Rural Management

Management Decision Making Tools

First Edition

Rural Management - Management Decision Making Tools

सत्यमेव जयते

MoE | Government of India
Ministry of Education

First Edition: 2020

ISBN: 978-93-89431-09-4

Price: ₹ 750/-

5-10-174, Shakkar Bhavan, Ground Floor, Fateh Maidan Road, Hyderabad - 500 004

Telangana State. Tel: 040-23422112, 23212120, Fax: 040-23212114

E-mail : editor@mgncre.org Website : www.mgncre.org

Published by: Mahatma Gandhi National Council of Rural Education (MGNCRE), Hyderabad

About the Book

Research is a scientific process and this definition holds true irrespective of disciplines. Research provides new domain of knowledge, new theories and laws, solution to a problem and open new area of thought. Further, it has the potential to question and analyses the existing theories in accordance to the present circumstances. Hence, it is important for educational/professional academicians to be acquainted with the process of research i.e. research methodology. It is worth mentioning that research is a systematic process and the knowledge of research methodology helps to understand what to do in specific situation.

Research Methodology is a way to systematically solve the research problem. A researcher must know the rationale for using specific methods or techniques and establish their relevance. Methodology may be understood as a science of studying how research is done scientifically. A research work must act as a catalyst for further researches to be undertaken in the said field. In the recent past, many new dimensions of research have emerged catering to the vast changes taking place in the present-day social situation. The quality of research depends on the understanding of the research problem, the research methodology and the appropriate tools used for the analysis. Research Methodology is the central and most crucial component, element of any research endeavour in social sciences. Any researcher has to be well-equipped and must have command over this tool. The course Material on Research Methodology is designed to impart knowledge for scientific research to integrate qualitative and quantitative approaches in research and orient the researchers to techniques for precise, rational and scientific research. It is particularly meant for those who are engaged in survey research and have some basic knowledge of data processing, analysis and interpretation.

The present text is a brief endeavor to explain the concepts of research and the various methods, tools and techniques for carrying out research in a rural scenario. The main focus of the book lies in the fact that it tries to explain the otherwise complicated terms such as research, statistics, probability and ethics in the most lucid manner with a lot of case studies to facilitate the user in understanding the same with quite an ease. The book is a designed as a textbook to cater to the requirements of the first course on Management Decision Systems to the undergraduate students of Bachelors of Business Administration course. The book is organized into 5 Units that provide a comprehensive coverage of all the relevant topics using simple and easy to understand language.

"Learning by Doing" is one of the best ways to understand a concept and get to know it. Keeping the same in mind, the book provides a number of Case studies and to-do-activities to supplement the text and enable the reader to explore on his own as well. References at the end of chapters and self-learning questions will surely kindle the minds of the readers.

I thank the contributors: Dr S Prasanna Devi, Professor and Head of the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Vadapalani Campus, Chennai and Dr Meenakshi S Arya, Professor of Computer Science at SRM Institute of Science and Technology, Vadapalani Campus, Chennai to this book for their outstanding insights. Also, I would like to thank MGNCRE Team members for extending their extreme support in completing this text book.

Dr W G Prasanna Kumar
Chairman MGNCRE

Contents

List of Tables

List of Figures

About the Book

Research is a scientific process and this definition holds true irrespective of disciplines. Research provides new domain of knowledge, new theories and laws, solution to a problem and open new area of thought. Further, it has the potential to question and analyses the existing theories in accordance to the present circumstances. Hence, it is important for educational/professional academicians to be acquainted with the process of research i.e. research methodology. It is worth mentioning that research is a systematic process and the knowledge of research methodology helps to understand what to do in specific situation.

Research Methodology is a way to systematically solve the research problem. A researcher must know the rationale for using specific methods or techniques and establish their relevance. Methodology may be understood as a science of studying how research is done scientifically. A research work must act as a catalyst for further researches to be undertaken in the said field. In the recent past, many new dimensions of research have emerged catering to the vast changes taking place in the present-day social situation. The quality of research depends on the understanding of the research problem, the research methodology and the appropriate tools used for the analysis. Research Methodology is the central and most crucial component, element of any research endeavour in social sciences. Any researcher has to be well-equipped and must have command over this tool. The course Material on Research Methodology is designed to impart knowledge for scientific research to integrate qualitative and quantitative approaches in research and orient the researchers to techniques for precise, rational and scientific research. It is particularly meant for those who are engaged in survey research and have some basic knowledge of data processing, analysis and interpretation.

The present text is a brief endeavor to explain the concepts of research and the various methods, tools and techniques for carrying out research in a rural scenario. The main focus of the book lies in the fact that it tries to explain the otherwise complicated terms such as research, statistics, probability and ethics in the most lucid manner with a lot of case studies to facilitate the user in understanding the same with quite an ease. The book is a designed as a textbook to cater to the requirements of the first course on Management Decision Systems to the undergraduate students of Bachelors of Business Administration course. The book is organized into 5 Units that provide a comprehensive coverage of all the relevant topics using simple and easy to understand language.

"Learning by Doing" is one of the best ways to understand a concept and get to know it. Keeping the same in mind, the book provides a number of Case studies and to-do-activities to supplement the text and enable the reader to explore oneself. References at the end of chapters and self-learning questions will surely kindle the minds of the readers.

I thank the contributors: Dr S Prasanna Devi, Professor and Head of the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Vadapalani Campus, Chennai and Dr Meenakshi S Arya, Professor of Computer Science at SRM Institute of Science and Technology, Vadapalani Campus, Chennai to this book for their outstanding insights. Also, I would like to thank MGNCRE Team members for extending their extreme support in completing this text book.

Dr W G Prasanna Kumar
Chairman MGNCRE

Chapter 1 Introduction to Research Methods

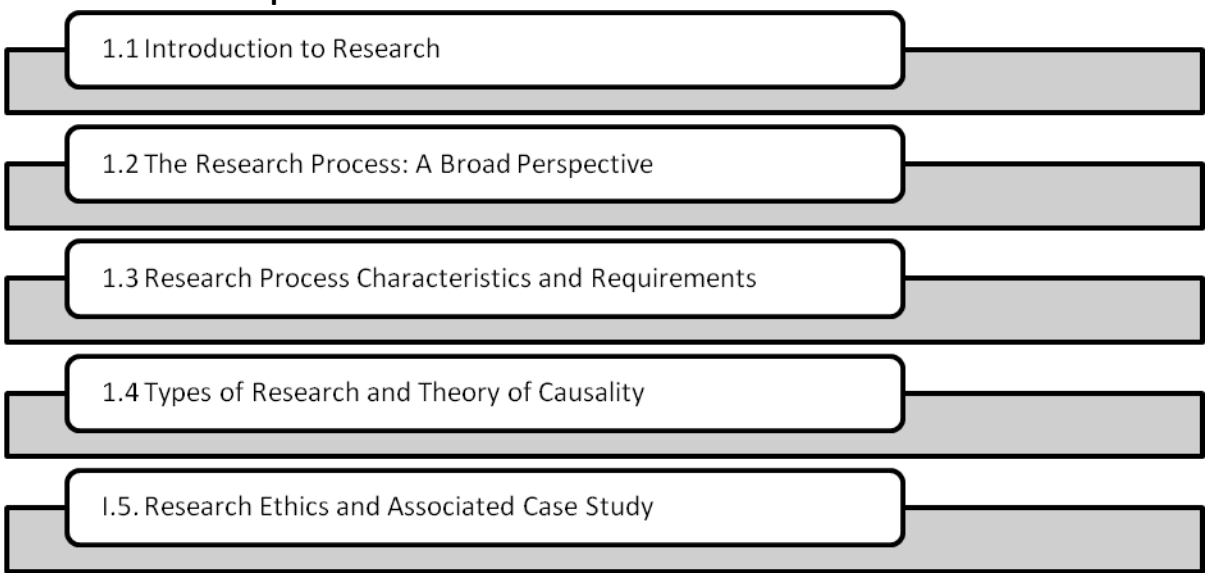
Introduction

Research is a word made up of two basic English words “Re” meaning again and “search” meaning to look around. This literal meaning sums up to a great extent what Research intends to do. It either searches for the explanation of existing phenomenon or tries to search for something which has not been searched for till now. In both the situations, it looks for the unknown, unexplored and unexploited.

Objectives

- To familiarise basic terminology related to research
- To familiarise linearity of research process
- To examine difference between various research approaches
- To familiarise ethics related to research

Structure of the Chapter



1.1 Introduction to Research

According to The Merriam Webster online Dictionary research refers to “A studious inquiry or examination, especially; investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or law in the light of new facts or practical application of such new or revised theories or law.”

This definition clearly emphasizes that research basically deals with a methodical, conventional, rigorous and detailed processes in order to gain solutions to problems or to discover and bring to light new facts and relationships. This can be achieved only through a lot of curiosity, systematic collection of facts and figures and their analysis for advancement of knowledge in any subject. Research attempts to find answers to intellectual and practical questions through application of systematic methods.

Research may also be considered as a movement, a movement from the known to the unknown. It is actually a voyage of discovery. The basic instinct of curiosity and inquisitiveness which is inherent to the human nature is what prompts us to probe and attain full and fuller understanding of the unknown whenever confronted with the same. All knowledge is an outcome of the thirst of humans for exploring

the unexplored; the formal tools, methods and techniques that are employed or deployed to quench this thirst is what essentially qualifies as can be termed as research.

Scientists at the laboratory working with scientific apparatus, or a genius like Stephen Hawking doing some complex mathematical calculation on some complex subject, or a data scientist glued to his screen trying to figure out the insights into yottabytes of data being generated every second are typical representative images of different aspects of research. To summarize and give a very simplistic view, research is simply the process of finding solutions to a problem after a thorough study and analysis of the situational factors. Let us try and look at research from different perspectives.

Research in Our Everyday Lives

We all encounter research in our daily lives. All of us in our daily lives explore, investigate, invent, solve problems at work, try out new recipes in the kitchen, find the best way to prune a plant, or simply play with the kids. Not restricting to this, results of research on many topics are presented to us in the form of newspaper articles, books, reports, and television programmes. For example, many of us who are actively involved in share market, stock market details are presented with detailed research about the stock prices, the trends and other information by television news reports. A few points, for example, diet and wellbeing, are exceptionally mainstream with magazines just as TV programs. In this manner, through different media we have gotten familiar with seeing, perusing and catching wind of research and in spite of the fact that we may not know about it, we are accustomed to making our own decisions about research discoveries. If we take the example of stock market, many of us take into consideration information that is been presented on '' and we make our choice of buying, selling or holding onto a stock depending on the market trends presented and our investment target. We evaluate the data with which we are introduced, structure an assessment with regards to the legitimacy and importance of the exploration, and arrive at our very own decisions dependent on contemplations, for example, what we understand is the question or problem to be answered

- how the findings are presented
- why the research was conducted
- what we know about who conducted the research
- what other people think about the information
- how the research findings relate to us

We may decide to overlook the data, we may choose we don't comprehend what the discoveries mean, we may dissent, or we may concur with the discoveries and either modify our purchasing and selling propensities in a way that the expenses of modification are excessively high comparative with the benefits. For some, because of their life circumstances, such as lack of capital or liquidity, the information may be of little immediate value. As such, you may not know about it, however you as of now make them comprehend of the exploration procedure, the job of research, look into ideas, and research assessment.

1.2 The Research Process: A Broad Perspective

The Linearity of Research Process

Research can be viewed as a progression of connected exercises moving from a starting to an end. These activities are interlinked to each other in a waterfall model structure wherein the outcome of previous step acts as an input to the next stage which further processes it through methodologies, tools, techniques to get the desired outcome. Research is a continuous and cyclic process which keeps iterating till the time findings made do not point to a conclusive end.

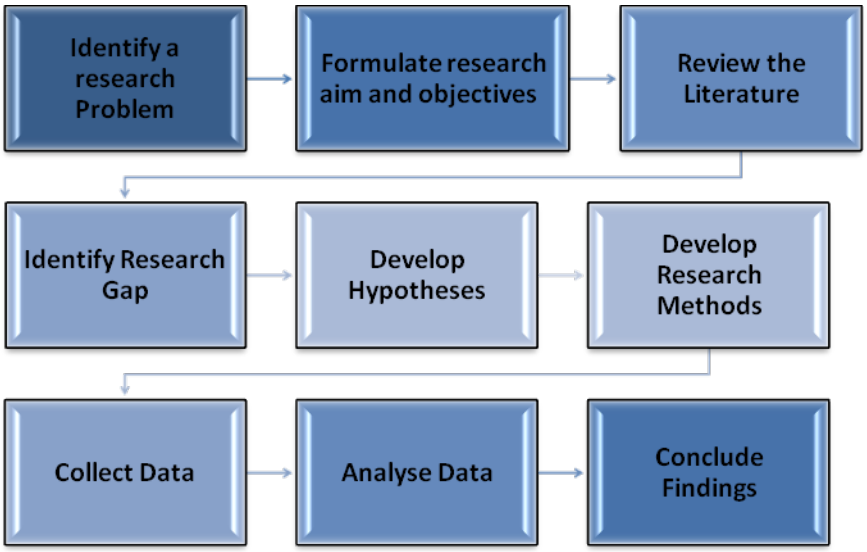


Fig 1.1: Linear Flow Depicting a Typical Research Process

Identify a Research Problem: Research generally starts with the recognizable proof of an issue. A research problem is a statement referring to a domain of concern, a specific state to be upgraded, a difficulty to be removed, or an intriguing question that occurs in scholarly literature, in theory, or in practice that calls for insightful understanding and thoughtful investigation. In some social science areas, the identification issue is commonly presented as a query. This step is most crucial and critical part of the entire research process as an incorrect or infeasible problem might lead to non- conclusive ends.

Formulate Research Aim and Objectives: This step performed in a suitable manner is one of the most important aspects of the entire process. This is because the research aim and the objectives determine the length(scope), breadth(depth) and the width (overall direction) of the research. These objectives become the fulcrum around which the entire research is centered.
Accomplishment of research point gives answer to the examination question. Research destinations isolate examine point into a few sections and address each part independently. Additionally, look into point indicates WHAT should be considered and explore goals contain various advances that address HOW research point will be accomplished.

Steps Involved in Writing Aims and Objectives

- State the overall intent of the identified problem.
- Emphasize on the motivation behind the research. The aim should focus solely on what is to be achieved rather than how it is to be achieved. The aim provides direction to the rest of the research.
- Elucidate all the key tasks that will help to achieve the complete intent.
- Enumerate the main tasks essential to achieve the research purpose. These tasks are the objectives.
- Objectives should be simple, actionable and measurable against a number of factors.
- It is very important not to include too less or too many objectives (avoid being over or under ambitious).
- Keep objectives SMART: Specific, Measurable, Achievable, Real and Time-bound

- Break down the main objectives to small, specific objectives: For example, if you plan to do research on the uptake of personal loans amongst subsistence farmers, you need to break it into tasks stating the financial institutions to visit and the questions to ask. Ensure that you are clear and brief.

Example
Title: A study of the student usage of e-books at Anna University.
Aim: Numerous scholarly libraries have extended their library arrangement by the obtaining of digital e-books.
In spite of this vital bearing, the writing uncovers that moderately little is thought about student observations and frames of mind towards e-books. Therefore, this exploration plans to limit this research gap and direct observational investigation into student's recognition towards digital books and their recurrence of utilization. The outcomes will be utilized to give proposals to library management with an aim to improve the nature of administration arrangement with respect to digital books.

Research Objectives

- Review the literature concerning the student uptake and experience of e-books in academic libraries
- Investigate perceptions and attitudes towards e-books and the usage of e-books at the Anna University
- Compare usage statistics between various user-groups, e.g. full-time, part-time, course type, etc.
- Identify if any improvements or alterations are required to facilitate a high service quality provision in relation to the e-books service at Bolton University library

As a rule of thumb, there would be one research aim and several research objectives to facilitate the achievement of this aim.

Review the Literature

A literature review is an all-inclusive summary of previous research on a topic. The literature review surveys published articles, books, news clipping, blogs and other sources relevant to a particular area of research. Generally, the purpose of a review is to analyze critically a segment of a published body of knowledge through enlisting, classifying, defining, abridging, quantitatively evaluating and comparing prior research studies, reviews of literature, and theoretical articles. It must give a theoretical base to the proposed research problem and help the author determine the nature of research. The literature review acknowledges the work of previous researchers, and in so doing, assures the reader that his work has been well conceived. It is assumed that by mentioning a previous work in the field of study, that the author has read, evaluated, and assimilated that work into the work at hand.
A literature review basically forms a "landscape" for the reader, giving her or him a complete understanding of the advances in the field. The organization of a survey of writing may change from discipline to discipline and from task to task. A review may be a self-contained unit — an end in itself — or an introduction to and basis for engaging in primary research.

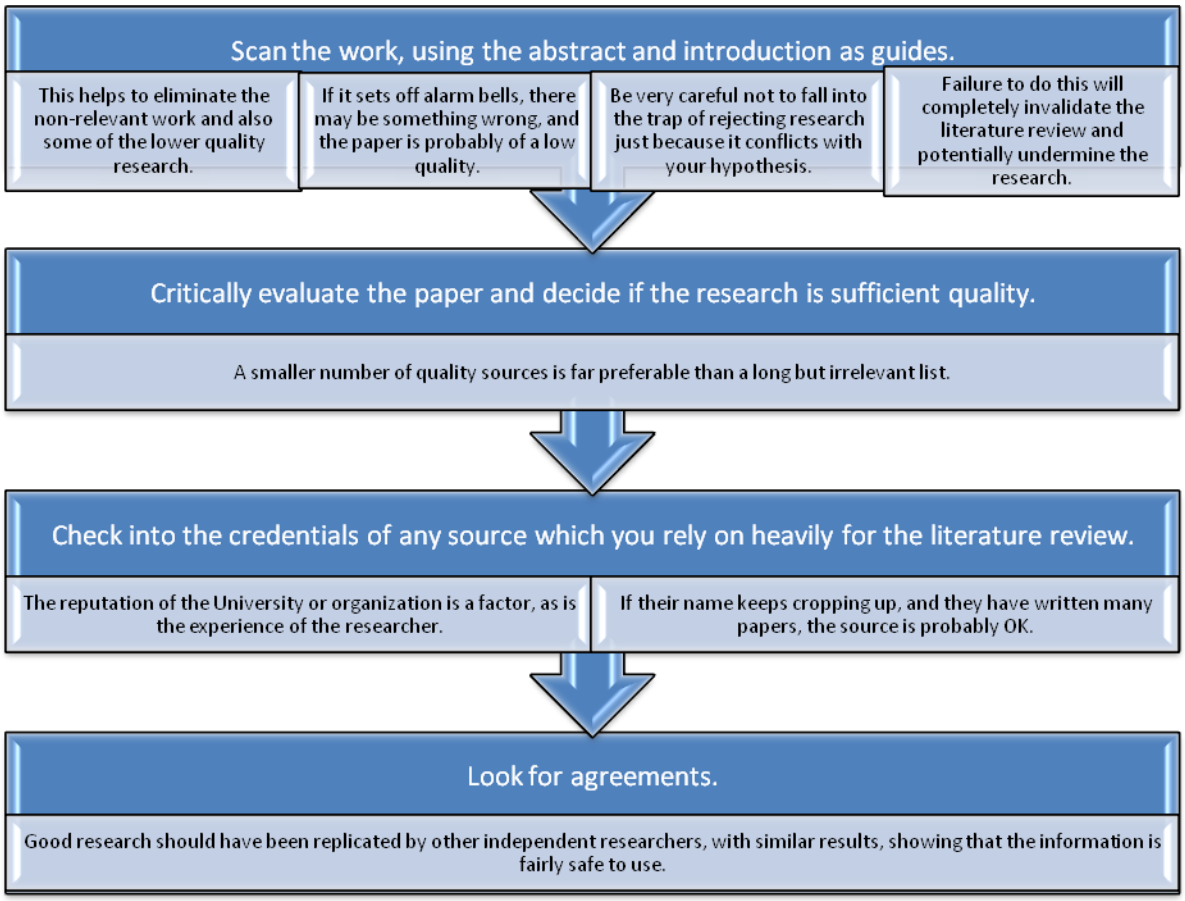


Fig 1.2: The Literature Review Process

Identify Research Gap

The outcome of the literature review is the identification of the research gap. A research gap may be considered as the missing piece or pieces in the research literature, which may correspond to the areas that has not yet been explored at all or is under-explored. It may also refer to something that remains to be done or learned in an area of research; it's a gap in the knowledge of the scientists in the field of research of study. Every research must, in some way, address a gap–that is, attempt to fill in some piece of information missing in the scientific literature. Otherwise, it is not novel research and is therefore not contributing to the overall goals of science This could be as straightforward as a populace or test (size, type, area, and so on.), explore technique, information assortment and additionally examination, or other research factors or conditions. A definitive objective of the distinguishing gap ought to be to discover a 'space' or opening for contributing new research.

Develop Hypotheses

When you distinguish the gap in the literature, you should advise your intended audience how you endeavor to address to some degree this absence of information or comprehension about the given research theme. The speculation is legitimately identified with a hypothesis however contains operationally characterized factors and is in testable structure. Theories enable us to decide, through research, if our hypothesis is right. A hypothesis is a basically statement that presents a research question and suggests a predictable outcome. It is an essential part of the scientific method that forms the foundation of scientific experiments. Therefore, you need to be cautious and exhaustive when

building your hypothesis. A slight flaw in the building of hypothesis could have an adverse effect on your experiment. More importantly, you need to build a robust testable hypothesis for your scientific experiments. A testable hypothesis is a speculation that can be demonstrated or negated because of experimentation.

Importance of a Testable Hypothesis

For a hypothesis to be viewed are testable it must meet the following basic criteria:

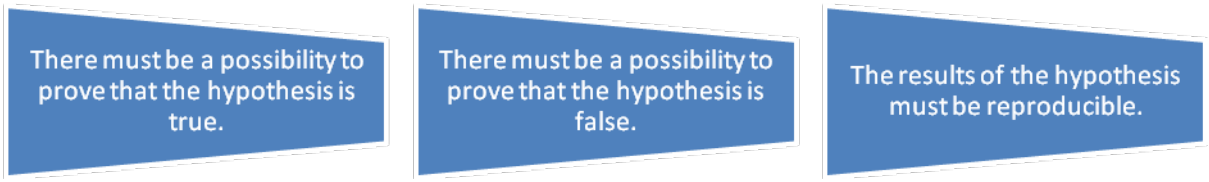


Fig 1.3: Hypothesis Criteria

Without these criteria, the hypothesis and the results will be vague. As a result, the experiment will not prove or disprove anything significant.

Formulation of an Effective Hypothesis

A testable hypothesis is essentially an intricate statement that must offer a precise introduction to a scientific experiment, its purposes, and the likely results. The following are some of the important considerations which must be kept into mind when building a convincing hypothesis.

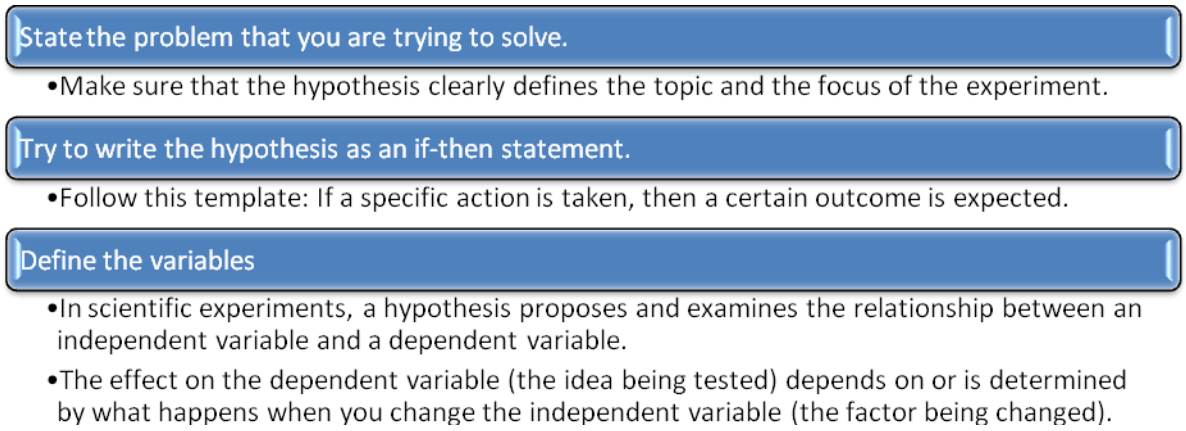


Fig 1.4: The Literature Review Process

The legitimacy of an experiment and its results depend on a robust testable hypothesis. Developing a strong testable hypothesis has few advantages, it forces us to think deeply and precisely about the results of a study. It enables us to gain an insight into the consequence of the question and the different variables involved in the study, thereby enabling to make exact predictions based on previous research. Hence, forming a hypothesis would be of significant value to the research.

Develop Research Methods

With well-defined goals, suggestions for metrics and methods are easier to make. In the event that quantitative strategies are utilized, published tests and surveys are suggested. These instruments have the upside of establishing legitimacy and unwavering quality and can be talked about as far as past uses

and discoveries. It is also possible to custom tailor and/or create measures for the needs of specific research questions or educational goals. In many cases, qualitative measures are appropriate. These incorporate focus groups, interviews, classroom perceptions, and "think-aloud" conventions.
A mixed method (quantitative and qualitative) approach is often the most useful, and both direct and indirect approaches can be employed to assess learning. The direct methodology utilizes quantitative presentation evaluations, for example,

- Portfolios
- Oral presentations
- Exams
- Problem sets
- Pretest-posttest comparisons of learning

The indirect approach uses qualitative and quantitative approaches, such as:

- Naturalistic descriptive observations
- Focus groups
- Journals
- Structured and open-ended interviews
- Surveys

Where it is conceivable, authentic or coordinated examination gatherings and test systems are utilized.

