION

Data preparation involves a series of steps to make raw data usable for analysis and modelling. These steps include cleaning the data to remove errors and inconsistencies, transforming the data into a format suitable for analysis, integrating data from different sources, selecting a relevant subset of the data, adjusting the data to meet assumptions, and converting continuous data into categorical data. It is a crucial step in the data analysis process that can greatly affect the accuracy and reliability of the results.



Data cleaning: Identifying and correcting errors and inconsistencies in the data, such as missing values, duplicate records, and outliers.

Data cleaning is a critical process in data analysis that involves identifying and correcting errors and inconsistencies in the data to ensure its quality and reliability. The primary goal of data cleaning is to remove any problems or issues that may affect the accuracy and validity of the analysis results. Some common examples of errors and inconsistencies that need to be addressed during data cleaning include missing values, duplicate records, and outliers.

Missing values are data points that are absent from the dataset. These can occur due to a variety of reasons, such as incomplete data collection or data entry errors. In data cleaning, missing values can be either deleted or imputed, depending on the specific circumstances.



Duplicate records occur when two or more entries in the dataset represent the same observation. In data cleaning, these duplicates can be identified and removed to prevent them from distorting the analysis results.

Outliers are data points that lie far outside the range of the other data points in the dataset. Outliers can occur due to measurement errors, data entry errors, or other factors. In data cleaning, outliers can be identified and either corrected, removed, or analyzed separately from the rest of the data, depending on the specific research question.

Overall, data cleaning is an essential step in the data analysis process that helps ensure the accuracy and validity of the analysis results. By identifying and correcting errors and inconsistencies in the data, data cleaning helps researchers make informed decisions based on reliable and trustworthy data.

Data transformation: Converting the data into a format that is suitable for analysis, such as converting text data into numerical data.

Data transformation is a key step in the data analysis process that involves converting the data into a format that is suitable for analysis. This transformation process may involve converting data from one type or format to another, such as converting text data into numerical data or converting a continuous variable into a categorical variable.

One common reason for data transformation is to make the data compatible with a particular statistical method or analysis technique. For example, many statistical analyses require numerical data, so text data may need to be converted into a numerical format before it can be analyzed. Similarly, some statistical models may require categorical data, so continuous data may need to be converted into a categorical format.

Another reason for data transformation is to make the data more meaningful or easier to interpret. For example, transforming a continuous variable into a categorical variable with predefined categories can make the data easier to understand and analyze.

Overall, data transformation is an essential step in the data analysis process that enables researchers to make sense of complex data and draw meaningful conclusions. By converting the data into a format that is suitable for analysis, researchers can use a variety of statistical methods and techniques to extract insights and uncover patterns in the data.

Data integration: Combining data from different sources, such as merging multiple datasets or integrating external data.

Data integration is a process of combining data from different sources into a single, unified view that provides a comprehensive understanding of the data. This process may involve merging multiple datasets or integrating external data into an existing dataset.



One of the primary goals of data integration is to provide a more complete and accurate picture of the data by combining information from different sources. This can be especially valuable in fields such as marketing research, where data may be collected from multiple sources such as surveys, social media, and transactional data. By integrating this data, researchers can gain a more holistic understanding of consumer behavior and preferences.

Data integration can be challenging, however, as different datasets may use different formats, variable names, and other characteristics. Data cleaning and standardization are often necessary to ensure that the data can be integrated effectively. Data integration can also require sophisticated analytical tools and techniques to ensure that the integrated data is accurate, reliable, and useful. Overall, data integration is an essential process in the data analysis process that helps researchers make informed decisions based on comprehensive and accurate data. By combining data from different sources, researchers can gain a deeper understanding of complex phenomena and develop more effective strategies for addressing real-world problems.

Data reduction: Selecting a subset of the data that is relevant to the analysis, such as removing irrelevant variables or observations.

Data reduction is a process of selecting a subset of data that is relevant to the analysis, while removing irrelevant variables or observations. The primary goal of data reduction is to simplify the data while retaining the most important and meaningful information for analysis.

Data reduction can be achieved through several techniques, including variable selection, dimensionality reduction, and sampling. Variable selection involves selecting only the most important variables for analysis and removing less relevant variables. Dimensionality reduction techniques, such as principal component analysis, involve transforming the data to a lower-dimensional space while retaining as much of the original variation as possible. Sampling techniques, such as stratified sampling, involve selecting a representative subset of the data for analysis.

Data reduction can be especially valuable in situations where the dataset is large and complex. By removing irrelevant data, researchers can focus on the most important and meaningful information for analysis, which can help simplify the analysis process and lead to more accurate and reliable results. Data reduction can also help reduce the computational requirements of the analysis, which can be especially valuable when working with large datasets. Overall, data reduction is an important step in the data analysis process that helps researchers simplify the data while retaining the most important and meaningful information for analysis. By selecting a subset of data that is relevant to the analysis, researchers can focus their efforts on the most important aspects of the data and develop more accurate and reliable insights.



Data normalization: Adjusting the data to meet certain assumptions, such as scaling the data to a common range.

Data normalization is a process of adjusting the data to meet certain assumptions or criteria. The primary goal of data normalization is to make the data comparable and consistent across different units, scales, or variables. This process may involve scaling the data to a common range or transforming the data to conform to a specific distribution.

One common method of data normalization is scaling the data to a common range. This involves transforming the data so that it falls within a specific range, such as between 0 and 1 or -1 and 1. Scaling can be especially useful when working with data that has different units of measurement or scales. By scaling the data to a common range, researchers can compare and analyze the data more easily and accurately.

Another common method of data normalization is transforming the data to conform to a specific distribution, such as a normal distribution. This can be useful in situations where the data does not follow a normal distribution, which is often assumed in statistical analyses. By transforming the data to conform to a normal distribution, researchers can use a wider range of statistical techniques and methods.

Data normalization can be essential in ensuring the accuracy and reliability of data analysis. By adjusting the data to meet certain assumptions or criteria, researchers can develop more accurate and meaningful insights from the data. Normalized data can also be more useful in making predictions or developing models, as it enables researchers to identify patterns and trends that may not be apparent in raw data.

Overall, data normalization is an important step in the data analysis process that helps researchers make accurate and reliable conclusions from the data. By adjusting the data to meet certain assumptions or criteria, researchers can ensure that the data is comparable, consistent, and meaningful for analysis.

Data discretization: Converting continuous data into categorical data, such as grouping age into specific ranges.

Data discretization is a process of converting continuous data into categorical data. The primary goal of data discretization is to simplify the data and make it more manageable for analysis, while still retaining as much meaningful information as possible.

One common example of data discretization is grouping age into specific ranges or categories. For example, researchers might group ages into categories such as 18-24, 25-34, 35-44, and so on. This can be useful in situations where age is not as important as age ranges for the analysis. Similarly, data discretization can also be used for other continuous variables such as income, temperature, or time. There are several methods for data discretization,



including equal width binning, equal frequency binning, and clustering. Equal width binning involves dividing the range of values into equal-sized intervals or bins, such as grouping ages into 5-year intervals. Equal frequency binning involves dividing the data into equal-sized groups based on frequency, such as grouping income into quartiles. Clustering involves grouping similar data points together based on their similarity in value or distance.

Data discretization can be especially useful in situations where there is a large amount of data that is difficult to analyze in its raw form. By converting continuous data into categorical data, researchers can simplify the analysis process and identify patterns and trends more easily. Discretization can also help reduce the computational requirements of the analysis, which can be especially valuable when working with large datasets.

Overall, data discretization is an important step in the data analysis process that helps researchers simplify the data while retaining as much meaningful information as possible. By converting continuous data into categorical data, researchers can develop more accurate and reliable insights from the data.

IN-TEXT QUESTIONS

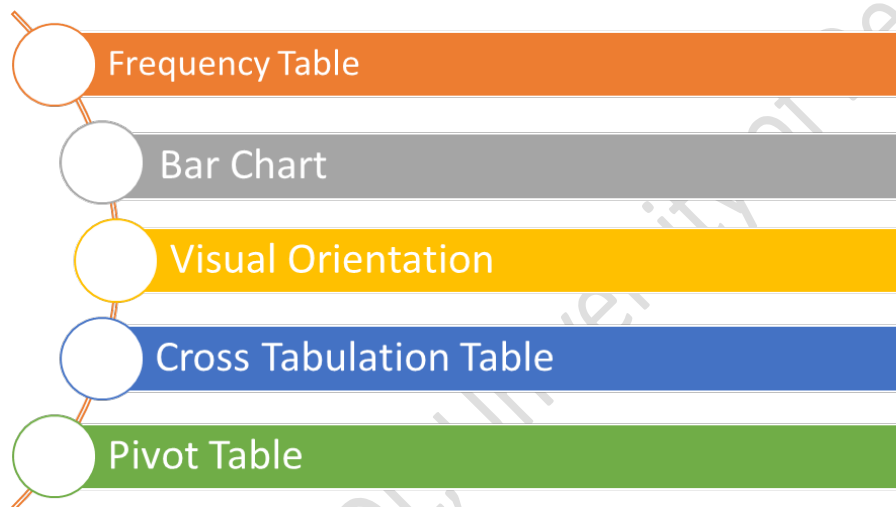
5. What step in data preparation involves converting the data into a format that is suitable for analysis?
 - a) Data cleaning
 - b) Data transformation
 - c) Data integration
 - d) Data reduction
 - e) Data normalization

Answer: b) Data transformation
6. What step in data preparation involves combining data from different sources?
 - a) Data cleaning
 - b) Data transformation
 - c) Data integration
 - d) Data reduction
 - e) Data normalization
7. What step in data preparation involves identifying and correcting errors and inconsistencies in the data?
 - a) Data cleaning
 - b) Data transformation
 - c) Data integration
 - d) Data reduction
 - e) Data normalization



6.6 TABULATION & GRAPHICAL ANALYSIS

Tabulation is the process of organizing data into tables or charts to make it more easily understood and interpreted. Common types of tabulation include frequency tables, cross-tabulation tables, visual representations like bar charts, line charts, and histograms, and pivot tables. Tabulation is an important step in data analysis as it helps identify patterns and trends in the data and makes it more presentable and understandable to the audience.



Consistency and accuracy must be maintained while tabulating to ensure data reliability and validity, and tables and charts should be clearly labelled and described.

Frequency tables: A frequency table is a way to organize data by displaying the number of occurrences or frequency of each category or attribute in a dataset. It's a simple way to understand the distribution of the data.

Cross-tabulation tables: A cross-tabulation table is a type of tabulation that shows the relationship between two or more variables. It's also known as a contingency table. It helps in understanding how the values of one variable are distributed across the levels of another variable.

Visual representation: Bar charts, line charts, and histograms are visual representation of data. They are used to show the distribution of data, such as the frequency of different categories or attributes. They are used to easily identify patterns and trends in the data.

Bar charts: Bar charts are used to compare the size of different categories or attributes. They are useful for comparing the frequency of different categories or attributes.



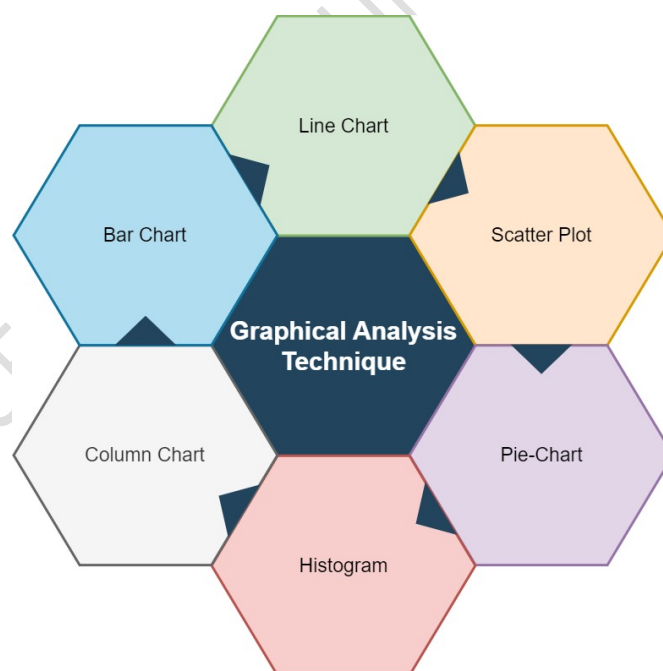
Pivot tables: Pivot tables are a powerful tool for summarizing large datasets by creating a table that shows the relationship between different variables. It allows users to organize and summarize data in a way that makes it easy to identify patterns and trends in the data.

It's important to note that these tabulation methods are used to make the data more presentable and understandable, but the data itself should be accurate and reliable before tabulation so that the final results are also accurate.

Graphical Representation

Graphical representation or data visualization involves using visual elements such as charts, graphs, and plots to represent and communicate data. It is an effective way to present data in a way that is easy to understand and interpret. Different types of visualizations like bar charts, line plots, scatter plots, pie charts and heat maps are used to represent different types of data and patterns. It helps in identifying patterns and trends that might not be obvious from raw data and also makes it easy to compare and contrast different aspects of the data. It is an important step in the data analysis process as it helps to make complex data more accessible and understandable to a wide range of audiences, including non-technical people.

There are many different types of graphical representation, including:



Bar charts: Represent data across different categories or attributes and are used to show the frequency of different categories or attributes.

Data visualization: also known as graphical representation, is a powerful tool for communicating data by using visual elements such as charts, graphs, and plots. It helps in identifying patterns, trends and outliers in the data and makes complex data more presentable and understandable to a wider audience. It's important to use appropriate visualization technique for the type of data and the purpose of the analysis. The graph or chart should be clearly labelled, described and the scale used should be appropriate. It's important to make sure that the visualization effectively communicates the insights and findings of the data analysis.

8. What type of graph would be best to show the trend of stock prices over time?
a) Bar chart b) Line chart
c) Scatter plot d) Pie chart e) Histogram
9. What type of graph would be best to show the relationship between education level and income?
a) Bar chart b) Line chart c) Scatter plot
d) Pie chart e) Histogram
10. What type of graph would be best to show the distribution of sales across different products?
a) Bar chart b) Line chart
c) Scatter plot d) Pie chart e) Histogram
11. What type of graph would be best to show the distribution of ages within a population?
a) Bar chart b) Line chart
c) Scatter plot d) Pie chart e) Histogram



6.7 SAMPLING ERRORS AND NON-SAMPLING ERRORS

Sampling error refers to the difference between the estimate obtained from a sample and the true population value. This is because a sample is only a small subset of the population and therefore the estimate may not be entirely accurate. To reduce sampling error, the sample size can be increased or a more representative sample can be used. For example, in a study aiming to estimate the percentage of people in a city who support a new policy, if a random sample of 1000 people is selected, and 55% of the sample supports the policy, then it is estimated that the true population value could be anywhere between 52% and 58%. The study can state that the estimated proportion of people who support the new policy is 55% with a certain level of confidence.

Non-sampling errors are errors that happen during the data collection and analysis process, which are not related to the sample. These errors can occur due to issues such as measurement errors, data entry errors, and bias. To reduce non-sampling errors, standardized procedures for data collection and analysis should be used and the data should be regularly checked for errors. For instance, in a study that aims to estimate the average income of people in a city, if the study uses a random sample of 1000 people and asks them to report their income, but some participants may not report their income accurately due to embarrassment or fear of sharing personal information, this would lead to measurement errors and bias in the data. To reduce non-sampling errors, the study could use anonymous surveys or other methods of data collection such as government records to ensure accurate income reporting.



CASE STUDY

A case study on sampling and non-sampling errors could be a research project on customer satisfaction with a new product. The population in this case would be all customers who have purchased the product. The research team decided to use simple random sampling to select their sample. They determined the sample size needed to accurately represent the population by using a sample size calculator and determined that a sample of 500 customers would be sufficient. Using a random number generator, the team selected 500 customers from the list of all customers and conducted a survey to gather information on their satisfaction with the product. After analyzing the data, the team found that 80% of the sample was satisfied with the product. However, they realized that there were errors in the data collection process, as some of the survey questions were not clearly worded and some of the customers were unsure of how to answer. This led to measurement errors in the data and affected the accuracy of the satisfaction estimate. To reduce non-sampling errors in the future, the research team decided to revise the survey questions, pilot test them before administering the survey, and increase the sample size to 600 customers to increase the precision of the estimate.

Questions:

1. What type of sampling method was used in this study? Answer: Simple random sampling
2. What type of errors occurred in this study? Answer: Non-sampling errors
3. How could the research team reduce non-sampling errors in the future?
Answer: Revising the survey questions, pilot

6.8 SUMMARY

The above-mentioned topics are all related to the process of collecting and analysing data for research. Sample size determination is the process of determining the number of observations or participants needed to achieve a specified level of precision or accuracy. Data coding is the process of assigning numerical or categorical values to data in order to prepare it for analysis, and data preparation is the process of cleaning, transforming, and organizing data in a format suitable for analysis. Tabulation is the process of organizing data into tables in order to present and summarize it, and graphical presentation is the use of visual aids such as charts



and graphs to represent data and make it easier to understand. Sampling error is the difference between a sample statistic and the corresponding population parameter, due to the fact that the sample is only a subset of the population, whereas non-sampling error are errors that arise from sources other than sampling, such as measurement error, data entry errors, and bias in the sample selection process.

6.9 GLOSSARY

Sample Size Determination: The process of determining the number of observations or participants needed for a study to achieve a specified level of precision or accuracy.

Data Coding: The process of assigning numerical or categorical values to data in order to prepare it for analysis.

Data Preparation: The process of cleaning, transforming, and organizing data in a format that is suitable for analysis.

Tabulation: The process of organizing data into tables in order to present and summarize it.

Graphical Presentation: The use of visual aids such as charts and graphs to represent data and make it easier to understand.

Sampling Error: The difference between a sample statistic and the corresponding population parameter, due to the fact that the sample is only a subset of the population.

Non-sampling Error: Errors that arise from sources other than sampling, such as measurement error, data entry errors, and bias in the sample selection process.

6.10 ANSWERS TO IN-TEXT QUESTIONS

Answer key:

- | | |
|----------------------------------|--------------------------|
| 1. d) None of the above | 6. c) Data integration |
| 2. b) Interval or ratio coding | 7. a) Data cleaning |
| 3. b) Stratified random sampling | 8. b) Line chart |
| 4. d) Convenience sampling | 9. c) Scatter plot |
| 5. b) Data transformation | 10. a) Bar chart |
| | 11. Answer: e) Histogram |

6.11 SELF-ASSESSMENT QUESTIONS

1. What is the purpose of determining a sample size in a study?
2. How does sample size affect the precision and accuracy of a sample estimate?



3. How does sample size affect the level of confidence in a sample estimate?
4. What factors should be considered when determining the sample size for a study?
5. How does the population size and the expected proportion of a characteristic within the population affect the sample size?
6. How does the desired level of precision and margin of error affect the sample size?
7. How can a sample size calculator be used to determine the appropriate sample size for a study?
8. Explain the process of data coding and the role it plays in data analysis?
9. Explain the best practices for data coding, including documentation and quality control measures?

6.12 REFERENCES

- Neuman, W. L. (2014). Social research methods: Qualitative and quantitative approaches. Boston, MA: Pearson.
- Kelle, U. (2017). Data preparation for data mining using SAS. Hoboken, NJ: Wiley.
- Field, A. (2013). Discovering statistics using R. London, UK: Sage Publications.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate data analysis (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Saldana, J. (2015). The coding manual for qualitative researchers. Thousand Oaks, CA: Sage Publications.

6.13 SUGGESTED READINGS

- Kelle, U. (1999). The data of grounded theory. Forum: Qualitative Social Research, 1(1), Art. 8. Retrieved from <https://www.qualitative-research.net/index.php/fqs/article/view/1036/2248>
- Creswell, J. W., & Clark, V. L. P. (2017). Designing and conducting mixed methods research (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Leech, N. L., Barrett, K. C., & Morgan, G. A. (2015). SPSS for intermediate statistics: Use and interpretation (5th ed.). New York, NY: Routledge.
- Onwuegbuzie, A. J., & Leech, N. L. (2006). The importance of sample size in qualitative research. The qualitative report, 11(2), 365-382.
- Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. Journal of mixed methods research, 1 (1), 77-100.



LESSON-7

PARAMETRIC HYPOTHESIS

TESTING AND ANOVA

Dr. Gurjeet Kaur
Assistant Professor
Department of Operational Research
University of Delhi
gkaur@or.du.ac.in

STRUCTURE

- 7.1 Learning Objectives
- 7.2 Introduction
- 7.3 Inferential Data Analysis
- 7.4 Characteristics and Types of Hypothesis
- 7.5 Parametric Hypothesis Testing
 - 7.5.1 Conducting Hypothesis Testing
- 7.6 Testing hypothesis about a population mean using the z statistic (σ known)
- 7.7 Testing hypothesis about a population mean using the t statistic (σ unknown)
- 7.8 Testing hypothesis for the difference of means
 - 7.8.1 Z-test of Difference of Means
 - 7.8.2 t-test for Difference of Means
 - 7.8.3 Paired t test
- 7.9 Design of experiments
 - 7.9.1 ANOVA: ANALYSIS OF VARIANCE (Completely Randomized Design)
 - 7.9.2 Randomized block design (Two Way ANOVA)
- 7.7 Glossary
- 7.8 Answers to In-text Questions
- 7.9 Self-Assessment Questions
- 7.10 References
- 7.11 Suggested Readings



7.1 LEARNING OBJECTIVES

The objective of this lesson is to impart knowledge of concepts related to Research and Statistical Hypothesis.

After completing this course, you will be able to:

- Elucidate the basic concepts related to inferential data analysis.
- Formulate Null and Alternative Hypothesis in alignment with the research objective.
- Understand Type I and Type II error.
- To perform parametric hypothesis testing of the mean.
- Understanding of the design of Experiments and ANOVA

1.2 INTRODUCTION

After collection, the data must be processed and analyzed following the guidelines established while creating the research strategy. This is crucial for a scientific investigation and for ensuring that we have all the pertinent information for the comparisons and analyses we intend to make.