Data Collection
Data Collection is a procedure of gathering data from all the important sources to discover answers to the examination issue, test the speculation and assess the results. The foundation to a good research is laid with the collection of the right data. Ample time and efforts should be dedicated to gather and verify relevant information that will construct the basis of your research. By mixing different techniques, the research will be more reliable. Relevant collection techniques are all about gathering reliable information to find answers and prove the hypothesis you are focusing on in your research. You need to ensure you gather information from reliable and relevant sources. Relevant information can be either quantitative or qualitative. Quantitative facts represent numbers or statistical results you can easily compare. Qualitative methods have nothing to do with numbers; instead, you focus on words, emotions, and feelings.

There are two methods of collecting data for a project – primary and secondary. As discussed above, primary quantitative collection methods focus on obtaining numbers from mathematical formulas. The methods include calculating the mean, mode, average, using questionnaires, and closed-ended questions. These methods are typically cheaper and quite easy to apply. Accurate results don't require spending a lot of time because they can be easily compared.

Primary qualitative collection methods are more complicated. They aim to provide a greater level of understanding of numbers and figures. They involve using open-ended questions, observations, and interviews. Analyzing and understanding qualitative data is a complicated task.
Secondary data collection methods involve using facts and statistics that have already been published in newspapers, journals, and other sources. You need to make sure you use reliable resources to guarantee the validity of your research project. A further detailed description of these methods is provided in the subsequent sections.

Data Analysis

Once the data has been collected, the biggest challenge remains to gain meaningful insights from it. For instance, if a dress brand is attempting to recognize the most recent patterns among young ladies, the brand will initially connect with young ladies and ask them inquiries applicable to the exploration objective. In the wake of gathering this data, the brand will investigate that information to recognize designs — for instance, it might find that most young ladies might want to see more assortments of pants. Data analysis is mining meaningful insights from the pile of data. There are a wide range of information examination strategies, contingent upon the kind of research. Here are a couple of strategies you can use to break down quantitative and subjective information.

Data Analysis in Qualitative Research

There are a wide range of information examination strategies, contingent upon the kind of research. Here are a couple of strategies you can use to break down quantitative and subjective information.

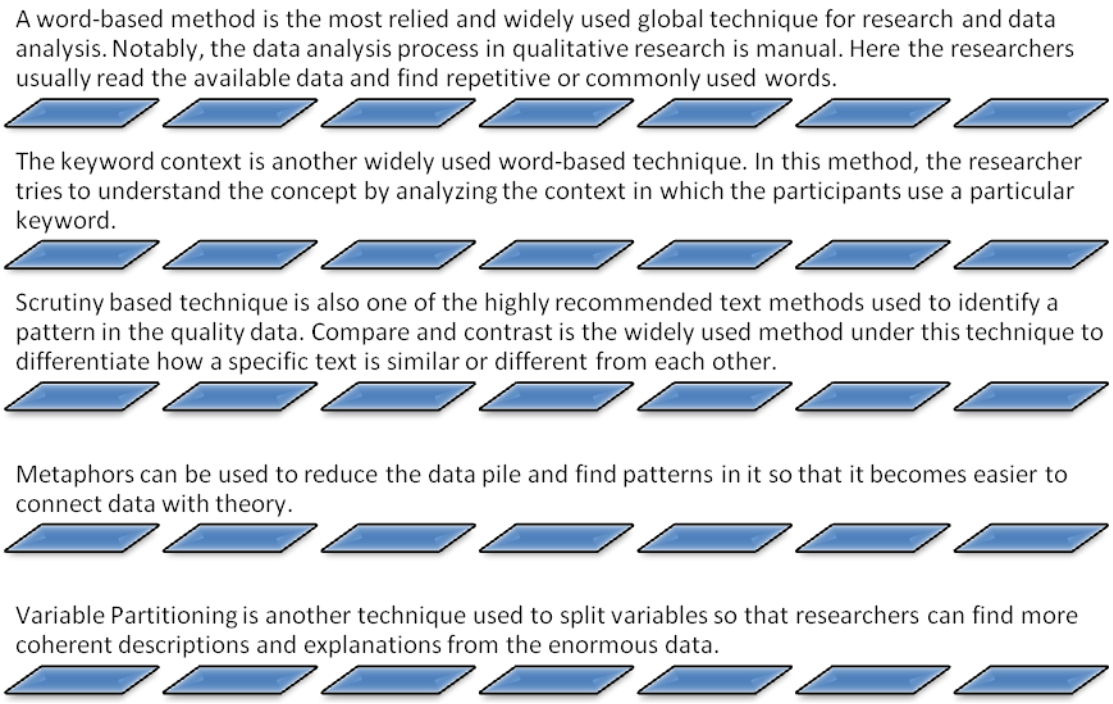


Fig 1.5: Finding patterns in the qualitative data

Data Analysis in Qualitative Research

The following are a few of the methods used for data analysis in qualitative research such as :

Content Analysis

It is widely accepted and the most frequently employed technique for data analysis in research methodology. It can be used to analyze the documented information from text, images, and sometimes from the physical items also. It depends on the research questions to predict when and where to use this method.

Narrative Analysis

This is a method used to analyze content gathered from various sources. Here, the source can be personal interviews, field observation, and surveys. The majority of times, stories or opinions shared by people are focused on finding answers to the research questions.

Discourse Analysis

Similar to narrative analysis, discourse analysis is used to analyze the interactions with people. Nevertheless, this particular method takes into consideration the social context under which or within which the communication between the researcher and respondent takes place.In addition to that, discourse analysis also focuses on the lifestyle and day-to-day environment while deriving any conclusion.

Grounded Theory

When you want to explain why a particular phenomenon happened, then using grounded theory for analyzing quality data is the best resort. Grounded theory is applied to study data about the host of similar cases occurring in different settings. When researchers are using this method, they might alter explanations or produce new ones until they arrive at some conclusion.

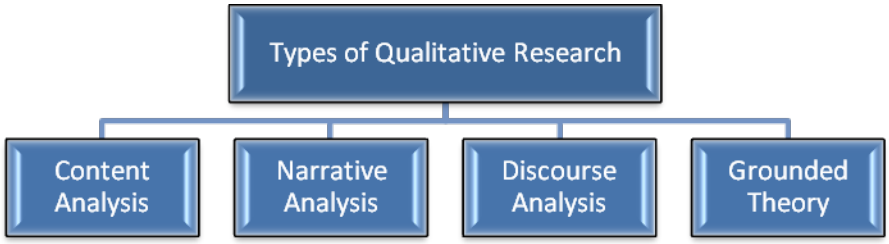


Fig 1.6: Methods used for data analysis in qualitative research

Data Analysis in Quantitative Research

There are a wide range of information examination strategies, contingent upon the kind of research. Here are a couple of strategies you can use to break down quantitative and subjective information.

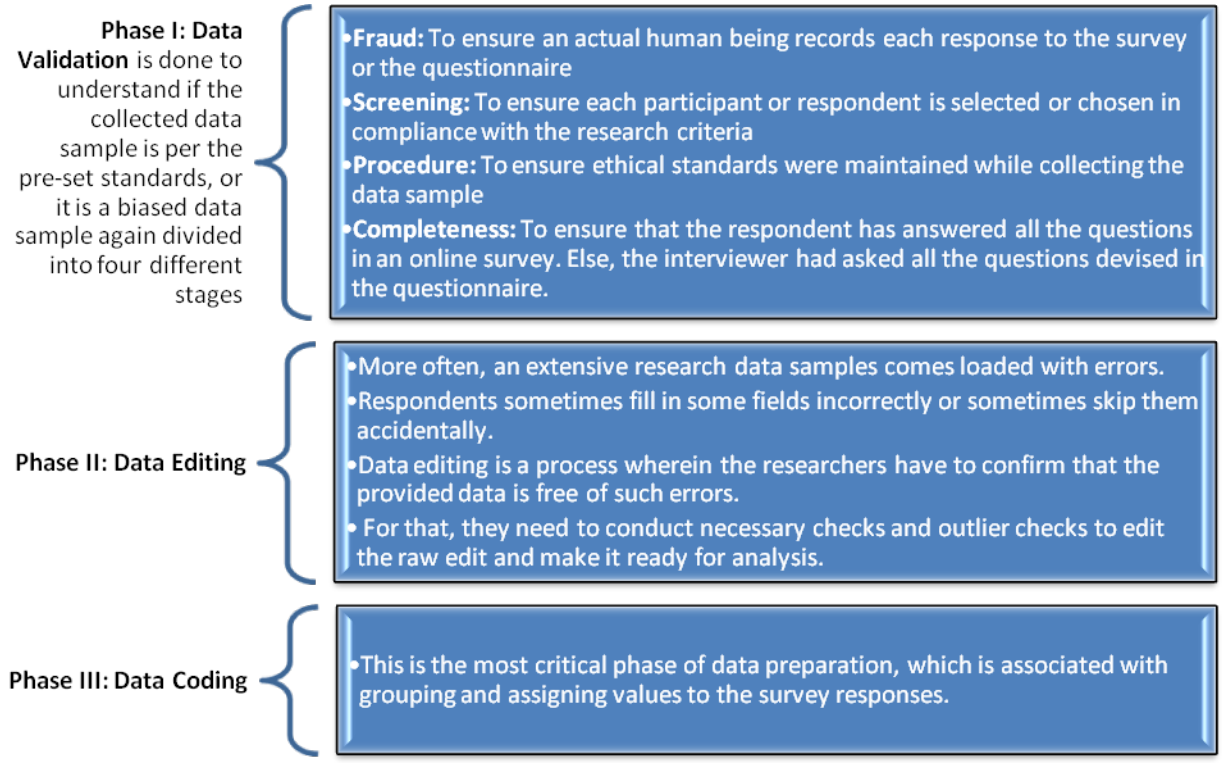


Fig 1.7: Preparing data for analysis

Methods Used for Data Analysis in Quantitative Research

Descriptive Statistics: After the data is set up for examination, researchers are available to utilizing diverse research and information investigation techniques to infer important bits of knowledge. Without a doubt, statistical techniques are generally supported to dissect and analyze the numerical data. The strategy is again grouped into two classification groups. First, 'Descriptive Statistics' used to describe data. Second, 'Inferential statistics' that helps in comparing the data.

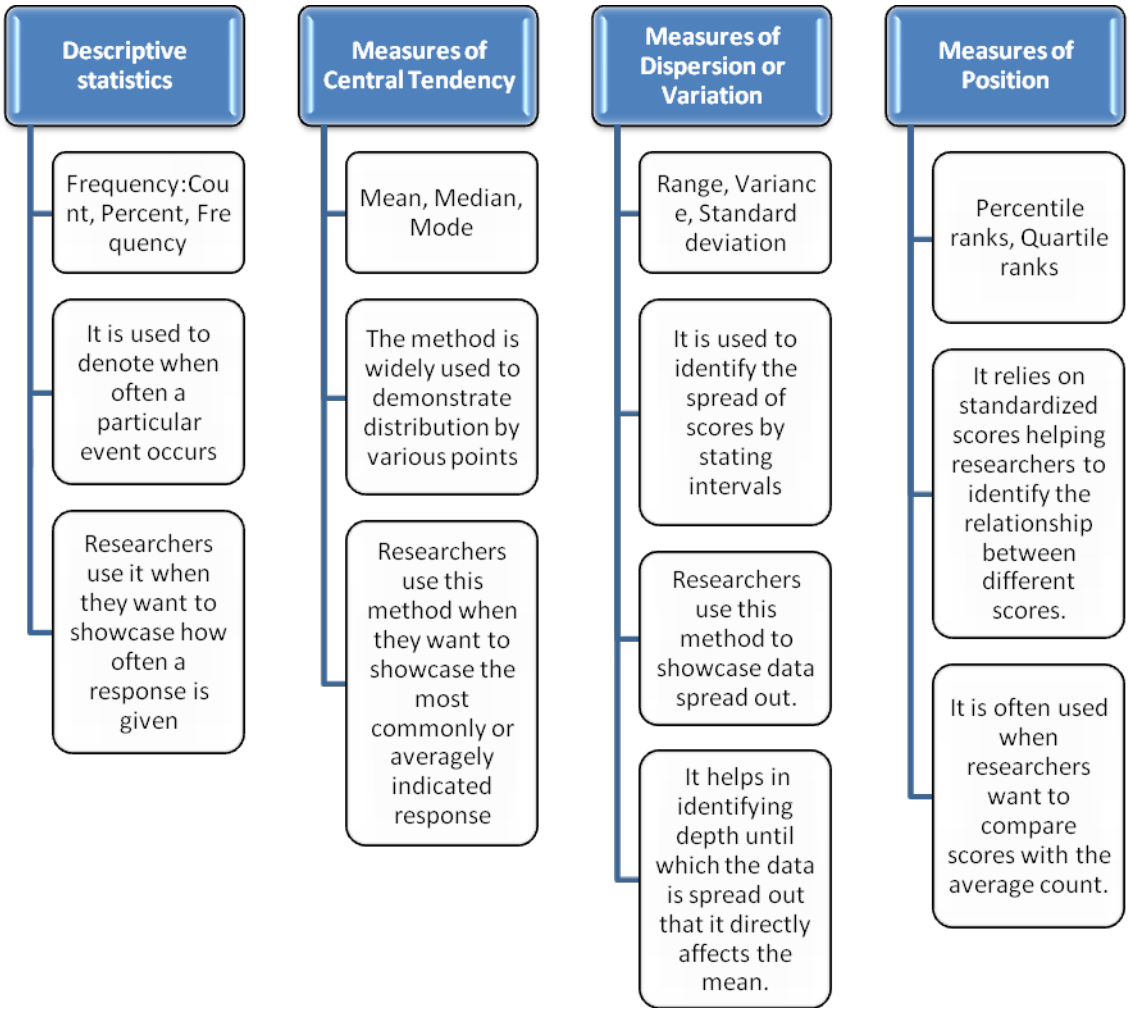


Fig 1.8:Descriptive Statistics

For quantitative market research use of descriptive analysis frequently gives complete numbers, but the analysis is never enough to establish the justification behind those numbers. The most important emphasis should be on thinking of the best method to be used for research and data analysis suiting your survey questionnaire which should suit the story that you want to tell through your research. For example, the best metric to best way to determine the average scores of the students in school is mean. The synonym used for Descriptive analysis is 'univariate analysis' since it is typically used to evaluate a single variable.

Inferential Statistics: They are used to make predictions about a bigger population after research and data analysis of the collected sample of the representing population. For example, at a movie theater, you can ask some odd 100 audiences if they like the movie they are watching. Researchers at this point

utilize inferential insights on the gathered example to reason that around 80-90% of individuals like the movie they are viewing.

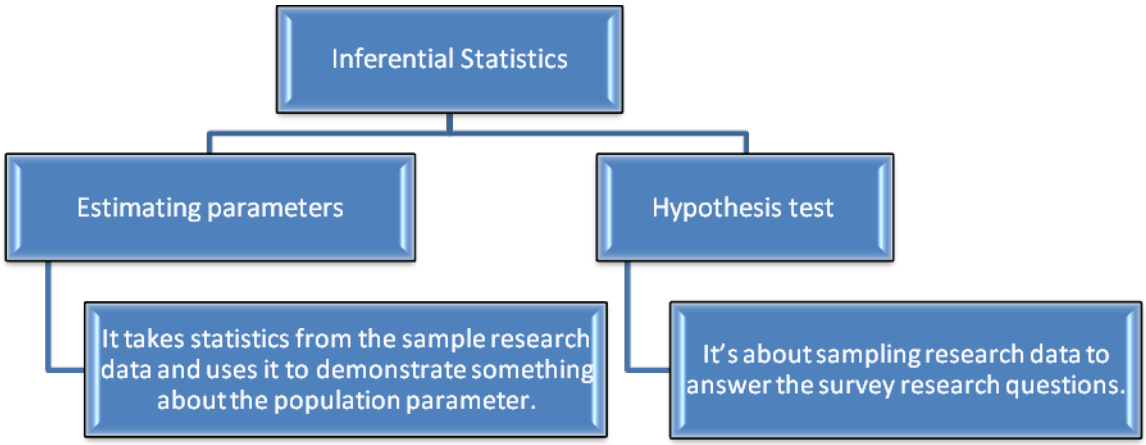


Fig 1.9: Inferential Statistics

These are classy methods of analysis used to display the relationship between different variables instead of describing a single variable.

Table 1.1: Some Commonly Used Methods for Data Analysis in Research

Correlation	**When researchers are not conducting experimental research wherein the researchers are interested to understand the relationship between two or more variables, they opt for correlation research methods.**
Cross-tabulation	Also called as contingency tables, cross-tabulation is a method used to analyze the relationship between multiple variables.
Regression analysis	For understanding the strong relationship between two variables, researchers do not look beyond the primary and commonly used regression analysis method, which is also a type of predictive analysis used.
Frequency tables	The statistical procedure is used for testing the degree to which two or more vary or differ in an experiment.
Analysis of variance	The statistical procedure is used for testing the degree to which two or more vary or differ in an experiment. A considerable degree of variation means research findings were significant. In many contexts, ANOVA testing and variance analysis are similar.

To Do Activity
It was claimed that rural women have longer and stronger hair than urban women. A more aggressive statistician wants to prove the assumed cause and effect relationship between the associated variables. How would you proceed with this type of research?

Findings and Conclusion
The conclusion is proposed to assist the reader in understanding the significance of your research and how it is relevant to them. A conclusion is not just a summary of the main topics covered or a re-statement of your research problem, but a summation of the major points and, if applicable, recommendation of new areas for future research. An elegantly composed conclusion furnishes you with significant chances to exhibit to the reader your comprehension of the research issue.

Presenting the last word on the issues you raised in the research problem.	Just as the introduction gives a first impression to your reader, the conclusion offers a chance to leave a lasting impression. Do this, for example, by highlighting key findings in your analysis or result section or by noting important or unexpected implications applied to practice.
Summarizing your thoughts and conveying the larger significance of your study.	The conclusion is an opportunity to succinctly answer [or in some cases, to re-emphasize] the "So What?" question by placing the study within the context of how your research advances past research about the topic.
Identifying how a gap in the literature has been addressed.	The conclusion can be where you describe how a previously identified gap in the literature [described in your literature review section] has been filled by your research.
Demonstrating the importance of your ideas.	Don't be shy. The conclusion offers you the opportunity to elaborate on the impact and significance of your findings.
Introducing possible new or expanded ways of thinking about the research problem.	This does not refer to introducing new information [which should be avoided], but to offer new insight and creative approaches for framing or contextualizing the research problem based on the results of your study.

Fig 1.10: Concluding Findings

Cyclical or Iterative Research Process
Depicted along these lines the research procedure is given the impression of linearity, yet research investigation is essentially an iterative procedure whereby the way toward directing the research will pave way to new thoughts which, in turn, provide feedback into the data collection and investigation stage. Decisions made early in the research process are often revisited in the light of fresh insights or practical problems met along the way. Despite the course taken accordingly, research should begin with the issue and the research questions. In the event that the aim of research is to respond to your

inquiries, it pursues that decision of strategy ought to create from your inquiry: pick the technique that can best give the data you have to address your research question given the resources accessible to you. This is one reason why it is very important to be clear as to what you are asking. As should be obvious there are various decisions to be made inside the research procedure. Arranging your research includes due deliberation to the four overlapping themes.

- The conceptual approach – the philosophical underpinnings of research
- Research design – how data collection is organised
- Data collection techniques – how data are collected
- Sampling – from whom data are collected

These aspects of research planning can be represented as forming different layers of a research 'onion'. The research onion was developed by Saunders et al. (2007) in order to define the stages through which the researcher must pass when framing an effective methodology.

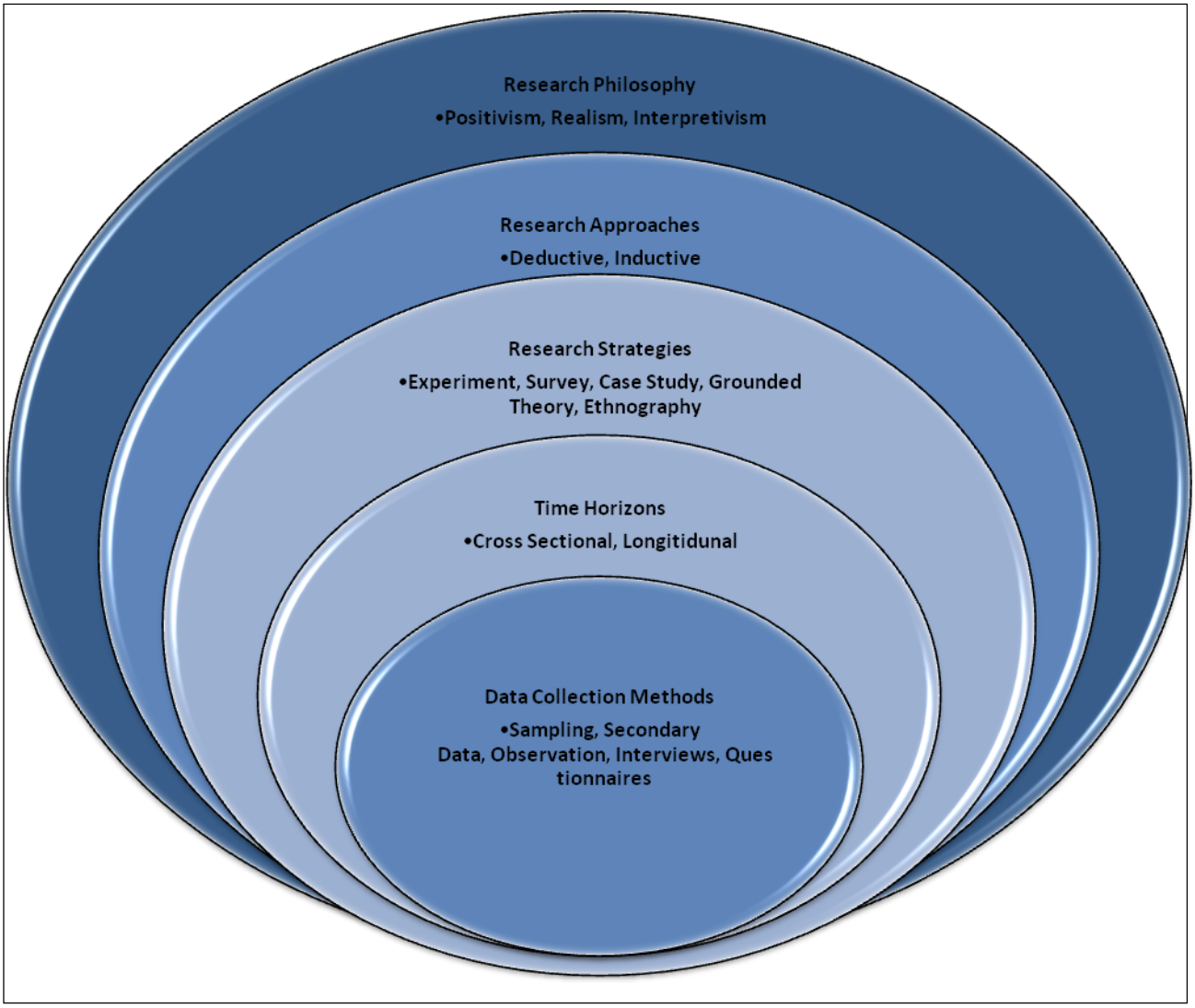


Fig 1.11: The 'onion' diagram of research choices

Subjective and Objective Research Perspectives

As discussed in detail in the previous subsection, research basically deals with affirmation or confirmation of the proposed hypothesis or understanding of certain phenomenon through analysis of the collected data. The method adopted for carrying out research largely depends on the existing knowledge of the chosen field or area of research. The researcher can choose either subjective or objective research methodology for carrying out research pursuit.

Subjective Research

Subjective research is based on the personal opinions, interpretations, points of view, emotions and judgment of the researcher on the basis of his interactions with the various stake holders involved in the proposed hypothesis or the proposed problem definition. It is not considered as a good choice for the research problems dealing with scenarios like news reporting or decision making in business or politics. On the other hand, objective information or analysis is more quantitative based on factual information and factors which are measurable and observable.

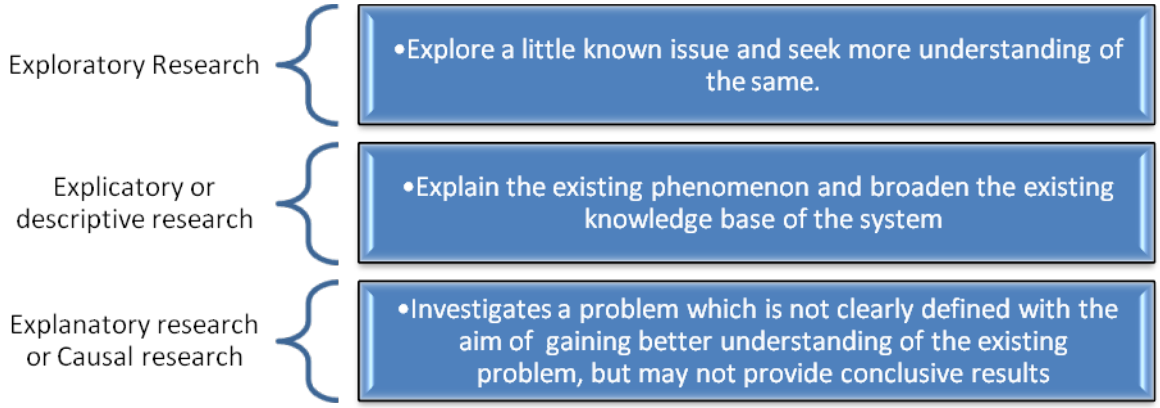


Fig 1.12: Subjective Research perspectives

Exploratory Research

Exploratory research is a kind of research directed towards a lesser explored problem which aims at establishing priorities, building newer definitions and expanding the final research design. The below figure gives broad classification of exploratory research design methodologies.

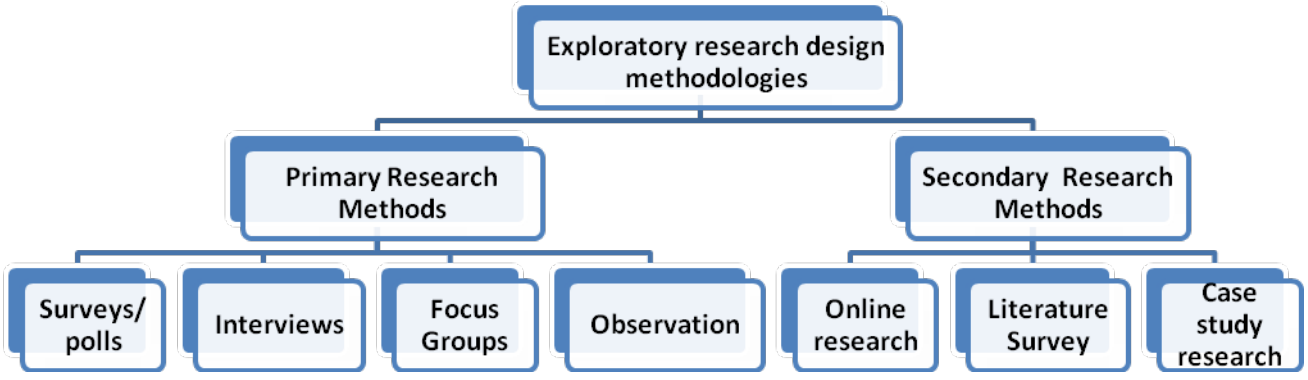


Fig 1.13: Exploratory Research Design Methodologies

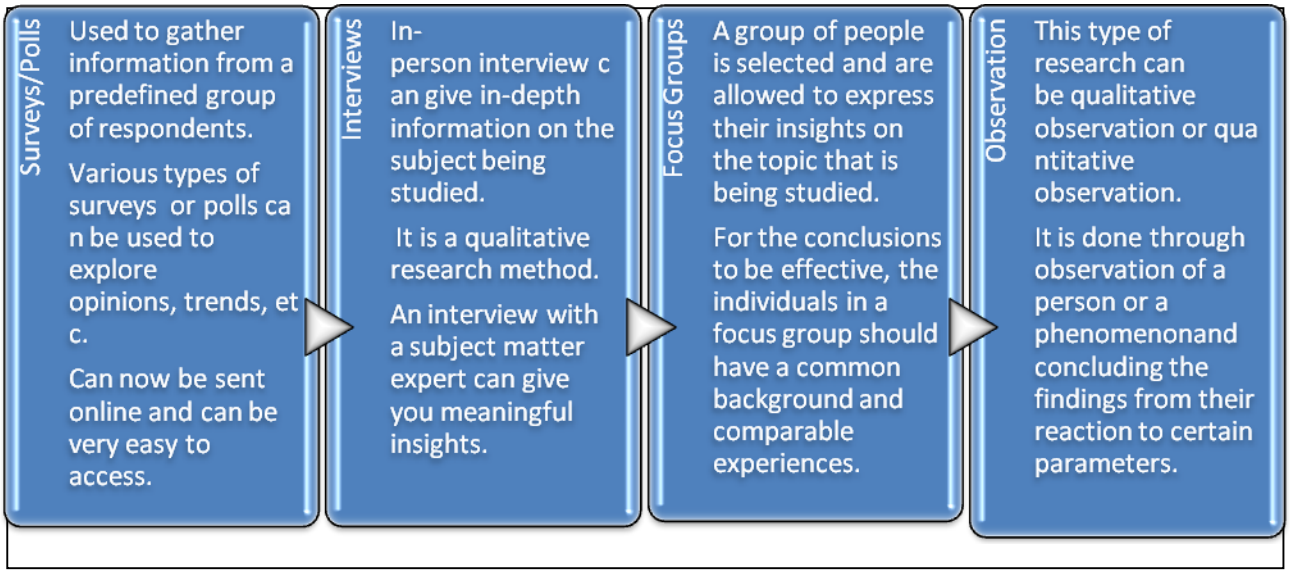


Fig 1.14: Primary Research methodology

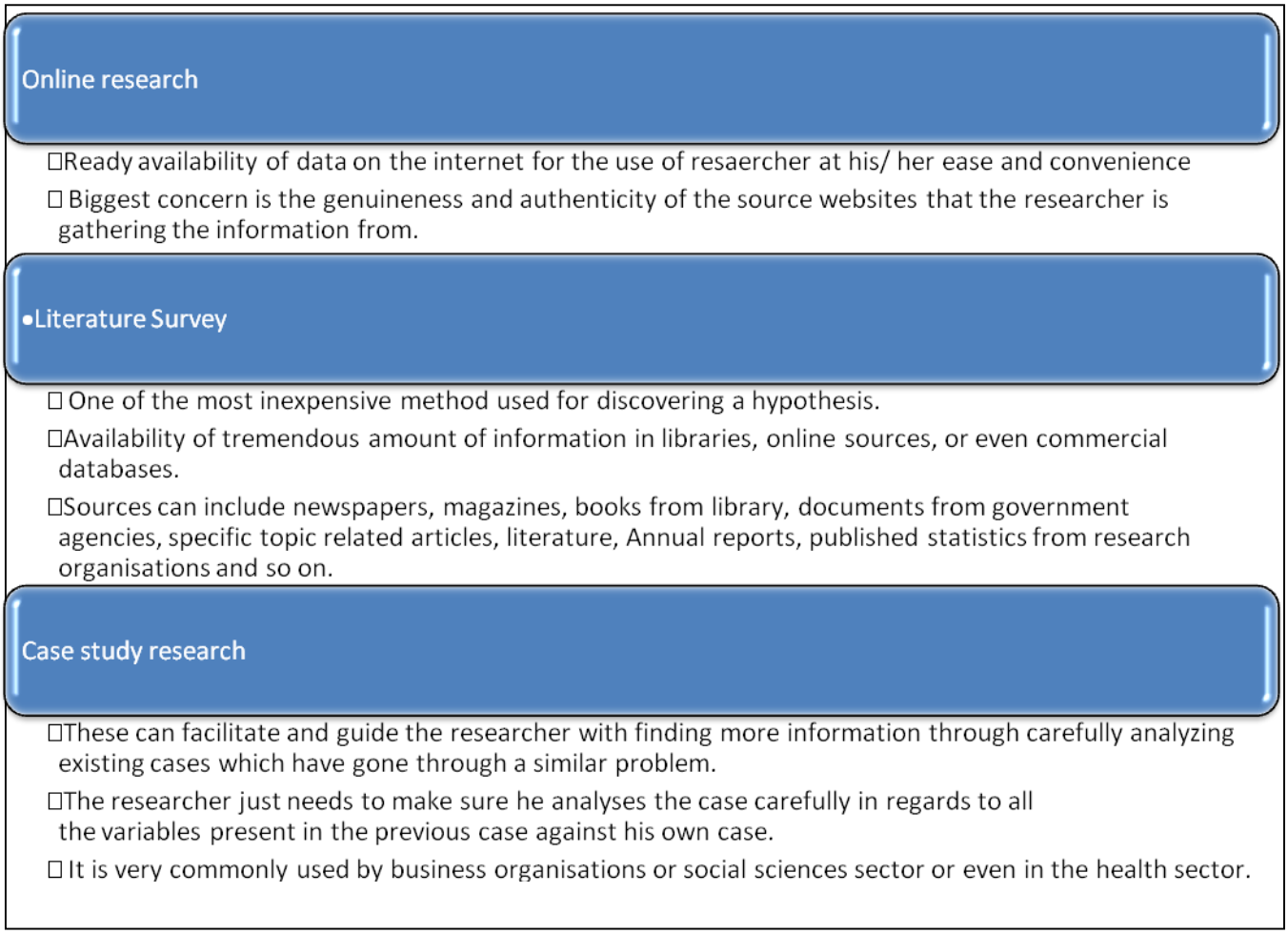


Fig 1.15: Secondary Research methodology

Explanatory or Causal Research

Explanatory research is based on the cause and effect relationship wherein the researcher builds the hypothesis on the basis of the literature surveys done through exploratory research and mostly tests hypotheses about cause and effect relationships. The most popular primary data collection methods in explanatory research design are experiments.

The main difference between exploratory and explanatory research is their correlation with inductive and deductive research. Explanatory research tends to highly correlated to deductive results whereas exploratory research has a strong correlation with induction. This is not always the case but, it can be treated as a generalization.

Descriptive or Explicatory Research

Descriptive research may be backed by an explanatory or exploratory study. On its own, descriptive research is not sufficient for an academic project. Academic research is aimed at progressing current knowledge.The table below compares the main characteristics of causal research to exploratory and descriptive research designs:

Table 1.2: Main Characteristics of Research Designs

	Causal research	Exploratory research	Descriptive research
Amountof uncertainty characterising decision situation	Clearly defined	Highly ambiguous	Partially defined
Key research statement	Research hypotheses	Research question	Research question
When conducted?	Later stages of decision making	Early stage of decision making	Later stages of decision making
Usual research approach	Highly structured	Unstructured	Structured
Examples	'Will consumers buy more products in a blue package?' 'Which of two advertising campaigns will be more effective?	'Our sales are declining for no apparent reason' 'What kinds of new products are fast-food consumers interested in?	'What kind of people patronize our stores compared to our primary competitor?' 'What product features are the most important to our customers?

Objective Research

Objective research looks into cases to depict a genuine and right reality, which is autonomous of those engaged with the exploration procedure. In spite of the fact that this is a disentangled perspective on the manner by which research can be drawn nearer, it is a significant differentiation to consider. Regardless of whether you consider your exploration theme in objective or emotional terms will decide the improvement of the research questions, the sort of information gathered, the strategies for information assortment and investigation you embrace and the determinations that you make. This is the reason it is critical to consider your very own point of view when planning your research. objective research will in general be displayed on the strategies for the regular sciences, for example, analyses or huge scale reviews. It also looks to set up law-like speculations which can be applied to a similar

phenomenon in various settings. This point of view, which benefits objectivity, is called positivism and depends on information that can be dependent upon measurable examination and speculation. Positivist analysts utilize quantitative systems, which depend on estimation and numbers, to gather and investigate information. Interpretivists are progressively worried about language and different types of subjective information, which depend on words or pictures. Having said that, scientists utilizing objectivist and positivist suppositions once in a while utilize subjective information while interpretivists some of the time utilize quantitative information. (Quantitative and subjective approaches will be discussed about in more detail in the last piece of this course.) The key is to comprehend the point of view you expect to receive and understand the constraints and openings it offers.

Case Study Related to the Research Process Steps

The case study given below elaborates the steps in the research process with reference to a rural setting.

Step 1: The goal of the chosen problem statement is to study the popularity and usage of software widget in rural areas.

Step 2: To collect possible background information, the students are expected to know the different Applications available for the rural citizens. To mention a few are apps like the ones listed below.

1. Meri Sadak: Pradhan Mantri Gram SadakYojna is an app that provides a platform to the citizens through a comment or complaint box. Anyone troubled by any government road can post a query or complaint using this app. Citizens can also share photographs in the feedback column.
2. Self-Reliant Initiatives through Joint Action (SRIJAN) is a mobile-based application for the monitoring of soy crops. This app empowers woman farmers to increase crop productivity and profitability.
3. OpAsha is an app created for active detection of TB contact patients and case finding. This app is essential as it increases detection rates of patients suffering from tuberculosis.
4. Swayam ShikshanPrayog is a mobile app that empowers women to become leaders and entrepreneurs. The app provides rural women and communities with a variety of opportunities including skill building, livelihood, and health-enhancing.
5. JayalaxmiAgrotech is a mobile application that aims at providing crop specific information to illiterate farmers through tools such as audio and video.
6. Swachh App for citizens serves as a tool to monitor the improvements in sanitation coverage in rural India, on a real time basis. Source: https://apps.mgov.gov.in/

Step 3: Further researching and consolidation of facts about every app, it was noted that the number of users of for many apps listed above was very meagre since the date of launch. For example, “Swachh App for citizens” had around 1000+ downloads till date since the year of launch in 2016 as shown in Fig 1.17.

Further to this study, the research question was formulated as why is the usage of the mobile apps very less despite the apps being exclusively launched for the benefit of the rural population.

Step 4: To study the above-mentioned research question, a descriptive research survey must be planned with a target population chosen from a village to study their attitude towards the usage of the App and to understand their awareness about the existence of such an App. A simple questionnaire-based analysis must be carried out.

Step 5: Further to the survey, based on the quantitative results obtained, the reasons for the less usage and least awareness of the availability and usage of the apps will be studied.

Step 6: The results of this study will be submitted as a project report so as to recommend the ways and means to promote the apps available for the rural population in a better way.

Step 7: The draft will be edited, refined and finalized as the final report

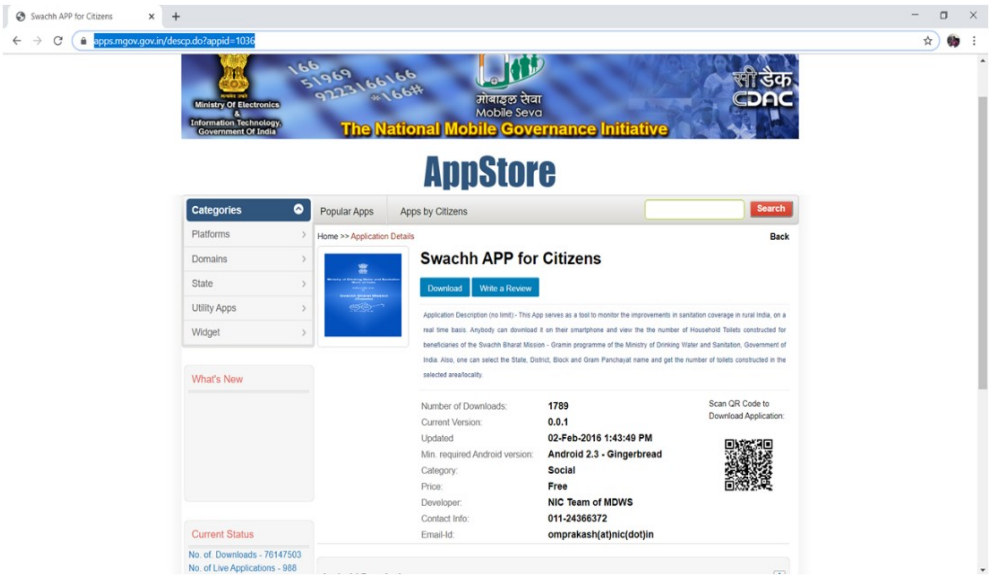


Fig 1.16: Depicting low usage (downloads) of the Swachh App widget

1.3 Research Process Characteristics and Requirements

As discussed in depth in sections 1.1 and 1.2, research is the process of arriving at a solution to a problem through the planned and systematic collection, analysis and interpretation of data. The main features of research are:

- Research is empirical in nature as it is built on observations and experimentation of proposed theories.
- Every research process follows a systematic, orderly and sequential procedures.
- Research is carried out under controlled conditions wherein all variables excluding those that are tested/experimented upon are set aside as constants.
- Every research employs some hypothesis which serves as a guide to the investigation process.
- Research tends to be analytical, wherein critical analysis of all the data used is done to avoid any error in their interpretation
- Research should be Objective, Unbiased, & Logical so as to base all the findings logically and empirically.
- Research employs quantitative or statistical methods to transform the data into numerical measures and treat it statistically.

Research Process Characteristics

Certain terms are ordinarily utilized in research and the accomplishment of any research relies upon these terms. These terms decide if a problem at hand is liberated from predispositions, biases, and emotional mistakes or not. They are known as the qualities of research.

Reliability is an emotional term which can't be estimated exactly, yet today there are instruments which can evaluate the dependability of any exploration. Reliability is the repeatability of any exploration,

examine instrument, apparatus or methodology. In the event that any research yields comparable outcomes each time it is undertaken with comparable populace and with comparable techniques, it is called to be a solid and reliable research. Assume a research is led on the impacts of single parenting on the performance of students in the class. In the event that the outcomes reason that it causes low evaluations in class, these outcomes ought to must be solid for another example taken from a comparable populace. More the outcomes are comparable; greater dependability is available in the exploration.

Validity is the quality with which we can substantiate the research conclusions, suppositions or suggestions genuine or bogus. Validity decides the relevance of the examination. Validity of the research instrument can be characterized as the appropriateness of the research instrument to the examination issue or how precisely the instrument gauges the issue. A few analysts state that validity and unwavering quality are co-related, yet the validity is significantly more significant than dependability. Without validity, research may get diverted in the wrong way and may not lend the desired conclusions. To keep the research on-track, characterize your ideas in the most ideal way with the goal that no blunder happens during the estimation.

Accuracy is likewise how much each research procedure, instrument, and device is identified with one another. Accuracy additionally gauges whether research tools have been chosen in most ideal way and research methods suits the examination issue or not. For instance, if a research must be conducted on the trans-sexual orientation individuals, a few data collection tools can be utilized relying upon the exploration issues. Yet on the off chance that you find that populace less agreeable the most ideal path is to watch them instead of submitting poll on the grounds that in survey it is possible that they will give one-sided reactions or they won't restore the surveys by any stretch of the imagination. So picking the best data collection tools to improve the precision of research is vital.

Credibility accompanies the utilization of the best source of data and best methodology in research. In the event that you are utilizing recycled data in your research because of any reason, your examination may finish in less time yet its credibility will be in question since secondary data has been controlled by people and is along the lines not substantial to use in investigation. A specific percentage of secondary data can be utilized if the essential source isn't accessible however putting together a research totally with respect to secondary data when essential information can be accumulated is least solid. At the point when analyst gives precise references in the research the validity of the research increments yet counterfeit references additionally decline the credibility of the exploration.

Generalization is the degree to which the research findings can be applied to bigger populace. At the point when a researcher directs an examination, he/she picks an objective populace and from this populace he takes a little example to lead the research. This example is illustrative of the entire populace so the research findings and observations ought to be likewise. In the likelihood of research findings be applied to any example from the populace, the consequences of the exploration are said to be generalizable.

Empirical nature of research implies that the research has been directed after thorough logical strategies and methodology. Each progression in the exploration has been tried for precision and depends on genuine encounters. Quantitative research is easier to prove scientifically than qualitative research. In qualitative research predispositions and bias are anything but difficult to happen.

Systematic approach is the main way to with carry on a research. No research can be led erratically. Each progression must pursue other. There are set of strategies that have been tried over a period of

time frame and are found suitable to be used in research. Each research, in this way, ought to pursue a system.

Controlled all things considered, understanding there are numerous variables that influence a result. A single occasion is regularly an aftereffect of a few components. At the point when similar occasions are tried in a research, because of the more extensive nature of components that impact that specific occasion, a few elements are taken as controlled variables while others are tried for a potential impact. The controlled factors, elements or variables should have to be controlled rigorously. In pure sciences, it is anything but difficult to control such components since tests are led in the lab however in social sciences it gets hard to control these variables in view of the idea of research.

1.4 Types of Research and Theory of Causality

Case Studies Explaining Each Research Characteristic with an Application in the Rural Research Examples

Empirical Research Case Study

Research Problem: Study about the "happiness quotient of rural adults".

Background: Happiness has been defined as the degree to which an individual judge the overall quality of his or her life as favourable. The happiness of a society, depends on its individual members (Rao et al ,2017). So, the study was designed to assess happiness among rural adults and to identify various socio-demographic, family and health determinants affecting happiness.

Methods: It was a cross sectional study which included permanent residents of the field practice area, aged ≥20 years, of both gender and willing to participate in the empirical study. Information pertaining to socio-demographic characteristics, details pertaining to co-morbidities, habits and family particulars were collected by personal interviews using a pre-designed questionnaire. Happiness was assessed using the Oxford happiness questionnaire.

Results of the study would specify what % of the rural community studied were happy and what % was not, as per the Oxford happiness questionnaire.

The study concluded that Good relationship with family members and neighbourhood, absence of co-morbidities was positive determinants of happiness, while hospitalization, concerns about employment and marriage of their children contributed to unhappiness.

Systematic Research Case Study

Research Problem: A study related to Two-wheeler purchase decision making in rural India.

Background: Rural markets have a socio-economic role to play in rural economy as the agents of rural developments. Study for the demand of two-wheelers in rural area and influence the factors of like family, friends, dealers, service and mileage for the process of purchasing a two-wheeler.

Methods: Survey method was used to collect the primary data from the respondents. A structured interview was administered to extract responses. Its methodology is made known to all concerned for critical scrutiny and for use in testing the hypothesis to derive at the conclusions.

Results: The research was systematic, not haphazard and was objective type to avoid the distorting effects of personal bias. Systematic research employs hypothesis to guide the investigation process and facilitate managerial decision making like:

- Identifying the factors influencing rural consumers, in brand selection while purchasing the two-wheelers.
- Identifying the rural consumer's choice of preference while purchasing the two wheelers etc.

Controlled Research Case Study

Research Problem: Study on disparities in maternal and neonatal outcomes due to poor maternal nutrition of rural women.

Background: Disparities in maternal and neonatal outcomes associated with poor maternal nutrition is a study of global importance (Ali et al, 2019).

Methods: An individually randomized controlled trial on pre-conceptionally maternal nutrition of women in rural areas. Eligible participants were identified, randomized and the participants of different groups were ensured compliance to the intervention, and measuring the primary outcome within the 24 hours of birth. Study's intervention included measured nutrient supplements and for underweightwomen, an additional protein supplement was provided.

Results: The primary hypothesis of the study was that in women living in poor, food insecure populations, commencing a maternal nutrition supplement at least 3 months prior to pregnancy (Group 1) would result in significantly greater fetal linear growth than starting the same nutrition supplement at 12 to 14 weeks gestation (Group 2) or than not providing this supplement (Control Group).

Analytical Research Case Study

Analytical research is a specific type of research that makes use of facts and information collected and application of the critical thinking skills to evaluate the same.

Research Problem: An analytical study challenges and strategies of rural marketing potential in India

Background: The Indian rural markets are opened with great opportunities to the FMCG marketers.

Methods/Statistical Analysis: The responses from rural customers were collected using structured questionnaire was used to collect the response from rural customers. Random samples of rural customers were selected. Analytical models like Structural Equation Model, etc. were used to analyse the relationship between the variables and also analyse effectiveness of rural marketing strategies.

Results: This study found that the availability, awareness, affordability and acceptably are major challenges to the rural marketers. The existing marketing strategies were not so effective. So, the marketers have to resign the rural marketing strategies and the approach, strategies and marketing mix need to be formulated and examined from rural consumer's prospective to be successful in the rural market.

To Do Activity

Rural migrant workers contribute a lot towards urbanization of our nation, but they get unequal treatment compared to urban residents, especially in housing security. Investigate on the rural migrants in a nearby town and present your report of investigation in terms of study of housing conditions, government support to those rural migrants in terms of providing urban migrant worker apartments, etc.

Theory of Causality

The Nature of Causality

We can say two variables are related if certain values of one variable coincide with certain variables of another variable. To put it another way, two variables are related if they vary together in a pattern. When the values of one variable produce the values of the other variable, the relationship is a causal relationship. E.g. Cutting of trees results in more soil erosion, hence cutting of trees is causal with soil erosion. In theoretical research, the causal relationships are predominantly examined. Examples: changes in party identification cause changes in vote choice; changes in military spending cause changes in the level of tension between states.