Statistics play a crucial part in research by serving as a tool for conducting studies, analyzing data, and coming to informed conclusions. The majority of research investigations produce a considerable amount of raw data, which must be appropriately reduced to make it easier to interpret and use for further analysis.

Even while the researcher may not have the opportunity to employ statistical procedures in all of their complexity and consequences, it is obvious that any researcher cannot disregard the science of statistics. And as mentioned in previous chapters, classification and tabulation partially accomplish this goal, but we still need to take additional steps.

The use of statistics by researchers involves more than just manipulating data; statistical techniques have their roots in the primary goals of the analysis. Two different applications of statistical data may be the subject of research.

- (i). Descriptive Statistical Analysis, and
- (ii). Inferential Statistical Analysis.



1. Descriptive Statistical Analysis

Any resemblance to individuals outside the group cannot be assumed because descriptive statistical analysis focuses on the numerical description of a specific group observed. The information solely describes that one group.

Descriptive statistics is frequently used in straightforward educational research and offers important details about the makeup of a given group or class.

The important statistical measures that are used to summarise the survey/research data are:

- (1) measures of central tendency or statistical averages; (2) measures of dispersion;
- (3) measures of asymmetry (skewness); and (4) measures of relationship

2. Inferential Statistical Analysis

Inferential statistical analysis involves sampling, which is the process of choosing a small group for the study that is presumptively connected to the larger group from which it is selected. A sample is a small group, and a population or universe is a big group. The statistic is measurements that are based on samples. A statistic that is calculated using a sample may be used to estimate a parameter's value in the population from which it was drawn.

7.3 INFERENCE DATA ANALYSIS

Finding concepts with broad applicability is the main goal of research. But it would be difficult, if not impossible, to generalize from a survey of a full community. A statistic is a calculated value based on sample data. A parameter is a population value calculated from a statistic. A sample is a discrete segment of the population chosen for examination. The sample can be used to conclude about the population. Samples are chosen using different sampling techniques to quantify the impact of chance or probability. In earlier chapters, a variety of sampling techniques were covered, each of which is best suited for a specific set of conditions.

Hypothesis testing is used to analyze the sample data for the population after sampling.

(ii) Hypothesis Testing

Often, the primary tool used in research is the hypothesis. Its primary purpose is to propose a research objective in form of a statement intended to validate based on the sample gathered. In actuality, many experiments are conducted specifically to test hypothesis.



DEFINITION: A hypothesis may be defined simply as a statement about one or more populations.

The parameters of the populations for which the statement is made are frequently a concern of the hypothesis. A public health nurse may be interested that a specific educational program will improve communication between nurse and patient; a hospital administrator may hypothesize that patients admitted to the hospital stay an average of 10 days; a doctor may hypothesize that a particular drug will be effective in 95% of the cases for which it is used. One can establish whether or not such statements are compatible with the existing evidence collected from the sample by using hypothesis testing.

Typically, when someone mentions a hypothesis, they merely mean a simple presumption or any assertion that needs to be supported or refuted. Nonetheless, a hypothesis for a researcher is a formal query that he hopes to answer. As a result, a hypothesis can be described as a claim or set of claims made as an explanation for the occurrence of a particular group of phenomena. These claims may be made solely as a provisional conjecture to direct an investigation, or they may be accepted as highly probable in light of existing facts. A research hypothesis is frequently a statement of prediction that links an independent variable to a dependent variable and may be investigated using scientific procedures. For example, consider statements like the following ones:

“Students who receive psychotherapy will show a greater increase in creativity than students not receiving psychotherapy” Or “the automobile X is performing as well as automobile Y.”

These are theories that can be tested and objectively proven. As a result, we can conclude that a hypothesis expresses what we are seeking and is a claim that can be tested to ascertain its veracity.

7.4 CHARACTERISTICS AND TYPES OF HYPOTHESIS

Hypothesis must possess the following characteristics:

- (i) The hypothesis must be specific and explicit. The conclusions reached based on an unclear or vague theory cannot be trusted.
- (ii) A hypothesis should allow for testing. Many times, research programs have stalled in a mire of unprovable theories. To make a theory testable, the researcher may conduct some preceding research.



- (iii) If additional inferences from the hypothesis can be made, which can then be verified or refuted by observation, the hypothesis is testable. If the hypothesis is a relational one, it must specify the relationships between the variables.
- (iv) The hypothesis must be precise and have a narrow scope. A researcher must keep in mind that more specific hypothesis are typically easier to test and he should develop such hypothesis.
- (v) The hypothesis should be presented in as few words as possible so that it is clear to all parties involved. However one must keep in mind that a simple hypothesis has nothing to do with its importance.
- (vi) The hypothesis must be consistent with the majority of established facts, or with a sizable body of known facts. In other words, it must be the option that judges believe to be most plausible.
- (vii) The hypothesis should be testable in a reasonable amount of time. Even the best hypothesis should not be used if it cannot be tested in a reasonable amount of time because one cannot spend a lifetime gathering data to test it.
- (viii) The requirement for an explanation must be explained by the hypothesis. This means that one should be able to determine the initial problem state by using the hypothesis along with other well-known and accepted generalizations. So, a hypothesis must have empirical support and must genuinely explain what it purports to explain.

Statistical and research hypothesis are the two categories of hypothesis that concern researchers.

DEFINITION: The hunch or presumption that spurs the investigation is known as the research hypothesis.

Research hypothesis is the starting point for statistical hypothesis, which are formulated in a way that allows them to be tested using the right statistical methods.

Further, depending on the data being analyzed the hypothesis testing can be broadly classified as:

(a) Parametric Hypothesis Testing

(b) Non- Parametric Hypothesis Testing

While this chapter will focus on Parametric Hypothesis Testing, Table 1 highlights the difference between Parametric and Non-Parametric Hypothesis Testing.



Basis	Parametric Test	Non-Parametric Test
Definition	Any test that can make assumptions about the parameters (defining characteristics) of the population distribution from which the samples are obtained is referred to as a parametric test.	A nonparametric test is one that is necessary to verify the population's hypothesis even when the researcher is oblivious to the population parameter and is unable to make any specific assumptions.
Information about population	Fully known	Unknown/Unavailable
Basis of test statistic	Makes use of the normal probability distribution	The distribution is asymmetric and arbitrary.
Measurement Level	Applied when a metric scale is being used for measuring e.g., ratio or interval	Applied for Nominal or Ordinal Scale
Applicability	Variables	Variables and attributes

Table 1: Parametric vs Non-Parametric Hypothesis Testing

Relationship Between Hypothesis and Research Design

Depending on the type of research design being employed, hypothesis can take many various shapes. Some hypothesis just explain the potential connections between two objects. For instance, in correlational research, a researcher can postulate that drinking too much alcohol is associated with making bad decisions. In other words, the researcher is speculating that drinking alcohol and having good decision-making skills are related (but not necessarily a causal relationship). But in a study with a randomized controlled setup, the researcher might contend that drinking alcohol results in bad judgement. It follows that the research design chosen by a researcher will have a significant impact on the hypothesis that is being examined.



7.5 PARAMETRIC HYPOTHESIS TESTING

The most powerful tests are parametric tests, which should be performed if the sample's selection criteria and the characteristics of the population underlie its fundamental presumptions:

1. The observations are independent and are not interdependent on one another.
2. The population values are normally distributed or if not, the nature of their distribution is known or the sample size is high enough that the Central Limit Theorem is relevant. In no way is the choice of one example reliant on the choice of any other case.
3. The variances of the population values are equal or their ratio is known.
4. Interval or ratio scales are used to express the variables being measured. Nominal and ordinal are not acceptable.

The utility of parametric hypothesis testing can be taken up by the following illustrations

1. A huge company uses tens of thousands of Bluetooth bulbs annually. The previously utilized brand has a 500-hour average lifespan with a 75-hour standard deviation. The corporation is presented with a new brand at a cost far less than what they are already paying for the old brand. Should the business adopt a new brand?
2. The army placed 12 new recruits through a demanding physical training regimen. Both before and after, their weights (in kg) were noted. Should one draw the inference that the program has an impact on the typical weight of a new recruit at a certain level of significance?
3. Assume that a public company has consented to advertise in a neighborhood newspaper if it can be proven that more than 70% of the company's consumers are reached by the newspaper's readership. Should the business run an ad in the neighborhood paper?

1.5.1 Conducting Hypothesis Testing

Hypothesis testing will be presented as a ten-step procedure.

Step 1: Data: Since this dictates the specific test to be used, it is important to understand the nature of the data that serve as the foundation for the testing methods. For instance, it must be established if the data are counts or measurements.



Step 2: Assumptions: Anytime we use statistical approaches to do research, it's important to identify the underlying assumptions that will guide the researcher's work. Depending on the presumptions, a common technique is altered. The normality of the population distribution, equality of variances, and independence of samples are a few examples of the assumptions.

Step 3: The next step in conducting the hypothesis test is defining the hypothesis(Fig 1).

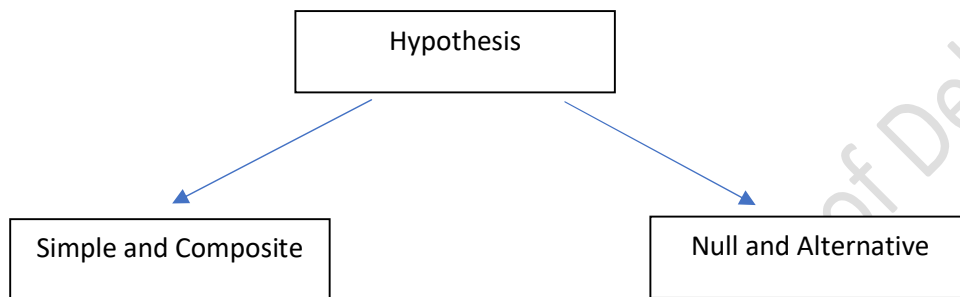


Fig. 1(Hypothesis Classification)

Simple Hypothesis vs Composite Hypothesis:

A simple hypothesis is one that, if true, entirely determines the population distribution, while a composite hypothesis does not.

A simple hypothesis specifies each distributional parameter. For instance, college students' heights are normally distributed with $\sigma^2=4$ and a simple null hypothesis of $H_0=62$.

Nonetheless, such a structure is known as a composite hypothesis if the null hypothesis is $H_0>62$.

Null and Alternative Hypothesis

In hypothesis testing, two statistical hypothesis must be stated explicitly. The hypothesis that will be examined is the null hypothesis. It is represented by the letter H_0 . The null hypothesis, which states that there is an agreement with (or no difference) conditions that are assumed to be true in the population of interest, is frequently referred to as a hypothesis of no difference. Typically, the null hypothesis is constructed with the intent of being shown as not true. As a result, the null hypothesis becomes the complement of the conclusion the researcher is trying to obtain. The null hypothesis is either rejected throughout the testing phase or is not.

If the null hypothesis is accepted, we will state that there is insufficient evidence to reject the null hypothesis based on the test's data. If the testing process results in rejection, we will state



that the available data do not support the null hypothesis but rather support the alternative hypothesis. If our sample results lead us to reject the null hypothesis, the alternative hypothesis states what we will assume to be true. The two terms are frequently used interchangeably and, in most cases, the alternative hypothesis and the research hypothesis are the same.

Suppose, for example, that we want to answer the question:

Can we say that, in this population, the mean is not 70? The null hypothesis is $H_0: \mu = 70$ and the alternative is $H_A: \mu \neq 70$

If the population mean is more than 70, how may we draw that conclusion?

Our hypothesis will be

$H_0: \mu = 70$

$H_A: \mu > 70$

The following hypothesis is presented to determine whether we can conclude that the population mean is less than 70, the hypothesis is

$H_0: \mu = 70$

$H_A: \mu < 70$

In this, it should be remembered that hypothesis can be non-directional (i.e., "reading more of this chapter could raise or decrease the reader's desire to appreciate its author") or directional (e.g., "the more someone reads this book, the more they want to explore the topic"). A one-tailed test is a statistical model that evaluates a directional hypothesis, whereas a two-tailed test evaluates a non-directional hypothesis.

In conclusion, the following guidelines can be used to determine which statements belong in the null hypothesis and which belong in the alternative hypothesis:

- (a) The alternative hypothesis should typically include the conclusion you hope or anticipate being able to draw from the test.
- (b) Either $\geq$ or $\leq$ or $=$ should be used in the null hypothesis.
- (c) The hypothesis under test is the null hypothesis.
- (d) The complementary nature of the null and alternative hypothesis. In other words, the two combined rule out all potential values for the postulated parameter.



A Caution: It is important to note that neither hypothesis testing nor statistical inference, in general, results in the evidence of a hypothesis; rather, they just indicate whether the hypothesis is supported or not by the given data. So, when we are unable to rule out a null hypothesis, we do not state that it is true, only that it may be true. We do not want to imply that accepting involves proof when we talk about accepting a null hypothesis because we are aware of this restriction.

Before the sample is drawn, the alternative hypothesis and the null hypothesis are selected (the researcher must avoid the error of deriving a hypothesis from the data that he collects and then testing the hypothesis from the same data). The following factors are typically taken into consideration when choosing a null hypothesis:

- (a) The null hypothesis is typically the one that one wishes to disprove, while the alternative hypothesis is the one that one typically wishes to prove. As a result, a null hypothesis is a theory we're seeking to rule out, and an alternative hypothesis is every other possibility.
- (b) If there is a high risk of rejecting a certain hypothesis when it is true, that hypothesis is regarded as the null hypothesis because there is a very tiny chance of doing so at the specified level of significance.
- (c) The null hypothesis should never refer to an approximate or about value; it should always be a particular hypothesis. In hypothesis testing, we typically move forward with the null hypothesis while keeping the alternative hypothesis in mind. Why so? Answer: If we proceed with the alternative hypothesis, we cannot assign probabilities to various possible sample results; however, if the null hypothesis is true, we can. Hence the use of the null hypothesis (at times also known as the statistical hypothesis) is quite frequent.

Step 4: Test statistic- The test statistic is a statistic that may be calculated using the sample's data. The test statistic typically has a wide range of potential values, with the actual value seen depending on the specific sample used. As we shall see, the test statistic plays the role of a decision maker since the magnitude of the test statistic determines whether or not the null hypothesis should be rejected.

Step 5: Test statistic distribution- It has been noted that sample distribution is the key to statistical inference. When it is necessary to specify the probability distribution of the test statistic, we are reminded of this once more.



Step 6: A rule of judgement. The test statistic's range of potential values is represented as two groups of points on the horizontal axis of the distribution graph; one group is referred to as the rejection region, while the other is known as the non-rejection region. The test statistic's values that make up the rejection region are those that are less likely to occur if the null hypothesis is correct, whereas the values that make up the acceptance region are more likely to do so.

If the test statistic value we calculate from our sample is one of the values in the rejection region, the decision rule instructs us to reject the null hypothesis; however, if the test statistic value is one of the values in the non-rejection region, we are instructed not to. Fig 2, gives the graphical illustration for a two-tail test decision rule.

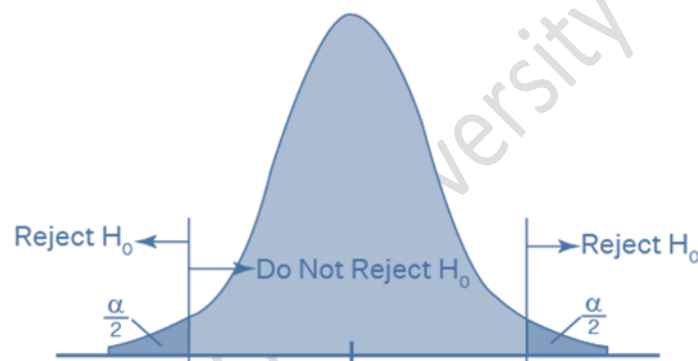


Fig2 (Two Tail Test Decision Rule)

Level of Significance:

The significance levels help us determine how confident we can be in our findings. The desired degree of significance, denoted by α , is used to determine which values belong in the rejection zone and which ones belong in the non-rejection region. A computed value of the test statistic that falls in the rejection zone is considered to be significant, and this is reflected in the phrase level of significance since hypothesis tests are also known as significance tests.

Usually, 5% is the level of significance that is taken in the research works. H_0 will be disregarded if we set the significance level at 5%, according to this assumption. When there is less than a 0.05 chance that the sample result, or the observable evidence, will occur if H_0 is correct. In other words, the 5% level of significance indicates that the researcher is willing to accept a 5% chance that the null hypothesis (H_0) will be true. As a result, the significance



level, which is typically chosen in advance before testing the hypothesis, is the highest value of the likelihood of rejecting H_0 when it is true.

Type I and Type II errors:

We have shown that test statistics can (to a certain degree of certainty) inform us about the real status of the world. We specifically want to determine if there is a population-level impact. In the real world, there are only two possibilities: either there is an effect on the population in reality or there isn't. Although we have no means of knowing which of these hypothesis is true, we may determine which is more likely by examining test results and the probability that goes along with them.

We must be as accurate as possible, which is why Fisher initially recommended that we be extremely cautious and only accept a result as real if we are 95% certain that it is, or if there is only a 5% possibility that the results might occur in the absence of an influence (the null hypothesis is true). Even if we are 95% certain, there remains a slight possibility that we will be mistaken. A Type I error and a Type II error are the two types of errors that we can make while carrying out hypothesis testing. When we think there is a real influence on our population when there isn't, we make a Type I error.

If we apply Fisher's criterion, the likelihood of this error, when there is no influence on the population, is .05 (or 5%); this probability is referred to as the α -level. If we repeated our data collection 100 times, assuming there is no effect in our population, we may anticipate that in five instances we would acquire a test statistic significant enough to lead us to believe that there is a real effect in the population even when there isn't. A Type II error, on the other hand, happens when we think there is no influence on the population when there is.

This would happen if we had a low test statistic (perhaps because there is a lot of natural variation between our samples). If there is an effect on the population, it is crucial that we can identify it; in an ideal world, we want the chance of this inaccuracy to be very low. According to Cohen (1992), the β -level, or maximum tolerable chance of a Type II error, would be .2 (or 20%). This implies that, if we collected 100 samples of data from a population with an effect, we would miss detecting the impact in 20 of those samples (missing 1 in 5 real effects).

It is clear that there is a trade-off between these two errors: if we reduce the likelihood that we will accept an effect as real (i.e., make it smaller), then we also increase the likelihood that we will reject an effect that is real (because we have been so rigid about the level at which we will accept that an effect is real). Because Type I and Type II errors are based on



distinct premises, it is difficult to determine their exact relationship. For a Type I error, there must be no effect in the population, whereas a Type II error requires that there be an effect that we have missed.

The table shown in this figure is an example of what is generally referred to as a confusion matrix. Also Type I error in statistical hypothesis is given by α and Type II error is denoted by β .

		Condition of Null Hypothesis	
		True	False
Possible Action	Fail to reject H_0	Correct action	Type II error
	Reject H_0	Type I error	Correct action

Step 7:

Calculating the test statistic- We compute the test statistic's value from the sample's data and compare it to the previously defined rejection and non-rejection zones.

Step 8:

Statistical judgement - Rejecting or not rejecting the null hypothesis is the statistical decision. When the computed value of the test statistic falls in the rejection region, it is rejected; when it falls in the non-rejection region, it is not rejected.

Step 9:

Finality. If H_0 is disproved, we deduce that H_A is accurate. If H_0 is not disproved, we deduce that H_0 might be accurate.

Step 10:

p-value, when the null hypothesis is true, the p-value is a number that indicates how odd our sample results are. If the null hypothesis is correct, the sample results are not probable to have occurred, according to the p-value, which supports doubting the validity of the null hypothesis.

**IN-TEXT QUESTIONS**

6. A _____ hypothesis is one that, if true, entirely determines the population distribution.
7. The probability of rejecting a true Null Hypothesis is a Type II error. True / False
8. The ____ value is a number that indicates how odd our sample results are.
9. When the computed value of the test statistic falls in the rejection region, the Null Hypothesis is _____
10. A two-tailed test is a statistical model that evaluates a directional hypothesis, whereas a one-tailed test evaluates a non-directional hypothesis. True / False

7.6 TESTING HYPOTHESIS ABOUT A POPULATION MEAN USING THE Z STATISTIC (σ KNOWN)

A test based on a population mean is one of the most fundamental hypothesis tests. A business researcher might be interested in evaluating a predicted mean value for a new theory or product or in testing to see if an existing or accepted mean value for an industry is still true. As an illustration, a corporation that sells computer products sets up a telephone service to help clients by offering technical support.

On weekdays, the wait takes an average of 55 minutes. The management, however, feels that the average wait time fell and they want to demonstrate it after a recent hiring drive added technical consultants to the system. A single mean hypothesis test may also be used in the following business scenarios:

- A monetary investment company wishes to evaluate whether the Dow Jones Average's average hourly change over ten years is +0.35.
- A manufacturing company wishes to conduct a test to discover whether 3.4 millimeters is the typical thickness of a plastic bottle.
- A retail establishment wants to run a test to see if the average age of its patrons is under 30. When the sample size is known to be big ($n > 30$) for any population and for small samples ($n < 30$) if it is known that x is normally distributed in the population, formula 1 can be used to test hypothesis about a single population mean.



Z-test FOR A SINGLE MEAN

This test is carried under the following assumptions

1. Data is at quantitative.
2. Sampling is from a normally distributed population
3. Variance of the population is known.

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Where $\bar{X}$ = Sample Mean

μ = Hypothesized mean value

σ = Population Standard Deviation

n = Sample Size

The average net income for sole owners, according to a study of CPAs nationwide, is 74,914. Since this survey was conducted more than 10 years ago, an accounting researcher wishes to examine this statistic by selecting 112 American single proprietor accountants at random to see if the net income number has changed. Let's say that a sample mean of 78,695 is produced by the 112 CPAs who respond.

The 10 steps of hypothesis testing could be used by the researcher to accomplish this. Assuming a 5% level of significance and a population standard deviation of net incomes for single proprietor CPAs of \$14,530.