Causality is all about the cause and effect. Theory of Causality governs the relationship between events. Formalizing this, the world consists of a collection of causal systems; in each causal system there is a set of observable causal variables. An example problem statement which studies the prevalence of chronic malnutrition and anaemia and probable associated factors in children under five using a multi-causal model in a rural community is presented here in Fig. 1.17

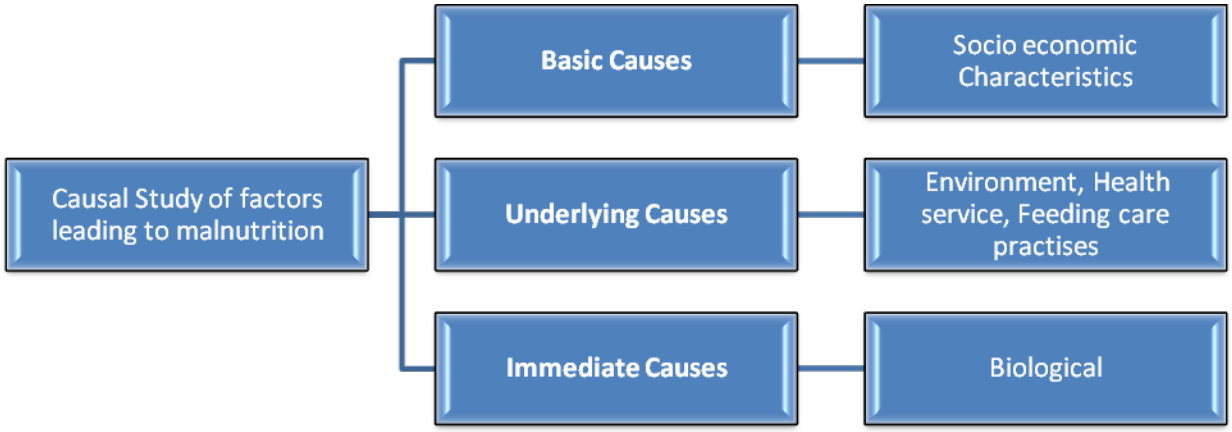


Fig 1.17 Causal Study of Factors Chronic Malnutrition/ Anaemia in Rural Children

From the above model, the independent variables were classified into four blocks or levels of analysis: Block 1 which includes the basic causes: the socioeconomic variables (education of mother and father, number of dependents in the home and family income); Block 2 and 3 which include the underlying causes. Block 2 includes the environmental characteristics (overcrowding, water supply, excreta disposal, source of water used for drinking) and variables related to health services; Block 3 includes feeding and care practices; Block 4, the immediate causes, which includes the biological characteristics etc. Thus, the diagram explains the hierarchical causal relationship between the 4 independent blocks and chronic malnutrition and anaemia in children.

Case Studies

Example 1: Causation Study of Rural Poverty

A study on causes of rural poverty may be attributed to less economic, educational and socio-economic-capabilities of the farmer, which in turn causes the qualitative growth of farmer to turn down from looking themselves as an entrepreneur in his/her village. This attitude of the farmer may be one of the predominant reasons for the farmers not in a position to directly sell their goods in the market at profit and as a result of which they are exploited and this leads to rural poverty.

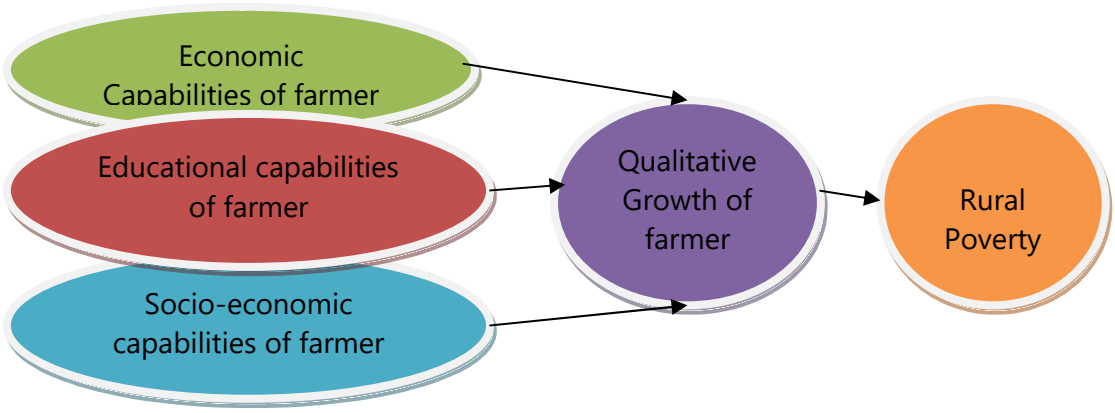


Fig 1.18: Causal diagram of "rural poverty"

Example 2

The causation theory generally helps to measure the output of interest either directly or indirectly. For example, in the example below, loyalty is the output we are interested to study and if we assume that there are two factors impacting loyalty namely, quality of the service and satisfaction of the customer, and then the Causation diagram is as given below.The indirect method of measuring loyalty shows the value as 0.2*0.1 = 0.02 and the direct method of measuring quality gives 0.3. Hence the total measure of the output is given as Loyalty = 0.3 + 0.02 = 0.32. Thus, we can understand that by means of calculating the impacts of each of the causes on the effects, the overall effect can be measured.

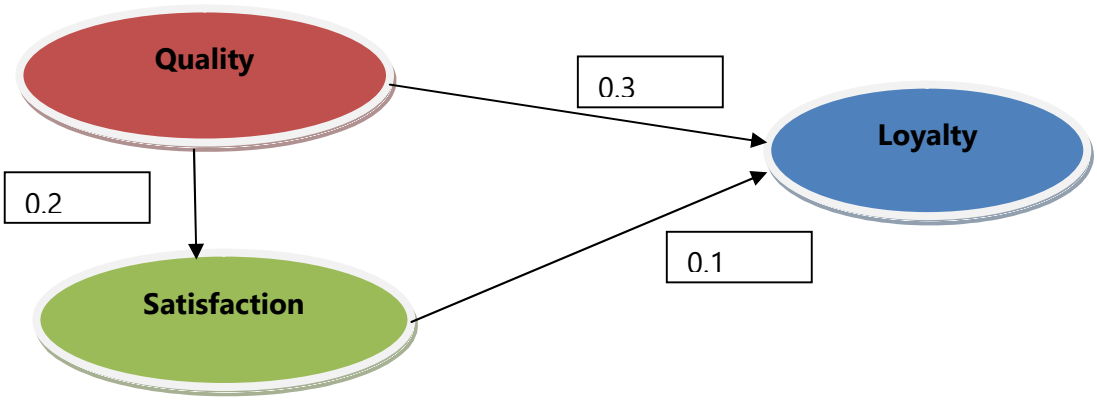


Fig 1.19: Cause and effect diagram of measuring loyalty

The above diagram can also be represented in another form known as “fish bone diagram”. Fishbone diagrams are so-called because they resemble fishbones as shown in Fig. They are also called "Ishikawa diagrams", named after Kaoru Ishikawa of Japan. A fishbone diagram is a graphical representation of the different factors that contribute to an effect.

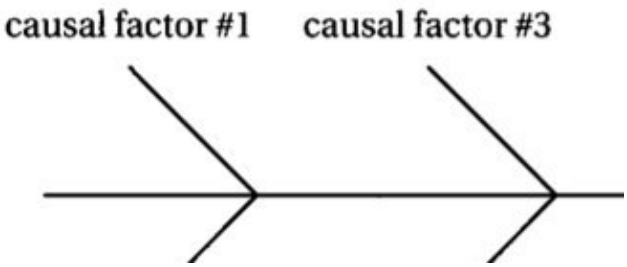


Fig 1.20: Representation of fish bone diagram

Using fishbone diagram, the loyalty of a customer can be studied as given in Figure below which depicts the four factors image, quality, trust and satisfaction as the causes impacting the loyalty of the customer.

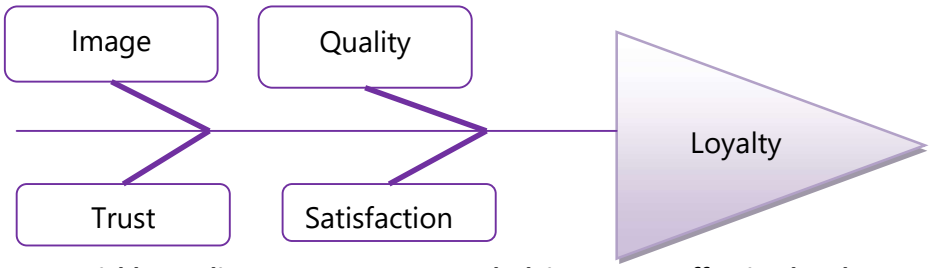


Fig 1.21: Fishbone diagram to measure underlying causes affecting loyalty

To do activity

The application of modern agricultural techniques among the population is at 25% and the aim of your program/project is for 75% of the population to use modern techniques (leaving you with a gap or need of 50%). With the help of a fishbone diagram, identify the categories of the causes.

1.5 Research Ethics

Research ethics is a field of research that systematically analyzes the legal and ethical questions raised in a research, which especially involves human subjects. It particularly aims in protecting the participants involved in the research. The fundamental principles of ethical research are owning responsibility, integrity, consensus, conscience and honesty, whilst respecting the participants autonomy , preserving privacy and confidentiality of the participants of research.

Fig 1.22: Principles of Research Ethics

Responsibility: in research is to clearly identify whether the research will benefit the society, especially in cases where it involves the public money.

Honesty: Dishonesty in research ranges from minor departure from good manners to more severe dishonesty and even criminal behaviours.

Privacy: Every individual has a right to privacy when participating in research, especially when it comes to healthcare research. One must ensure security of medical records and physical privacy. However, the privacy of patients in rural areas is more difficult to protect due to the characteristics of rural communities and lack of healthcare education. Protecting privacy and confidentiality of patient's data in a rural environment is especially important because majority of rural population is interested in knowing everybody else's job.

Confidentiality: Refers to a condition in which the researcher knows the identity of a research subject, but ensures steps are taken to protect the same.

Autonomy: The term autonomous means that a person can make his or her own decisions about what to do and what to agree to. An ethically done research must respect the decision of every individual to make an informed decision about whether to participate in research or deny.

Beneficence and Non-maleficence: Benefit involves balancing benefits of a research as against the associated risks and costs, whereas non-maliciousness means avoiding harm caused due to conduct or

non-conduct of research in the appropriate way.
Consensus and Conscience: Disagreements on any ethical research is a necessary part of ethics analysis. Confronting and reacting to counter-arguments makes a decision making stronger. Good reasoning is based on reliable facts and is backed by consideration for different values, the rating of competing values, and/or the least violation of key values. Identifying potential sources of disagreement is critical. The most difficult issue to address is the conflict over how to reconcile different views [William A. Nelson]. A thorough and cautious ethical decision-making process usually leads to consensus which results in conscience in fair practices of ethics, there may still be significant disagreement which is unavoidable.
Dignity: It is the right of a person to be valued and respected for their own sake.
Informed consent: means that a person knowingly, voluntarily and intelligently, and in a clear and manifest way, gives his consent to participate in a treatment/ research study.

Case Study to Understand Principles of Research Ethics in a Rural Environment

The case studies given below help us to understand the definitions of principles of ethics in the context of a rural environment.

Case Study1: Practicing Ethics in Rural Healthcare
Dr. X is a primary care physician who has been taking care of the ABC family since coming to a rural community five years ago. Mr.ABC is a farmer who presents to Dr.X with symptoms of coronary artery disease. Mr.ABC doesn't want to go to a multi-specialty hospital in the city for further assessment and tests, and without savings or medical insurance, Mr.ABC does not want to 'ruin the happiness of his family' and possibly leave his family destitute to pay for medical care, when he may die anyway. Mr.ABC wants Dr. X to keep the information only between the two of them.Dr. X's professional values include promoting Mr. ABC's well-being by offering him the standard of care in pursuing further evaluation and treatment of his suspected coronary artery disease (**Beneficence and Non-maleficence**).

Informed consent is the major ethical issue in conducting research. According to Armiger, "it means that a person gives his consent knowingly, willingly, intelligently, and in a clear manner".Other values include truth-telling **(honesty)—**making sure that Mr. ABC has enough information to make a reasoned decision to refuse further testing—and not lying to Mr. ABC's family if he is asked direct questions. It may also be a form of being dishonest if Dr.X accepts Mr. ABC telling his family something that is not true, and Dr. X does not correct the misinformation.**Trust** is at stake for the doctor's relationship with Mr. ABC, but also for his relationship with the farmer's family—who are also his patients.

Dr. X wants to respect Mr. ABC's decisions (respect his **autonomy**), but not at the expense of being untruthful and unfair to his family. Mr. ABC will need a lot of support during his illness, if he seeks treatment, but especially if he doesn't. It is not guaranteed that Mr. ABC will die suddenly of a heart attack. Is that what the farmer is assuming? Dr. X should discuss alternative futures with Mr. ABC so he clearly understands what's at stake by refusing further testing.
He may have a gradual decline, and if so, he may need a great deal of support. Is it fair to his family to keep them uninformed and unable to prepare to help Mr. ABC?

The farmer is sparing his wife a decision that he thinks she shouldn't have to make. But is that fair to

her? The trust of this rural community is also at stake. If people believe that Dr. X failed to diagnose Mr. ABC's problems, will they still trust Dr. X to provide their care? Wherein the dignity of the Doctor is also at stake. He is treating Mr. ABC, even though the farmer doesn't have insurance, and the doctor wants to offer the same level of care to Mr. ABC as to his other patients who do have insurance. The doctor has compassion for Mr. ABC and his dilemma—he wants to be able to get Mr. ABC any needed services without sacrificing the family farm. We can assume that Mr. ABC's wife values her husband's health and well-being. She is the one encouraging her husband to see Dr. X. We don't know anything about how this family makes important decisions, but it doesn't seem like shared decision-making and openness are priorities—at least for Mr. ABC, since he is asking to make this decision alone and wants to keep the information from his wife. But learning how this family usually makes tough decisions **(consensus which results in conscience)** would be important information to aid the professional in solving the ethics issues.

The conflict arises between the clinician's **responsibility** for truth-telling among the family, and promoting the well-being of Mr. ABC by securing further tests and possible treatment—versus respecting Mr. ABC's desire to make this decision alone **(autonomy/privacy/confidentiality)** and to not spend money on tests (to protect the family's financial well-being).

What could a provider do to deal with the issue of ethics; what are the options?
1. Dr. X could do as Mr. ABC requests, but with one limitation: he could tell Mr. ABC that if his family asks, he will direct them to talk with Mr. ABC. Dr. X will not lie to Mr. ABC's family.
2. The doctor could call Mr. ABC's wife and make her to help convince Mr. ABC to pursue the tests. 3. Dr. X could try to try to persuade Mr. ABC to allow a discussion about options between the doctor, Mr. ABC and Mr. ABC's wife.
4. Dr. X could try to coerce Mr. ABC into getting further tests by means other than taking help from someone other than Mr. ABC's wife **(consensus which results in conscience).**
5. Dr. X could discharge Mr. ABC from his practice, if the patient is not willing to get more tests.
How could this ethics issue have been prevented? One possible prevention strategy is for the provider to discuss the scope and limitations of confidentiality with all patients, including among family members. It might be easier for Dr. X to keep Mr. ABC's wishes private if he had previously had a conversation with Mr. ABC's wife, and she had understood that sometimes Dr. X would not share things with her about Mr. ABC's care.

Given the above options, Dr.X chose the 3rd option to persuade Mr.ABC to discuss the issue with his wife and convinced him to participate in the treatment procedures. Later, an **informed consent** was obtained from Mr.ABC before he was registered for further treatment of his coronary artery disease.

Case Study 2: Closure ofa Healthcare Centre ina Village
Some rural healthcare centres are aging and medical administrators made calls to close some of the very old healthcare centres. It was decided to close acute functioning, yet old healthcare centre at location A to location B, which was very far away from the village and close to a town. The administrators did a very primitive, unorganized and haphazard research on closing the existing medical centre and to build a

new hospital at location B.

Initially the physicians from the village had to travel a long distance due to which they quit. Later, the hospital decided to hire new physicians to cover the hospital atleast four days a week, but there was a great resistance to travel and there was acute shortage of doctors to run the new hospital. The earlier medical centre was also a place where physicians gathered, had coffee and communicated information about the individuals under their care. In the new hospital, there was no scope for such discussions. On the other hand, patients and families also had resistance to travel to the new facility for treatments. Transportation was also a bottleneck for the villagers to travel to the new hospital. The new doctors appointed in the new hospital did not know the patient's history and there was not much a bond for the patients to visit the new hospital for their medical care. Although the travel time for the people in town was less than 20 minutes it was seen as a normal occurrence but for the villagers this situation is less than ideal and many patients ended up dying due to progressed diseases and untimely treatment of diseases.

The decision-making process here was not supported by a consensus of the villagers as a result of which the decision has ignored some of the highest values held by these rural residents and the worst of all, the village residents must go to a strange place to die during their emergency medical crisis. Beneficence of establishing a new hospital and non-maleficence of closing the old healthcare centre was studied. This reality is even more emotional when many villagers had generously donated to establish a healthcare centre in their own community over their lifetime and had the expectation that the care would be there when they needed it. With very huge investment of public funds, the new hospital did not turn out to be useful. This is a typical case of irresponsible investing and irresponsible decision making.

Summary of the Chapter

Summary: The chapter presents a detailed overview of the process of research and throws light on the linearity of the research process. It presents a broad description of the various research methodologies and techniques. The research characteristics and features have been explained in depth and each unit is supplemented with appropriate case studies to gain thorough understanding of the same. The chapter concludes with a brief introduction to ethics in research and how essential it is for the conduct of good research.

Model Questions

1. What is the genesis of the word research?
2. How will you justify that listening to news every day is a research process?
3. Though research is a linear process, there's some cyclic nature to it. Do you agree with the statement? Justify.
4. How significant is the literature review in the entire research process?
5. Differentiate between Quantitative and Qualitative data analysis.
6. I have included Surveys, Polls and focus groups in my research. What type of research am I involved in? What are certain key characteristics of this kind of research?

7. In your opinion which characteristics are the most important for a research process?
8. Define research in your own words.
9. Recall the eight characteristics of research.
10. Ethics and research have a causal relationship. Do you agree with this statement? Explain your viewpoint.

To-Do- Activity

1. In your peer group, discuss on the various issues which can be treated as unethical in research.
2. Identify a prevalent problem in your area and try to formulate how a research definition can be created out of your problem. Classify the problem into its appropriate class and also mention the steps that you will take to do the literature survey to describe the problem in a statistical framework.

References

1. Ali, S. A., Khan, U., Abrejo, F., Saleem, S., Hambidge, M. K., Krebs, N. F., Pasha, O. (2019), Challenges of Implementing an Individual Randomized Controlled Trial (Women First: Preconception Maternal Nutrition Study) in a Rural Study Site: A Case Study From Pakistan. Nutrition and Metabolic Insights. https://doi.org/10.1177/1178638819852059
2. Chythra R. Rao, Shashi J. Soans, Zari Anjum, Muthu Kumar, Asha Kamath, Avinash Shetty, (2017), Are rural adults happy? An observational study of happiness and its determinants, International Journal of Community Medicine and Public Health, Int J Community Med Public Health, 4(8):2810-2815.
3. Rivadeneira, M.F., Moncayo, A.L., Tello, B. (2019), Matern Child Health Jhttps://doi.org/10.1007/s10995-019-02837-x
4. Anil Kalotra,(2013)Rural Marketing Potential in India – An Analytical Study, Volume 3, Issue 1, 1-8.
5. William A. Nelson, 2009, Handbook for Rural Health Care Ethics: A Practical Guide for Professionals Dartmouth College Press Published by University Press of New England.
6. Zikmund, W.G., Babin, J., Carr, J. & Griffin, M. (2012) “Business Research Methods: with Qualtrics Printed Access Card” Cengage Learning.

Chapter 2 Statistics as a Tool for Research

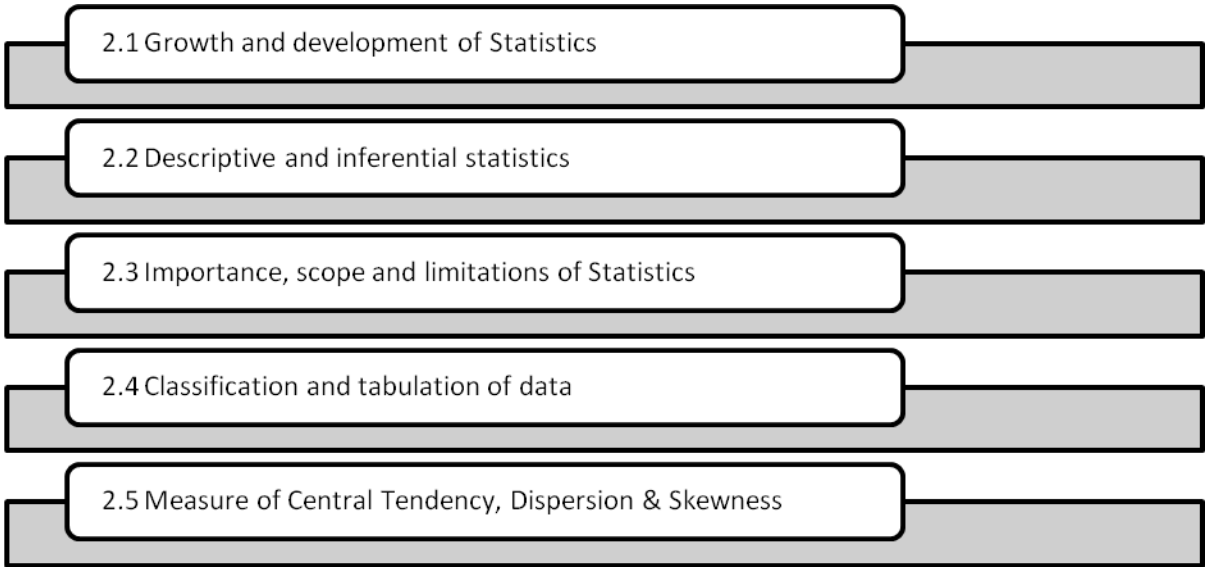
Introduction

Research and statistics go hand in hand. As mentioned in Chapter 1, the main aim of research is to find solution to an existing problem or develop new solutions for the problems which have been previously addresses. The knowledge of statistics is a great boon to the researcher as it enables them to uncover the huge amount of information available in the raw data collected. The application of right statistical tools to the data can give extremely beneficial insights into the data and facilitate the researcher in deducing his conclusions.

Objectives of the Chapter

- To familiarize the students with the history of the development of statistics
- To illustrate the current scope of statistics
- To differentiate between Differential and Inferential Statistics
- To discuss techniques for classifying data
- To familiarize the student with the techniques for tabulating data
- To describe Measures of Central Tendency, Dispersion and Skewness

Structure of the Chapter



2.1 Growth and Development of Statistics

The percolation of statistics in the good books of people can be attributed to the happy marriage of statistics and computer science. The world of statistics and software engineering have impacted and merged together due to the fact that statistics has moved onto the realm of electronic gadgets through programming. Programming Languages like R and Python rank as the quickest developing and most utilized programming languages over the most recent 5 years. The utilization of R has developed especially in academic circles for computing using statistics and now a days has become the very much searched out aptitude and capability by numerous businesses particularly for the individuals who are seeking after professions in STEM. Statistical tests have made some amazing progress since the start and bridling the force and utility of computers will just observe it advance and impact others all the more quickly and productively. The Roman Empire began to give too much of importance to data collection. It was the first government to collect widespread data about the wealth of the territories that it

controlled, population and area. During the Middle Ages in Europe few detailed censuses were made. The Carolingian kings Pepin-the-Short ordered surveys of ecclesiastical holdings in 758 whereas the similar kind of surveys were ordered by and Charlemagne ordered in 762. William I, king of England, ordered a census to be conducted post the Norman Conquest of England in 1066. This census was further completed in 1086 and the information collected was recorded in the DOMESDAY BOOK.

Another school of thought believes that the origin of statistics dates back to 1662, with the publication of Natural and Political Observations upon the Bills of Mortality by John Graunt. Early utilizations of statistical thinking rotated around the necessities of states to put together arrangement with respect to demographic and monetary information, thus its detail derivation. Its significance expanded in the mid nineteenth century to embrace the gathering and examination of data. Today statistics is broadly utilized in social sciences, government and business.

Due to the empirical roots of statistics and its focus on applications, it is generally considered to be a distinct mathematical science rather than a branch of mathematics. The development of probability theory by Blaise Pascal and Pierre de Fermat laid the mathematical foundations of statistics in the 17th century. The method of least squares was initially defined by Carl Friedrich Gauss around 1794.The utilization of present-day computers and their tremendous computation power has assisted in huge statistical calculation, and has likewise made conceivable new strategies that are unrealistic to perform physically

The coining of the term “Statistics” is accredited to G. Achenwall however, statistics, in the modern sense of the word, came into existence with the publication in 1761 by J. P. Sussmilch, through a published research paper titled “Die glittlicheOrdnung in den Veranderungen des menschlichenGeschlechtsaus der Geburt, dem Tode, and der Fortpflanzungdesselbenerwiesen”. The book attempted to present a systematic approach to use a class of facts which up to that time had been regarded as belonging to "political arithmetic," under which depiction probably the most significant issues of what current scholars term "vital statistics" had been examined, particularly in England.

The scientific techniques for statistics rose up out of probability hypothesis, which can be dated to 1654 to the correspondence of Pierre de Fermat and Blaise Pascal. In 1657 Christiaan Huygens gave the initial known logical treatment of the subject. After his death in 1713, Jakob Bernoulli's Ars Conjectandi and Abraham de Moivre's Doctrine of Chances in 1718, regarded the subject as a part of mathematics. In the present times, the work by Kolmogorov has been instrumental in defining the crucial model of Probability Theory, which is utilized all through statistics.

The theory of errors might be followed back to 1722(after his demise) to Roger Cotes' Opera Miscellanea. However, a diary arranged by Thomas Simpson in 1755 which was printed in 1756 used the theory for application to the deliberation on errors of observation. In 1757, the republished version of this journal sets out the maxims that positive and negative mistakes are similarly plausible or equally probable, and that there are sure assignable cutoff points inside which all errors should fall; consistent errors are talked about and a probability distribution curve is given.

In 1774, Pierre-Simon Laplace made the primary endeavor to reason a standard for the combination of observations from the principles of the theory of probabilities. He spoke about the law of probability of errors through a curve. He found a recipe for the mean of three perceptions. He likewise in 1781 gave a recipe for the law of facility of error (a term coined by Joseph Louis Lagrange in 1774), yet one which prompted unmanageable equations. In 1778 Daniel Bernoulli presented the principle of the maximum product of the probabilities of a system of concurrent errors.