The null hypothesis is that the mean still equals \$74,914. These hypothesis follow.

$H_0: \mu = 74914$

$H_1: \mu \neq 74914$

As per the formula given above

$$Z \text{ (calculated)} = \frac{78,695 - 74,914}{\frac{14,530}{\sqrt{112}}} = 2.75$$

$Z(\text{tabulated, at 5\% level of significance, two tail}) = 1.96$

As this test statistic, $z = 2.75$, is greater than the critical value of z in the uppertail of the distribution, $z = +1.96$, the statistical conclusion reached is to reject the null hypothesis.

Statistically, the researcher has enough statistical support to disprove the assertion that the national average net income for sole-owner CPAs is 74,914.



7.7 TESTING HYPOTHESIS ABOUT A POPULATION MEAN USING THE T STATISTIC (UNKNOWN)

We will look at the t-test for a single population mean in this section. If the population standard deviation is unknown, the population is normally distributed for the measurement of interest, and the researcher is testing the value of a population mean μ , then this t-test is generally appropriate.

The following is the formula for testing such hypothesis:

t-test FOR A SINGLE MEAN

This test is carried under the following assumptions

1. Data is at quantitative.
2. Sampling is from a normally distributed population
3. Variance of the population is unknown.
4. Sample Size is preferably less than 30.

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

Where $\bar{X}$ = Sample Mean

μ = Hypothesized mean value

s = Population Standard Deviation

n = Sample Size

This test is carried out at $n-1$ degree of freedom.

Note: The largest number of logically independent values—that is, values with the freedom to change—in the data sample is referred to as the degree of freedom.

7.8 TESTING HYPOTHESIS FOR DIFFERENCE OF MEANS

7.8.1 Z-test of Difference of Means

1. Data is at quantitative.
2. Sampling is from two independent normally distributed populations.
3. Variance of the populations is known.



Likewise, when comparing two independent populations (given by Null Hypothesis stated below), with known sigma we can employ the formula-

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

(Note this is two tail hypothesis, similarly, we can make one-tail hypothesis testing too for a difference of means)

Formula :

$$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

7.8.2 t-test for Difference of Means

However when the variance of the population is unknown but equal (assumption underlying this test) then we can use the following t-statistic formula-

Formula 4:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

$$\text{Where } s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

This is carried at $n_1 + n_2 - 2$ df.

7.8.3 Paired t-test

Also when we wish to study the impact of certain training on people (say),

$$H_0: \mu_{\text{after}} - \mu_{\text{before}} = \mu_d(\text{say}) = 0 \text{ or } H_0: \mu_{\text{before}} - \mu_{\text{after}} = \mu_d(\text{say}) = 0$$

$$H_A: \mu_d \neq 0$$

then we perform t-paired test. The formula for the same is-

Formula 4:



$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

This is carried at n-1 df.

$$\text{Where } s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$$

This is carried at n_1+n_2-2 df.

This test is carried out under the assumption the sample observations are related.

Problem statement: A diet plan has just been added by the health ministry to the government school pupils covered by the mid-day meal program that will assist reduce the harmful sugar levels in their blood. Before and after the diet plan was introduced, the body cholesterol levels of 11 students were measured. At a 90% confidence level, you must decide whether or not the diet plan causes a considerable drop in blood sugar levels.

S.No.	Before	After	d(before-after)
1	88.4	84.4	4
2	93.9	87.3	6.6
3	84.6	83.7	0.9
4	82.5	80.2	2.3
5	90.7	82.3	8.4
6	101.5	93.6	7.9
7	99.6	89.1	10.5
8	89.8	86.2	3.6
9	113.4	103.6	9.8
10	88.6	87.4	1.2
11	93.8	90.7	3.1

t_{cal}	5.07563
T_{tab}	1.37218
p-value	0.00024

Conclusion: At 90% confidence level, t_{cal} is greater than t_{tab} and also p-value is less than alpha. Therefore, we will be rejecting the null hypothesis as we have enough evidence. So, there is a decrease in cholesterol levels in the blood of kids who are taking the new diet plan

**CASE QUESTION**

Large harvesters are made by a Production Company. A 25-pound plate is mounted on the side of a harvester for it to be correctly balanced when it is operating. The equipment used to make these plates is programmed to create plates weighing 25 pounds on average. The machine produces plates in a typical distribution. The shop manager is concerned that the machine may not be adjusted correctly because it is producing plates that are not averaging 25 pounds. He weights 20 plates that were manufactured the day before at random to investigate this worry. The weights acquired, as well as the calculated sample mean and sample standard deviation, are displayed in Table.

The shop supervisor has not said whether he thinks the machine is creating plates that are too heavy or too light, and the test is meant to assess whether the machine is out of control.

WEIGHTS IN POUNDS OF A SAMPLE OF 20 PLATES

21.6	23.2	23.2	27.4	24.5
26.0	25.6	28.1	26.9	24.9
25.2	25.3	23.1	24.2	26.1
26.8	30.4	28.6	23.5	25.6

IN-TEXT QUESTIONS

16. If the population standard deviation is unknown, the population is normally distributed for the measurement of interest, and the researcher is testing the value of a population mean μ , then this t-test is generally appropriate. True/False
17. Z-test for one sample is given by statistic _____.



7.9 DESIGN OF EXPERMENTS

A structure and plan for testing hypothesis known as an experimental design involve the researcher controlling or modifying one or more variables. Both independent and dependent variables are present. A treatment variable or a classification variable can both be independent variables in an experimental design. In an experiment, a treatment variable is a variable that the experimenter adjusts or controls. A classification variable is a feature of the experimental subject that existed before the experiment and was not brought about by manipulations or controls by the researcher. Sometimes independent variables are referred to as factors.

Internal research to analyze daily sales volumes for a particular size shop in four distinct demographic situations might be approved by Wal-Mart executives:

- (1) Stores in a large city's inner city,
- (2) Stores in its suburbs,
- (3) Stores in a medium-sized city, and
- (4) Stores in a village. Also, managers may decide to compare sales throughout the five distinct weekdays (Monday through Friday). The day of the week and store demographics are the independent factors in this study.

Analysis of variance, or ANOVA, is a statistical tool used to evaluate the experimental designs.

7.9.1 ANOVA: ANALYSIS OF VARIANCE (Completely Randomized Design)

Many hypothesis tests are run using the F distribution. For instance, the analysis of variance (ANOVA) approach and the F-statistic is used to test the equality of three or more population means.

ANOVA Assumptions

Three presumptions must be true for an ANOVA to evaluate the equality of three or more population means:

1. Populations generally follow a normal distribution.
2. The standard deviations σ of the populations are equal.
3. Each population is autonomous.



F is taken as the distribution of the test statistic when these criteria are satisfied. Why is ANOVA research necessary? Why can't we just apply the population mean difference test that was covered in the previous chapter? Two at a time, we could compare the population means. The unsatisfactory accumulation of Type I errors is the main cause. Let's use the example of four different firefighter training programs (A, B, C, and D) to better clarify. Each of the 40 recruits in this year's class is randomly assigned to one of the four techniques. We give the four groups a test after the training course to gauge their understanding of firefighting procedures.

Is there a difference between the four groups in terms of mean test scores? This question's resolution will enable us to contrast the four training approaches.

We would need to run six different t-tests to compare the four population means using the t-distribution. Specifically, we would have to compare the mean scores for the four approaches using the following comparisons: A versus B, A versus C, A versus D, B versus C, B versus D, and C versus D. Let's assume that we select $\alpha = .05$. Consequently, there is a 95 percent chance that we will not reject the null hypothesis when it is true. We do six independent tests, and the probability that the results of all six tests are right is α . All correct: $(.95)(.95)(.95)(.95)(.95)(.95) = .735$.

We deduct this value from 1 to determine the likelihood of at least one sample error. As a result, the likelihood of making at least one mistaken judgment as a result of sampling is

$1 - .735 = .265$.

In conclusion, the likelihood of rejecting a correct null hypothesis due to sampling error is unacceptable if we perform six independent tests using the t distribution. 265. We can compare population means simultaneously using the ANOVA technique at a chosen level of significance. It prevents the accumulation of Type I errors brought on by testing several hypothesis.

Several of the words from that context are still used today because ANOVA was originally created for agricultural uses. The various populations being studied are specifically identified by the term treatment. Treatment, for instance, describes how a piece of ground was treated with a certain fertilizer. The example (ANOVA1) will define the term "treatment" and show how to use ANOVA.

The ANOVA Test

The fundamental approach is to calculate the ratio of these two estimations after estimating the population variance (standard deviation squared) twice. If this ratio is close to 1, then the



two estimates are logically equivalent, and we conclude that the population means are also equivalent. If the ratio deviates significantly from 1, we can infer that the population means are different. When the ratio of the sample variances is too big to have occurred by coincidence, the F distribution acts as a referee by indicating it.

TOTAL VARIATION: The sum of the squared differences between each observation and the overall mean.

TREATMENT VARIATION: The sum of the squared differences between each treatment mean and the grand or overall mean.

RANDOM VARIATION: The sum of the squared differences between each observation and its treatment mean.

We determine the test statistic, which is the ratio of the two estimates of the population variance, from the following equation.

$F = \frac{\text{Estimate of the population variance based on the differences between the treatment means}}{\text{Estimate of the population variance based on the variation within the treatments}}$

Hypothesis	
Assumption	The population is normally Distributed
Hypothesis	H ₀ Prices of a commodity (data provided below) do not differ significantly
	H ₁ Prices of commodities do differ significantly
Decision Rule	Reject H ₀ if the computed value of F is greater than the critical values of F at 5% level of significance
Calculation	F_Cal 5.406342914
	F_Tab 3.098391212
	p-value 0.006875948
Statistical Decision	Since, $F_{cal} > F_{tab}$, therefore we reject the null hypothesis. Also, the p-value is less than the level of significance ($0.006 < 0.05$), therefore we reject the null hypothesis
Conclusion	There is a significant difference in the price of a commodity in the cities



Cities			
C1	C2	C3	C4
264	278	275	255
272	291	293	266
268	297	278	249
277	282	271	264
256	285	263	270
295	277	276	268

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
C1	6	1632	272	178
C2	6	1710	285	60.4
C3	6	1656	276	97.6
C4	6	1572	262	67.6

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1636.5	3	545.5	5.40634291	0.006875948	3.09839121
Within Groups	2018	20	100.9			
Total	3654.5	23				

7.9.2 Randomized block design (Two Way ANOVA)

The randomized block design is a second type of research methodology. Similar to the completely randomized design, the randomized block design concentrates on one important independent variable (the treatment variable). The blocking variable, which is a second variable in the randomized block design, can be used to account for confounding or contemporaneous variables.



Confounding variables, also known as concurrent variables, are those that the researcher cannot control during the experiment but which may nevertheless have an impact on how the treatment under study performs.

Sir Ronald A. Fisher was one of the first users of the randomized block design. To analyze the growth patterns of various seed kinds for a particular type of plant, he applied the design to the field of agriculture. His independent variable was the seed variety. Yet as he conducted his experiments on various pieces of land, he concluded that the "block" of land might have some bearing on the results. Blocks were defined by Fisher as a variety of distinct pieces of land that he controlled as a second variable. On each of the blocks, a different seed variety was sown. The primary goal of his research was to contrast the seed variants (independent variable). He only wished to account for the variation in ground plots (blocking variable).

The sum of squares in a completely randomized design is

$$SST \text{ (Total)} = SSG \text{ (Between Groups)} + SSW \text{ (Within Groups)}$$

In a randomized block design, the sum of squares is

$$SST \text{ (Total)} = SSG \text{ (Between Groups)} + SSB \text{ (Between Blocks)} + SSW \text{ (Within Groups)}$$

The illustration for Two Way ANOVA is as follows-

Hypothesis	
Assumption	The population is normally Distributed
Hypothesis 1	H ₀ There is no significant difference between consignments
	H ₁ There is a significant difference between consignments
Hypothesis 2	H ₀ There is no significant difference between observers
	H ₁ There is a significant difference between observers
Decision Rule	Reject H ₀ if the computed value of F is greater than the critical values of F at 5% level of significance
Calculation 1	F _{Cal} 2.219047619
	F _{Ta} 2.901294536
	b
	p-value 0.10636051
Calculation 2	F _{Cal} 5
	F _{Ta} 3.287382105
	b
	p-value 0.013373471



Statistical Decision 1	Since, $F_{cal} < F_{tab}$, therefore we do not reject the null hypothesis. Also, the p-value is greater than the level of significance ($0.106 > 0.05$), therefore we do not reject the null hypothesis
Statistical Decision 2	Since, $F_{cal} > F_{tab}$, therefore we reject the null hypothesis. Also, the p-value is less than the level of significance ($0.013 < 0.05$), therefore we reject the null hypothesis
Conclusion 1	There is no significant difference between the consignments
Conclusion 2	There is a significant difference between the observers

Data:

Observer	1	2	3	4	5	6
1	9	10	9	10	11	11
2	12	11	9	11	10	10
3	11	10	10	12	11	10
4	12	13	11	14	12	10

Anova: Two-Factor
Without Replication

SUMMARY	Count	Sum	Average	Variance
Row 1	6	60	10	0.8
Row 2	6	63	10.5	1.1
Row 3	6	64	10.66666667	0.666666667
Row 4	6	72	12	2
Column 1	4	44	11	2
Column 2	4	44	11	2
Column 3	4	39	9.75	0.916666667
Column 4	4	47	11.75	2.916666667
Column 5	4	44	11	0.666666667
Column 6	4	41	10.25	0.25

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	13.125	3	4.375	5	0.013	3.287
Columns	9.708	5	1.942	2.219	0.106	2.901
Error	13.125	15	0.875			
Total	35.958	23				



7.6 SUMMARY

In this chapter, you have learned the concepts related to parametric hypothesis testing. After reading this chapter you can classify the hypothesis into Null and Alternate; Simple and Composite, type I and type II error. The chapter covered testing of mean of mean of one population and difference of means using Z and t test. Further paired t test was also discussed in the chapter. The Chapter extended the parametric hypothesis testing to design of experiments, thereby you will be able to carry ANOVA.

7.7 GLOSSARY

Inferential statistical analysis : Inferential statistical analysis involves sampling, which is the process of choosing a small group for the study that is presumptively connected to the larger group from which it is selected.

Research Hypothesis: The hunch or presumption that spurs the investigation is known as the research hypothesis.

Simple Hypothesis: A simple hypothesis is one that, if true, entirely determines the population distribution

Null Hypothesis: Hypothesis of no difference\

Type I error: The probability of rejecting a true Null Hypothesis.

ANOVA: Analysis of variance, or ANOVA, is a statistical tool used to evaluate the experimental designs in this chapter.

7.8 ANSWERS TO IN-TEXT QUESTIONS

1. Simple
2. False
3. p
4. rejected
5. False
6. True
7. $\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$



7.9 SELF-ASSESSMENT QUESTIONS

7. Sort out the differences between the following:
(i) simple hypothesis and composite hypothesis; (ii) null hypothesis and alternative hypothesis; (iii) one-tailed test and two-tailed test; (iv) type I error and type II error; and (v) acceptance region and rejection region.
 8. The technique of testing hypothesis involves a researcher to adopt various steps. Briefly describe each of these steps.
 9. Provide a brief description of the key parametric tests that are utilised when testing hypothesis. What distinguishes these tests from non-parametric tests? Explain.
 10. Describe the t-test. When and for what purposes is it used? Describe with examples.
 11. The heights of ten randomly chosen pupils from a school are 40, 42, 42, 43, 45, 46, 47, 48, 48, and 49 in inches. Discuss the claim that the mean height of the school's students is 44 inches in light of these statistics. Conduct the testing at a 5% level of significance.
 12. Six tailors were chosen at random, and their heights are 63, 65, 58, 69, 71, and 72 inches, respectively. The heights of the ten soldiers who were chosen at random are 61, 62, 65, 66, 69, 69, 70, 71, 72, and 73 inches. Do these data suggest that, on average, soldiers are shorter than tailors? Test at 5% level of significance.
 13. Describe what the term "analysis of variance" means. Provide a succinct explanation of the one-way and two-way classifications variance analysis technique.
 14. The sample psychological health ratings of corporate executives working in the different industries are shown in the following table-
- | | |
|-------------------|------------------------|
| Banking | 51, 63, 54, 65, 54, 56 |
| Manufacturing | 45, 67, 56, 59, 54 |
| Fashion retailing | 44, 44, 46, 45, 51 |

9.10 REFERENCES

1. Cohen, J. (1992). Statistical power analysis. *Current directions in psychological science*, 1(3), 98-101.
2. Kothari, C. (2017). research methodology methods and techniques by CR Kothari. Published by New Age International (P) Ltd., Publishers, 91.



3. Marczyk, G. R., DeMatteo, D., & Festinger, D. (2010). *Essentials of research design and methodology* (Vol. 2). John Wiley & Sons.
4. Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: a foundation for analysis in the health sciences*. Wiley.

9.11 SUGGESTED READINGS

1. Kothari, C. (2017). research methodology methods and techniques by CR Kothari. *Published by New Age International (P) Ltd., Publishers, 91.*
2. Daniel, W. W., & Cross, C. L. (2018). *Biostatistics: a foundation for analysis in the health sciences*. Wiley.



LESSON-8

NON PARAMETRIC TEST

Jigmet Wangdus

Assistant Professor

DDCE, Campus of Open Learning

University of Delhi

jigmet@sol-du.ac.in

STRUCTURE

- 8.1 Learning Objectives
- 8.2 Introduction
- 8.3 Chi-square test
 - 8.3.1 Chi-square goodness- of-fit test.
 - 8.3.2 Chi-square test of independence.
- 8.4 Sign test
 - 8.4.1 One sample sign test
 - 8.4.1 Two sample sign test
- 8.5 Mann-Whitney U test
- 8.6 Wilcoxon signed-rank test
- 8.7 Kruskal Wallis test
- 8.8 Summary
- 8.8 Answers to In-text Questions
- 8.9 Self-Assessment Questions
- 8.10 References
- 8.11 Suggested Readings

8.1 LEARNING OBJECTIVES

After studying this lesson, you should be able to:

1. Understand the concept of non-parametric test
2. Identify different types of non-parametric test
3. Hypothesis testing using chi-square



4. Hypothesis testing using Sign test
5. Hypothesis testing using the Mann-Whitney U test
6. Hypothesis testing using Wilcoxon signed-rank test
7. Hypothesis testing using the Kruskal-Wallis test

8.2 INTRODUCTION

Parametric tests are generally used to draw inferences about one or more population based upon sample information from those populations by making certain assumptions about the population distribution like t-distribution or normal distribution. When the assumption of normality is not satisfied along with small sample size, then we can use non-parametric statistical test which is free from the assumption of normality.

The most important and popular statistical test for nominal and ordinal data is chi-square. A nominal scale can be constructed for affiliation to political ideology by grouping individual according to party loyalties – so many Right-wing, so many Left-wing, so many independent and so on. From the nominal scale we cannot characterize the strength of any individual affiliation to political ideology; we can only know what the affiliation is. From nominal scale we get nose-counting data which tells us in how many cases a certain trait occurs. There is no scope of shades of Gray in nominal data ; an observation either has the trait or not.

Nominal data are generated by sorting and counting – sorting the data into discrete, mutually exclusive categories and then counting the frequency of occurrence within each category.

The non-parametric test is an alternative to parametric tests such as T-Test, ANOVA, Linear Regression which are based upon the certain assumptions of the data.

Assumption of Non-parametric Statistics:

- **There are no assumptions about the data's normality.** Non-parametric statistics are also defined as distribution-free since no assumptions are made regarding the distribution of the data. Moreover, the sample size in this situation may be tiny; for instance, N maybe 4 or 6. (Mangal, 2010).
- **Only data with nominal and ordinal scales are measured as variables.** Unlike parametric tests where the variables are measured as interval and ratio scales, the non-parametric test is used on nominal and ordinal scales. Ordinal scales involve the ranking of data based on magnitude whereas nominal scale measure variables that are qualitative as well as exclusive in nature like Country, Department, Branch, etc.



- **No assumption of Independence of observations:** There is no assumption regarding the independence of the relationship between the observations.
- **No need for homogeneity of Variance:** In the non-parametric test the sample and the variance can be heterogeneous.

Some of the non-parametric tests are: The chi-square test, Sign test, Mann Whitney U test, Wilcoxon signed-rank test, Kruskal-Wallis test, etc.

8.3 CHI-SQUARE TEST

One of the most versatile and extensively used statistical tests is chi-square (χ^2). The chi-square is a non-parametric test applied to test any hypothesis relating to categorical data.

$$\chi^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i}$$

Where O_i = Observed frequency

E_i = Expected frequency

Chi-square distribution is defined for a certain number of degrees of freedom. The degree of freedom ν (nu) is equal to the number of classes or categories to be compared as reduced by the number of restrictions imposed on the comparison.

- If we do not have to estimate population parameter from the sample statistic for calculation of expected frequency, then $\nu = k - 1$
- If we have to estimate (m) population parameter from the sample statistic for calculation of expected frequency, then $\nu = k - m - 1$
- When the total frequency is small and $\nu = 1$, then we need to introduce a continuity correction factor equal to 0.5 in computing chi-square value.

$$\chi^2 = \sum_{i=1}^2 \frac{(|O_i - E_i| - 0.5)^2}{E_i}$$

Chi-square distribution properties:

- The chi-square χ^2 value can never be negative as chi-square χ^2 value is a squared value.



- The standard normal variable (Z) lies between -1 and 1 is approximately 68 percent, hence $\chi^2(Z)^2$ will be between 0 and 1 approximately 68 percent of time. Similarly, 27 percent of the time the value lie between 1 and 4.
- The shape of chi-square distribution depends on the degree of freedom. With the increase in the degree of freedom, the distribution trends to be normal.
- For the application of the chi-square test, the data should be given in proportions or percentages which should be converted into frequencies.
- The height of the density function decreases as the value of the random variables increases.
- As the degree of freedom ν (read as nu) becomes large the distribution becomes more symmetrical. In fact, as ν become infinite, its limit is the normal distribution
- The distribution of chi-square χ^2 can be presented graphically as in Figure given below

. Types of chi-square tests:

- Chi-square test of goodness-of-fit
- Chi-square test of independence/ homogeneity
- Chi-square test for difference in proportions

8.3.1 Goodness-Of-Fit Tests:

The chi-square goodness-of-fit test is also known as chi-square test for single sample, is employed in hypothesis testing situation involving single sample. It determines the association between the two variables.

Hypothesis evaluated : In the underlaying population represented by a sample are the observed cell frequencies different form the expected cell frequencies?

The chi-square goodness-of-fit test is used to test the hypothesis that a random variable follows a specified (i.e., hypothesized) distribution. The procedure employs a criterion proposed by Karl Pearson in 1899.

The chi-square goodness-of-fit test is based upon following assumptions

- The data is categorical/nominal. The test data should represent frequencies for k mutually exclusive categories
- The data evaluated consist of a random sample of n independent observations.
- The expected frequency of each cell is 5 or greater.



The following are the steps of computation:

1. Formulate the hypothesis:
 - a. Null Hypothesis: Data is consistent with specified population proportions
 - b. Alternative Hypothesis: Data is not consistent with the specified population proportions
2. Select an appropriate level of significance (α) for testing the null hypothesis (H_0). (E.g., 1%, 5%, 10% etc)
3. Compute the expected frequencies (E_i) of the occurrence of events. The expected frequency counts at each level of the categorical variables are equal to the sample size times that hypothesized proportion from the null hypothesis.

$$E_i = n \times p_i$$

Where E_i = expected frequency count at the level i ,

n = sample size

p_i = probability or proportion of observation in level i .

4. The formula used to calculate chi-square is

$$\chi^2 = \sum_{i=1}^J \frac{(O_i - E_i)^2}{E_i}$$

Suppose we are given a set of J categories and from sample data the observed number of occurrences of the random variable in each category: O_j where $j = 1, 2, \dots, j$. Under the null hypothesis we assume that a set of expected frequencies from the J categories can be computed: E_j where $j = 1, 2, \dots, j$. The data could be presented as in Table below

Category	1	2	3		J
Observed frequency	O_1	O_2	O_3		O_J
Expected frequency	E_1	E_2	E_3		E_J

Using Pearson's criterion, we compute a statistic known as Pearson's χ^2 .



$$\chi^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i}$$

For the goodness of fit test, the number of degrees of freedom (ν) is always $i-1$. One degree of freedom is lost because the sum of the expected frequencies is forced to equal the sum of the observed frequencies.

Null Versus alternative hypotheses

Null hypothesis H_0 : $O_i = E_i$ for all cells

Where O_i = observed frequency for i^{th} cell

E_i = Expected frequency for i^{th} cell

Alternative hypothesis H_1 : $O_i \neq E_i$ for at least one cell.

Where O_i = observed frequency for i^{th} cell

E_i = Expected frequency for i^{th} cell

Illustration 1

We are given a die that we hypothesize to be fair: i.e., all faces equally likely to appear. Then, if p_j equals the probability that the k^{th} face appears, suppose that the die is thrown 54 times and the one-through-six faces are observed 5, 8, 9, 14, 12 and 6 times respectively. If the die is fair, we would expect to see each face 9 times. The data are presented in table given below.

Null Hypothesis

H_0 : $p_j = 1/6$ (The die is fair)

Alternate Hypothesis

H_1 : $p_j \neq 1/6$ (The die is not fair)

Level of Significance, $\alpha = 0.01$

Calculation of chi-square $\chi^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i}$

Die	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	5	9	-4	16	1.78
2	8	9	-1	1	0.11
3	9	9	0	0	0
4	14	9	5	25	2.78



5	12	9	3	9	1
6	6	9	-3	9	1
	$\sum O_i$ =54	$\sum E_i$ 54	$\sum (O_i - E_i) = 0$		$\sum \chi^2 = 6.67$

The critical value of chi-square χ^2 with 5 degree of freedom ($v = k-1$) and α of 0.01 is 15.09 (given in chi-square table). Since our computed χ^2 value of 6.67 is less than the critical χ^2 (15.09) we fail to reject the null hypothesis. Thus, we can conclude that the die is fair. We could have rejected the null hypothesis if the computed χ^2 exceeds the critical value of χ^2 .

Illustration 2

An automobile company routinely purchases a certain type of raw material in lots of 1,000. The purchase department spread the orders among suppliers M,N,O,P in the ratio of 2:2:1:1. Data of 48 purchase orders are randomly selected from the last one-year orders. Supplier M,N,O and P have received 26,8,8 and 6 orders respectively. Does the purchases were done as per the given ratio? Test at $\alpha = 0.01$.

Ho: purchases are distributed in the ratio of 2:2:1:1

H1: purchases are not distributed in the ratio of 2:2:1:1

Under Ho:

Supplier M is expected to receive = $2/(2+2+1+1) = 1/3$ of the purchase orders. (16)

Supplier N is expected to receive = $2/(2+2+1+1) = 1/3$ of the purchase orders. (16)

Supplier O is expected to receive = $1/(2+2+1+1) = 1/6$ of the purchase orders. (8)

Supplier P is expected to receive = $1/(2+2+1+1) = 1/6$ of the purchase orders. (8)

Calculation of chi-square $\chi^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i}$

Supplier	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
M	26	16	10	100	6.25
N	8	16	-8	64	4



O	8	8	0	0	0
P	6	8	-2	4	0.5
	$\sum O_i$ =48	$\sum E_i = 48$	0		$\sum = 10.75$

The critical value of chi-square χ^2 with 3 degree of freedom ($v = k-1$) and α of 0.10 is 6.25 (given in chi-square table). Since our computed χ^2 value of 10.75 is more than the critical χ^2 (6.25) so we reject the null hypothesis. Thus, we can conclude that the purchases are not distributed in the ratio of 2:2:1:1.

Illustration 3

ABC Sports company manufactures cricket bat of 3 different sizes. The company claims 60% of bats are large, 30% Medium and 10 % small. Suppose a random sample of 100 cricket bats has 45 large size, 50 medium and 5 small sizes. Is this consistent with the company's claim? Use a 0.05 level of significance.

Solution:

Let p_l = the proportion of large
 p_m = the proportion of medium
 p_s = the proportion of small

Null hypothesis: $H_0 : p_l = 0.60 \quad p_m = 0.30 \quad p_r = 0.10$

Alternate Hypothesis: $H_a : \text{at least one of the proportions in the null hypothesis is false.}$

Calculation of expected frequency:

Expected frequency large size = $100 \times 0.6 = 60$

Expected frequency medium size = $100 \times 0.3 = 30$

Expected frequency small size = $100 \times 0.1 = 10$

The formula for calculation of chi-square $\chi^2 = \sum_{i=1}^2 \frac{(O_i - E_i)^2}{E_i}$



Bat Size	Observed (O _i)	Expected (E _i)	(O _i -E _i)	(O _i -E _i) ²	$\frac{(O_i - E_i)^2}{E_i}$
Large	45	60	-15	225	3.75
Medium	50	30	20	400	13.33
Small	5	10	-5	25	2.5
$\sum O_i = 100$		$\sum E_i = 100$	0		$\Sigma = 19.58$

The computed value of the chi-square test is 19.58.

The critical value of chi-square χ^2 with 2 degree of freedom ($v = k-1$) and α of 0.00 is 5.99 (given in chi-square table). Since our computed χ^2 value of 19.58 is more than the critical χ^2 (6.25) so we reject the null hypothesis. Thus, we can conclude that the cricket manufactured by ABC Sport company is not in the proportion of 6:3:1

8.3.2 Test of Independence of Variables/ Test of association

Chi-square test of independence is applied when you have two categorical variables form a single population. It is used to determine whether there is a significant association between the two variables. Here the data are represented in tabular form in row and column. These tests are also termed contingency-table tests because the tables representing cross-classified data are called contingency tables.

For example: if you want to understand whether there is any association between gender (Male & Female) and profession (Teacher & Doctor), then we employ test of independence to determine whether the variables are statistically independent. Suppose the response obtained are given below

Gender	Teacher	Doctor	Total
Male	32	48	80
Female	16	24	40
Total	48	72	120



Here the respondents are classified into four categories: Male teacher, Female teacher, Male Doctor and Female doctor.

The following are the steps of computation

- Formulate the hypothesis:
 - **Null Hypothesis:** There is no association between the two variables
 - **Alternate Hypothesis:** There is an association between the two variables
- or
- **Null Hypothesis:** The two variables are independent of each other
 - **Alternate Hypothesis:** The two variables are not independent of each other
 - Select an appropriate level of significance α for testing the null hypothesis.
 - Compute the expected frequency (E_i). The expected frequency counts at each level of the categorical variable are equal to the sample size times the hypothesized proportion from the null hypothesis

$$E_{ij} = \frac{R_i \times C_j}{n} = \frac{\text{Row total} \times \text{Column total}}{\text{Grand total}}$$

Where E_i = expected frequency count at level i

N = sample size

R_i = Total observation of i^{th} row

C_j = Total observation of j^{th} column

- The formula used to calculate chi-square test statistics is

$$\chi^2 = \sum_{i,j=1}^{rc} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Where, O_{ij} = Observed frequency counts at ij^{th} cell

E_{ij} = expected frequency

r = number of rows

c = number of columns

- Estimate the table value of chi-square test statistics by using chi-square distribution table with a degree of freedom = $(r-1)(c-1)$ and α percent of level of significance.



- Decision rule: Compare the computed chi-square with critical value of chi-square. If the computed value is greater than critical value (from chi-square table) then rejects the null hypothesis. And if computed chi-square value is smaller than or equal ($\leq$) to critical value of chi-square then we fail to reject the null hypothesis.

Illustration 4

Apple a smartphone manufacturing company provides you with the following information about age groups and the liking for a particular model of smartphones which it plans to launch.

On the basis of above data, can we conclude that the model appeal is independent of the age of the respondents? Use $\alpha = 0.05$ percent.

Customer who	Age (< 25)	Age (25-50)	Age (>50)
Like	90	60	50
Unlike	110	40	50

Null Hypothesis H_0 : the model appeal is independent of age.

Alternate hypothesis: H_1 : The model appeal is not independent of age.

Level of significance = $\alpha = 0.05$

Decision rule: $\chi^2_{2,0.05} > 5.9915$ then reject the null hypothesis.

Calculation of Expected frequency

Customer who	Age (< 25)	Age (25-50)	Age (>50)	Total
Like	90	60	50	200
	100	50	50	
Unlike	110	40	50	200
	100	50	50	
Total	200	100	100	400



Calculation of Chi-square value

Observed frequency	Expected frequency	(O _i -E _i)	(O _i -E _i) ²	(O _i -E _i) ² /E _i
90	100	-10	100	1
60	50	10	100	2
50	50	0	0	0
110	100	10	100	1
40	50	-10	100	2
50	50	0	0	0
Total				6

Conclusion: The computed value of the chi-square test is 6.

The critical value of chi-square χ^2 with 2 degrees of freedom ($v = (c-1)(r-1) = (3-1)*(2-1)$) and α of 0.05 is 5.99 (given in chi-square table). Since our computed χ^2 value 6 is more than the critical χ^2 (5.99), so we reject the null hypothesis. Thus, we can conclude that the model appeal is not independent of age.

IN-TEXT QUESTIONS**True or False**

11. Non-parametric test deal with cases in which the assumption of normality of the population is not satisfied.
12. Non-parametric tests are also called distribution-free tests because they do not involve tests related to probability distribution
13. The observed frequency of every class should be at least equal to 5.
14. The O and E used to calculate χ^2 must be absolute frequency
15. In goodness-of-fit test, the rejection region lies in both tails of the distribution.



8.4 SIGN TEST

The sign test is used to test the null hypothesis that the median of a distribution is equal to some value. It is applicable when the objective is to compare samples from two populations and the samples are matched pairs. It can be used a) in place of a one sample t-test b) paired sample t-test, c) for ordered categorical data (ordinal data) where a numerical scale is inappropriate but where it is possible to rank the observations.

Types of sign tests:

- One-sample sign test
- Two sample sign tests.

8.4.1 One sample sign test

The sign test can test if the median of a collection of numbers is significantly greater than or less than a specified hypothesized value.

The following steps are required to calculate the sign test:

- Calculate the median value. The value greater than the median value is treated with positive signs and values lesser than the median value is treated as a negative sign. If the value is equal to the median value, that observation is excluded from data, and the size of the sample gets reduced accordingly.
- For hypothesis testing, it is assumed that positive sign values are equal to negative sign values. Hence, the data for the small sample is assumed binomial distribution. The data for a large sample is assumed normal distribution. Data is considered as large sample when np and nq are at least 5. In such a situation Z-statistic is used. Formulate null (H_0) and Alternate (H_1) hypothesis:

Null hypothesis: $p = \frac{1}{2}$

Alternative hypothesis: $p \neq \frac{1}{2}$ (Two-tail) or $p > \frac{1}{2}$ or $p < \frac{1}{2}$ (One-tail)

- Select an appropriate level of significance (α) for testing the null (H_0) hypothesis. (Usually, 0.05/0.01/0.10 is used).
- The formula used to estimate Z-test is:

$$Z = \frac{X - \mu_p}{\sigma_X}$$



- Where X = observed number of positive signs in sample. For large sample (np and nq are at least 5) X value is increased by 0.5 if $X < np$ and decreased by 0.5 if $X > np$, μp = mean of the binomial distribution, σ_x = standard deviation of the binomial distribution

$$\mu p = np$$

$$\sigma_x = \sqrt{np(1-p)} = \sqrt{npq}$$

Where n = sample size, p = probability of occurrence = $\frac{1}{2}$, $q = 1 - p = \frac{1}{2}$

- Estimate the table value for critical regions by using the Z-test statistic table.
- Decision Rule: Compare Z- test computed value and critical value (given in chi-square table) (Decision Rule of the run test).

Illustration 5

In a clinical trial, survival time (weeks) is collected for 10 subjects with non Hodgkin's lymphoma. The exact survival time was not known for one subject who was still alive after 362 weeks when the study ended. The subjects' survival times of 10 subjects are

49, 58, 75, 110, 112, 132, 151, 276, 281, 362

The researcher wished to determine if the median survival time was less than or greater than 200 weeks. Use 5% significance level for testing.

Solution

The median value is given as 200 weeks.

Number of observations below 200 weeks = 7 observations Negative (-)
49, 58, 75, 110, 112, 132, 151

Number of observations above 200 weeks = 3 observations (+) 276, 281, 362

Hypothesis: The null hypothesis is that the median survival is 200 weeks

$H_0: p = \frac{1}{2}$

$H_a: p \neq \frac{1}{2}$

The formula used to estimate one sample Z-test is

$$Z = \frac{X - \mu p}{\sigma X} = \frac{6.5 - 5}{1.581} = 0.949$$



Where $X = 7$. For large sample (np and nq are at least 5) X value is increased by 0.5 if $X < np$ and decreased by 0.5 if $X > np$. Here, $np = 10 \times \frac{1}{2} = 5$, $nq = 5$. Hence, consider the sample as large sample size. Now, $X > np$; X value is decreased by 0.5 which will be $7 - 0.5 = 6.5$

μp = mean of the binomial distribution, σX = standard deviation of the binomial distribution

$$\mu p = np = 10 \times \frac{1}{2} = 5$$

$$\sigma x = \sqrt{np(1-p)} = \sqrt{npq} = \sqrt{10\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = 1.581$$

Where $n = 10$, p = probability of occurrence = $\frac{1}{2}$, $q = 1 - p = \frac{1}{2}$

The critical value of the Z-test at 5% level of significance lies between (± 1.96). As the calculated value (0.949) of Z is lies between the table values (± 1.96), the null hypothesis (H_0) is accepted. Therefore, the median survival is 200 weeks.

8.4.2 Two-sample sign test

The sign test is used when two observation of the same sample is done, one before treatment another one after the treatment. It uses only sign between two values let's say x and y . Sign test will only observe whether $x > y$ or $x = y$ or $x < y$. The values are not used in the computation process.

The following steps are required to calculate the sign test:

- calculate the difference between the two observation's value. The difference between two value is given positive or negative sign. Hence two samples are taken, one with a plus and another with a minus sign. If the values have no difference, the values are deleted from the data and the sample size gets reduced accordingly.
- For hypothesis testing, it is assumed that positive sign values are equal to negative sign value. Hence, the data for the small sample is assumed as binomial distribution. The data for a large sample is assumed as normal distribution. Data is considered as large sample size when np and nq are at least 5. In such a situation Z-statistic is used.

Formulate null (H_0) and Alternate (H_1) hypothesis:

Null hypothesis: H_0 = There is no difference before and after observations

Alternative hypothesis: H_1 = There is a difference in before and after observations

- After that, the third (3) step to the sixth (6) step process is the same as in the one-sample sign test.

**Illustration 6:**

Two professors have developed height measurement instrument for students. A researcher has administered 10 students. The results are presented in the below table. Use a 5% level of significance, test the claim that there is no difference between two instruments.

Students	Researcher1	Researcher2
1	142	138
2	140	136
3	144	147
4	144	139
5	142	143
6	146	141
7	149	143
8	150	145
9	142	136
10	148	146

Solution

Students	Researcher1	Researcher2	Difference (R1-R2)
1	142	138	+
2	140	136	+
3	144	147	-
4	144	139	+
5	142	143	-
6	146	141	+
7	149	143	+



8	150	145	+
9	142	136	+
10	148	146	+

There are 8 positive differences and 2 negative differences.

Formulate hypothesis: The null hypothesis is that there is no difference between the instruments used for height measurement.

Ho: There is no difference between the two instruments

Ha: There is a difference in two instruments

$$z = \frac{X - \mu_p}{\sigma_X} = \frac{7.5 - 5}{1.581} = 1.582$$

Where $X=8$. For large sample (np and nq are at least 5) X value is increased by 0.5 if $X < np$ and decreased by 0.5 if $X > np$. Here, $np = 10 \times \frac{1}{2} = 5$, $nq = 5$. Hence, consider the sample as large sample size. Now, $X > np$ ($8 > 5$); X value is decreased by 0.5 which will be $8 - 0.5 = 7.5$

μ_p = mean of the binomial distribution, σ_X = standard deviation of the binomial distribution

$$\mu_p = np = 10 \times \frac{1}{2} = 5$$

$$\sigma_X = \sqrt{np(1-p)} = \sqrt{npq} = \sqrt{10 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right)} = 1.581$$

Where $n=10$, p = probability of occurrence = $\frac{1}{2}$, $q=1-p = \frac{1}{2}$

The critical value of the Z-test at 5% level of significance lies between (± 1.96). As the calculated value (1.581) of Z lies between the critical values (± 1.96), the null hypothesis (H_0) is accepted. Therefore, there is no difference between the two instruments used for measuring height.

8.5 MANN- WHITNEY U TEST

The Mann-Whitney U test is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed. The test is used to examine whether two samples have been drawn from population with the same means. This test is an alternative to a parametric independent sample test.



The following steps are required to calculate the Mann-Whitney U test:

- 1) **Formulate hypothesis:** The two independent samples are equal or not.
Null Hypothesis: Two samples have an equal distribution from the mean
Alternative Hypothesis: Two samples do not have equal distribution

Or

The sample I has the higher or lower distribution of mean as compared to the II sample

- 2) The two independent samples are pooled together into a single sample to provide a rank to each observation in the pooled sample. Arrange all the observations in the ascending order and assign ranks all. If two or more observations have the same value then the rank is equal to the average of the ranks that otherwise be allotted in the pooled sample.
- 3) Now, Segregate the sample and determine the sum of the ranks of each sample (R_1 and R_2). R_1 and R_2 , is the sum of the ranks to sample one and two
- 4) The formula used to estimate U-test is:

$$U_1 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_2$$

$$\text{And note } U_1 + U_2 = n_1 n_2$$

U_1 = U-test with the sample 1,

U_2 = U-test with sample II,

n_1 = Sample size I,

n_2 = sample size II

R_1 = Sum of ranks of I,

R_2 = sum of ranks of II

A large sample is considered when n_1 or n_2 is greater than 10. Then, Z-test used is:

$$Z = \frac{U - \mu_U}{\sigma_U}$$



Where, $\mu_1 = \frac{n_1 n_2}{2}$ $\sigma_1 = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}$

μ_1 = mean of sample I, σ_1 = standard deviation of sample I

- 5) Estimate the table value for critical regions by using the Z-test statistic table. (Refer run test step 4 for computation of z-table critical value).
- 6) Decision Rule: Compare Z- test computed value (given in step 4) and table value (given in step 5)

Illustration 7

The following data shows the age at which type II diabetes was diagnosed in young adults. Use the 5% level of significance to check whether age of diagnosing is different for males and females?

Females: 20 11 17 12 18

Males: 19 22 16 29 24 20 23 25 28 10

Solution:

The two independent sample of females (n_1) and males (n_2) age at the time of diagnose of type II diabetes is given.

Null Hypothesis: H_0 : Two groups of males and females have an equal distribution of age mean

Alternative Hypothesis: H_1 : Two groups of males and females do not have an equal distribution of age mean

Ist step is to pool the data in a large sample and arrange it in ascending order to assign rank.