The strategy for least squares, which was utilized to minimize errors in data measurement, was

discussed in the research publication independently published by Adrien-Marie Legendre in1805, Robert Adrain in 1808, and Carl Friedrich Gauss in 1809. Gauss had utilized the technique in his renowned 1801 to forecast of the location of Ceres- the dwarf planet. Further evidences were provided by Laplace in 1810 and 1812, Gauss in 1823, Ivory in 1825 and 1826, Hagen in 1837, Bessel in 1838, Donkin in 1844 and 1856, Herschel in 1850, Crofton in 1870, and Thiele in 1880 and 1889.

Further contribution was done by Ellis in 1844, De Morgan in 1864, Glaisher in 1872, and Giovanni Schiaparelli in 1875. The formula for r which gives the "probable error" of a single observation given by Peter in 1856 was widely used and encouraged initial robust statistics which was resistant to outliers.
From 1839 through 1914, Charles S. Peirce articulated frequentist theories of estimation and hypothesis-testing in which the concept of "confidence" was introduced. He also presented blinded, controlled randomized experiments with a recurrent measures design and invented an optimal design for experiments on gravity.

At present, statistics is a dependable method for depicting precisely the estimations of financial, political, social, mental, natural, and physical data and fills in as an instrument to connect(correlate) and break-down(analyze) such data. The role of a statistician is not anymore restricted to collecting and classifying data, however is essentially a procedure of deciphering the data and gaining the valuable insights into it. The development of the theory of PROBABILITY expanded the extent of statistical applications. A large amount of available data can be approximated precisely by certain probability distributions, and the consequences of these can be utilized in analyzing and deducing inferences from the statistical data. Probability can be used to test the reliability of statistical inferences and to indicate the kind and amount of data required for a particular problem.

2.2 Descriptive and Inferential Statistics

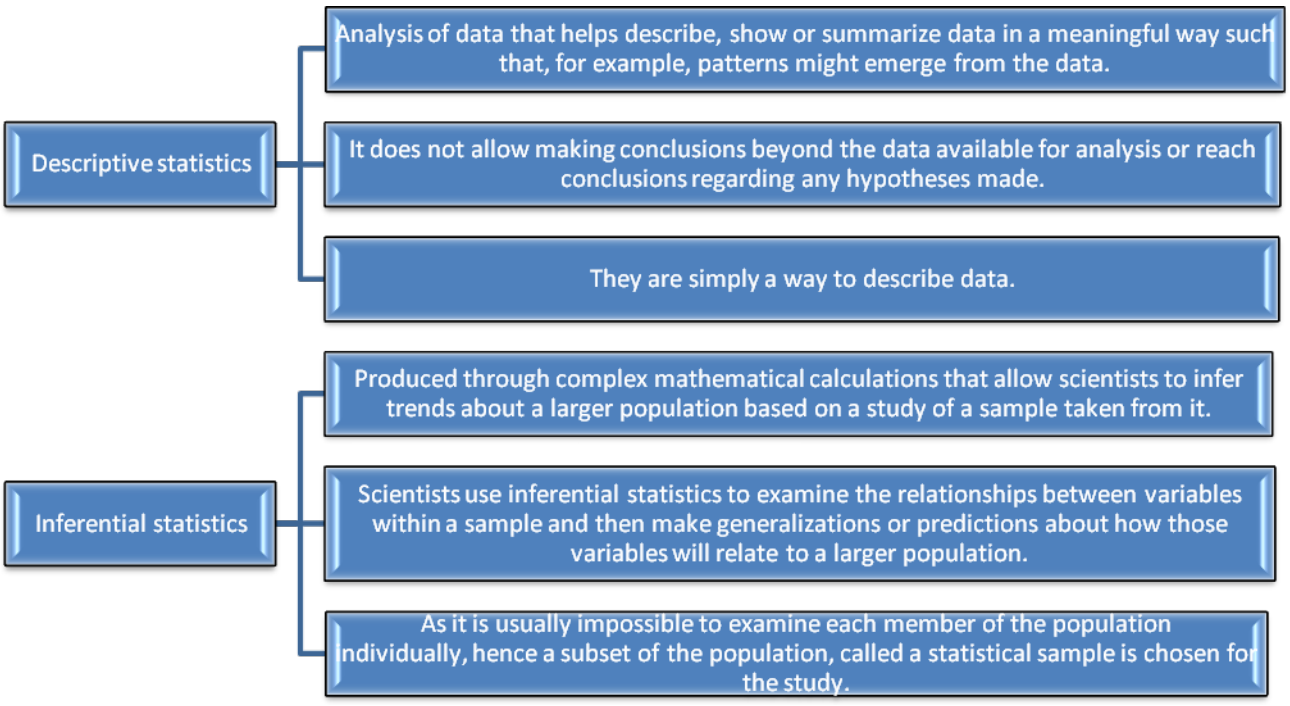


Fig 2.1: Two General Types of Statistics used to Describe Data

Prior to collection of data for any statistical study, a population should initially be characterized. 'Population' refers to a sub-group that has been assigned for collecting data from. The information collected from this population becomes the data as it will require further processing. A population does not essentially mean people; it could be a group of individuals, estimations of precipitation in a specific

region or a record of student's achievements. During data analysis, for example, the scores obtained by 100 students for an assignment, it is possible to use both descriptive and inferential statistics in the analysis of the scores. Normally, in most research directed on groups of individuals, you will use descriptive as well as inferential statistics to analyze outcomes and derive conclusions.

Descriptive Statistics

The essence of descriptive statistics can be understood from the fact that simple presentation of raw data will not allow visualization of the insights hidden in that data, particularly if there was a great deal of it. Descriptive statistics in this manner empowers us to show the present the data in an increasingly important manner, which permits less difficult understanding of the information. For instance, in the example of scores of students described above, we might be keen on the overall scores or performance of the students. We would likewise be keen on the distribution or spread of their scores. Descriptive statistics enable us to do this. Instructions to appropriately depict information through measurements and charts are a significant subject and talked about in many reference literatures.

Descriptive statistics helps to provide information that portrays and projects the hidden insights in the data in some way. For instance, assume a grocery shop sells rice, dal and potatoes. On the off chance that 100 kgs worth goods were sold and 40 kg out of the 100 kg was rice, at that point one portrayal of the information on the goods sold would be that 40% of the goods sold was rice. This same pet shop may conduct a study on the quantity of potatoes sold every day for one month and discover that a normal of 10kg potatoes were sold every day. The average calculated is a typical example of descriptive statistics

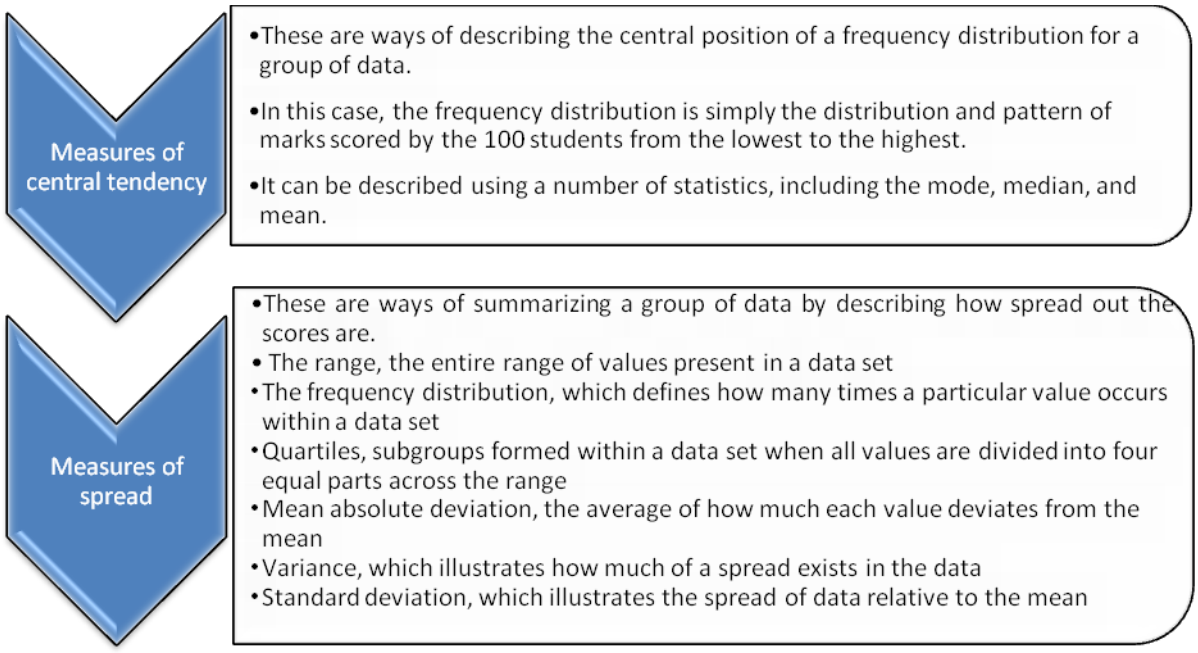


Fig 2.2: Two kinds of descriptive statistics

At the point when we utilize descriptive statistics, it is helpful to use a combination of tools such as tabulation (i.e. tables), statistical commentary (i.e., an analysis and summarization of the results), graphical depiction (i.e., charts and graphs) to summarize the data.

Inferential Statistics

Descriptive statistics is a means for obtaining information about our immediate group of data, whereas, the numerical techniques whereby we convert data about the example into intelligent speculations about the population fall under the purview of inferential statistics.

There are two major divisions of inferential statistics:

- Tests of significance or hypothesis testing metric is used where researchers analyze the population and make a claim about it. There is some amount of inherent vulnerability in this procedure which can be stated in terms of a level of significance.
- Confidence interval is a metric which gives a range of values by measuring the statistical sample of the population. Within the interval that a specific parameter is existing, confidence measure is expressed in terms of degree and interval.

Techniques that social scientists use to examine the relationships between variables, and thereby to create inferential statistics, include linear regression analyses, logistic regression analyses, ANOVA, correlation analyses, structural equation modeling, and survival analysis. When conducting research using inferential statistics, scientists conduct a test of significance to determine whether they can generalize their results to a larger population. Common significance measurement tests include the t-test and the chi-square test. These tell scientists the probability that the results of their analysis of the sample are representative of the population as a whole.

Continuing with the example previously described of the 100 students, we could calculate the mean and standard deviation of the scores obtained by the 100 students and this could provide meaningful insights about the performance of this population. The size of the population may vary from small to large, as long as it is inclusive of the data that is of interest in the present scenario. For example, in the current context if the area of concern or interest is the scores of 100 students, the 100 students would represent your population. Descriptive statistics are generally applied to populations and the properties of populations (like the mean or standard deviation) are referred to as parameters due to their representation of the entire population.

More often than not, you don't approach the entire population you are keen on researching, however only a limited chunk of it. For instance, you may be interested in the scores obtained by students all across Tamil Nadu. It is impractical to measure all exam scores of all the students across entire Tamil Nadu, so you need to quantify a relatively smaller sample of students (e.g., 100 students), which can be used to represent the larger population. Properties of samples, for instance, the mean or standard deviation, in this case are not called parameters but statistics. Inferential statistics are strategies that enable us to utilize these examples to make generalized conclusions about the population from which these samples were taken. It is, accordingly, significant that the sample is a precise representation of the population. The way toward accomplishing this is called sampling. Inferential statistics emerge out of the way that sampling normally causes errors and along these lines an example isn't relied upon to flawlessly represent the population. The strategies for inferential measurements are

- The testing of statistical hypotheses
- The estimation of parameter(s)

Table 2.1 Comparison between Descriptive and Inferential Statistics

BASISFORCOMPARISON	DESCRIPTIVE STATISTICS	INFERENTIAL STATISTICS
Main Focus	Description of population under study.	Drawing conclusions pertaining to the population, on the basis of sample analysis and observation.
Purpose	Organizing, analyzing and presenting data in a more meaningful way.	Comparing, testing and predicting data.
Outcome	Charts, Graphs and Tables	Probability/ Probability Distribution Functions
Usage	Description of a situation.	Explanation of the chances of occurrence of an event.
Utility	Explains the known characteristics of the data, and summarizes the sample.	Uncover the hidden patterns in the data which extend beyond the data available.

2.3 Importance, Scope and Limitations of Statistics

Importance of Statistics in the Field of Research

The field of statistics is the study of gaining from information. Statistical knowledge causes you utilize the best possible techniques to gather the information, utilize the right examinations, and adequately present the outcomes. It is a critical procedure behind how we make revelations in science, settle on choices dependent on information, and make forecasts. Insights enables you to comprehend a subject considerably more profoundly.

Majority of the research studies result in a huge volume of raw data which must be appropriately reduced with the goal that the equivalent can be perused effectively and can be utilized for additional examination. Plainly the study of statistics can't be disregarded by any researcher, despite the fact that he might not have any opportunity to utilize statistical methods in the entirety of their details and repercussions in his field of study. Classification and Tabulation, as expressed prior, accomplish this goal somewhat, however we need to go above and beyond and build up certain lists or measures to summarize the gathered/characterized data. Simply after this we can embrace the procedure of generalization from test data to the entire populace.

There are two principle reasons why concentrating on the field of measurements is vital in current society. To begin with, statisticians are guides for gaining from information and exploring regular issues that can lead you to inappropriate conclusions. Second, given the increasing essence of choices and sentiments which are information-dependent, it becomes imperative that you fundamentally evaluate the nature of analysis that others present to you.

Scope of Statistics in the Field of Research

Statistics is an energizing field about the rush of disclosure, learning, and testing your hypothesis and assumptions. A tiny bit at a time, we push back the outskirts of what is known. The following examples illustrate the scope of statistics in our everyday life and how it is becoming critical for simple decisions as well as complex issues.

- Everyone watches weather prediction or forecasting. Have you ever figured out how would you get that data? There are a few computer models built on these statistical ideas. These models draw a contrast between the earlier weather and the present weather and foresee future climate.
- Researchers generally utilize their statistical skills to gather the significant data and apply statistical techniques on the same. If it is not done this way, it results in a lot of loss of time, data as well as money.
- What do you comprehend by insurance? Everyone has some sort of insurance, regardless of whether it is medical, home or some other kind of insurance. In view of an individual application a few organizations utilize statistical models to predict the risk factor of providing/ selling insurance to the customers.
- Statistics have a crucial and critical role to play in the financial markets as well. Statistic techniques provide valuable insights and are the key of how merchants and businessmen invest and make profit.
- Statistical ideas are utilized in quality testing. Organizations make numerous items consistently and each organization should ensure that they sold the best quality things. Yet, organizations can't test every one of the items, so they use statistical test.
- A lot of prediction goes into our everyday lives. For example, irrespective of the fact that whether a person will wake up in the morning or not, he/she sets an alarm every day before going off to bed. Here we use the basic concept of statistics to make a prediction.
- Disease prediction is done by the doctors based on the concepts of statistics. Assume a study shows that 75%-80% individuals have malignant growth and they are not able pin point the reason for the same. At the point when the statistics gets included, the patient can have a better understanding of how the disease may influence his/her body or is smoking the significant reason behind it.
- Journalists make prediction of election outcomes and the winners depending on the political campaigning done by various parties. Here statistics have a solid influence on who will be our administrators.
- Statistics information enables us to gather the data around the globe. The web is a devise which help us to gather the data. The underlying concept behind the web depends on statistics and arithmetic ideas.

Having mentioned these day-to-day examples, it is imperative to mentions that statistics has slipped into the lives of people. This has happened due to the amalgamation of the branches of computer science and statistics. These two fields have impacted and merged together as the statisticians have moved onto electronic gadgets through programming. Programming Languages like R and Python rank as probably the quickest developing and most utilized programming languages over the most recent 5 years. The utilization of R has developed especially in academic circles for performing statistical computations. R and Python now stands as an all-around searched out skills and capability in them is currently an important requirement by majority of the employers particularly for the individuals who are pursuing careers in STEM. Statistical tests have progressed significantly since the start and outfitting the force and utility of PCs will just observe it advance and impact others all the more quickly and proficiently.

2.4 Classification and Tabulation of data

Collection of Data

Collection of Data is the most essential step in statistical investigation. If the data are inaccurate and inadequate, the whole study purpose becomes wrong. Along these lines, care must be taken while gathering the necessary data.

There are two types of data, namely:

- **Primary Data:** Data, which are collected for the first time, are called primary data.
- **Secondary data:** Data, which are already collected, are called secondary data to others.

Methods to Collect Primary Data are as follows:

- **Direct Personal Interview:** According to this method, the investigator/interviewer collects the desired data by the direct personal contact with the people from whom data is required. So, there is a face-to-face contact with the interviewees.
- **Indirect Oral Interview:** According to this method, the investigator collects the required information/data from third person. This method is generally used by the enquiry committee and this method saves time, money and labor.
- **Information from Correspondents:** According to this method, the investigator appoints local agents or correspondents in different places to collect information. Such correspondents collect and transmit the information to the central office, where the data are processed. Newspaper, Television, and Radio Agencies generally adopt this method.
- **Mailed Questionnaire:** According to this method, a list of questions called questionnaire, is prepared and sent to the various informants. The questionnaire contains questions with space for answer. Informants are requested to answer the questions and send back within a specified period of time.
- **Schedule to be filled by Enumerators:** According to this method, the questionnaire is sent through trained enumerator. The enumerator goes to the informants with schedule and meets them personally. He explains the purpose of his visit and asks them questions contained in the schedule and fills the schedule with his own handwriting. This method is suitable in that case when the informants are illiterate.

Methods to Collect Secondary Data are as follows

Secondary Data can be collected mainly through many sources, which are as follows:

- Information collected through newspapers and periodicals
- Information obtained through published academic research papers
- Information obtained through the official publications of the central, state and the local governments dealing with crop statistics, industrial statistics, trade and transport statistics, etc.
- Information from the official publications of international organizations.

Case Study: Planning for Rural Employment

Problem statement: If a researcher working on the area of “Planning for rural employment” intends to study the usefulness of two major schemes: (i) Swarnjayanti Gram Swarozgar Yojana and (ii) Mahatma Gandhi National Rural Employment Guarantee Act run by the Government to remove unemployment and to provide employment and self-employment in the rural areas. The research intends to:

- To find out whether this scheme provides self-employment among rural labourers.
- To find out how many individuals got self-employment in some chosen district for a particular period under this scheme.
- To access the impact on/problems faced by the beneficiaries under this yojana.

Extensive data must be collected to answer the above research questions. The data collection methods used are elaborated as follows:

- After making the contact with respondent, firstly researcher has to develop good relationship with respondent to gather required **primary data** through **direct interviews,questionnaires** (from literate respondents) and through **schedules** (from illiterate respondents). Researcher collects important information from the beneficiaries/labourers through direct interview in the rural area of the district under survey.
- To complete the survey **secondary data** also play an important role. In survey, researcher use both published and unpublished data, like collection of data and information from the chosen District's Statistical Office, Irrigation Office, Employment Office, District Rural Development Agency, Village Panchayat, etc.

Classification of Data

Classification of Data is the process of arranging collected data in different groups or classes on the basis of common characteristics. Collected data are not suitable for immediate analysis because they are in jumbled and raw form. So, collected data are to be classified.

Objective of Classification

- To present the data in a comprehensible (easily understandable) form by removing unnecessary things.
- To arrange the scattered data in an organized form.
- To facilitate comparison.
- To make data fit for tabulation.

Types of Classification

There are 5 major types of classification of data:

- **Quantitative Classification:** When the basis of classification is according to differences in quantity, the classification is called quantitative. It is made on the basis of such characteristics, which can be measured, such as, height, weight, income, production, etc. Such a classification can be represented using simple frequency distribution of the measured attribute. There are two types of frequency distributions, namely discrete and continuous frequency distribution.
    - **Formation of a discrete frequency distribution**

      The process of preparing these types of distribution is very simple. We have just to count the number of times a particular value is repeated which is called the frequency of that class. In order to facilitate counting, prepare a column of "tallies'' and in another column, place all possible values of variable from the lowest to the highest. Then put a bar (vertical line) opposite the particular value to which it relates. To facilitate counting, blocks of five bars are prepared and some space is left in between each block. We finally count the number of bars and get frequency.
    - **Formation of a continuous frequency distribution**

      Continuous data are obtained through measurements, whilediscontinuous data are derived by counting. Measures which are represented by a continuous variable are called continuous series.

A simple formula to obtain the estimate of appropriate class interval, i.e. i is:
i=(L – S)/k
Where, L = largest item
S = smallest item
k = the number of classes

Example

In a survey of 35 families in a village, the number of children per family was recorded and the following data obtained

1	0	2	3	4	5	6
7	2	3	4	0	2	5
8	4	5	12	6	3	2
7	6	5	3	3	7	8
9	7	9	4	5	4	3

Represent the data in the form of a discrete frequency distribution.

Solution:

Table 2.2: Discrete Frequency distribution of the number of children*

No. of children	Tallies	Frequency
0	II	2
1	I	1
2	IIII	4
3	𝍸 I	6
4	𝍸 II	5
5	𝍸	5
6	III	3
7	IIII	4
8	II	2
9	II	2
10	-	0
11	-	0
12	I	1
Total		**35**

It is clear from the table that the number of children varied from 0 to 12. There were 2 families with no child, 5 families with 4 children and only one family with 12 children.

Example

In a survey of families in a village, the weight (Kg) of around 18 children (under the age of 10 years) was recorded and the following data obtained. Represent the data as continuous frequency distribution.

30	**25**	**35**	**27**	**22**	**36**
21	**18**	**38**	**24**	**25**	**21**
26	**31**	**33**	**24**	**19**	**28**

If we want to form 5 classes, then the class interval is calculated as follows:

L = 38 Kg (Largest of all weights recorded)

S = 18Kg (Smallest of all weights)

No of classes required, k=5

Hence, class interval, i= (L – S)/k = (38 – 18)/5 = 4.

Hence the continuous frequency distribution derived for the above given data is as given in Table below:

Table 2.3: Continuous frequency distribution table of weight

Weight of Children (Kg)	Frequency
18-22	4
22-26	5
26-30	3
30-34	3
34-38	3

- **Temporal Classification:** When the basis of classification is with reference to time, the classification is called temporal or chronological classification.
  Example: Classification of Rural population in Tamil Nadu (millions) based on 2001 & 2011 census (Source: Director of Census Operations, Tamil Nadu).

Table 2.4: Example of Temporal classification

Category	2001	2011
Total	34.92	37.23
Males	17.54	18.68
Females	17.39	18.55
Child [0-6 years]	4.20	3.91

- **Spatial/Geographical Classification:** According to geographical classification, data are classified on the basis of geographical location or division.
  Example: Ranking of top 10 states/ union territories of India based on the reports of annual reports of Reserve Bank of India, Government of India. 2013. The rank is calculated according to the percentage of people below poverty-line.

Table 2.5: Example of Spatial classification

Rank	State/U.T.	Poverty (% of people below poverty line)
1	Chhattisgarh	39.93
2	Dadra and Nagar Haveli	39.31
3	Jharkhand	36.96
4	Manipur	36.89
5	Arunachal Pradesh	34.67
6	Bihar	33.74
7	Odisha	32.59
8	Assam	31.98
9	Madhya Pradesh	31.65
10	Uttar Pradesh	29.43
All India Average		**21.92**

- **Qualitative Classification:** When the basis of classification is according to characteristics or attributes like social status, gender, marital status, religion, etc. which cannot be quantitatively measured.

Example: Classification of Indian population based on their religion (Source: 2011 Census) is as given below:

Table 2.6: Qualitative classification of Indian population based on religion

Religion	Percent	Estimated	State Majority
All Religion	100.00 %	121 Crores	35
Hindu	79.80 %	96.62 Crores	28
Muslim	14.23 %	17.22 Crores	2
Christian	2.30 %	2.78 Crores	4
Sikh	1.72 %	2.08 Crores	1
Buddhist	0.70 %	84.43 Lakhs	-
Jain	0.37 %	44.52 Lakhs	-
Other Religion	0.66 %	79.38 Lakhs	-
Not Stated	0.24 %	28.67 Lakhs	-

- **Classification according to number of attributes** is of two types, viz. Simple or two-fold classification or manifold classification.