Next step is to segregate the sample and determine the sum of ranks of each sample (R_1 and R_2)

The formula used to estimate U-test is:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$= (5)(10) + \frac{(5)(10)}{2} - 24.5 = 50.5$$

Here, n_1 (number of females) = 5, n_2 (number of males) = 10, R_1 = 24.5, R_2 = 95.5



A large sample is considered when n_1 or n_2 is at least 10. Then, Z-test used is:

$$Z = \frac{\bar{x}_1 - \mu_1}{\sigma_{\bar{x}_1}} = \frac{50.5 - 25}{8.165} = 3.123$$

Where, $\mu_1 = \frac{n_1 n_2}{2} = \frac{(5)(10)}{2} = 25$

$$\sigma_{\bar{x}_1} = \sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}} = \sqrt{\frac{(5)(10)(5+10+1)}{12}} = 8.165$$

μ_1 = mean of sample I = 25, σ_x = standard deviation of sample I = 8.165

Age	M/F	Rank	Ri Females	R2 Male
10	M	1		2
11	F	2	2	
12	F	3	3	
16	M	4		4
17	F	5	5	
18	F	6	6	
19	M	7		7
20	F	8.5	8.5	
20	M	8.5		8.5
22	M	10		10
23	M	11		11
24	M	12		12
25	M	13		13
28	M	14		14
29	M	15		15
	Total	120	24.5	95.5



The critical value of the Z-test at 5% level of significance lies between (± 1.96). As the calculated value (3.123) of Z is higher than the table values (1.96), the null hypothesis (H_0) is not accepted. Therefore, there is a difference in two groups of males and females as they do not have an equal distribution of age mean.

8.6 WILCOXON SIGNED-RANK TEST

Wilcoxon signed-rank test is used to test the paired or two related samples means. It is used to examine whether their population mean ranks are different in two related samples or not. It can be used as an alternative to paired t-test. The assumptions of the normal distribution are not required in the Wilcoxon signed-rank test. It is a non-parametric test that can be used to determine whether two dependent samples were selected from populations having the same distribution or not.

The following steps are required to calculate the Wilcoxon signed-rank test:

1. Formulate the hypothesis

Null Hypothesis; There is no difference in two related sample mean

Alternate Hypothesis: There is a difference in two related sample mean

2. Compute the difference between the two paired samples. Retain the positive and negative values but remove the value with no difference (zero) which reduces sample size accordingly.
3. Convert the difference in values into absolute values by removing their signs and assign the rank to the absolute difference values. If two or more observations have the same value then the rank is equal to the average of the ranks.
4. Segregate the ranks of absolute differences into positive and negative ranks. The division is possible by re-looking at the signs given in first step difference columns.
5. Calculate the sum of ranks of positive and negative values which is represented as T^+ and T^- respectively. Select T be the smaller of T^+ / T^- .
6. The formula used to estimate Z-test is:

$$Z = \frac{T - \mu T}{\sigma T}$$

$$\text{Where, } \mu T = \frac{n(N+1)}{4} \quad \sigma T = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$



N = sample size, T^+ = sum of positive absolute ranks differences, T^- = sum of negative absolute ranks differences, T = smaller of T^+ or T^- ,

μ_1 = mean of sample I, σ_x = standard deviation of sample T

7) Estimate the table value for critical regions by using the Z-test statistic table.

test step 4 for computation of z-table critical value)

8) Decision Rule: Compare Z- test computed value (given in step 6) and table value (given in step 7)

Illustration 8

The below table provides details of people who lost weight due to medicine -X. The table provides weight details of 12 people before and after taking medicine - X. The weight (in kgs) is given below

Before	7	5	8	9	4	6	8	10	10	6	7	8
After	6	5	3	8	4	9	12	14	13	7	5	11

Does the medicine significantly (5%) impact reduction in weight?

Solution:

The paired samples are given before and after medicine 'X.

Null Hypothesis: H_0 : There is no impact on weight reduction due to medicine X*

Alternative Hypothesis: H_1 : There is an impact on weight reduction due to medicine X

$N = 11$, $T^+ = 47$, $T^- = 19$, T (smaller value of T^+ / T^-) = 19

$$Z = \frac{T - \mu_T}{\sigma_T} = \frac{19 - 33}{11.247} = -1.244$$

$$\text{Where, } \mu_T = \frac{n(n+1)}{4} = \frac{11(11+1)}{4}$$

$$\sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = \sqrt{\frac{11(11+1)(2(11)+1)}{24}} = 11.247$$

μ_1 = mean of sample T, σ_x = standard deviation of sample T



Before	After	d (Before After)	- The value d	absolute The rank of absolute d	Negative rank	Positive rank
7	6	-1	1	2	2	
5	5	0	0	6.5		6.5
8	3	-5	5	11	11	
9	8	-1	1	2	2	
4	4	0	0			
6	9	3	3	6.5		6.5
8	12	4	4	9.5		9.5
10	14	4	4	9.5		9.5
10	13	3	3	6.5		6.5
6	7	1	1	2		2
7	5	-2	2	4	4	
8	11	3	3	6.5		6.5
Total				66	19	47

The critical value of the Z-test at 5% level of significance lies between (± 1.96). As the calculated value (-1.244) of Z lies between the table values (± 1.96), the null hypothesis (H_0) is accepted. Therefore, there is no impact on weight reduction due to medicine X.

8.7 KRUSKAL- WALLIS TEST

The Kruskal -Wallis test is used to compare differences of more than two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed. It is also used when the sample size is very small. The test is used to examine whether the 'n' number of samples has been drawn from population with the same mean or median. This test is an alternative to a parametric one-way ANOVA test.

The following steps are required to calculate the Kruskal-Wallis test:



1) Formulate the hypothesis

Null Hypothesis: the samples (groups) come from populations with the same distribution.

Alternative Hypothesis: At least one of the samples (group) comes from different populations

2) The 'n' independent samples are pooled together into a single sample to provide a rank to each observation in the pooled sample. Arrange all the observations in the ascending order and assign rank to all. If two or more observations have the same value then the rank is equal to the average of the ranks that otherwise be allotted in the pooled sample.

3) Now, segregate the sample and determine the sum of ranks of each sample ($r_1, r_2, \dots$)

Let r_i be the sum of all ranks $r_1 + r_2 + \dots$ till i^{th} sample.

4) The formula used to estimate H-test and follows the chi-square distribution:

$$H = \frac{12}{n(n+1)} \sum_i^k \frac{r_i^2}{n_i} - 3(n+1)$$

Where k = number of groups, n = total number of elements in k groups

5) Estimate the table value for critical regions by using chi-square distribution to test the null hypothesis with a degree of freedom, $df=k-1$ and at alpha percent level of significance.

6) Decision Rule: If the Computed value > Critical Table value, the null hypothesis is rejected. In other words, if the computed value $\leq$ Critical Table value, the null hypothesis is accepted.

Illustration 9

Does physical exercise alleviate depression? The researcher finds some depressed

people and check that they are all equivalently depressed. The allocation of each person is done randomly to one of three groups: no exercise; doing jogging every day or doing gym every day. At the end of the month, each participant needs to rate how depressed they feel now on a Likert scale that runs from 1 ("totally miserable") to 100 (ecstatically happy"). The table is given below:



NE	Jog	Gym
23	22	59
26	27	66
51	39	38
49	29	49
58	46	56
37	48	60
29	49	56
44	65	62

Use 5% level of significance to check whether physical exercise significantly effects in the alleviation of depression or not?

Solution:

The researcher has three sample groups with no exercise, doing jogging and gymming.

The hypothesis is:

Null Hypothesis: H_0 : Alleviation of depression level is the same from all types of physical exercises

Alternative Hypothesis: H_1 : Alleviation of depression level in at least two is different from all

types of physical exercises

The next step is to pool the data of all the samples and arrange them in ascending order. After that assign rank and segregate based on type of physical exercise. Then do some of the ranks of each sample and label it 'r1', 'r2', 'r3'.



Pooled data	K sample	Ascending order		ri Rank	R1 (NE)	R2 (Jog)	R3 (Gym)
23	NE	22	Jog	1		1	
26	NE	23	NE	2	2		
51	NE	26	NE	3	3		
49	NE	27	Jog	4		4	
58	NE	29	NE	5.5	5.5		
37	NE	29	Jog	5.5		5.5	
29	NE	37	NE	7	7		
44	NE	38	Gym	8			8
22	Jog	39	Jog	9		9	
27	Jog	44	NE	10	10		
39	Jog	46	Jog	11		11	
29	Jog	48	Jog	12		12	
46	Jog	49	NE	14	14		
48	Jog	49	Jog	14		14	
49	Jog	49	Gym	14			14
65	Jog	51	NE	16	16		
59	Gym	56	Gym	17.5			17.5
66	Gym	56	Gym	17.5			17.5
38	Gym	58	NE	19	19		
49	Gym	59	Gym	20			20
56	Gym	60	Gym	21			21
60	Gym	62	Gym	22			22
56	Gym	65	Jog	23		23	
62	Gym	66	Gym	24			24
Total				300	76.5	79.5	144



The formula used to estimate H-test and follows the y distribution:

$$H = \frac{12}{n(n+1)} \sum_i^k \frac{r_i^2}{n_i} - 3(n+1)$$

$$H = \frac{12}{24(24+1)} \left[\frac{76.5^2}{8} + \frac{79.5^2}{8} + \frac{144^2}{8} \right] - 3(24+1)$$

$$= 7.27$$

Where $k = 3$ and $n = 8+8+8 = 24$

Estimate table value by using chi-square distribution with degree of freedom, $df = k - 1 = 3 - 1 = 2$. And at 5% level of significance. The table value is 5.991

Here, computed value (7.27) > Critical Table value (5.991). The null hypothesis (H_0) is rejected. That means alleviation of depression level in at least two is different from all types of physical exercises. There is a significant effect of exercise on depression levels.

Illustration: In a college 4 (Four) different groups of students were randomly assigned to be taught by 4 different teachers, and their achievement test scores were recorded. Are the distributions of test scores the same, or do they differ?

1	2	3	4
67	75	59	94
87	69	78	89
73	83	67	80
79	81	62	88

Solution: Rank the 16 measurements from 1 to 16, and calculate the four rank sums.

1	2	3	4
67 (3)	75 (7)	59 (1)	94 (16)
87 (13)	69 (5)	78 (8)	89 (15)
73 (6)	83 (12)	67 (4)	80 (10)
79 (9)	81 (11)	62 (2)	88 (14)



T_i 31 35 15 55

$$\begin{aligned} H &= \frac{12}{n(n+1)} \sum_i^k \frac{r_i^2}{n_i} - 3(n+1) \\ &= \frac{12}{16(16+1)} \left[\frac{31^2 + 35^2 + 15^2 + 55^2}{4} \right] - 3(16+1) \\ &= 8.96 \end{aligned}$$

H_0 : the distributions of scores are the same

H_a : the distributions of scores are not same

Rejection region: For a right tailed chi-square test with $\alpha = .05$ and $df = 4-1 = 3$, reject H_0 if $H \geq 7.81$.

Reject H_0 . There is sufficient evidence to indicate that there is a difference in test scores for the four teaching techniques.

8.8 KOLMOGOROV-SMIRNOV TEST

Kolmogorov- Smirnov test or K-S Test is a nonparametric test named after Andrey Kolmogorov and Nikolai Smirnov which is used as a goodness of fit and is suitable for small sample size. It is the test of the equality of continuous or discontinuous, one-dimensional probability distributions that can be used to compare a sample with a reference probability distribution (one-sample K–S test), or to compare two samples (two-sample K–S test). In essence, the test answers the question "What is the probability that this collection of samples could have been drawn from that probability distribution?" or, in the second case, "What is the probability that these two sets of samples were drawn from the same (but unknown) probability distribution?". It is named after Andrey Kolmogorov and Nikolai Smirnov.

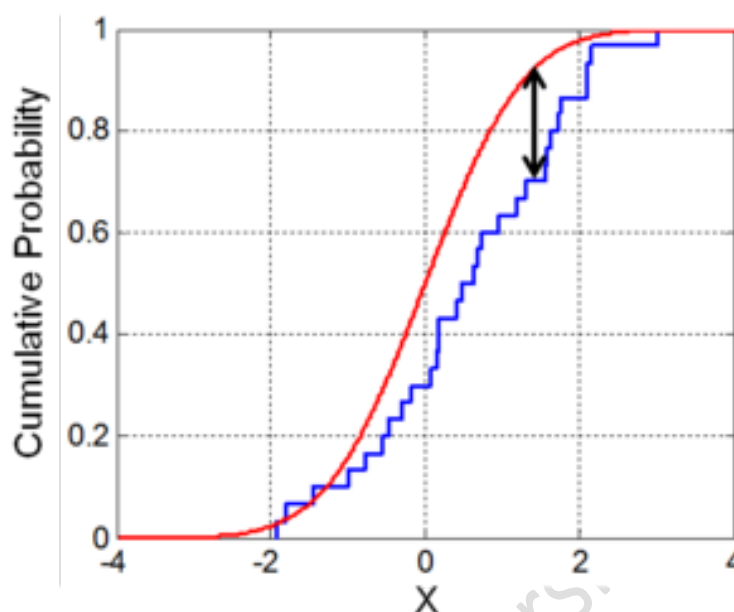


Figure 6.1: Illustration of the Kolmogorov–Smirnov statistic. The red line is a model CDF, the blue line is an empirical CDF, and the black arrow is the K–S statistic

The Kolmogorov–Smirnov statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The null distribution of this statistic is calculated under the null hypothesis that the sample is drawn from the reference distribution (in the one-sample case) or that the samples are drawn from the same distribution (in the two-sample case). In the one-sample case, the distribution considered under the null hypothesis may be continuous, purely discrete or mixed. In the two-sample case, the distribution considered under the null hypothesis is a continuous distribution but is otherwise unrestricted. However, the two-sample test can also be performed under more general conditions that allow for discontinuity, heterogeneity and dependence across samples.

Let's assume that we have observation $X_1, X_2, \dots, X_n$, which come from a distribution P . The KS Test is used here to test whether the sample belong to P distribution.

Null Hypothesis H_0 = the sample come from P distribution

Alternate Hypothesis H_1 = the sample do not come from P distribution

The value of test statistics D is calculated as:

$$D = \text{Maximum } |F_o(X) - F_r(X)|$$



Where:

- $F_0(X)$ = Observed cumulative frequency distribution of a random sample of n observations.
- and $F_0(X) = k/n = (\text{No. of observations} \leq X)/(\text{Total no. of observations})$.
- $F_r(X)$ = The theoretical frequency distribution.

The critical value of D is found from the K-S table values for one sample test.

Acceptance Criteria: If calculated value is less than critical value accept null hypothesis.

Rejection Criteria: If calculated value is greater than table value reject null hypothesis.

Illustration: Data was collected from 60 students (12 from each program) enrolled in different programs to understand the intention of student to join the NCC.

Course	B.Sc. Maths	B.A. Eco	B.Com	MBA	BMS
Number of students	5	9	11	19	16

It was expected that 12 students from each class would join the NCC. Using the K-S test to find if there is any difference among student classes with regard to their intention of joining the NCC.

Solution:

H_0 : There is no difference among students of different program with respect to their intention of joining the NCC.

We develop the cumulative frequencies for observed and theoretical distributions.

Streams	No. of students interested in joining		$F_0(X) = O/N$	$F_T(X) = T/N$	$ F_0(X) - F_T(X) $
	Observed (O)	Theoretical (T)			
B.Sc Maths	10	12	0.1667	0.2	0.17
B.A. Eco	11	12	0.1833	0.2	0.18



B.Com	15	12	0.25	0.2	0.25
BMS	10	12	0.1667	0.2	0.17
MBA	14	12	0.2333	0.2	0.23

Test statistic $|D|$ is calculated as:

$$D = \text{Maximum}|F_0(X) - F_T(X)| = 0.25$$

The table value of D at 5% significance level is given by

$$D_{0.05} = 1.36/\sqrt{n} = 1.36/\sqrt{60} = 0.175$$

Since the calculated value (0.25) is greater than the critical value (0.175) hence we reject the null hypothesis and conclude that there is a difference among students of different streams in their intention of joining the Club.

K-S Two Sample Test

In case of two independent samples, we can use the K-S two sample test. The test is used to test the agreement between two cumulative distributions. The null hypothesis (H_0) states that there is no difference between the two distributions. The D-statistic is calculated in the same manner as the K-S One Sample Test

Formula

$$D = \text{Maximum}|F_{n_1}(X) - F_{n_2}(X)|$$

Where –

- n_1 = Observations from first sample.
- n_2 = Observations from second sample.

It has been seen that when the cumulative distributions show large maximum deviation $|D|$ it is indicating towards a difference between the two sample distributions.

The critical value of D for samples where $n_1 = n_2$ and is ≤ 40 , the K-S table for two sample case is used. When n_1 and/or $n_2 > 40$ then the K-S table for large samples of two sample test should be used. The null hypothesis is accepted if the calculated value is less than the table value and vice-versa.



8.9 SUMMARY

- Non-parametric tests make fewer and less stringent assumptions like data is not requirement to be normally distributed, data is drawn by random sampling, variables association can be non-linear and observations are independent of each other. The test is mainly used non-metric data i.e., ordinal or nominal.
- The Chi-square test is a non-parametric test used for mainly two purposes. First it determines whether a sample data matches the population. This is a test of goodness of fit. Secondly, it determines the association between the two variables. The comparison of two contingency (categorical) variables are checked to find whether they are related or not. This is a test of the independence of variables. Chi-square test has certain properties like data is unsymmetrical values greater than or equal to zero and categorical variables.
- The sign test is used to test the null hypothesis that the median of a distribution is equal to some value. It can be used a) in place of a one-sample t-test b) in place of a paired t t-test some value. or c) for ordered categorical data where a numerical scale is inappropriate but where it is possible to rank the observations.
- The Mann-Whitney U test is used to examine whether two samples have been drawn from population with the same mean. This test is an alternative to a parametric independent sample test.
- The Wilcoxon signed-rank test is used to test the paired or two related samples means. It is used to examine whether their populations mean ranks are different in two related samples or not. It can be used as an alternative to paired t-test.
- The Kruskal-Wallis test is used to examine whether the 'n' number of samples has been drawn from populations with the same mean or median. This test is an alternative to a parametric one-way ANOVA test.
- Kolmogorov- Smirnov test or K-S Test is a nonparametric test which is used as a goodness of fit and is suitable for small sample size.



8.10 ANSWERS TO IN-TEXT QUESTIONS

- | |
|------|
| 1. T |
| 2. F |
| 3. F |
| 4. T |
| 5. F |

8.11 SELF-ASSESSMENT QUESTIONS

1. The nonparametric equivalent of an unpaired samples t-test is
 - a) Sign test
 - b) Wilcoxon signed-rank test
 - c) Mann-Whitney U test
 - d) Kruskal-Wallis test
2. The Wilcoxon rank-sum test compares
 - a) Two populations
 - b) Three or more populations
 - c) A sample mean to the population mean
 - d) None of these
3. The Wilcoxon rank-sum test can be
 - a) Upper tailed
 - b) Lower tailed
 - c) Either of upper tailed or lower tailed
 - d) None of these
4. If a Chi-square goodness of fit test has 6 categories and an $N=30$, then the correct number of degrees of freedom is
 - a) 5
 - b) 6
 - c) 28
 - d) 29



5. Comparing the times-to-failure of radar transponders made by firms A, B, and C, based on an airline's sample experience with the three types of instruments, one may well call for:
 - a) A Kolmogorov-Smirnov test
 - b) A Kruskal-Wallis test
 - c) A Wilcoxon rank-sum test
 - d) A Spearman rank-correlation test
6. Which of the following test use rank sums?
 - a) F test
 - b) Chi-square and Sign tests
 - c) Runs test
 - d) Kruskal-Wallis and Wilcoxon test
7. In the Kruskal-Wallis test of samples, the appropriate number of degrees of freedom is
 - a) K
 - b) K-1
 - c) K-2
 - d) n-K
8. When using the Sign test, if two scores are tied, then we
 - a) Count them
 - b) Discard them
 - c) Depends upon the scores
 - d) None of these
9. The Mann-Whitney U test is preferred to a t-test when
 - a) Data are paired
 - b) Sample sizes are small
 - c) The assumption of normality is not met
 - d) Samples are dependent
10. The sign test assumes that the samples are
 - a) Independent
 - b) Dependent
 - c) Have the same mean
 - d) None of these



11. The sign test is
- Less powerful than that of the Wilcoxon signed-rank test
 - More powerful than the paired sample t-test
 - More powerful than the Wilcoxon signed-rank test
 - Equivalent to the Mann-Whitney test
12. Which of the following test is most likely assessing the null hypothesis of “the number of violations per apartment in the population of all city apartments is binomially distributed with a probability of success in any one trial of $P=0.4$ ”
- The Kolmogorov-Smirnov test
 - The Kruskal-Wallis test
 - The Mann-Whitney test
 - The Wilcoxon signed-rank test
13. The Wilcoxon signed-rank is used
- Only with independent samples
 - Only in matched pairs samples
 - As an alternative to the Kruskal-Wallis test
 - To test for randomness
14. Which of the following tests must be two-sided?
- Kruskal-Wallis test
 - Wilcoxon Signed rank test
 - Runs test
 - Sign test
15. In testing for the difference between two populations, it is possible to use
- The Wilcoxon rank-sum test
 - The sign test
 - Either of the above
 - None of the these

Answers MCQ

- d) Kruskal-Wallis test
- a) Two populations
- c) Either of upper tailed or lower tailed
- a) 5



- 5. b) A Kruskal-Wallis test
- 6. d) Kruskal-Wallis and Wilcoxon test
- 7. b) K-1
- 8. b) Discard them
- 9. c) The assumption of normality is not met
- 10. b) Dependent
- 11. a) Less powerful than that of the Wilcoxon signed-rank test
- 12. a. The Kolmogorov-Smirnov test
- 13. b) Only in matched pairs samples
- 14. d) Sign test
- 15. c) Either of the above

Questions

- 15. A poker-dealing machine is supposed to deal cards at random, as if from an infinite deck. In a test, you counted 1600 cards, and observed the following: Spades 404 Hearts 420 Diamonds 400 Clubs 376 Could it be that the suits are equally likely? Or are these discrepancies too much to be random?.
- 16. A genetics engineer was attempting to cross a tiger and a cheetah. She predicted a phenotypic outcome of the traits she was observing to be in the following ratio 4 stripes only: 3 spots only: 9 both stripes and spots. When the cross was performed and she counted the individuals she found 50 with stripes only, 41 with spots only and 85 with both. According to the Chi-square test, did she get the predicted outcome?
- 17. The personnel department of a large company wanted to see if there is any pattern in the casual leave taken by the employees on weekdays. A random sample of 132 cases is taken and classified by the day of the week the leave was taken. The observed frequencies are given below:

Day	Mon	Tue	Wed	Thu	Fri	Sat
Observed frequency:	21	26	12	15	23	35

Can the personnel department conclude that there is no difference among days of week insofar as casual leave is concerned? Test at 5 percent level of significance.



18. The divisional manager of a retail chain believes that the average number of customers entering each of the five stores in his division weekly is the same. In a given week, the manager reports the following number of customers in the five stores: 3,000; 2,960; 3,100; 2,780; and 3,160. Test the divisional manager's belief at 5 percent level of significance.
19. In setting sales targets, the marketing manager makes the assumption that order potentials are the same for each of the four sales territories. A sample of 200 sales data is given below:

Sales Territories

I	II	III	IV
120	95	118	72

Should the manager's assumption be rejected? Use $\alpha = 0.05$.

20. A sample analysis of examination results of 1,000 students was made. It was found that 260 students had failed, 110 students secured first division, 420 secured second division and the rest obtained third division. Are these figures commensurate with the general examination result which is in the ratio of 2:1:4:3, respectively, for the various categories? You are given that the critical value of chi-square for 3 degrees of freedom at 5% level of significance is 7.81
21. The following table gives the number of car accidents in Delhi that occurred during various days of the week.

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat
No. of accidents:	14	16	8	12	11	9	14

Test whether the accidents are uniformly distributed over the week.

22. 7. A die is tossed 180 times with the following results:

No. Turned Up:	1	2	3	4	5	6
Frequency:	25	35	40	22	32	26

Test the hypothesis that the die is unbiased. Take $\alpha = 0.05$.

23. A survey of 300 families, with three children each, selected at random gave the following results:



No. of Male Children:	0	1	2	3
No. of Families:	60	87	93	60

Are these data consistent with the hypothesis that male and female births are equally likely, at a lev of significance of 0.01?

24. The first proof of 200 pages of a book containing 560 pages revealed the number of printing errors: following distribution of the

No. of errors in a page:	0	1	2	3	4	5
Total						
No. of pages:	112	63	20	3	1	200

Fit a Poisson distribution corresponding to these data and test the goodness of fit.

8.12 REFERENCES

1. Aczel, Amir D., Sounderpandian, J., & Saravanan P. (2012). *Complete Business Statistics* (latest ed.). India: McGraw Hill Education India Private Limited.
2. Albright, S.C., & Winston, W.L. (2015). *Business Analytics: Data Analysis and Decision Making* (Latest ed.). Cengage Publications.
3. Anderson, D., Sweeney, D., Williams, T., Camm, J., & Cochran, J. (2014). *Statistics for Business & Economics* (Latest ed.). India: Cengage Learning.
4. Anderson, D., Sweeney, D., Williams, T., Martin, R.K. (2012). *An Introduction to Management Science: Quantitative Approaches to Decision Making* (13th ed.). India: Cengage Learning.
5. Balakrishnan, N., Render, B., Stair, R. M., & Munson, C. (2017). *Managerial decision modeling*. Upper Saddle River, Pearson Education.
6. Powell, S. G., & Baker, K. R. (2017). *Business analytics: The art of modeling with spreadsheets*. Wiley.
7. Stine, R.E., & Foster, D. (2013). *Statistics for Business: Decision Making and Analysis* (2nd ed.). India: Pearson Education



LESSON-9

REPORT WRITING

Dr. Supreet Kaur
Assistant Professor
Department of Commerce
University of Delhi
Email-Id: supreetkaur@commerce.du.ac.in

STRUCTURE

- 9.1 Learning Objectives
- 9.2 Introduction
- 9.3 Importance of Report Writing
- 9.4 Types of Research Reports
- 9.5 Structure of Report
- 9.6 Guidelines for effective report writing
- 9.7 Summary
- 9.8 Glossary
- 9.9 In-Text Questions
- 9.10 Self-Assessment Questions
- 9.11 References
- 9.12 Suggested Readings

9.1 LEARNING OBJECTIVES

- classify the various types of reports.
- understand the relevance of a good research report.
- understand the process of report writing.
- understand the key features of a report.

9.2 INTRODUCTION

Abhiraj Desai, a Project Leader at Vayu Research Centre, Chennai was worried about the results of the in-depth study of the rural communities conducted by his team. They measured



the impact of organic cultivation on the lives of the farmers. The team conducted an extensive field study using a standardised questionnaire, formulated a few hypotheses, and went through various secondary reports and surveys but found that the impact of the cultivation on the lives of the farmers was negligible and the hypothesis was also rejected. Therefore, Abhiraj was worried that the entire effort of the team has gone waste and that they have nothing concrete to show in the board meeting. Their entire efforts of the past 6-7 months have turned futile. he was anxious about the results of the study.

He narrated the entire scenario to Shantanu, the senior consultant. He could not believe his ears when Shantanu remarked that there is nothing to worry about and why is he looking so miserable. Shantanu explained to Abhiraj that every fact whether having a positive impact or a negative needs to be documented. Whatever has been done has to be recorded in a logical and sequential manner in the form of a report. The report will explain the underlying reasons for the negligible impact of organic cultivation on the farmers and will throw light on undiscovered variables. He explained that a concise and succinct report will better indicate their findings and can explain to their clients the work the company is capable of doing. He explained that it does not matter whether the results are not as per expectation what matters is the verity, consistency, applicability and validity of the results.

It is difficult to imagine that a research report which is assumed to be rudimentary will hold such significance in the research. We all at times ignore the relevance of writing a well-defined and structured research report. Just like every other step in the research process, concerted efforts should be directed towards the documentation of the study. It is quintessential that all the elements pertaining to the investigated topic should be well documented and presented in the form of a good research report.

9.3 IMPORTANCE OF REPORT WRITING

Writing a well-structured and integrated report is an art. It should be documented in such a manner that the analysis and interpretation of the major findings of the study can be linked with the hypothesis formulated at the beginning of the study. Proper sequencing and careful noting of the research process adopted should be documented in a well-defined report. Further, a good report is a one-way communication therefore it should be free from any sort of subjectivity and biases.

Writing a good comprehensive report is critical on account of many reasons as under;

- a) A good report serves as a documentary and concrete proof of the efforts taken during a study.



- b) A good report provides a basic framework for further studies in related areas.
- c) A good report adds to the existing body of literature in an advanced manner.
- d) A good report helps to reach a wider audience and since it entails all the crucial steps followed in a study therefore, it validates the quality of the work carried out and supports and substantiates the findings obtained.
- e) A good report can serve as insight and guidance for various practitioners, marketers, academicians and researchers.

9.4 TYPES OF RESEARCH REPORTS

A research report is a well-documented and complete study that records the detailed investigation of identification of a problem, formulation of hypothesis, posed research questions, methods of data collection, analysis and interpretation of data and proposes recommendations and suggestions based on the findings of the study. There are various types of research reports which differ in terms of relevance, length, format and type of research (Qualitative and Quantitative).

- 1) **Brief Reports:** These reports are loosely structured and comparatively provide a synoptic view of the study in a concise manner. The scope of such reports is limited on account of the limited information given in such reports.
 - a) **Survey Reports:** These reports are easy to comprehend and may or may not have academic orientation. They are very lucid and easy to understand and include charts, diagrams and tables.
 - b) **Basic Reports:** These are basic reports entailing the basic framework, scope and significance of the study. It gives mention of the methodology adopted in the study. The key results and findings are well presented in such reports as the emphasis is given primarily on the present study rather than analysing the past literature.
- 2) **Detailed Reports:** These are more formal, structured, academic and research-oriented reports. Such reports can be prepared for both academic and business purposes and therefore the format of these varies depending on the type of audience.
- 3) **Technical Reports:** As the name suggests, such reports are way more structured and include each and every element of the research process. Apart from recording the elements of a basic report, detailed notes of the past sets of information, data sets, interpretation of the data and the summarised conclusion are also recorded. All the mapped sources of data collection and instruments used for the collection of data variables are documented in a logical and sequential manner.

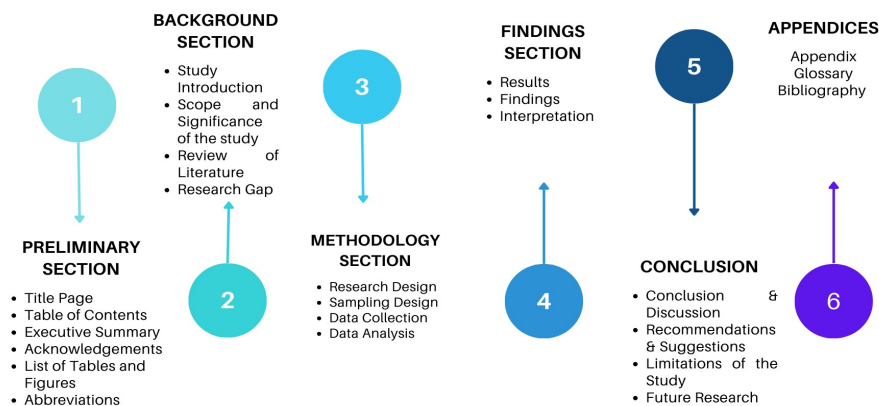


- 4) **Business Reports:** Such reports are prepared keeping in mind the requirement of business managers. More emphasis is paid towards infographics, pictograms and tables and charts. These are more focused on the managerial implications of the study. They are prepared keeping in mind the manager's and industry perspective.

9.5 STRUCTURE OF REPORT

For the purpose of dissemination of information to a wider audience, the report (be it any type of research report) needs to be formal and structured. There is a standardised structure for report writing that needs to be followed. The use of jargon, technicalities, and language may differ depending on the purpose (Business or Academic) but the main theme runs apparently across the report structure. The use of pictograms, and infographics is more in a business report whereas technical jargon would be perfunctory in a technical report. Providing information pertaining to questionnaire design is one of the prefatory questions addressed in a survey report whereas the methodology used is key in an academic report.

Figure 1: The Stages of Report Writing



The Stages of Report Writing

Source: Author's Own Work

The above-mentioned figure 1 clearly gives a brief of the stages involved in the process of report writing. The main components of a research report majorly include:

A) Preliminary Section

The preliminary section includes the prefatory elements of a research report which includes a title page, table of contents, list of tables and figures, list of abbreviations, executive



summary and acknowledgements (if any). The preliminary section is arranged in the manner as given under;

- **Title Page**
- The title page includes all the details regarding the author(s), their name, institutional affiliation, official mail Ids (if any to be substantiated in case of working papers), Organising Agency (if any), Sponsored funding agency (if any).
- The title of the study should be mentioned in a clear and lucid manner.
- The intended audience of the study is also mentioned in the report in certain instances.
- The year of publication is also mentioned on the title page so as to indicate relevance in the current pretext.
- Some of the examples of the title on the title page are:
- Impact of financial inclusion on industrial workers
- Analysing consumer behaviour for organic food market

Table 1: Title page of a Thesis

Analysing the impact of live streaming purchase on the consumer's buying behaviour

A

Thesis

**Presented to the Faculty of Business Studies of the
Rajpur University, India**

in Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

IN

COMMERCE

by

Radhey Shyam

DEPARTMENT OF COMMERCE

RAJPUR UNIVERSITY, INDIA

FEBRUARY, 2023

Source: Author's Own Work



- **Table of Contents**
- The table of contents clearly delineates different sections and subsections of the study.
- Different sections and sub-sections are sequentially paginated.
- Chapterisation of the sections is clearly defined in the table of contents segment.
- An illustrative example of a table of content is given under;

Table 2: Table of Contents

S.No.	Description	Page No.
CHAPTER-I:	Introduction	
1.1	Introduction	1-5
1.2	Types of Consumers	7-10
1.3	Evolution and Conceptual Analysis of Theory of consumer behaviour	11-30
1.4	Objectives of behavioural analysis	30-36
1.5	Review of Initiatives	36-50
1.6	Role of other theories	51-65
1.7	Global Scenario	65-69
1.8	Current Status	70-72
1.9	Challenges	73
	References	
CHAPTER-II:	Review of Literature	
2.1	Literature Review	74-115
2.2	Research Gap	116-117
	References	



CHAPTER-III: Research Methodology		
3.1	Introduction	118-119
3.2	Steps in Research Methodology	120-145
3.2.1	Nature and Scope of Study	146
3.2.2	Need and objectives of Study	147
3.2.3	Significance of Study	148
3.2.4	Limitations of the Study	149
	References	
CHAPTER-IV: Summary, Findings and Suggestions		
4.1	Conclusion	150-151
4.2	Suggestions	152
4.3	Areas for Future Research	153
	Bibliography	
	Annexure	

Source: Author's Own Work

- **List of Tables and Figures**
- The list of tables, figures, graphs and exhibits (if any) follows the table of contents in the structure of a report.
- The list of tables and figures should be clearly numbered.
- The list of tables and figures should be logically paginated.
- The list of tables and figures must match the captions given in the chapter.
- The tables and figures should be logically arranged to avoid any ambiguity.

Further, the list of tables and figures is followed with the abbreviations (if any) used in the report so as to provide clarity of context to the readers.


Table 3: List of Tables

S.No.	Description of Items	Page No.
1.1	Evolution of FinTech Industry	5
1.2	Working Definition of FinTech	8
1.3	Region Wise Scores for Various Dimensions of Financial Inclusion	26
1.4	Summary of Financial Inclusion Progress (1960-2023)	38
4.3	Summary of Year Wise Fund Transfer	44
5.1	Distribution of Total Indian Workforce across Organised-Unorganised Sectors	86
5.3	Cross Tabulation of Schemes Aailed and Type of Industry	124
5.4	KMO and Bartlett's Test	155
5.5	Output from Exploratory Factor Analysis using RCM	156
5.8	Reliability and Convergent Validity	164

Source: Author's Own Work

Table 4: Abbreviations

ABBREVIATIONS	
AGFI	Adjusted Goodness of Fit Index
AMOS	Analysis of Moment Structure
ANOVA	Analysis of Variance
APY	Atal Pension Yojana
ATM	Automated Teller Machines
EFA	Exploratory Factor Analysis
EVA	Economic Value Added
GFI	Goodness of Fit Index



KMO	Kaiser Meyer Olkin
RBI	Reserve Bank of India
RMSEA	Root-Mean Square Error of Approximation
SPSS	Statistical Package for Social Sciences

Source: Author's Own Work

● ***Executive Summary***

This is one of the crucial parts of the preliminary section. The summary of the entire study which is presented in the report is presented in a succinct and coherent manner. The summary runs through the scope, significance, and research objectives of the study to the methodology engaged and results obtained in the study.

- The executive summary is divided into four to five sections in a condensed and concise manner.
- The first section discusses the introduction, background and scope of the study.
- The second section provides a brief of the extant literature and highlights the research gap identified in the study.
- The next section discusses the methodology deployed and the data collection techniques used in the study.
- The fourth section centres around the analysis of the data and focuses on the key findings of the study.
- The last section gives a concluding remark with the major recommendations and suggestions in the study.
- The executive summary is just one to two pages long as it provides a synoptic view of the full-fledged report of the study.
- As the word count limit is applicable in the case of executive summary therefore clear and concise language should be used.
- The use of jargon and technical language should be clearly selected as this element of the preliminary section plays a major role in engaging readers.

● ***Acknowledgment***

- Acknowledgement is a form of a token of appreciation for the respondents, experts, academicians, researchers, field experts, industry analysts and many more for providing relevant and critical inputs throughout the conduct of the study.



B) Background Section

The background section deals with the most crucial element of the entire report. It entails a detailed set of information which academically is the robust part of the report. This section includes the introductory part of the report and formally identifies and defines the problem in the study. It has the following segments:

- ***Introduction***

This is the foremost and significant part of a research report which begins with an introductory part of the study and clearly delineates and defines the problem identified for further investigation in the study. This section highlights the relevance of the study in the present context and links how the study contributes to society through the investigated topic. This section defines the key research questions posed in the study and corresponding research objectives to find the underlying solutions. In line with the proposed research objectives, propositions/hypotheses are framed and their significance is tested using suitable statistical analysis tools. The introductory segment should be written in clear and effective language using appropriate jargon and technical terms so that it engages the readers with the background of the study.

- ***Review of Literature***

This section provides the research background of the study in terms of preliminary conceptualisation. It provides a detailed account of the groundwork done previously by various other researchers, academicians and research organisations. Critical analysis of the past literature serves as a guiding path for the topic under investigation as it helps document the various variables and constructs previously worked upon and identify the unexplored areas for further study. In the case of an academic report, there is a dedicated segment focusing on the review of literature as all the previously studied documents, reports and research are duly reported and it clarifies how the present study bridges the research gap and adds to the extant literature in an advanced manner. The researcher arranges the literature in a sequenced manner wherein various reports and studies reporting similar findings are quoted using in-text citations. Furthermore, the studies pertaining to common themes are logically arranged and henceforth the unexplored area of the major problem is highlighted and a reference is established as to how the present study bridges the research gap. A review of the literature is generally followed by the scope and significance of the study which gives a detailed vision of the purpose of the study. The significance of the study briefs the main reason for choosing the topic of research and how the topic holds prominence in the present context.



C) Methodology Section

The methodology section provides the details of the main research design used throughout the study and the various instruments used for data collection. There are majorly two sources of data collection: Primary and Secondary, which are used as per the study's purpose. In business reports, not much emphasis is placed on the methodology segment rather a short description is provided concerning the research design and sample collection (if any) and appendices are attached at the end for detailed insights. However, for academic and technical reports, methodology forms a critical element of the study. Therefore, a detailed description of the research design and sampling frame is provided.

This section is a detailed one and provides the specification of the sampling technique used and various sources of information utilised for data collection for example, in the case of the primary study the data is collected using questionnaires/schedules and are primarily on-field studies whereas compared to this secondary data originates from published reports, previous studies from various organisations, research agencies, corporate surveys and many more. The variables to be studied are clearly defined in the methodology section and in cases where casual and effect relationships are to be studied the experimental models for the same are also enlisted in the methodology section. The researcher must carefully present the step-wise process of the methodology undertaken and present the facts in a simple and lucid language for easy understandability.

The sampling design should also be carefully presented and the execution details concerning the same should be carefully documented. The calculations concerning sample size, sampling frame and the notations used in the study should be presented in a concise and clear manner. The researcher must be careful of the fact that this section needs to be framed in a reader-friendly manner and not too much technical jargon should be used. In case there are too many technical equations or notations involved then the same may be attached as appendices in the end.

Data analysis tools used in the study should be clearly mentioned in the study. The tools used for analysis must justify the framed research objectives. The assumptions with respect to the tools used for analysis must be clearly met. The methodology section should be sequentially and logically arranged to avoid any ambiguity. The use of repetitive calculations, and explanations should be as much as possible be avoided.



D) Findings Section

This part of the report requires due attention and care as it provides the key findings/results of the investigated topic. It is one of the longest parts of the report which is why it has various sub-chapters or sub-sections. In the findings section, the report highlights the key results obtained either through primary or secondary data collection. Findings in case of qualitative study present the excerpts gathered from various subjects under study and in case the questionnaire method is utilised then the responses of the respondents are presented using infographic bars, charts, diagrams, etc. Using various software as a medium of analysis, output results are presented in the form of descriptive or multivariate statistics. In case open-ended responses were elicited from the respondents then the detailed section on the same is also presented in the study.

The findings section is classified in such a manner that corresponding to each research objective the main findings of the study whether having a positive impact or the negative one are presented in a reader-friendly manner. The results are interpreted in a compendious way so that the readers are easily able to comprehend the key findings and results of the study. The hypotheses are tested to check whether they have a significant impact or not and thereby based on the same they are either accepted or rejected.

For example, The title of the topic is “Analysing the impact of social media on small business entrepreneurs” and the researcher has utilised the questionnaire method to elicit responses from the respondents then the findings corresponding to the same will be presented in the findings section wherein firstly the demographic details of the respondents will be shared as age, gender, occupation, income slab, region, etc. then the descriptive analysis pertaining to variables on the scale either on a five-point or a seven-point sub-scale will be presented. Finally, if the causal and effect relationship is studied then the predictive analysis related to the same will be presented in the report. Furthermore, if the researcher has asked a few open-ended questions then a segment related to the same will also be given in the findings section. It is advisable to provide the key parameters highlighted in the study in form of charts, bars, graphs or figures.

E) Conclusion Section

This section provides a summarised view of the study in a more analytical manner. The researcher has to be skilful to present the main gist of the study in a meaningful way. This section incorporates the sub-sections as discussion, recommendations and suggestions, limitations of the study and the future scope of the study. Moreover, in case the findings draw similarities with the past studies the same has to be carefully presented and documented.



Sometimes the results drawn from the study are in line or sync with the past studies for example *“The results of the structural equation modelling indicates that the various social media platforms such as Facebook, Instagram, LinkedIn are effective in boosting the sales of khadi products (Kaur, 2021; Rajesh, 2006; Tseon et al., 2013).”*

whereas sometimes the results inferred from the main study are contradictory to the past studies therefore in such cases the researcher has to be diligent enough to provide a direction to the results obtained. For example *“The results from the confirmatory factor analysis shows that none of the factors was significant enough to push the sales of abnormally shaped food products which is in contrast to the findings of Joseph (1998) wherein the price turned to be a significant driver pushing the sales of abnormally shaped products. However, this has to be understood in a manner that the market area for both studies differs significantly in terms of taste, preferences, and attitude which is one of the reasons that people in India avoid the purchase of abnormally shaped food products.”*

Henceforth, the researcher has to be skilful to provide a direction to the study.