Example: Two fold classification of Indian Population into urban and rural population as per 2017 statistics (data source: censusindia.gov.in) is classified as under:

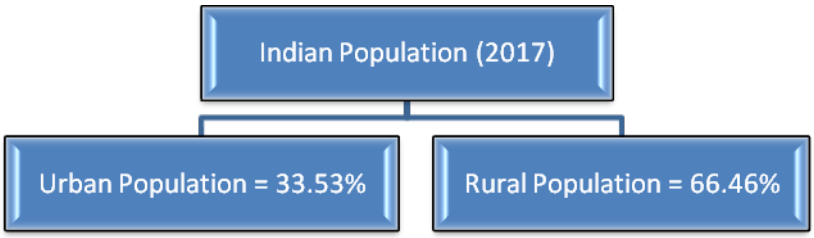


Fig2.3: Two-fold classification of Indian population

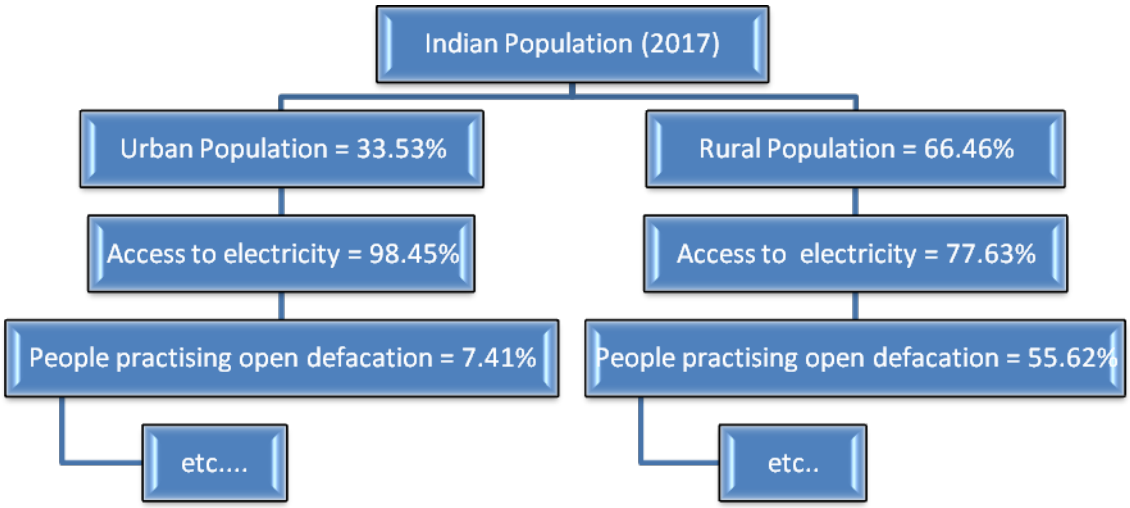


Fig 2.4: Multi-fold classification of Indian population

Tabulation of Data

Tabulation of Data is the systematic arrangement of statistical data in rows and columns. Rows are horizontal arrangement, whereas columns are vertical arrangement. Statistical data are tabulated after classification. The main purpose of tabulation is to simplify the presentation and to facilitate comparison.

Importance of Tabulation of Data

Tabulation of data helps to present a huge mass of data in a systematic manner within minimum space. The importance of tabulation can be explained under following grounds:

- **It simplifies complex data:** When data are tabulated, unnecessary details and repetitions are removed. Data are presented in table easily within a short period of time.
- **It facilitates comparison:** Tabulation facilitates comparison. A table is divided into various parts and for each part, there are total and sub-total. Data of different parts can be compared easily.
- **It gives identity to the data:** When data are arranged in the table with table number and title, they can be clearly identified. Such arranged data help to interpret the problems.
- **It gives patterns:** Tabulation reveals patterns within the figures, which cannot be seen in raw form of the data. It may also facilitate to sum up the figures if the reader desires to check the totals.

General Rules of Tabulation of Data

There is no hard and fast rule for tabulation of data. Experience and common senses are required to construct good table. Following points should be kept in mind while tabulating data:

- Size of table should be suitable with available size of paper.
- Captions and Stubs of table should be arranged in alphabetical order or in geographical order or in chronological order.
- Unit of measurement, such as, Price in Rs., Weight in kg,etc. should be given in the table.
- Figures should be written in **rounded form.**
- Table should not be overloaded with unnecessary details.
- A **'Miscellaneous Column'** should be added for those data, which do not fit in the classification.
- Items related to each other should be placed nearabout. **Total, Percentage, and Ratio** should be computed if necessary and they should be kept close to the tabulated data.
- Don't use **Zero** for the information, which is not available. Use **N.A.** (Not Available) or **– (Dash)** to indicate it. Use **'0'** to indicate zero value only.
- If any number is repeated, write again that number, but don't use **Ditto Mark (")** and avoid using **etc.** and **e.g.** in the table.
- Source should be given for secondary data.

Parts of Table

The major parts of a table are as follows:

- Table Number
- Title of Table
- Caption (Column Heading)
- Stub (Row Heading)
- Body (Numerical Information)
- Head Note (Enclosed in Brackets Like **'In lakh'**, **'In kg'**, etc.)
- Foot Note (To clarify anything of the Table)
- Source

Table 2.7: Population of Tamil Nadu(A Sample of Table)

Regions	No. of Population		Total
	Female*	Male*	
1. Terrain	-----	-----	-----
2. Sea shore	-----	-----	-----
3. Mountain	-----	-----	-----

* Both literate and illiterate Source: Central Bureau of Statistics (CBS)

Note : It is not necessary that all information will be available to frame the table. At times some data has to be derived from the given data (wherever possible) to present the tabulated data.

Example : Given a district is divided into two areas: urban area and rural area. Total population of the district is 351,076 out of which only 266,492 live in urban area. Total male population of the district is 139699 out of which 23083 live in rural area. Prepare a table showing the population of the district by residence and gender.

Table 2.8: Growth population in a district

Location	Male	Female	Total
Rural area	23083	61,501	84584
Urban area	116,616	149,876	266,492
Total	139699	211,377	351,076

- Step 1: Represent the given data in the table (shaded cell).
- Step 2: Wherever possible, derive the data to fill the data cell which is not given (white cell) and complete the table.

2.5 Measures of Central Tendency, Dispersion and Skewness

The Measures of Central tendency provides typical information about the set of values that gather around the middle of the set. Three common measures of central tendency are the arithmetic mean (mean), median, and mode.

Mean is the average of all values of the set.

Median is the middle value of ordered samples of numerical values.

Mode is the value that occurs most frequently. It is generally applied to the categorical data.

To better understand the above definitions, let us look at an example.

How do you choose which measure of central tendency to use?

Mean: Use when the data is fairly **close together**

Ungrouped Sample Mean : $\bar{x} = \frac{\sum x}{n}$, **where x is the sample value and n is the size of ungrouped sample population.**

We use $\bar{x}$ to denote sample mean

Grouped Sample Mean :

$$Mean(\bar{x}) = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{f_1 + f_2 + f_3 + \ldots + f_n}$$

where 'f' is the frequency of each group

Median: Useful when there is an **outlier** (extreme value that is **far away** from other values) which would **skew** the data.

The median of a set of ungrouped data (observations) is defined to be the middle value when the observations are sorted from lowest to highest.

If the number of observations is odd, then the median is the observation exactly in the middle of the data set.

Median = Middle data value

If the number of observations is even, then the median is the mean of the two middle observations in the ordered list.

Median = Sum of middle two values/2.

Median of grouped data,

Step 1: Draw the cumulative frequency polygon.

Step 2: The median is the datum corresponding to the middle value of the cumulative frequency.

Mode: Good when the **value** of the data is the most important information. Shows **consistency** and is the only choice with **categorical** type data.

The mode is the value that occurs most frequently in a data set.

For ungrouped data, mode is the datum that has the highest frequency.
Mode of grouped data, modal class is the class that has the highest frequency.
Find the modal class. The modal class is the class interval that has the largest frequency.
Find the lower class boundary of the modal class ($= Lb$)
Find the difference of frequency between the modal class to its upper class ($= a$).
Find the difference of frequency between the modal class to its lower class ($= b$).
Add the Lb to products $\frac{a}{a+b}$ by C, then add it to A.
Mode = $Mo = Lb_{Mo} + \frac{a}{a+b}.C$
Note:
There can be more than one mode.
When no value is repeated, we say there is no mode.

Example 1: The following are the income for four randomly selected small scale traders in a village. Here, the random selection of few traders is the sample from the set of all traders in the village, which is termed as population.

Sample representation	X1	X2	X3	X4
Annual Income of small scale traders	30,000	35,000	32,500	36,000

(a) Mean, $\bar{X}$ = (30,000 + 35,000+ 32,500+36,000)/4 = 33,500.

Hence, the mean income of the small scale traders is Rs.33,500.

(b) To find the median for the sample of four salaries of the small scale traders,

Step 1: First put the salaries in order:

To find median, sort the salaries	**30,000**	**32,500**	**35,000**	**36,000**

Step 2: Since the sample size n= 4 (Even), median is the mean of the two middle observations in the ordered list.

$$\text{Median} = \frac{32{,}500 + 35{,}000}{2} = \text{Rs. } 33{,}750$$

To find mode:
Step1 : Compute the frequency of occurrence of the values is listed as given below:

Annual Income of small scale traders	30,000	35,000	32,500	36,000
Frequency of occurrence of the value	1	1	1	1

Step 2 : Calculate mode based on the frequency of occurrence of the values.
Since, there is no repeated sample value, there is no mode for this problem.

Example 2: Find the mean from the set of data given below:

Class mark	10.5	30.5	50.5	70.5	90.5	110.5
Frequency	19	6	3	2	1	2

Solution:

x	*F*	*x f*
10.5	19	199.5
30.5	6	183
50.5	3	151.5
70.5	2	141
90.5	1	90.5
110.5	2	221
Sum	33	986.5

Mean = 986.5/33 = 29.89

Example 3: Find the **median class of grouped data** for the following table which shows the ages of 40 participants of the kabaddi game conducted in a village fest.Find the median of ages of the participants.

Age	11 – 20	21 – 30	31 – 40	41 – 50	51 – 60
Frequency	9	17	6	4	4

Solution:
Step 1: Construct the cumulative frequency distribution.

Age	11 – 20	21 – 30	31 – 40	41 – 50	51 – 60
Frequency	9	17	6	4	4
Cumulative Frequency	9	26	32	36	40

The median is the datum corresponding to the middle value of the cumulative frequency.
n/2 = 40/2 = 20. This implies the class median is the second class [21-30].

Example 4: Find the **mode of ungrouped data** for the following test scores of 6 students in the subject 'Management decision making' for rural studies?

Samples	X1	X2	X3	X4	X5	X6
Test Score	84	89	82	91	86	84
Frequency	1	1	1	1	1	2

The mode is 84, since it has the highest frequency of occurrence.
Example 5: Find the **mode of grouped data** for the following table which shows the ages of 40 participants of Kabaddi game in a village fest.Find the mode of ages of the participants.

Age	11 – 20	21 – 30	31 – 40	41 – 50	51 – 60
Frequency	9	17	6	4	4

Solution:

Step 1: Find the modal class.

The modal age class of the grouped participants is 21-30, since it has the highest frequency of occurrence. Class difference , C = 10.

Step 2: Find the lower class boundary of the modal class ($= Lb$)

Lb = 21

Step 3: Find the difference of frequency between the modal class to its upper class ($= a$).

a= 17-9 = 8

Step 4: Find the difference of frequency between the modal class to its lower class ($= b$).

b=17-6= 11.

Add the Lb to products $\frac{a}{a+b}$ by C, then add it to A.

Mode = $Mo = Lb_{Mo} + \frac{a}{a+b}.C$; Mo = 21 + $\frac{8}{8+11}$.10 = 21 + 4.21 = 25.21

Example 6 :

Consider the numbers: 2, 3, 4, 5, 5

(a) Compute the mode, median, and mean.
Answer: Mode = 5; median = 4; mean = 3.8.

(b) If the numbers represented codes for the colors of T-shirts ordered from a catalog, which average(s) would make sense?
Answer: Mode.

(c) If the numbers represented one-way mileages for trails to different lakes, which average(s) would make sense?
Answer: Mean, median, and mode.

(d) Suppose the numbers represent survey responses from 1 to 5, with 1 disagree strongly, 2 disagree, 3 agree, 4 agree strongly, and 5 agree very strongly. Which averages make sense?
Answer: Mode, median.

Measure of Dispersion

Measures of central tendency provide only a partial description of a quantitative data set. The description is incomplete without a measure of the variability, or spread, of the data set. The **measure of dispersion** helps us to study the variability of the items.

Three common measures of variability are the range, variance, and standard deviation.

The range of a data set is equal to the largest measurement minus the smallest measurement.

The sample variance for a sample of n **ungrouped** measurements is equal to the sum of the squared distances from the mean divided by (n - 1). In symbols, using s^2 to represent the sample variance

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

In calculating s^2, we divide by (n – 1) instead of n to get a better estimate of the population variance.

The sample standard deviation, s, is defined as the positive square root of the sample variance, s^2.

$$s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

The sample variance for a sample of n **grouped** measurements is given as.

$$Var=\frac{\sum f_i.(x_i-\bar{x})^2}{\sum f_i},$$

where $\bar{x}=mean$, f_i = frequency of 'i' th interval, x_i is the midpoint of the 'i'th class. The standard deviation σ for the grouped data is given as below:

$$\sigma=\sqrt{Var}=\sqrt{\frac{\sum f_i.(x_i-\bar{x})^2}{\sum f_i}}$$

Note:

- Variance and Standard deviation are useful for comparing variability in two data sets. The data set with the larger variance, or standard deviation, exhibits more variation in the data.
- Standard deviation has the advantage over variance in that it provides a measure of variability in the same units as the original data.

Example 7: Consider the following three data sets, which contain tests scores for students in three different classes. Calculate :

(a) The range for class 1 scores.

(b) The Sample Variance, s^2 and Std. Deviation, s for the class 1 scores

	Test Scores						Mean	Median
C1	50	60	70	80	90	100	75	75
C2	50	73	74	76	77	100	75	75
C3	70	73	74	76	77	80	75	75

Solution:

The three data sets all have the same center but where they differ is in terms of variation in the scores.

(a) The range for class 1 scores.

Range = 100 – 50 = 50

(b) Calculation of s^2 and s for the class 1 scores

x	$x-\bar{x}$	$(x-\bar{x})^2$
50	-25	625
60	-15	225
70	-5	25
80	5	25
90	15	225
100	25	625

	$\Sigma(x-\bar{x})=0$	$\Sigma(x-\bar{x})^2=1{,}750$

$$s^2=\frac{\Sigma(x-\bar{x})^2}{n-1}=\frac{1750}{5}=350\,;\ s=\sqrt{350}=18.708$$

Data Transformation and Its Impact on the Variability in the Measures of Central Tendency:

In this section, we explore the effect on the standard deviation of changing each data value in a data set by the same constant. In general, multiplying each data value by the same constant c results in the standard deviation being |c| times large.

To do activity

Survey high school students in your locality and their parents to find out their heights.

(i) Calculate the mean and standard deviation of height of the group.

(ii) Calculate the coefficient of variation of the teenagers

(iii) Calculate the coefficient of variation of their parents

(iv) Which group (teenagers/parent) shows higher variability.

Example 8: Consider the data set 5, 9, 10, 11, 15 to be the distance travelled (Km) by a farmer in his bicycle to go to visit his different farms.

a) Use the computation formula, compute s.

b) Given, the farmer's son travels 5 times the distance travelled by his father to visit another farmland , the new data set is obtained as 25, 45, 50, 55, 75. Compute s.

c) Compare the results of parts (a) and (b). In general, how does the standard deviation change if each data value is multiplied by a constant c?

x	$x-\bar{x}$	$(x-\bar{x})^2$
5	-5	25
9	-1	1
10	0	0
11	1	1
15	5	25

(a) $s^2=\frac{\Sigma(x-\bar{x})^2}{n-1}$ = 52/4 = 13, s = 3.6055

(b) When each data value is multiplied by 5, calculate the standard deviation.

x	$x-\bar{x}$	$(x-\bar{x})^2$
25	-25	625
45	-5	25
50	0	0

55	5	25
75	25	625

$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$ = 1300/4 = 325, s = 18.02

(c) SD of set b (s= 18.02) is five times greater than that of the original data set a (s=3.6055).

Example 9: Find out standard deviation of the given series of grouped data.

Size of the holdings (Hectares)		**2.5-3.5**	**3.5-4.5**	**4.5-5.5**	**5.5-6.5**	**6.5-7.5**	**7.5-8.5**	**8.5-9.5**
No of farmers	fi	1000	2300	3600	2400	1700	3000	500

Size of the holdings (Hectares)	**No of farmers f_i**	**Mid-point (x_i)**	$(x_i - \bar{x}), \bar{x} = mean = 6$	$(x_i - \bar{x})^2$	$f_i.(x_i - \bar{x})^2$
2.5-3.5	1000	3	-3	9	**9000**
3.5-4.5	2300	4	-2	4	**9200**
4.5-5.5	3600	5	-1	1	**3600**
5.5-6.5	2400	6	0	0	**0**
6.5-7.5	1700	7	1	1	**1700**
7.5-8.5	3000	8	2	4	**12000**
8.5-9.5	500	9	3	9	**4500**
	$\sum f_i$ = **14,500**				$\sum f_i.(x_i - \bar{x})^2$ = **40,000**

$$Var = \frac{\sum f_i.(x_i - \bar{x})^2}{\sum f_i} = 40{,}000/14{,}500 = 2.758$$

$$\sigma = \sqrt{Var} = \sqrt{\frac{\sum f_i.(x_i - \bar{x})^2}{\sum f_i}} = \sqrt{2.758} = 1.6609$$

Skewness

Skewness is a measure of symmetry. A distribution is said to be symmetric if the frequencies are symmetrically distributed about the mean. Histograms are used to represent the frequency distribution. A distribution is said to be symmetrical when its mean, median and mode are identical. i.e. Mean = Median = Mode.A distribution which is non-symmetrical is said to be an asymmetrical distribution. It is also known as skew distribution. In a skewed distribution, Mean ≠ Median ≠ Mode, i.e. the values are pulled apart.If the histogram has a longer tail towards the left, it is said to be **negatively skewed** and if the histogram distribution has a longer tail towards the right , it is said to be **positively skewed.**

Example 10: Consider the following frequency distribution of the average monthly income of small scale traders in a village as given below, find the mean, median, mode and plot the histogram for the given data.

X (Rs)	1000	2000	3000	4000	5000	6000	7000
F	3	5	6	11	6	5	3

Solution:
Mean is calculated as given below:

X (Rs)	1000	2000	3000	4000	5000	6000	7000	Sum
F	3	5	6	11	6	5	3	39
Xf	3000	10000	18000	44000	30000	30000	21000	156000

Mean = 156000/39 = Rs. 4000.
Mode = Rs.4000, since that "x" value has the highest frequency.
Median: Since the sample size n= 7 (Odd), median is the middle observations in the ordered list, which corresponds to n=4, i.e. Rs.4000.

Thus, Mean = Median = Mode. Hence, the distribution is symmetric.
The histogram plot of the above problem is as given below, which is perfectly symmetrical.

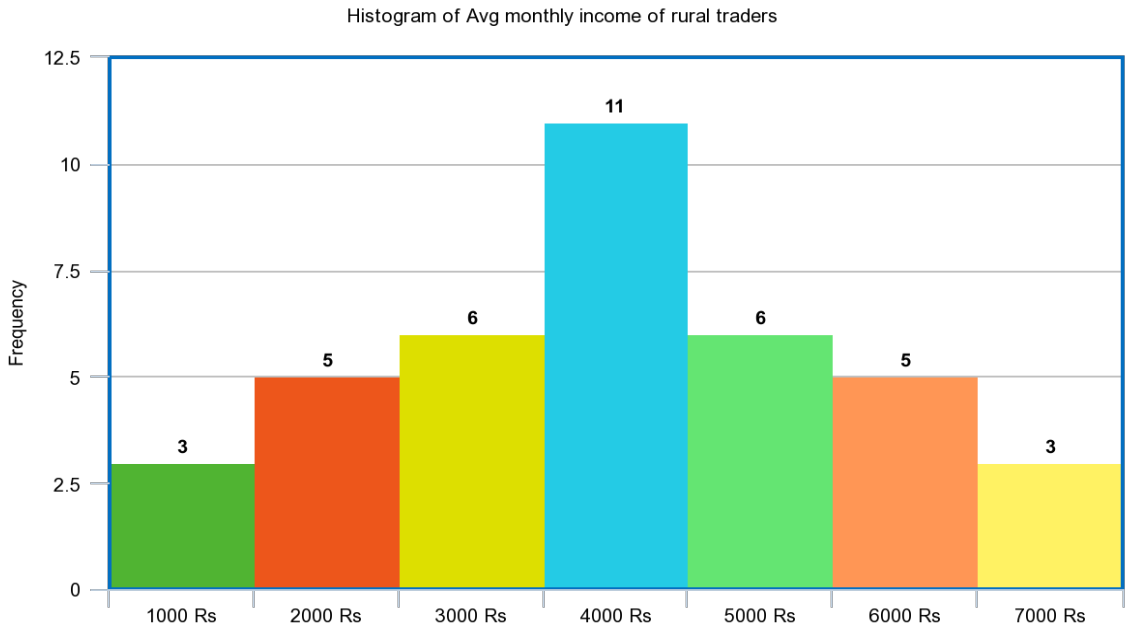


Fig 2.5: Histogram Plot

Example 11: Given the mortality (death) rates of a small village community with respect to age (after ignoring the accidental deaths) observed in one month is as follows. Draw the histogram for the data. Is it symmetrically skewed or not?

Age (Sorted), Years	Newborn -20	20-40	40-60	60- 80	80-100
Frequency	11	17	58	61	21

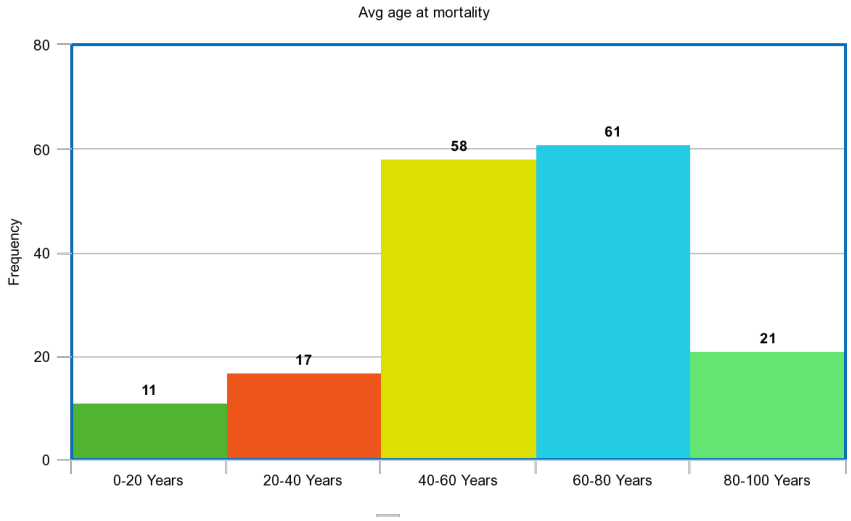


Fig 2.6: Histogram of average age at mortality in a village/month

The histogram distribution having a longer tail towards the right, it is said to be **positively skewed.**

Example 12: Given the distribution of the purchase of onion in a shop observed with respect to its cost is given below. Illustrate the histogram for the purchase pattern. Is the distribution symmetrically skewed or not?

Cost of Onion/Kg (Rs)	**<30**	**30-60**	**60-90**	**90-120**	**>120**
Frequency of customers purchase	80	72	54	28	11

The histogram distribution having a longer tail towards the left, it is said to be **negatively skewed.**

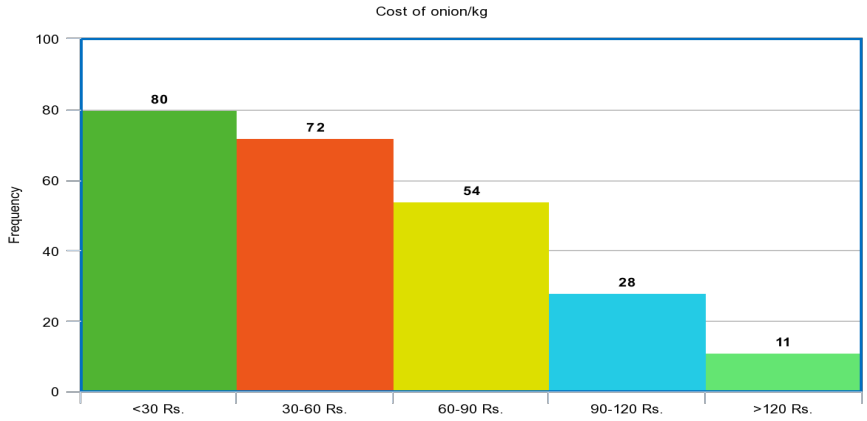


Fig 2.7: Histogram of cost of onion/kg and frequency of purchase by customers

Measure of Skewness

The measure of asymmetry is called as measure of skewness. Measure of skewness is used to find out the extent of skewness and whether it is positive or negative. These measures can be absolute or relative. The absolute skewness measured as 'Sk' is expressed in units of original data and that cannot be used for comparison of two different distributions. In that case, relative skewness is used.

Absolute skewness :Sk = Mean – Mode.

Relative skewness or coefficient of skewness is measured using Karl Pearson's coefficient of skewness.

Karl Pearson's coefficient of skewness: Sk= (Mean – Mode)/Standard Deviation

$$sk = \frac{\mu - \text{mode}}{\sigma}$$ **.....(1)**

For moderately skewed distribution, Mode = 3 Median – 2 Mean.

Substituting in equation 1, we get Sk = (Mean – (3Median – 2 Mean))/ Std Deviation.

i.e. Sk= 3(Mean – Median)/Std deviation = $\frac{3(\bar{x} - \mu)}{\sigma}$

Interpretation of Karlson Pearson Coefficient of skewness

i. Sk value lies between [-1,+1]
ii. When its value is [0], then there is no skewness. i.e. Distribution is symmetrical.
iii. When its value is negative, then the distribution is negatively skewed.
iv. When its value is positive, then the distribution is positively skewed.

Example 13: The data given below corresponds to the No of hours spent/week in watching TV serials by a sample of 5 villagers in a particular street. Is it symmetrically distributed? If not, estimate the measure of skewness.

#Villagers	**1**	**2**	**3**	**4**	**5**
Hours spent on TV serial/week	12	18	35	22	18

Solution:

Mean : $\bar{x}$ = [12+18+35+22+18]/5 = 21 Hours.

Median : Since the sample size n= 5 (Odd), median is the middle observations, which corresponds to n=3, i.e. 35 Hours.