Recommendations proposed in the study should be realistic and logical and must have practical relevance for the segment they are specific to. There may be general or specific recommendations proposed in the study. The recommendations can also be supplemented with the practical and managerial implications of the study. Apart from the workable recommendations this section limelight the limitations of the study in terms of the problems encountered during the conduct of the study. The constraints in terms of time, resources, finance, and region can also be limiting factors in a number of studies. The limitations are clearly mentioned in the study so that the reader can keep in mind all the limiting factors while going through the analysis. Along with the limitations of the study, it is advisable to provide a brief of the future areas of research concerning the study so that those interested can extend the study either in longitudinal or regional dimensions.

F) Appendices

This is the final section of a report which includes all the addendum whether related to statistical calculations or the equations used in the study. These are the end notes provided in the report in terms of the appendix, bibliography and glossary used in the study.

- All sort of supportive material is presented in this section.
- In the case of primary information, a questionnaire or schedule is provided as an addendum in the appendices section.



- Raw data sheets and the final data sheet used for analysis can be provided in the appendices section.
- In the case of the secondary study, all the published reports, case studies used, log tables, graphs and original data sheets can be supplemented in the appendices section.

A bibliography section is also provided in this final section which provides a complete list of all the information sources. The bibliography is a complete list of all the journals, reports, studies, theses and research papers which the researchers used or cited in the study. Even the sources which have not been cited in the study but were used as a reference point should also be cited. The method of citation should be standardised all throughout the study. There are various methods of reference styling as provided by MLA, APA or Chicago. The latest and updated citation styles should be used in the study.

A glossary of terms for the technical terms or jargon used in the study should be provided at the end of the section to provide ease and convenience to the readers.

Table 5: Glossary of Terms

- **Avatar:** A customised image representation of a person to depict the social media or any other forum account of the person. For example, recently WhatsApp has started providing the Avatar creation icon on its application.
- **ChatBots:** A computer-trained robot which is capable of having an organic conversation with the user based on preset algorithm rules.
- **Cost Per Click Model:** The price paid by an advertiser for every click that is made by the user on the advertisement. Akin to the traditional advertising models, it is the per-click cost of the advertisement on digital platforms.
- **Engagement Rate:** It is a metric used to measure user engagement with a post or the content posted by the marketer.
- **Hashtags:** A short keyword or phrase followed by the symbol hash (#) to create a series of events, and links associated with the same keyword.
- **Key Performance Indicator (KPIs):** One of the crucial metric tools to analyse the key performing metrics towards the achievement of organisational goals. It helps in tracking the total number of visitors on a webpage, lead rate, engagement ratio and many more.
- **Organic Reach:** It refers to the number of users who watched or liked the content without being routed through a paid medium.



- **Search Engine Optimisation (SEO):** It is the technique to optimise the website with search engines. various long-tail or short-tail keywords can be used for search engine optimisation so that the website is able to appear on the first few SERP results.
- **URL Shortener:** Unique Reference Link shortener is an application through which the long URLs can be shortened to provide ease in sharing and accessing the link across different channels.

Source: Author's Own Work

9.6 GUIDELINES FOR EFFECTIVE REPORT WRITING

It is of utmost significance that each and every component of a research report is well documented. There are a number of points that the researcher should be mindful of while writing a report.

- An effective report should not be influenced by any biases, subjectivity or persuasion.
- It should be concise, succinct and unambiguous.
- The report should be indicative, simple and comprehensive in nature.
- The writing style should be clear and statements should be direct and precise instead of long phrases. The use of long paragraphs should as much as possible be minimum and rather the information should be in clear chunks for reader-friendly versions.
- The dataset should be free from any errors, therefore, a thorough check with respect to repetitions and omissions should be done.
- The language of a good research report should be structured and formal in character.
- While writing the study problem, the researcher has to be focussed and very explicit in terms of the problem defined.
- The background of the study should be logically presented to provide impetus to the investigated topic.
- As the report is a one-way communication therefore a clear and incisive mandate of the report should be written.
- The report writing should preferably be done in the past tense however certain exceptions in the case of recommendations and suggestions are always there.



- When using in-text citations, researchers should be cautious to avoid prefixes and the last names are to be provided as in-text citations instead of the use of first names.
- The use of too many abbreviations should be avoided and full terminologies should be used when mentioning an abbreviation for the first time in the report.
- Numerical notations/equations should be carefully written and the use of numerical digits in case of giving a reference to dates should be restricted.
- A thorough check of the omissions, lexical usage and grammatical errors should be done.
- The version of the software used for data analysis should be mentioned in the report.
- The use of personal pronouns should be avoided as it indicates the subjectivity of the researcher towards the investigated topic. Phrasing protocols should be carefully kept in mind while framing a research report.
- Irrespective of the diagrammatic representation of the facts, a rigorous explanation of the statistics should be provided in the study.
- A good report needs to be crisp and as much as possible technical jargon and terminologies should be avoided as readers may find it difficult to consult a thesaurus along with the main report. Though, these technical terms may be explained in the glossary section wherever required.
- Proper formatting and standardisation should be adopted throughout the study so that the report sounds presentable to the audience rather than appearing to be dull.

Apart from the above-mentioned basic elements, a researcher should always make sure that the research report is formulated in such a manner wherein due attention is paid towards its main components. For effective documentation, the report should always have

- a) Clear, focussed, organised and structured research objectives. Furthermore, keeping too many objectives and delivering nothing approach should be avoided. There should be a line of clarity and order.
- b) Simple and organised background of the study in terms of the problem defined, study introduction, scope and significance of the study.
- c) Clearly defined methodology with a focused orientation.
- d) Findings of the study should be presented in a logical and sequential manner. A clear representation of the findings of the study should be done. The findings should have



wide applicability. Standard protocols and procedures should be adopted throughout the study.

9.7 SUMMARY

Writing a good research report is an art wherein the adept skills of the researcher are a must. Writing an effective, concise and succinct report requires much deliberation and discussion. It is considered an essential component of research. A good report helps link the present study with the extant literature and adds value to the existing body of knowledge in a logical and rational manner. There are various types of research reports based on their usage, applicability and purpose. Every kind of report holds great significance in its defined areas. The structure of the report consists of six main sections: Preliminary, Background, Methodology, Findings, Conclusion and Appendices. Each section contributes to the main report in explicit terms. It is of utmost relevance that effective writing skills and standardised protocols should be adopted throughout the study.

9.8 KEY TERMS

- Research Report
- Business Report
- Technical Report
- Survey Report
- Executive Summary
- Study Background
- Appendices
- Bibliography
- Glossary
- Acknowledgement

9.9 IN-TEXT QUESTIONS

- a) Business Reports pay more attention towards
- i) Technical Jargon
 - ii) Glossary



- iii) Managerial Implications
- iv) Surveys
- b) What all are included in the preliminary section of a good report?
 - i) Acknowledgement
 - ii) List of Tables and Figures
 - iii) Abbreviations
 - iv) Table of Contents
 - v) All of the Above
- c) Collection of data using primary surveys forms part of a
 - i) Findings Section
 - ii) Methodology Section
 - iii) Appendices
 - iv) Preliminary Section
- d) Review of literature provides insight about
 - i) Past studies
 - ii) Sampling design
 - iii) Questionnaire Structure
 - iv) Limitation of study

9.10 SELF-ASSESSMENT QUESTIONS

- a) What is the difference between a survey report and a business report?
- b) What are the steps that are to be followed in the preparation for a good research report? Does the criterion differ for different types of reports?
- c) What is the relevance of writing a good research report?
- d) What are the guidelines that are to be kept in mind while writing an effective report? Illustrate with the help of examples.
- e) How do the findings section of a good research report be documented? what do you think are the key elements of a findings section?
- f) Examine a technical report and a thesis report and write down the key features/elements of these reports.



9.11 REFERENCES

- Chawla, Deepak & Sondhi, Neena (2011). *Research Methodology Concepts and Cases*. Vikas Publishing House Pvt. Ltd. ISBN: 978-81-259-5205-3.
- Creswell, J.W. (2022). *Educational Research: Planning, Conducting and Evaluating Quantitative and Qualitative Research*. Pearson India. ISBN: 978-93-325-4947-0.

9.12 SUGGESTED READINGS

- American Psychological Association (2020). *Publication Manual*. 7th ed. Washington, DC: American Psychological Association. Retrieved from <https://apastyle.apa.org/products/publication-manual-7th-edition-introduction.pdf>
- Kothari, C.R. (2004). *Research Methodology: Methods and Techniques*. 2nd Ed. New Delhi: New Age International Publications.



LESSON-10

MARKETING RESEARCH AND ITS APPLICATIONS

Mr. Animesh Minz

Animeshminz07@gmail.com

STRUCTURE

- 10.1 Learning Objectives
- 10.2 Introduction
- 10.3 Applications of Marketing Research
 - 10.3.1 Sales and Market Analysis
 - 10.3.2 Product Research
 - 10.3.3 Business Research
 - 10.3.4 Advertising Research
 - 10.3.5 Consumer Research
- 10.4 Attitudes Measurement and Scaling Techniques
 - 10.4.1 Attitude
 - 10.4.2 Attitude Measurement
 - 10.4.3 Scaling Techniques
- 10.5 Challenges in Conducting Marketing Research in India
- 10.6 Ethical Issues in Marketing Research
- 10.7 Summary
- 10.8 Glossary
- 10.9 Answers to In-text Questions
- 10.10 Self-Assessment Questions
- 10.11 References
- 10.12 Suggested Readings

10.1 LEARNING OBJECTIVES

The learning objectives of this unit is as follows:

- (i) to understand the significance of marketing research in the context of marketing decisions,
- (ii) to identify the applications of marketing research,



- (iii) to discuss the scaling tools and techniques used to measure attitude, and
- (iv) to develop a holistic understanding of the entire process of marketing research
- (v) to comprehend the challenges faced by researchers while conducting marketing research in India
- (vi) to understand the ethical issues involved in marketing research

10.2 INTRODUCTION

It is critical for a marketing professional to grasp the applications of marketing research. By far, you must be theoretically aware of the concept of marketing research. This chapter will provide you with a quantitative grasp of marketing research. As a marketing manager, you are in charge of making managerial decisions about your product, service, or consumer on a regular basis. Every decision has a risk factor attached to it; efficient marketing research helps decrease and, in some cases, eliminate the risk involved in such decisions. One method of marketing research assists you in gathering data and information that will assist you in reducing risk while making current judgements. But, another technique to marketing research will assist you in anticipating the risk involved in future decisions.

As a manager, you must not only make decisions but also keep in mind the ethical issues and ramifications of those decisions after they are executed. For example, if you are performing advertising research, you must be aware of the distinction between influencing and deceiving or manipulating. Before making any advertising decisions, the marketer is also expected to learn about his or her target audience's expectations and anticipate the audience's reaction to the commercial or the product in general. This is precisely what marketing research attempts to accomplish.

This subject will introduce you to the broader uses of marketing research in everyday life. Scaling and forecasting strategies for conducting market research will be presented in a much broader context.

10.3 APPLICATIONS OF MARKETING RESEARCH APPLICATIONS

Sales and market analysis, product research, advertising, consumer research or market research, consumer behaviour research, business and corporate research are all examples of marketing research applications. According to industry, the various applications of marketing research are discussed below.

**10.3.1 Sales and Market Analysis:**

- a) **Determination of market potential:** Marketing research is performed to determine a company's or organization's market potential. The total amount of a product or product group that could be sold to a market in a specific time period and under certain conditions is referred to as market potential. When introducing a new product, changing an existing product, or expanding an existing product into a new geographical market, market potential is important.
- b) **Market share determination:** In the case of an existing product, a corporation may be curious about the market share that their brand commands. Marketing research can assist in determining the same.
- c) **Forecasting sales:** Sales forecasting is one of the most common applications of marketing research. Based on available data, sales forecasting attempts to predict the level of sales at a certain moment in the future. Forecasting sales for existing and future products is possible. Units or dollars can be used to compute sales. Predictions are divided into two categories: short-term and long-term. The short-term forecast considers seasonal fluctuations, seasonal trends, and cycles. The long-term prediction is based on the growth trend of the industry to which the product belongs, as well as the industry's business cycle.
- d) **Market segmentation:** Marketing research also aids in market segmentation. A market is a collection of potential customers who share certain traits. A geographical area, gender (only men use aftershave lotion), age (toys for children under the age of 5, between the ages of 5-7, etc.), physical trait (poor vision, obesity), wealth, and lifestyle are all prevalent characteristics. Children dominate the toy business. However, within this broad category, the market can be divided into several smaller markets or segments, such as those for pre-school children, school-age children, instructional toys, mechanical toys, electrical toys, indoor games, and so on. The marketing manager must determine whether to target a broad toy market or a niche toy market.
- e) **Market Testing:** A controlled experiment attempting to anticipate the effects of various marketing methods on sales or profit. It refers to experimenting with something in a small market before developing it on a larger scale. You may have seen adverts for soaps or snack foods that proclaim 'only in Hyderabad' or 'only in Calcutta.' Very likely, the company offering these products is testing the product. The market test results provide research data that can be used to decide whether to expand



marketing to additional areas or discard the idea entirely. Test marketing also produces data that may be utilised to improve the product and marketing plan in order to raise the product's chances of success. Test marketing is used not only for new items, but also to study the influence of retail level promotional displays and promotional programmes such as coupons and discounts on sales.

- f) **Channel Distribution studies:** Market research can be done to determine the most successful and profitable distribution channels for various product kinds.
- g) **Determination of market characteristics:** Surveys of market characteristics can be done to assist a new entrant in planning his entry or an existing company in improving its strategy for increasing market share. It is possible to gather information on the number of brands competing in the market, the state of technology prevalent in the market, the geographical concentration and dispersal of clients, the nature of retail outlets selling the items, the number of such retail outlets, and so on.
- h) **Determination of competitor information:** Research can provide information on competing brands' marketing efforts as well as each brand's "unique selling proposition."

ACTIVITY 1

Imagine you are the marketing manager of a new firm that sells customized sweatshirts for college students and working professionals. You are expected to launch an advertising campaign promoting your products. How would you identify your target audience? What would be the basis of segmentation that you would consider before launching the campaign?

10.3.2 Product Research:

Product research is primarily adopted for the following purposes:

- Assessing new product concepts
- Acceptance testing for new products
- Assessing the need for product formulation changes
- Testing package design for aesthetic appeal, product protection, and the ability to withstand transportation and stocking ordeals.



- Product positioning testing. Should a new tea brand be positioned based on its fragrance and taste, colour and strength, or price, etc?

10.3.3 Business and Corporate Research:

Business research

Business research is how a business improves its profits and organizational performance by gathering and analysing relevant data regarding its operations. It's a continuous process that involves all aspects of an organization's business environment, like the markets the company is engaged in, its competitors, local and global economic trends, technological breakthroughs in the field, new business practices and any other elements that may affect the organization's operational efficiency. Top-level executives use the information gathered through business research to make informed decisions regarding the company's activities.

- Business research is done largely to:
- Analyze business trends in order to identify industries with significant development potential and those with a grim future.
- Research product pricing to assess the level of demand at various prices. Such studies demonstrate how price sensitive clients are and provide significant market information, as well as measuring the impact of price hikes or cuts on sales.
- Diversification studies: These provide information on profitable new business growth options that a company should consider diversifying into. Diversification can occur in wholly new and undiscovered regions, or it can occur in connected areas.
- Product-mix studies: If a company is considering diversification into related product categories, it may want to determine what product-mix combinations will optimise its existing resources while also providing synergy for growth. A cooking oil manufacturer want to do research on one or more of the following Butter, Vanaspati ghee, spices, dried foods, frozen meals, instant food mixes, custard powder, branded wheat flour, and rice were used to create a synergistic product mix.
- Plant and warehouse site research: Further research is needed to discover the ideal area for developing a new facility. A corporation should study aspects such as raw material and labour availability, accessibility to the market, telecommunication and transportation infrastructure, financial, taxation, and other incentives available in each site before making a selection. In the case of warehouse location, you would look into goods movement patterns to different cities, high and low sale potential areas, the



number of quality checks required en route to the final customer, the benefit of conducting these checks versus the cost of acquiring and maintaining a warehouse, and convenient rail/road connections.

10.3.4 Advertising Research:

Advertising Research

Advertising is part of your marketing plan that uses paid, non-personal messages. These messages are designed to reach large groups of customers and sell your products or services. Like all forms of marketing, your advertising will be most successful if it's based on high-quality research. This should be a regular part of your market research strategy.

Research in advertising is something which is inevitable and of extreme importance. Starting from communicating the right message, identifying the target audience, designing the message, choosing the right media option, choosing the appropriate media mix and lastly to even measure advertising effectiveness, research plays a fundamental part in every step of the process. Some of the practical applications of advertising research are as follows:

- i. Audience measurement for advertisements in a variety of media, including newspapers, magazines, journals, radio, television, outdoor billboards, kiosks, bus side panels, and so on. The purpose of this type of research is to assess the size of each media channel (for example, press) and, within that, each media vehicle's audience (India Today, Readers Digest, The Indian Express, etc.). Given the audience's size, you'd want to know its age, gender, financial status, and cultural makeup in order to personalise your advertising campaign.
- ii. Choose the least expensive media strategy: Each media channel has unique benefits and drawbacks, and each media vehicle has its own cost structure. By matching your product's qualities with the audience demographics of various media vehicles and the related expenses of advertising in them, research may be used to discover the optimal media vehicle for your product.
- iii. Advertising Copy examination: Testing the following elements is one way for analysing the efficiency of copy:
 - i. Headlines, photos, jingle, and story sequence
 - ii. Pre-testing entire advertising in rough or finished form
 - iii. Pre-testing the effect of repetition to simulate a campaign (all of the above can be tested under simulated conditions)



iv. Post-testing individual advertisements in their typical media after they have been released. The other approach for conducting research is to assess the copy or the entire advertisement/campaign for the following:

- Assessing for its attention value, interest value and arousal,
- Testing for communication clarity
- Testing for their effect on consumer attitudes
- Testing for their effect on purchase behaviour

Assessing the success of advertising: Once the commercials are issued, it is vital to track their impact in terms of attaining the stated aim. To what extent has advertising succeeded in creating brand awareness, building corporate image, educating customers on product usage, and so on? The effectiveness is always weighed against the expense.

10.3.5 Consumer Research:

Consumer research:

Consumer research is a subset of market research that identifies the preferences, driving forces, and purchasing tendencies of the target audience. Customer psychology is better understood and thorough profiles of purchase behaviour are created because of consumer research. It offers organised information on what clients need using research methodologies. With the use of this data, businesses may alter their goods and services to become more centred around the needs of the consumer and raise customer happiness. In turn, this will support an increase in business. An organisation that has a thorough understanding of how customers make decisions is more likely to design a product, charge a given price for it, set up distribution facilities, and advertise a product based on findings from consumer research in order to enhance consumer interest.

- Marketing research is most commonly used to investigate and study consumer behaviour. Consumer behaviour is a distinct domain that allows for a more in-depth study of buying behaviour of persons via the use of research tools and procedures. Real-world examples of consumer research applications include:
- Identify and profile product clients (men, women, children, working women, housewives, and retirees) based on socioeconomic background, age, religion, and occupation.
- Determine the clients' location.
- Determine their motivations for acquiring your brand of product.



- Determining their buying behaviour pattern by identifying sources of knowledge and impact, as well as the sequence of purchase decisions.
- Determine the amount of consumer post-purchase satisfaction.

Application	Uses
Product Research	• Product planning and development. • Designing the right product by providing all information about new ideas in the market, testing the new product, and evaluating the current product mix to determine what changes need to be made.
Demand Forecasting	• Estimating customer demands. • Predicting the right quantity of goods needed and then producing it accordingly using various forecasting techniques.
Sales Analysis	• Analysing their sales by examining sales reports. • Identifies which goods are selling well • Whether the sales force is working effectively • Which inventory should be stocked more, and • What the production capacity should be.
Advertising Research	• Determining the most effective and appropriate advertising medium. • Recall Measurement • Attitude Measurement • Brand Usage • Pre-Testing, Post Testing, Concurrent Testing • Copy Testing: Analysing various aspects of advertising such as themes, appeals, headlines, communication clarity, attention value, and so on.



Application	Uses
	Media Research
Positioning Research	- Developing a brand's or service's optimal market position. - Collecting and distributing all relevant information about potential customers to businesses, assisting in better positioning products among customers.
Market Segmentation	- Collecting and communicating all market information to businesses. - The medium through which businesses can acquire key attributes about their potential customers, allowing them to create various target market groups.
Pricing Research	- Developing appropriate pricing strategies for their products. - Gathering all information about competitors' pricing strategies, customer price expectations, and other factors influencing pricing decisions.
Distribution Research	- Identifying the best distribution channel for delivering products to customers in the shortest amount of time. - Analysing and identifying potential distribution channels, selects efficient market intermediaries, reduces distribution costs, and evaluates distribution channel performance.
Customer Satisfaction Research	- Interacting with customers and gathering information about their brand shopping experience. - Incorporating customer feedback and focusing on their perceptions of the company's products.

Table 1.1: Applications of Marketing Research and their uses in real-life (*Reference Source: www.commercemates.com*)



IN-TEXT QUESTIONS

1. Which of the following activities are not covered by the marketing research?
(a) Pricing policies (b) Promotional policies
(c) Distribution Policies (d) primitive policies
2. Product research covers:
(a) Ascertaining extent of price modification
(b) Studying the price policy and strategies or the competitor
(c) Research on causes and extent of customer dissatisfaction
(d) Measuring advertising effectiveness
3. How would you describe the consumer buying behaviour for the following products?
(a) Apple iPhone (b) Pepsi
(c) Bajaj Air-Cooler (d) Mango Ice cream

10.4 ATTITUDE MEASUREMENT AND SCALING TECHNIQUES

A number of management choices are made on an irregular basis in an organisation, such as the hiring or dismissing of employees, the manufacture or marketing of products, and so on. Some of these decisions may be based on data whose units of measurement can be statistically modified. This information usually refers to the numerical or quantitative characteristics of a certain sample unit. Some decisions, on the other hand, may be based solely on behavioural data or information that cannot be changed or controlled statistically. Such data is difficult to measure and cannot be submitted to a thorough statistical analysis. The principal application of such data is in marketing, as management wants to know how it's current and potential customers feel about their product or service. Understanding attitudes could lead to sensible and practical decisions. A few examples of managerial actions that rely on attitude measurement are market segmentation, advertising message selection, product positioning, and advertising campaign selection.