Mode = 18 hours, since it occurs maximum number of times in the series.

Mean ≠ Median ≠ Mode. Hence, distribution is not symmetrical.

Computation of Standard deviation:

x	$x - \bar{x}$	$(x - \bar{x})^2$
12	-9	81
18	-3	9
35	14	196
22	1	1
18	-3	9
	Sum = 0	Sum = 296

$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$ = 296/4 = 74.

$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{74}$ = 8.602

Karl Pearson's coefficient of skewness: Sk= (Mean – Mode)/Standard Deviation

Sk = (21 – 18)/8.602 = 0.348 > 0.

Since Sk> 0, the distribution is positively skewed.

To Do Activity

1. Compute the Karl Pearson's coefficient of skewness from the following data :

Daily Expenditure of villagers (Rs)	0-20	20-40	40-60	60-80	80-100
No of families	13	25	27	19	16

2. Similarly, survey a known street and get their average expenses per family and tabulate. From the tabulated data, find the Karl Pearson's coefficient of skewness.

Example 14 : Non-agricultural jobs in rural areas include jobs like beedi making, sweeping, cleaning, tractor driving, weaving, loading/unloading. The average hourly wages for the listed jobs and the number of workers for each wage category is as given below:

Hourly Wages (Rs)	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No of workers	8	16	30	45	62	32	15	6

Draw the histogram chart and check if the distribution is symmetrical. If not, calculate the coefficient of skewness.

Solution:

(a) Histogram is as given below

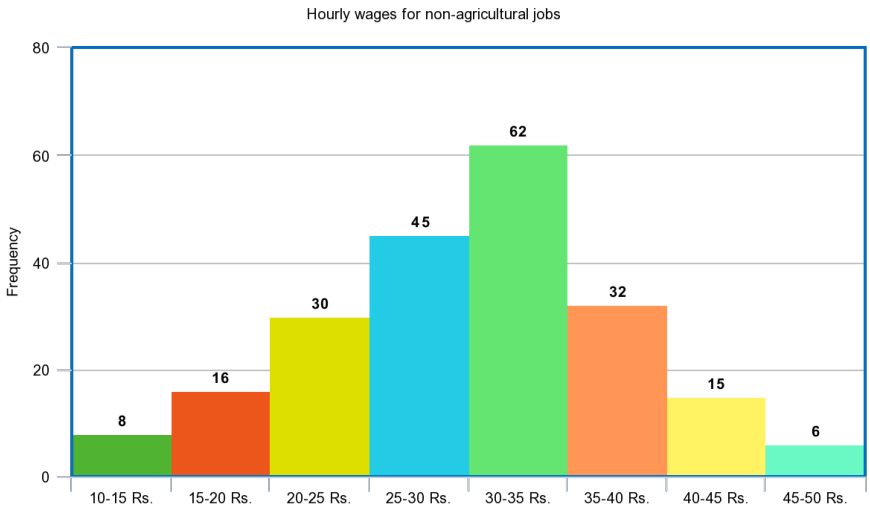


Fig 2.8: Distribution of Hourly wages of non-agricultural jobs in a village

(b) To check for skewness, compute mean, median and mode of the given grouped data.

Mean : = = **30**

Median :

Step 1: Construct the cumulative frequency distribution.

Hourly Wages (Rs)	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No of workers	8	16	30	45	62	32	15	6

Cumulative frequency	8	24	54	99	161	193	208	214

The median is the datum corresponding to the middle value of the cumulative frequency.

n/2 = 214/2 = 107. This implies the class median is the fifth class [30-35].

Mode = Fifth Class = [30-35] , since it has the highest frequency.

(i) The modal class is the class interval that has the largest frequency = [30-35]

(ii) The lower class boundary of the modal class $= Lb$ = 30

(iii) The difference of frequency between the modal class to its upper class

$= a$ =62-45=17

(iv) The difference of frequency between the modal class to its lower class

$= b$ = 62-32= 30

Mode = $Mo = Lb_{Mo} + \frac{a}{a+b}.C$ = 30 + $(\frac{17}{17+30}) * 5$ = 31.808

(c) Compute coefficient of skewness:

Size of the holdings (Hectares)	**No of farmers f_i**	**Mid-point**	$(x_i - \bar{x})$, $\bar{x} = mean$ **= 30**	$(x_i - \bar{x})^2$	$f_i.(x_i - \bar{x})^2$
10-15	8	12.5	-17.5	306.25	**2450**
15-20	16	17.5	-12.5	156.25	**2500**
20-25	30	22.5	-7.5	56.25	**1687.5**
25-30	45	27.5	-2.5	6.25	**281.25**
30-35	62	32.5	2.5	6.25	**387.5**
35-40	32	37.5	7.5	56.25	**1800**
40-45	15	42.5	12.5	156.25	**2343.75**
45-50	6	47.5	17.5	306.25	**1837.5**
N= $\sum f_i$ **=214**					$\sum f_i.(x_i - \bar{x})^2$ = **13287.5**

$s^2 = \frac{\sum (x - \bar{x})^2}{n-1}$ = 13287.5/213 = 63.38

$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{63.38}$ = 7.89

Karl Pearson's coefficient of skewness: Sk= (Mean – Mode)/Standard Deviation

Sk = (30 –31.808)/7.89 = -0.22915 < 0.
Since Sk< 0, the distribution is negatively skewed.

To do activity

The following series gives the height of the trees in a small forest area surveyed by the forest authorities. From these data, calculate the Karl Pearson's coefficient of skewness and also find the coefficient of variation.

ight (Ft)	low 7	low 14	low 21	low 28	low 35	low 42	low 49	low 56
of trees	26	57	92	134	216	287	341	360

Summary of the Chapter

Summary: The chapter describes the essence of stastistics, it's genesis and how does it facilitiate in decision making. Various statistical tools such as mean, median, mode, standard deviation, skewness etc. as tools for doing statsitical analysis are introduced and explained using suitable examples. The chapter also throws light on inferential and descriptive statistical methods and aims at giving a clear diffrence between both the methods to the reader.

Model Questions

1. What is the primary reason for the growing importance of statistics? Throw some light on the factors responsible for its growth.
2. What do you understand by population? How is the usage of the population different in Descriptive and inferential Statistics?
3. Which average, mean, median, or mode, is associated with the standard deviation?
4. What is the relationship between the variance and the standard deviation for a sample data set?
5. When computing the standard deviation, does it matter whether the data are sample data or data comprising the entire population?
6. What symbol is used for the standard deviation when it is a sample statistic?
7. Consider the data set 5, 9, 10, 11, 15. (a) Use the computation formula to compute s. (b) Add 5 to each data value to get the new data set 10, 14, 15, 16, 20. Compute s. Explore the effect on the standard deviation of adding the same constant to each data value in a data set.
8. The weights of 20 oranges (in grams) picked at random from a harvest are given as follows. Construct a frequency table and a cumulative frequency table for the same.

60	65	45	80	65	70	55	85	90	75
40	80	100	70	50	55	75	95	50	70

9. What is the relationship between the variance and the standard deviation for a sample data set?
10. When computing the standard deviation, does it matter whether the data are sample data or data comprising the entire population?
11. What symbol is used for the standard deviation when it is a sample statistic?
12. Consider the data set 5, 9, 10, 11, 15. (a) Use the computation formula to compute s. (b) Add 5 to each data value to get the new data set 10, 14, 15, 16, 20. Compute s. Explore the effect on the standard deviation of adding the same constant to each data value in a data set.

13. The table below gives the percentage distribution of female teachers in the primary schools of rural areas of various states and union territories (U.T.) of India. Find the mean, median and mode of percentage of female teachers.

Percentage of female teachers	15-25	25-35	35-45	45-55	55-65	65-75	75-85
Number	6	11	7	4	4	2	1

14. The data below are from an investigation of severe vomiting by children less than age of 10 years in a rural village due to contamination of wheat flour with lead dust. Summarise the blood lead level data with a frequency distribution and calculate mean, median and standard deviation.

ID	1	2	3	4	5	6	7	8	9	10
Age (Years)	3	4	6	7	9	10	11	12	13	14
BLL†	69	45	49	84	48	58	17	76	61	78

15. The milk purchased (in Litres) by 15 families in a village street is as given below. Tabulate the frequency distribution table of the given data and calculate mean, median and standard deviation.

0.5	1.5	2.0	0	0.25
0.25	1	0.5	0.25	1.25
1	0.5	0	0.25	1.0

To-Do- Activity

Indian vedas and upanishads are considered as the repository of all the tools and techniques being used today. Can you find out which veda contains the description of Statistics and its usage in old times?

References:

1. Arora, P.N and Arora, S (2007), Comprehensive statistical methods, S.Chand& Co.
2. Pocket Book of Agricultural Statistics, 2017, Directorate of Economics & Statistics New Delhi.

Chapter 3 Probability and Sampling

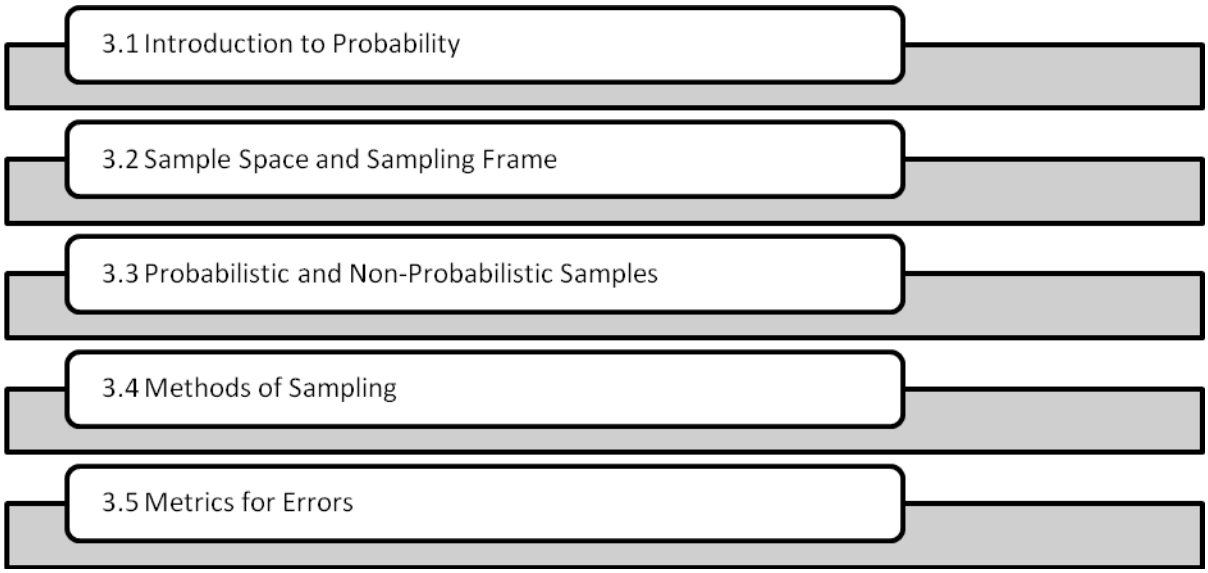
Introduction

Probability is a tool which mankind has been using for ages to predict the outcome of any situation. Be it the game of dice played in Mahabharata or an IPL match being played in today's world. Probability plays a major role in determining the odds in favour or against an event. The application of probability over the years has grown form merely a statistical tool to a tool useful in data science and data analytics. The knowledge of probability and its application in the right places is considered an important skill in today's world. In this chapter let us make an endeavour to understand certain concepts related to this extremely vast field of probability and sampling.

Objectives of the Chapter

- To acquaint the students with the importance of Probability
- To illustrate through examples the concept of sample space and sample frame
- To differentiate between Probabilistic and Non-Probabilistic Samples
- To discuss various Methods of Sampling
- To familiarize the student with the Metrics for errors

Structure of the Chapter



3.1 Introduction to Probability

Consider the scenario that you are given this book to read, there are two possible outcomes to this event, either you will read this book or not. Once you have finished reading the book, there are two further outcomes, whether you have liked the book or not. A scientific study of how likely an occasion or phenomenon will occur from rolling of a dice to winning or losing a game in cards to the likelihood of it raining on a particular day.

Probability is utilized, in all of these diverse areas to detect the chance of a thing happening or not**.** As described in the previous chapter, Probability theory began in seventeenth century France when the

two incredible French mathematicians, Blaise Pascal and Pierre de Fermat, compared more than two issues from rounds of possibility. Issues like those Pascal and Fermat understood proceeded to impact such early scientists as Huygens, Bernoulli, and DeMoivre in setting up a numerical Probability Theory. Today, Probability theory is a well-established part of mathematics that discovers applications in each zone of insightful action from music to material science, and in day by day experience from climate forecast to foreseeing the dangers of new medicinal medications. Probability is defined as the ratio of the number of outcomes favorable for a particular event to the total number of outcomes.

$$bability = \frac{Number\ of\ outcomes\ favourable\ for\ the\ event}{Total\ number\ of\ outcomes} \quad \text{(i)}$$

Assigning Probabilities

- $probability\ of\ any\ event\ must\ be\ nonnegative, e.g., P\ (P_i)\ \geq\ 0\ for\ each\ i.$
- $probability\ of\ the\ entire\ sample\ space\ must\ be\ 1, i.e., P\ (S)\ =\ 1.$
- $two\ disjoint\ events\ A\ and\ B, the\ probability\ of\ \ the\ union\ of\ A\ and\ B\ is\ equal\ to\ the\ sum\ of\ the$

$$probabilities\ of\ A\ and\ B, i.e., P\ (A\ \cup\ B)\ =\ P\ (A)\ +\ P\ (B)\ .$$

Example: Suppose that you are going to roll a standard dice, and you want to predict the chances of getting a 6.

Solution: Total possible available outcomes are: 6

p_6= chance of rolling a 6

Applying equation (i) $Probability = \frac{1}{6}$

Approaches to Assigning Probability

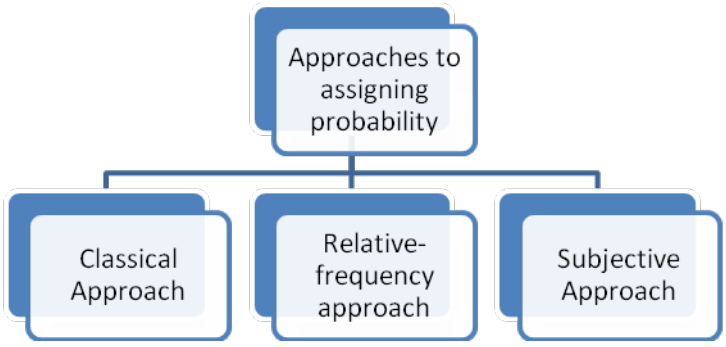


Fig 3.1: Approaches to Assigning Probability

Classical Approach: The Classical Approach for determining probability is utilized if the probable outcomes are known in their entirety ahead of time and all the likelihood of each outcome is equal. In this approach, the probabilities are based on the actual experiments. The best case of the old-style technique for likelihood is flipping/ tossing a coin. With a two-faced coin, the two potential outcomes are known to you ahead of time and almost certainly, you will flip a Head as it is that you will flip a Tail.
If an experiment has n simple outcomes, this method would assign a probability of $1/n$to each outcome. In other words, each outcome is assumed to have an equal probability of occurrence. This method is also called the axiomatic approach.

Example: Roll of a Die

$Possible\ outcomes = \{1, 2, \cdots, 6\}$
Probabilities: Each simple event has a $1/6$ chance of occurring.

Example: Two Rolls of a Die
$Possible\ outcomes$= {(1, 1), (1, 2), · · ·,(6, 6)}
Assumption: The two rolls are "independent."
Probabilities: Each simple event has a $(1/6) \cdot (1/6) = (\frac{1}{36})$ chance of occurring.

Statistical Approach: The observed frequency approach is utilized when every single plausible result is not known ahead of time and the entirety of the likelihood of all the outcomes is not equally likely. This strategy utilizes comparative insights from a past however comparable example. A case of how the observed frequency approach would be utilized would be a storekeeper setting orders dependent on the earlier year's deals. The information required to apply the classical technique isn't accessible however the availability of similarly reliable information is. Probabilities are allocated based on experimentation or authentic information.
Let us assume A to be an event under consideration, and expect that you have played out a similar test n times with the goal that n is the event A could have happened. Furthermore, let $n_A$be the occasions that A actually took place. Presently, think about the relative frequency n_A/n. At this point, in this strategy, we "attempt" to characterize $P\ (A_n)$ as:

$$P\ (A) = \lim_{n \to \infty} \frac{n_A}{n}$$

The above must be seen as an endeavor since it isn't physically achievable to repeat an experiment an infinite number of times. Another significant issue with this definition is that two arrangements of n examinations will commonly bring about two unique proportions. Be that as it may, we anticipate that the error should converge to 0 for large values of n. Subsequently, for large values of, the proportion $\frac{n_A}{n}$ might be taken as a sensible approximation for $P\ (A)$.
Example: Roll of a Die
$Probability\ of\ the\ event\ = \{1, 2, \cdots, 6\}$
Probabilities: Roll the given die 1000 times (say) and assume the number of times the outcome 1 is observed is 25. Thus, $A = \{1\}, n_A = 25, and\ n = 1000.$ Therefore, we say that $P\ (A)$ is approximately equal to $25/1000 = 0.025.$
Example: Grocery Sales
A grocery store tracks the daily sales of rice (in kgs) in the past 30 days.
The resulting data is:

Quantity Sold(in kgs)	No. of Days
0	1
1000	2
2000	10
3000	12
4000	5
5000 or more	0

The approximate probabilities are:

Quantity Sold(in kgs)	No. of Days	Probability
0	1	**1/30 = 0.03**
1000	2	**2/30 = 0.07**
2000	10	**10/30 = 0.33**
3000	12	**12/30 = 0.40**
4000	5	**5/30 = 0.17**
5000 or more	**0**	**0**

Thus, for example, there is a 40% chance that the store will sell 3000 kgs of milk on any given day.

Subjective Approach: In the subjective approach, we characterize probability as the level of conviction that we hold in the happening of an event. In this manner, judgment is utilized as the metric for allotting probabilities. Notice that the classical styled approach of allocating equivalent probabilities to basic events is, in reality, likewise dependent on judgment. What is to some degree distinctive here is that the utilization of the subjective approach is normally constrained to tests that are unrepeatable.

Example: Village Kabadi
Consider a Kabadi match organized for 15 villages with 80 teams competing. What is the probability for a particular team to win? Is it practical to assume that the probability is 1/80? Take a note that the relative-frequency approach cannot be applied to this scenario. Judges of the Kabadi matches may place their bets on the outcomes of such "one- time" experiments based on their judgment as to how good players are skilledina specific team. Definitely, having dissimilar judgments is what makes betting probable!

Example: Game of Dhoni
What is the probability for a Dhoni to hit a century in the match being played tomorrow? Again, this "experiment" cannnot be repetitive, and it is not feasible to apply the relative-frequency approach.
Models relying on past data are frequently used to make such predictions, as thoughtlessly following unreliable judgments is often hazardous.

Basic Rules of Probability
The above described three definitions of probability must adhere to the same rules. In this section, a description of some basic concepts and rules is given.

Complement of an event:
Let Abe an event. It's complement is represented by A^c which denotes the event that $A\ does\ not\ occur$. By definition, we have $A \cap A^c = \emptyset$ (the empty set) and $A \cup A^c = S$. Here, the intersection operation $\cap\ corresponds\ to\ "and"$; and the union operation $\cup\ corresponds\ to\ "or"$.
As per the relation that all probabilities must sum upto 1,

$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c), we\ have$
$P(A^c) = 1 - P(A).$ (ii)

Example: Picking a card from a pack of cards
$P(\{6\ of\ spades\}) = 1/52$
$P(the\ outcome\ is\ not\ a\ 6\ of\ spades) = 1 - 1/52 = 51/52$

Union/Addition of Events:
Let *A* and *B* be two events. Then,
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (iii)
As it is assumed that A and B are non- disjoint sets and thus may overlap, the subtraction of $P(A \cap B)$ is essential. However, in case*A* and *B* are *mutually exclusive*, i.e., $A \cap B = \emptyset$, then
$P(A \cup B) = P(A) + P(B).$

Example: Picking a card from a pack of cards
P (red) = 26/52 and *P* (6) = 4/52
P (red and 6) = 2/52
P (red or 6) = 26/52 + 4/52 − 2/52 = 28/52

Conditional Probability
Let $A\ and\ B$be two events. Then, the *conditional* probability refers to the probability of occurrence of$A$ provided Bhas already happened. It is read as Probability of A given B and is defined as:

$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$ (iv)

The logic behind this definition is that if and only if*B*has occurred, then only the "portion" of *A* that is contained within *B*, i.e., $A \cap B$, could occur; besides, the original probability of $A \cap B$must be recalculated to reflect the fact that the "new" sample space is B.

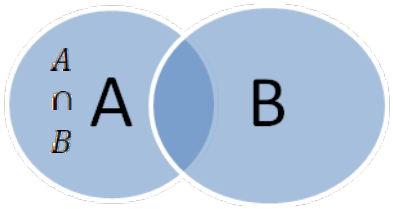


Fig 3.2: Venn Diagram Depicting Conditional Probability

Example: Picking a card from a pack of cards
Suppose a card is drawn randomly from a deck and found to be an Ace. What is the conditional probability for this card to be Spade Ace?
$A = Spade\ Ace$
$B = an\ Ace$
$A \cap B = Spade\ Ace$
$P(A) = \frac{1}{52};\ P(B) = \frac{4}{52};\ and$
$P(A \cap B) = 1/52$

$$Hence, \quad P(A \mid B) = \frac{\frac{1}{52}}{\frac{4}{52}} = \frac{1}{4}$$

Multiplication

Joint probability of two events is calculated using the multiplication rule. It is merely a reorganization of the conditional probability formula; see (iv). Formally,

$P(A \cap B) = P(B|A)\, P(B)$

$P(A \cap B) = P(A|B)\, P(A)$

Example: Drawing an Ace of Diamonds from a pack of cards

$A = an\ Ace;\ B = a\ Diamond\ ;\ A \cap B = the\ Ace\ of\ Diamonds$

$P(B) = 13/52;\ P(A \mid B) = 1/13$

$$Hence, P(A \cap B) = P(A \mid B)P(B) = \frac{1}{13}\ x\ \frac{13}{52} = \frac{1}{52}$$

Example: Choosing Athletes from a Pool of Players

A University has seventy male and thirty female athletes. The Sports In-charge wants to select two students at random to play in the Inter-University Championship. What is the probability that the two students chosen are female?

$A = the\ first\ student\ selected\ is\ female$

$B = the\ second\ student\ selected\ is\ female$

$A \cap B = both\ chosen\ students\ are\ female\ P(A) = 30/100;\ P(B \mid A) = 29/99$

$$Hence, P(A \cap B) = P(B \mid A)P(A) = \frac{29}{99}\ x\ \frac{30}{100} = \frac{870}{9900} = 0.087$$

Though the approaches and the basic principles have been explained in detail, all the theories discussed above have some serious limitations and shortcomings. For instance, these approaches cannot be applied to the experiments which have a large number of outcomes. The classical definition of probability cannot be applied whenever it is no possible to make simple enumerations of cases which can be considered equally likely. Take for example the question of traffic congestion in an area, how does it apply to that? What are the likely outcomes? We might think two outcomes “congested” or “not congested”. But at any given place, it will not generally be agreed that they are equally likely. In reality, Classical definition is difficult to apply in scenarios where there is a deviation from the experiments pertaining to cards, coins, dice or other simple examples.

The Statistical definition has limitations from the mathematical viewpoint as a true limiting factor or number might not at all exist in reality. Due to these limitations, the modern probability has been developed axiomatically. A. N. Kolmogrov developed this Theory of probability in the year 1933 in his book titled “Foundations of Probability”. He discussed certain axioms to interpret probability. To To understand this approach, we must familiarize ourselves with some basic terms:

- Random Experiments
- Sample Space
- Elementary Events
- Compound Events

Random Experiments

The term experiment implies an activity which can create some well-characterized outcome(s). In probability an experiment is characterized as the system which can be iterated over many times and have a well-characterized set of any number of potential outcomes.
There are two kinds of experiments viz:

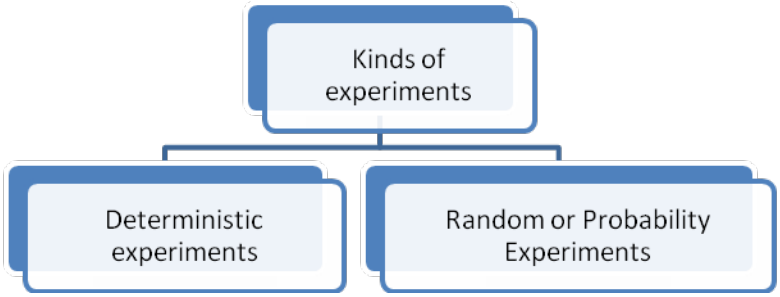


Fig 3.3: Kinds of Experiments

Deterministic Experiment: A deterministic experiment is the subset of experiments wherein there is just a single conceivable outcome. i.e., the outcomes can be anticipated before the examination is led. Deterministic experiments are those experiments which when repeated under identical conditions produce the same result for outcome when experiments in Science and Engineering repeated under identical conditions, we obtain almost the same result every time. Take for example the case of combining Sodium (Na) and Chlorine (Cl) in the proportion of 1:1(under controlled conditions, the outcome is known that it will form salt. Similarly adding two numbers such as 6+3 will yield 9.In the event of a stone being dropped from a window, it is known without a doubt, which the stone will go down. However, in case it is thrown up in the air, it is known that it will travel upwards over a certain time interval; but afterwards will come down.