Marketing research should preferably focus on understanding the typical consumer's purchasing behaviour or just understanding the consumer. The goal of this technique is to build a positive image of the firm in the eyes of the consumer and other potential customers. Consumers must have a favourable opinion of a company in order for it to sell its goods and services. As a result, it is critical for the firm to comprehend and evaluate the consumer



feelings towards its goods and services. This is due to the fact that if there is an unfavourable image or a negative perception of the brand or organization, the organization will not be able to last for a very long time in the business.

The firm must not only ensure that customers like its goods, but also anticipate their future desires. As a result, the question of how to quantify consumer attitudes towards a company's products arises. Because attitude is a very qualitative and subjective concept, translating it to quantifiable terms is difficult, making measuring attitude a difficult task. In other words, attitude is suggested rather than explicitly stated. Knowing consumer behaviour and what they like and dislike about items is critical for predicting how they will react to them. This could be beneficial in determining how much of a product a company can sell in the future.

10.4.1 Attitude

The term "attitude" refers to a person's or a user's proclivity or mental state towards an item, an idea, or any of its characteristics. Furthermore, it implies that a person is mentally prepared to behave in a certain way, which influences how they conduct towards the person, group, organisation, or thing in question. The following factors influence an individual's overall attitude towards an object: a) his or her beliefs about the qualities that the object possesses; b) his or her preference for those qualities; and c) the relative weight that each quality has in the person's decision-making process.

Attitude measuring would necessitate an awareness of the following essential components of attitude:

- (i) **Cognitive Component** – It pertains to the person's thinking or mental component. For example, an appealing advertisement entices the audience to learn more about the product. It raises awareness, offers information about the product/service or brand, and develops interest in the product. These are the characteristics that influence a person's cognition or mental state.
- (ii) **Affective Component** – It relates to the person's emotional side. Using the same example as before, an appealing commercial creates desire, as well as strong conviction and liking for the product/service or brand. These are the characteristics that influence a person's emotional component or feelings.
- (iii) **Behavioral Component** – is assigned to the action component or that affects the behaviour of the person. Continuing with the same example, an appealing advertisement that leads to the audience purchasing the goods or using the service is



said to affect the behavioural component of the individual because it requires some form of action from the target audience.

Since it is believed that there is a connection between attitude and behavior, the study and measurement of attitudes are crucial. But according to the findings, this relationship holds true at the abstract level than at the interpersonal basis. There may be other elements besides attitude that have a stronger influence on behavior. Attitude may merely be one of the factors determining behavior. For instance, despite having a positive attitude about a product, a person may decide against purchasing it due to financial constraints. The attitude-behavior link relates to evaluating cognitive and affective components and being able to anticipate future behavior for the purpose of marketing decisions.

10.4.2 Attitude Measurement

The term "measurement" refers to the process of gathering data that may be analyzed. The technique of determining a person's attitude toward a certain thing is known as attitude measurement. When measuring attitudes or any other parameter, it is important to distinguish between the following:

- What needs to be measured exactly?
- Who is being measured, exactly?
- The acceptable level of measurement accuracy
- The allowable cost
- The options for measurement/data gathering techniques.

While measuring attitudes, the researcher is primarily interested with gauging the respondent's "state of mind". It may include elements such as awareness, attitudes, and decision-making methods. The fact that these metrics are difficult to verify is an intriguing characteristic. There is no way of knowing if a respondent's answer to a question about how much they appreciate a new product—say, an energy drink—represents the "truth" or not. The researcher cannot actually see mental states such as preferences, likes and dislikes, and so on. Such things are only theoretically possible.

As previously stated, attitudes are impacted by qualities and ideas. The initial step before initiating an attitude-measurement activity is to select the suitable attributes of the item under study. For example, the pricing, storability, scent, packaging, durability, size, and other characteristics of a product such as "Lipstick" or any skin care product. The features or factors for a public transportation system can include the cost of travel, convenience, frequency, operating hours, and safety. There is no doubt that it is impossible to measure



every aspect of the process or object under investigation. The researcher should only select the most important ones. It is recommended that only characteristics related to respondents' actions be evaluated. Such qualities can be discovered through exploratory study. It is possible to apply both implicit and explicit procedures, such as projective techniques and depth interviews. Depth interviews are the most commonly used method. They do not use any form of coordinated data collection system. The investigator works hard to find the most noticeable features of the thing under investigation, while responders are encouraged to talk about it. This approach, however, necessitates the use of skilled research workers. Furthermore, research is a time-consuming and costly procedure, and the results are prone to bias. The projective techniques try to elicit information from the responder in an indirect manner.

The third major issue in attitude measurement is the alternatives for data collection and measurement methodologies. Data gathering procedures are classified into two types: (a) questionnaire methods and (b) observational approaches. Generally, attitudes are assessed via questionnaires. The following strategies are used to assess attitudes:

1. Self-Reported Personal inventories
2. Using psychological tools and devices like the pupillator that measures the pupillary response or psychogalvano meter that measures the galvanic skin response of the respondent
3. Projective methods, such as a perception test using themes.

Respondents are given a set of positive and negative statements about the research topics and asked whether they agree or disagree with each of them in the self-report inventories, commonly known as attitude scales. The self-report methodology is the most often used tool for measuring attitudes. They differ, however, in terms of size design and application. However, the self-report measurements have the following flaws:

- The results are only as accurate as what the subjects are willing to communicate about their attitudes,
- The validity of the expressed attitudes seems to be arguable

In this circumstance, the costs involved and the necessary precision in measurement are the most important concerns for a managerial researcher. These measures, as previously stated, are never totally error-free. Furthermore, in measuring, cost and precision are often mutually exclusive traits. A good understanding of a research tool can substantially benefit in the



accurate interpretation of data. In the following sections of this unit, we will discuss a number of these scales and their characteristics.

10.4.3 Scaling Techniques

Scales are highly structured and sophisticated methods used to measure attitudes. The fundamental principle behind attitude scaling is that attitudes influence how people feel and act toward a given good, service, or idea. Understanding the ideas and fundamental attitudes that underlie human behaviour is required in order to change it in a desired direction.

Scaling methods and scales have a wide range of options. We must choose the most appropriate technique for the research being done. Its assertions must be designed in such a way that the responses can be swiftly transformed into numerical values. The three most commonly used scales, which you should be familiar with from earlier chapters, are the Nominal scale, the Ordinal scale, the Interval scale, and the Ratio scale. As you will see in a later section, the type of scale used to collect the data influences which statistical approach should be used to analyse the data.

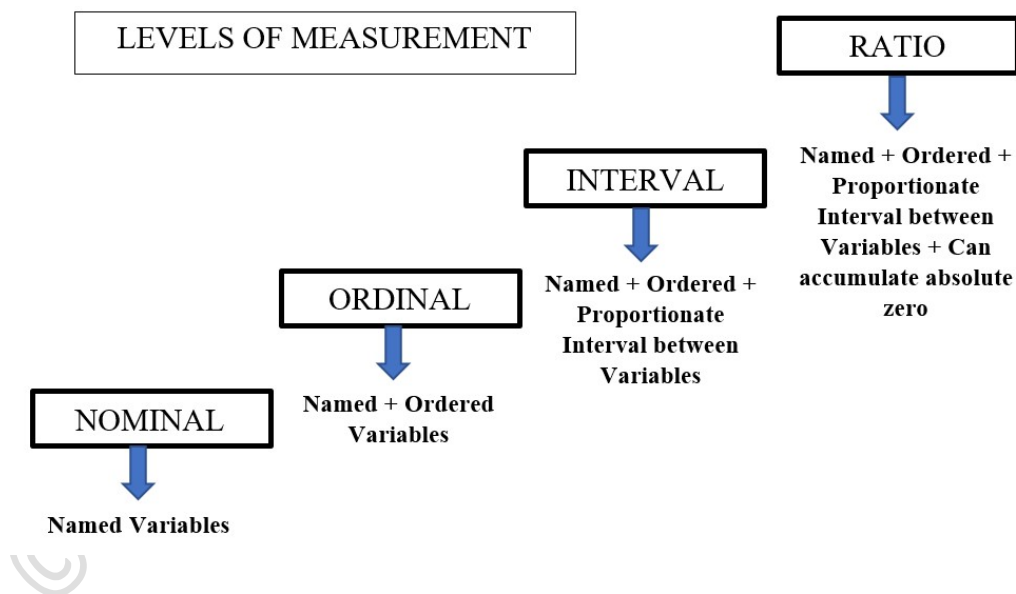


Fig 10.1: Different types of Scales of Measurement

The following table briefly summarizes the different types of scales and examples of its applications:


Table 1.2: Definitions and Examples of different scales of measurement

Scale	Characteristics	Example																												
Nominal Scale	➤ Classified into two or more categories. ➤ Data should fall into either of the categories. ➤ Qualitative Data. The numbers are used here to identify the objects.	What is your gender? M- Male F- Female																												
Ordinal Scale	➤ Represents the “order.” ➤ Qualitative data or categorical data that can be grouped, named and also ranked. ➤ Relative ranking of the variables without establishing the degree of variation between them. ➤ Researcher can effectively analyses the degree of agreement concerning the identified order of variables.	Customer Feedback Ratings The customer service was great at the restaurant Totally agree Agree Neutral Disagree Totally disagree																												
Ratio Scale	➤ Quantitative. ➤ It allows researchers to compare the differences or intervals. ➤ Has absolute zero ➤ The variables can be orderly added, subtracted, multiplied, divided. Mean, median, and mode can be calculated using the ratio scale.	What is your weight in Kgs? Less than 55 kgs 55 – 75 kgs 76 – 85 kgs 86 – 95 kgs More than 95 kgs																												
Interval Scale	➤ Quantitative measurement scale in which the difference between the two variables is quantifiable ➤ Variables are measured in an exact manner, not as in relatively ➤ The presence of zero is arbitrary. ➤ It allows calculating the mean and median of the variables ➤ Preferred scale in Statistics as it helps to assign any numerical values to arbitrary assessment such as feelings, calendar types, etc.	1. Likert Scale Customer satisfaction is key to organizational growth. 1- Completely agree 2- Somewhat agree 3- Neutral 4- Somewhat disagree 5- Completely disagree 2. Net Promotor Score (NPS) DETRACTORS PASSIVES PROMOTERS 1 2 3 4 5 6 7 8 9 10 3. Bipolar Matrix Table What do you think of the dress? <table><tr><th></th><th>Very Much</th><th>Somewhat</th><th>Neither</th><th>Somewhat</th><th>Very Much</th><th></th></tr><tr><td>Itchy</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>Super Soft</td></tr><tr><td>Dull</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>Vibrant</td></tr><tr><td>Casual</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>☐</td><td>Formal</td></tr></table>		Very Much	Somewhat	Neither	Somewhat	Very Much		Itchy	☐	☐	☐	☐	☐	Super Soft	Dull	☐	☐	☐	☐	☐	Vibrant	Casual	☐	☐	☐	☐	☐	Formal
	Very Much	Somewhat	Neither	Somewhat	Very Much																									
Itchy	☐	☐	☐	☐	☐	Super Soft																								
Dull	☐	☐	☐	☐	☐	Vibrant																								
Casual	☐	☐	☐	☐	☐	Formal																								



CASE STUDY

Consumer Research at Procter & Gamble

The Procter & Gamble Company (P&G), the largest manufacturer of fast-moving consumer goods (FMCG), was well known for its consumer research. Every year, the organization engaged with more than five million customers in about 100 different countries. It spent more than US\$350 million annually on customer knowledge and conducted more than 15,000 research projects annually. P&G established its market research (MR) department in 1924, making it one of the first corporations in the world to do formal research on consumer needs. The business created numerous ground-breaking MR techniques during the next nine decades. Its researchers had the know-how to ask customers for the data they required. P&G evaluated the feedback it received and came to the appropriate conclusions using cutting-edge technologies.

Marketing Research Tools Used by P&G

Market researchers were hired by P&G for each of its product lines. Additionally, the business maintained two independent internal research units with the responsibility of conducting market research and advertising research. Functional experts from each group, including MR managers, support experts (survey designers, statisticians, and behavioural scientists), and field representatives who coordinated consumer interviews, were present.

Tools for Qualitative Research

P&G used qualitative research to generate fresh concepts for the growth of markets and products. Customers and researchers engaged in lengthy discussions about many aspects of the company's products. As a result, there was a lot of direct communication between the researchers and the customers. Focus groups, in-home visits, in-context visits, and in-store interviews were some of the primary qualitative research methods used by P&G.

Tools for Quantitative Research

P&G utilises a lot of quantitative research instruments. These resources helped P&G come up with fresh product concepts. Although quantitative research typically used bigger sample numbers, there was little direct contact between the researchers and the respondents. P&G used blind tests, concept-aided usage tests, concept and use tests, behavioural patterns, and quality monitoring as its main quantitative research instruments.

**Conducting Marketing Research Online**

P&G conducted online idea tests to get user feedback. This feedback helped them introduce updated versions of current products as well as new products. 15% of its US market studies done in 1999 used the Internet. In 2000, it reached 40%, and in 2001, it reached 50%.

The Advantages

Online MR surveys from P&G saved the company a significant amount of time and money. A consumer study that was conducted using traditional methods cost the business close to \$50,000 and took at least three to four weeks. For US\$10,000, a ten-day online consumer survey was possible to be conducted. Similar to this, a traditional product or concept test cost close to US\$25,000 and could possibly take two months. For US\$2,500, a comparable test could be conducted online in one week.

Challenges

While P&G's market researchers were generally praised for their work that made P&G one of the top businesses in the CPG sector, they also drew criticism in 2018 for failing to alert the company's top management about changes in the US consumer market, which is its largest market.

Source: <https://www.icmrindia.org/casestudies/catalogue/Marketing/MKTG386.htm>

IN-TEXT QUESTIONS

1. Attitudes can be measured using _____
2. What are the two types of data collection procedures?
3. What are the three components of attitude measurement?

10.5 CHALLENGES IN CONDUCTING MARKETING RESEARCH IN INDIA

One of the primary issues with performing marketing research in India, particularly by smaller organisations, is that research is viewed as a large expense rather than an investment. Research is a lengthy process that necessitates a significant investment of money, time, effort,



and energy. Furthermore, it necessitates a significant number of employees with a strong eye for detail as well as the necessary skills and capacity to handle and analyse data effectively. Organizations frequently regard this work as tedious and proceed without devoting sufficient time and effort to investigation.

Furthermore, the fundamental problem for anyone attempting to conduct research in India is a lack of secondary data or, in certain circumstances, a poor interpretation or understanding of secondary information. The census, for example, takes years to collect and make public, yet the census itself is a massive source of consumer data. Numerous journal entries and handbook entries are normally two to three years old. Even research publications released by academics refer to a specific period of data that must be challenged on a regular basis with new data. As a result, the utility of any given piece of data is typically outmoded, and in most cases considerably reduced.

After discussing the challenges with secondary data, there are some important aspects to consider when working with primary data. In most circumstances, the researcher has little access to data. The population or sample is either widely distributed or poorly identifiable, there are communication challenges, and there are issues when the infrastructure does not permit access to the sample unit, such as in rural and distant locations. If the sampling units are industrial units in the unorganised sector, there is no map to identify them. Even today, given the digital transformations we face on a daily basis, data collection from the unorganised sector remains a time-consuming effort. Respondents are also prone to being hesitant, biased, and/or judgmental while answering questions. Respondents frequently view interviewers with suspicion and may even refuse to supply any information. Respondents' inability to comply when questioned stems from their utter ignorance of the nature of marketing research and its applications. These are only a few of the issues that contribute to the difficulty of gathering primary data.

The bulk of market research firms are headquartered in urban areas and have an urban bias, which precludes them from efficiently reaching out to or having a presence in rural areas. There, the majority of market research for commodities used by city people is conducted. Marketing research is not as advanced in developing countries as it is in developed countries. Marketing research approaches in India are still relatively modest, and there is a lot of room to incorporate more advanced methodologies.



10.6 ETHICAL ISSUES IN MARKETING RESEARCH

Some areas of marketing research have serious ethical considerations. The majority of marketing research is conducted for commercial clients. On occasion, the business motivation may induce researchers or clients to compromise the impartiality or professionalism of the marketing research process. If the researcher fails to follow approved marketing research techniques, or if the client misrepresents the findings in the company's advertising, ethical norms are broken. Ethical concerns can arise at any point during the marketing research process.

An Overview of Ethical Issues in the Marketing and Research Process	
Problem Definition	➤ Using surveys as a guide for selling or fundraising ➤ Personal agendas of the researcher or client ➤ Conducting unnecessary research
Developing an Approach	➤ Using findings and models developed for specific clients or projects for other projects
Research Design	➤ Formulating a research design more suited to the researcher's rather than the client's needs ➤ Using secondary data that are not applicable or have been gathered through questionable means ➤ Disguising the purpose of the research ➤ Not maintaining anonymity of respondents ➤ Disrespecting privacy of respondents ➤ Misleading respondents ➤ Disguising observation of respondents ➤ Embarrassing or putting stress on respondents ➤ Using measurement scales of questionable reliability and validity ➤ Designing overly long questionnaires, overly sensitive questions, piggybacking ➤ Using inappropriate sampling procedures and sample size
Data Preparation and Analysis	➤ Identifying and discarding unsatisfactory respondents ➤ Using statistical techniques when the underlying assumptions are violated ➤ Interpreting the results and making incorrect conclusions and recommendations
Report Preparation and Presentation	➤ Incomplete reporting ➤ Biased reporting ➤ Inaccurate reporting

Table 10.2: An Overview of Ethical Issues in Marketing Research



10.7 SUMMARY

Marketing research is a significant instrument for facilitating fundamental organisational managerial decisions. Every decision has a risk factor attached to it; efficient marketing research helps decrease and, in some cases, eliminate the risk involved in such decisions. Sales and market analysis, product research, advertising research, consumer research or market research, consumer behaviour research, business and corporate research are all examples of marketing research applications. Some administrative decisions that may rely solely on behavioural data or information that cannot be modified or controlled statistically are also examples. Assessing such data is difficult and cannot be subjected to a complete statistical investigation, which is why knowing attitudes could lead to reasonable and practical actions. Because it is considered that there is a link between attitude and conduct, attitudes must be studied and measured. Normally, attitudes are measured via surveys; however, self-reported attitudes are also used. Personal inventories, which employ psychological tools and devices, as well as projective procedures such as a perception test based on themes, can be utilised in conjunction with the appropriate scale techniques. It is the marketing researcher's obligation to keep ethical considerations and challenges in mind while conducting research.

10.8 GLOSSARY

Marketing Research: "The American Marketing Association defines marketing research as "the systematic gathering, recording and analysing of data about problems related to the marketing of goods and services". Crisp has defined marketing research as "...the systematic, objective and exhaustive search for and study of the facts relevant to any problem in the field of marketing".

Attitude: The word "attitude" describes a person's or a user's inclination or mental state toward a thing, an idea, or some of its qualities.

Scale: Scales are highly structured and sophisticated methods used to measure attitudes.

Primary data: Data which is collected originally for the current investigation.

Secondary data: Data which has already been collected by an agency or individual and is available in a published or unpublished form.

Questionnaire: An organised and written format which contains all the relevant questions for gathering data from the sample.



Respondents: An individual in his personal capacity or representing an institution who fulfils all the requirements of a sampling unit and is used for collecting data: He is known as respondent because he responds to the questionnaire.

10.9 ANSWERS TO IN-TEXT QUESTIONS (MCQ)

1. Primitive policies
2. Research on causes and extent of customer dissatisfaction
3. Self-Reported Personal Inventories, Psychological tools and projective methods

10.10 SELF-ASSESSMENT QUESTIONS

1. Discuss the different applications of marketing research from the industry perspective. Elaborate your answer with examples.
2. What are the four different levels of measurement?
3. Indicate giving reasons the scale of measurements for the following variables: (a) Geographical area (b) Costs (c) Index numbers (d) Preferences
4. What are the ethical issues a researcher should keep in mind while conducting research in India?

10.11 REFERENCES

Malhotra, N., & Dash, S. (2015). *Marketing Research: An Applied Orientation* (6th ed.). New Delhi: Pearson.

Parasuraman, A., Grewal, D., & Krishnan, R. (2006). *Marketing research*. Cengage Learning.

10.12 SUGGESTED READINGS

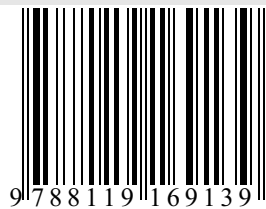
Malhotra, N., & Dash, S. (2015). *Marketing Research: An Applied Orientation* (6th ed.). New Delhi: Pearson.

Burns, A. C., Veeck, A. F. & Bush, R. F. (2017). *Marketing Research* (8th ed.). New Delhi: Pearson.



- Churchill, G., Iacobucci, D., & Israel, D. (2010). Marketing Research: A South Asian Perspective. Delhi: Cengage.
- Zikmund, W. G., & Babin, B. J. (2016). Exploring Marketing Research (11th ed.), USA: Cengage South-Western.
- Aaker, A. D. (2014). Marketing Research (10th ed.). New Delhi: Wiley.
- Cooper, D. R., & Schindler, P. S. (2005). Marketing Research. Tata McGraw Hill Education.
- Bradley, N. (2013). Marketing Research: Tools and Techniques (3rd ed.). New Delhi: Oxford University Press.
- Tull, D. S., & Hawkins, D. I. (1993). Marketing Research: Measurement and Method (6th ed.). New York: Collier MacMillan.
- Boyd, H., Westfall, R., & Stasch, S. (1989). Marketing Research: Text and Cases, Homewood, Ill. : R. D. Irwin

978-81-19169-13-9



Department of Distance and Continuing Education
University of Delhi