Random or Probability Experiments
In this section, we will initially consider the example of rolling a die which has a limited number of potential results:

= chance of rolling a 1, = chance of rolling a 2, = chance of rolling a 3, = chance of rolling a 4, = chance of rolling a 5, = chance of rolling a 6

$$\sum_{i=1}^{6} p_i = 1 \qquad \text{(sum of probability of all events is always equal to 1)}$$

Table 3.1: Examples of Random Experiments and their Outcomes

Experiment	Outcomes
Toss a Coin	Head or Tail
Roll a die	Numbers: 1, 2, 3, 4, 5, 6
Exam Marks	Numbers: (0, 100)
Course Grades	F, D, C, B, A
Task completion times	Nonnegative values

This leads us to the description of a random experiment. It corresponds to an activity or procedure that gives one of numerous possibleoutcomes. In the event that an examination when iterated under similar

conditions, don't create a similar result each time of iteration, however final outcome is one of the few potential outcomes, at that point it is known as random or probabilistic experiment.In other words, and experiment who outcomes cannot be predicted or determined in advance is called a random experiment for example, in tossing a coin one is not sure if a head or tail will be obtained so it is a random experiment similarly, rolling and unbiased die and drawing a card from a well shuffled pack of playing cards are examples of random experiments.

Remember that our possible objective in this course is to go from the arbitrary samples to the entire population. The theory that takes into consideration this progress is the theory of probability.The list of possible outcomes of a random experiment must be exhaustive and mutually exclusive.A few examples of random experimentation are given as below:Let us now discuss various terms associated with the random experiment. These terms and help us in introducing the theoretical concept of probability.

3.2 Sample Space and Sampling Frame

Associated to every random experiment there are two basic terms with outcomes for elementary events and sample space. In this section we will discuss about these two for different random experiments.

Elementary event: An outcome of a random experiment is called an elementary event.If a random experiment is performed then each of its outcomes is known as an elementary event. In other words, outcomes of a random experiment are known as elementary events associated to it. Elementary events are also known as simple events.

Example : Consider the random experiment of tossing a coin. The possible outcomes of this experiment areHead (H) or Tail (T). Thus, if we define

$$P1 = Getting\ head\ (H) on\ the\ upper\ face\ of\ the\ coin,$$

and,

$$P2 = Getting\ it\ on\ the\ upper\ face\ of\ the\ coin.$$

Then P1 and P2 are elementary events associated with the experiment of tossing a coin.

Let us now consider the random experiment of tossing two coins simultaneously or a coin twice in this experiment the possible outcomes are:

Heads on first and Head on second,

Head on first and Tail on second

Tail on first and head on second

Tail on first and Tail on second

If we define

$$HH = Getting\ head\ on\ both\ coins,$$
$$HT = Getting\ head\ on\ first\ and\ tail\ on\ second$$
$$TH = Getting\ tail\ on\ first\ and\ head\ on\ second$$
$$TT = Getting\ tail\ on\ both\ the\ coins$$

Then,

HH, HT, TH, TT are elementary events associated to the random experiment of tossing of two coins. Thesample space associated with the random experiment of tossing three coins simultaneously or tossing a coin 3 times is given by

$$HHH,\ HHT, HTH, THH,\ TTH, HTT, THT,\ TTT$$

Sample space the set of all possible outcomes of a random experiment is called the sample space

associated with it and is generally denoted by if $P_2, P_3, \ldots\ldots P_n$ are possible outcomes for elementary events of a random experiment, then $\{P_1, P_2, P_3, \ldots\ldots P_n\}$ is the sample space associated with it.

Example: Consider the random experiment of tossing a coin. The possible outcomes of this experiment are H and T. Thus, if we define

$P_1 = Getting\ head\ (H) on\ the\ upper\ face\ of\ the\ coin, and,$
$P_2 = Getting\ tail\ (T)\ on\ the\ upper\ face\ of\ the\ coin.$

then, P_1, P_2 are elementary events associated to the random experiment of tossing a coin. The sample space associated to this experiment is given by $S = \{P_1, P_2\}$.

Example:Consider the experiment of throwing a die let the six faces of a die be marked as 1, 2, 3, 4, 5, 6. If the dice thrown, then anyone of six faces may come upwards. So, there are six possible outcomes of this experiment, namely 1, 2, 3, 4, 5, 6. Does if we define

$P_i = Getting\ a\ face\ marked\ with\ number\ i, where\ i = 1,2,3,4,5,6$

then $P_1, P_2 \ldots . P_6$ are 6 elementary events associated to this experiment. The sample space associated to this space is $S = \{P_1, P_2 \ldots . P_6\}$.

The sample space for Example of tossing three coins simultaneously or one after the other is given by:

$$S = \{HHH,\ HHT, HTH, THH, TTH, HTT, THT,\ TTT\}.$$

In the preceding section, we learnt about sample spaces associated with several random experiments. In this section, we will introduce an important term associated with a random experiment.

Event: A subset of the sample space associated with the random experiment is called an event. Consider the random experiment of throwing a die. The sample space associated with the experiment is S= $S = \{1,2,3,4,5,6\}$. Clearly, S has $2^6 = 64$ subsets

Each one of these 64 subset is an event associated with the random experiment of throwing a die.
For example, $= \{2,4,6\}$, $B = \{1,3,5\}$, $C = \{3,4,5,6\}$, $D = \{1,2,6\}$ etc. are events as they are subsets of S.
These events $A, B\ and\ C$ can also be described in words as follows
$A =$ getting an even number
getting an odd number
$C = getting\ a\ number\ greater\ than\ 2$
However, there is no general description in words for the event D. Thus, we find that some events associated with the random experiment may be described in words. However, it is not possible for every

event.

Consider the experiment of tossing three coins at a time. The sample space as associated with this experiment is

$S = \{HHH, HHT, HTH, THH, TTH, HTT, THT, TTT\}.$

$A = \{HHT, HTH, THH\} B = \{HHH, HHT, HTH, THH\},$

$\{HHH, HHT, HTH, THH, TTH, HTT, THT, TTT\}$,$= \{HHH, HTH, TTT\}$

Clearly, A, B, C *and* D, being subsets of S, are events associated with random experiment of tossing three coins. These events can also be described in words as follows

$A = Getting\ 2\ heads$

$number\ of\ heads\ exceeds\ the\ number\ of\ tails$

$C = getting\ at\ least\ one\ head$

But event D cannot be expressed in words.

Occurrence of an Event: An event A associated to a random experiment is said to occur if any one of the elementary events associated to it is an outcome. Thus, if an elementary event P is an outcome of a random experiment and A is an event such that $P \in A$, then we say that event A has occurred.Let's take the random experiment of rolling an unbiased die. That A denote the event "$Getting\ an\ even\ number$". $2, 4, 6$ are the elementary events associated to this event. Now, suppose that in a trial the outcome is 4, then we say that event A has occurred. In the second roll of the die, let the outcome be 3, then we conveniently say that the event A has not occurred.

Let a die be rolled and outcome of the trial be 4. Then, it can be easily understood that the following events have occurred:

i. $ting\ a\ number \geq 2, represented\ by\ the\ set\ \{2,3,4,5,6\}$
ii. $ting\ a\ number \leq to\ 5,\ represented\ by\ the\ set\ \{1,2,3,4,5\}$
iii. $ting\ an\ even\ number, represented\ by\ the\ set\ \{2,4,6\}.$

On the basis of the same outcome, we can say that the following event have not occurred:

i. $ting\ an\ odd\ number$
ii. $ting\ a\ multiple\ of\ 3.$

Favorable Elementary Events: Let S be the sample space associated with the random experiment and$A$ be an event associated with the experiment. Then, elementary events belonging to A are known as favorable elementary events to the event A.In other words, an elementary event P is favourable to a compound event A, if we say that the event occurs when P is an outcome of a trial.

Let's take into consideration, the random experiment of rolling a pair of dice and the compound event a Defined by "$Getting\ 8\ as\ the\ sum$". We observe that the event occurs if we get any one of the following elementary events as outcome: $(2,6), (6,2), (3,5), (5,3), (4,4)$. Hence, there are 5 favourable elementary events to the event A.

Negation of an Event: Corresponding to every event A associated to a random experiment, we define an event "$not\ A$" which is said to occur when and only when it does not occur.

For example in a single throw of a die if A denotes the event that the outcome is 2 or 4 and B did not say when that the outcome is even. Then,$A = \{2,4\} and\ B = \{2,4,6\}$. Clearly, the occurrence of A implies the occurrence of event B.For if 2 or 4 occurs, we say that the outcome is an even number. Thus, if the occurrence of an event A implies the occurrence of event B, then we say that A implies B. clearly, If A implies B, then we have$A \subset B$

Types of Events

Let there be n elementary events associated with a random experiment. Then the corresponding sample space has n elements and hence 2^nsubsets. Each subset of S is an event associated to the random experiment and the sample space is the universal set of these events. These 2^nevents are divided into four different types on the basis of their nature of occurrence. In this section we shall learn about such types.

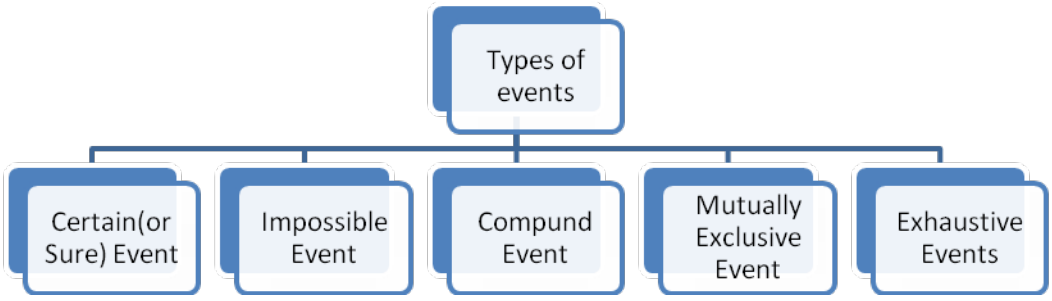


Fig 3.4: Types of Events

Certain (or sure) Event: An event associated with the random experiment is called a certain event if it is always occurring whenever the experiment is performed. For example, associated with the random experiment of Rolling a die, the event a getting an even number or an odd number is a certain event. Clearly, this event is represented by the set $\{1,2,3,4,5,6\}$ which is the sample space of the experiment.If S is the sample space associated with the random experiment. Then, S, being subset of itself, defines an event. Also, every outcome of the experiment is an element of S, so the event represented by S always occurs whenever we perform the experiment. Consequently, the event represented by S is a certain event. Thus, the sample space associated with the random experiment defines a certain event.

Impossible Event: An event associated with a random experiment is called an impossible event if it never occurs whenever the experiment is performed. Consider the experiment of rolling a die. Let A be the event *"The number turns up is divisible by 7"*. Clearly, none of the possible outcomes is divisible by 7. So, the event A cannot occur at all. In other words, there is no outcome belonging to set represent event a. So, the set$A$ is the null set.

Compound Event: An event associated to a random experiment is a compound event if it is obtained by combining two or more elementary events associated to the random experiment. A compound event is the disjoint Union of two or more elementary events. In other words, an event having more than one sample point is called a compound event. In fact, other than elementary events and impossible events associated with the random experiment, all events are compound events as they are obtained by combining two or more elementary events.

If a pair of dice is thrown together, then there are 36 elementary events associated to this experiment. The sample space associated to this experiment is;

$$S = \{(1,1),(1,2),\ldots..(1,6),(2,1)(2,2),\ldots\ldots\ (2,6),\ldots\ldots\ldots..(6,1)(6,2),\ldots\ldots\ (6,6)\}$$

if we define the event as “Getting a doublet” i.e. A is equal to
$S = \{(1,1), (2,2), \ldots \ldots .. (6,6)\}$

Clearly, it is a compound event obtained by combining 6 elementary events.
Consider the random experiment of tossing two coins simultaneously if we define the event "Getting exactly one head", then HT and TH are two elementary events associated to it. So, it is a compound event**.**

Mutually Exclusive Events: Two occasions are said to be mutually exclusive when both can't happen simultaneously. Events are the outcomes of an experiment in probability. A portion of these events have relations with several other events. At the end of the day, we state that a few events are influenced by the occurrence of other events. For example, tossing a coin is and getting a “Head” or a “Tail” are mutually exclusive events. Clearly, elementary events associated with the random experiment are always mutually exclusive, because elementary events are outcomes (results) of an experiment when it is performed and at a time only one outcome is possible. Consider the random experiment of rolling a die. Let $and\ C$ be three events associated with the experiment as given below:
$A = getting\ an\ even\ number,$
$B =\ getting\ an\ odd\ number,$
$C = getting\ a\ multiple\ of\ 3$

These events in set theoretical notations are:
$A = \{2,4,6\}, B = \{1,3,5\}, C = \{3,6\}$
Clearly, $A \cap B = \emptyset, A \cap C \neq \emptyset, B \cap C \neq \emptyset, A \cap B \cap C = \emptyset$
So, $A\ and\ B$ are mutually exclusive events but $A\ and\ C$as well as $B\ and\ C$ are non mutually exclusive events. however, $A, B\ and\ C$ taken all the three together are mutually exclusive events.In the experiment of throwing a pair of dice events$A =\ equal\ to\ getting\ 8\ as\ the\ sum\ and\ B = getting\ an\ even\ number\ on\ first\ die$are not mutually exclusive because, $A \cap B = \{(2,6), (6,2), (4,4)\} \neq \emptyset$

To Do Activity
Mr.Jayesh is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Jayesh's wedding?

Exhaustive Events: Two or more events associated with the random experiment are exhaustive if their union is the sample space i.e. events$A_1, A_2, \ldots \ldots A_n$associated with the random experiment with sample space s are exhaustive if$A_1 \cup A_2 \cup \ldots \ldots A_n = S$

Thus, a set of events associated with the random experiment is an exhaustive set of events if one of them necessarily occurs whenever the experiment is performed. It is evident from the above definition that all elementary events associated with the random experiment form a set of exhaustive events. Consider the experiment of drawing a card from a well shuffled deck of playing cards. Let A be the event card is red, B be the event card is black. Clearly, $A\ and\ B$ are exhaustive events because$A \cup B = S$.In a single throw of an ordinary die, let us consider the following events

$A_1 = Getting\ an\ even\ number = \{2,4,6\} A_2 = Getting\ an\ odd\ number = \{1,3,5\}$
$A_3 = Getting\ a\ multiple\ of\ 3 = \{3,6\} A_4 = Getting\ a\ number\ greater\ than\ 3 = \{4,5,6\}$
we observe that $A_1 \cup A_2 = S$. Also $A_1 \cup A_2 \cup A_3 \cup A_4 = S$. But $A_1 \cup A_3 \neq S$. So, $A_1\ and\ A_2$ are exhaustive events.Also $A_1, A_2, A_3, A_4,$ are exhaustive events but $A_1\ and\ A_3$ are not exhaustive events.

To Do Activity
In a village, the probability that a girl student continues education after puberty is 0.60, a girl student continues with sports is 0.20 and those who continue both is 0.10. What is the probability that the girl student continues education, continues sport or both?

3.3 Sampling Unit and Sampling Frame

Sampling Unit

The population separated into a limited number of particular and recognizable units is called examining units. Or

The people whose attributes are to be estimated in the investigation are called elementary or sampling units. Or

Before choosing the example, the populace must be isolated into parts called sampling units or simply sample units**.**

Sampling Frame

The set of all the sampling units with an appropriate distinguishing proof (which is representative of the population to be considered is called Sampling Frame). The frame may comprise of either a rundown of units or a map of region (in the event that the sample is obtained from a region), with the end goal that each component in the population has a place with one and only one unit. The list ought to be precise, without any omissions and duplication, sufficient, up-to-date data and the units must cover the entire population and ought to be all around recognized. With an aim to improve the sampling design, supplementary data for the problem under consideration by the sampling frame may be significant as well.

Examples: Sampling Frame and Sampling Unit

- List of households (and persons) enumerated in population census.
- A map of areas of a country showing the boundaries of area units.
- In sampling an agricultural crop, the unit might be a field, a farm or an area of land whose shape and dimensions are at our disposal.

An ideal sampling frame will have the subsequent qualities/characteristics:

- All sampling units are regularly found e.g. contact data, map area or other significant data about sampling units is available.
- No elements from outside the population of interest are present within the frame
- All sampling units have a rational and have numerical identifier
- The frame is organized in a logical and systematic manner
- Every element of the population of interest is present within the frame

- Every element of the population is present just one occasion within the frame
- The data is up-to-date

The sampling frame has some additional information about the units that allow the utilization of more advanced sampling frames. A sampling frame can be classified subject to several types of defect as follows:

A frame may be:	
	Inaccurate: where a portion of the sampling units of the population are recorded inaccurately or a few units which don't really exist are present in the list.
	Inadequate: when it excludes all classes of the population which are to be taken in the survey.
	Incomplete: when a portion of the sampling units of the population are either totally removed or incorporated more than once.
	Out of date: when a portion of the sampling units of the population are either totally excluded when it has not been refreshed by the interest of the event, in spite of the fact that it was exact, finished and satisfactory at the hour of development

3.4 Probability vs. Non-Probability Samples

Sampling methods are categorized into two methods termed as Non-probability sampling and probability sampling.

- **Non-probability sampling methods** do not guarantee the selection of each population element will be selected with a known probability value. But, the only advantage of non-probability sampling is the convenience and cost.
- **Probability samples**. With probability sampling methods, each population element has a known (non-zero) chance of being chosen for the sample. Probability sampling permits researcher to ensure that the sample statistics is more likely to be the same as population statistics.

Sampling Method refers to the way that observations are selected from a population to be in the sample for a sample survey.

Population Parameter vs. Sample Statistic

A sample survey is conducted to estimate the value of some attribute of a population.

- **Population parameter** is the holistic and true value of an attribute of the complete population.
- **Sample statistic** is an estimate of the sample data chosen from the population data.

Example: A public opinion researcher wants to know the percentage of voters in a village who pay flat-rate income tax. The *total* percentage of all the voters is the population parameter. The *estimate* obtained for the sample data is percentage of sample data and is termed as sample statistic.

The quality of a sample statistic is strongly influenced by the way that sample observations are chosen;

that is., by the chosen sampling method. The classification of sampling methods is as shown in Fig 3.5.

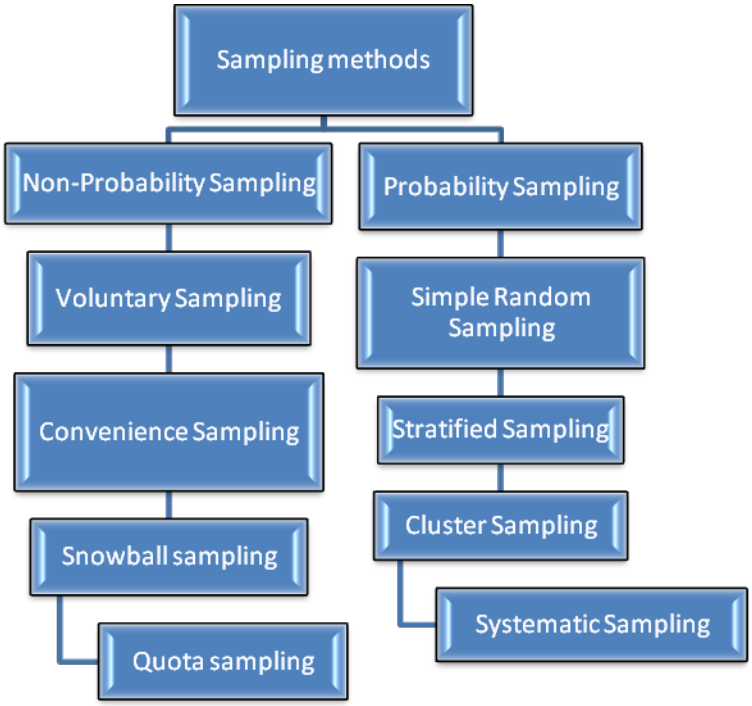


Fig 3.5: Classification of Sampling Methods

Non-Probability Sampling Methods

The main types of non-probability sampling methods are voluntary sampling, convenience sampling, snowball sampling and quota sampling. The diagrammatic representation of these four major types of non-probability sampling methods is shown in Fig 3.6.

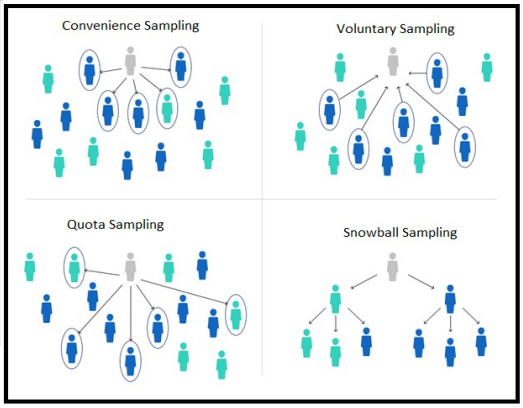


Fig 3.6: Types of Non-Probability Sampling Methods

Voluntary Sampling: Voluntary sample consists of individuals who pick themselves in the survey. These people often have a strong interest in the survey's main topic. This sampling method cannot be generalized because it is chosen from participative samples and not from representative samples.

Example: If a community radio channel intends to conduct a poll like "2 months from today is the 2019 Tamil Nadu local body elections. This would be a volunteer sample. The sample data is collected by the listeners of the radio who volunteer to answer the question. When a radio jockey announces the poll question and asks its listeners to call and vote on a particular issue, the researcher would capture the data from the participative samples. The sample participants may not be representative of the chosen issue and hence the results may be biased. The radio community reflects only the people who happen to be listening to the show and are motivated to call in.

Convenience Sampling: A convenience sample is made up of people who are easy to reach.
Example: National Rural Health Mission (NHRM) was launched by Government of India in 2005. Under this mission, if a researcher intends to study the implementation of "Best Healthy Village" scheme to improve the healthcare practices of women and children in villages. Some of the parameters in deciding the Best Healthy Village are – Children's immunization, Women health, Number of deliveries, Sex Ratio, Registration of births and deaths, Institutional deliveries, Nutritional status of children, etc. These parameters would also be periodically monitored and evaluated. Taking these parameters in to consideration a Best Healthy Village scheme can be promoted. To undertake this study, the above-mentioned parameters were collected using convenience sampling technique like selecting the conveniently reachable $m \ out \ of \ n$ (say 5/10) Taluk's in a rural district and choosing around m out of n (say 10/15) Primary Healthcare Centre's (PHC's). Once the accessible Taluks/ PHC's are selected based on convenience, the above-mentioned data of the samples may be accessed from the health records maintained in the respective PHC's.

Snowball Sampling: It is also called as chain or referral sampling which relies on the initial respondents to refer to the next respondents. Just as the snowball rolls and gains mass, as you move through the survey process, the sample constructed in this way will grow in size.In this technique, you rely on your initial respondents to refer you to the next respondents whom you may connect with for the purpose of your survey.
Example: To conduct a survey to find the average daily wages of unskilled laborers Average Daily Wage Rate (DC) in Rural in Non-Agricultural sector like Weavers and Beedi Makers in a village can be done by snowball sampling, which would be a quick, accurate way to create the sample. The researcher first finds a first unskilled labour and then through him gets the referral of the next person and so on. This sampling method is usually restricted to a small, homogenous section of population.

Quota Sampling: Quota Sample starts with characterizing the population based on certain desired features and assigns a quota to each subset of the population.
Example: If a survey requires a sample of fifty men and fifty women, a quota sample will survey respondents until the right number of each type has been surveyed. Unlike stratified sampling, the sample isn't necessarily randomized. If a researcher intends to survey a village population to collect information about almost forgotten plant uses, interviews were scheduled with about 100 elderly respondents of a chosen village, where in 50 chosen respondents were elderly men and 50 were elderly women.

Probability Sampling Methods

The main types of probability sampling methods are simple random sampling, stratified sampling, cluster sampling, and systematic random sampling. The key benefit of probability sampling methods is to ensure that the sample selected represents the population. It guarantees the integrity of the statistical results.

Simple Random Sampling

When: There is a very large population and each member of the population is hard to identify.

How: The entire sampling process is performed in a single step with each selected subject independentl y of the other population members. The term random has a very precise meaning and on the street, you cannot just collect answers and get a random sample. Generate a number that is random to select a sample from a given population.

Pros: Each member of the population has an equal chance of being chosen as a subject in this technique.

Cons: It is often difficult to identify every member of the population when there are very large populatio n and the sample of subjects is skewed.For example, dialing numbers from a phone book may not be co mpletely random as the numbers may correspond to a particular area, though random.

Example: Want to study and understand the rice consumption pattern across a very small village? While it might not be possible to cover every household in the village, you could draw meaningful insights by building your sample from different parts of the village. If there are about 16 households in the chosen village, the population is represented as R1 to R16. Each household is characterized as majority rice consumer “R” or majority wheat consumer “W”. The population from the set of all households of the chosen village comprises the population statistics (N=16 in the Figure given below), a sample statistic is chosen at random (n=5). The random number between 1 to 16 is generated to select a sample. The random number can be generated using a computer or each value [1,2,3…16] are written on a paper and randomly selected by picking up a value one at a time.

To Do Activity

In a small village of 1000 farmers, if a Taluk officer decides on providing free training to the villagers on effective farming methods, it is highly likely that they would prefer picking chits out of a bowl. In this case, each of the 1000 farmers has an equal opportunity of being selected. What sampling method would be used in this case?

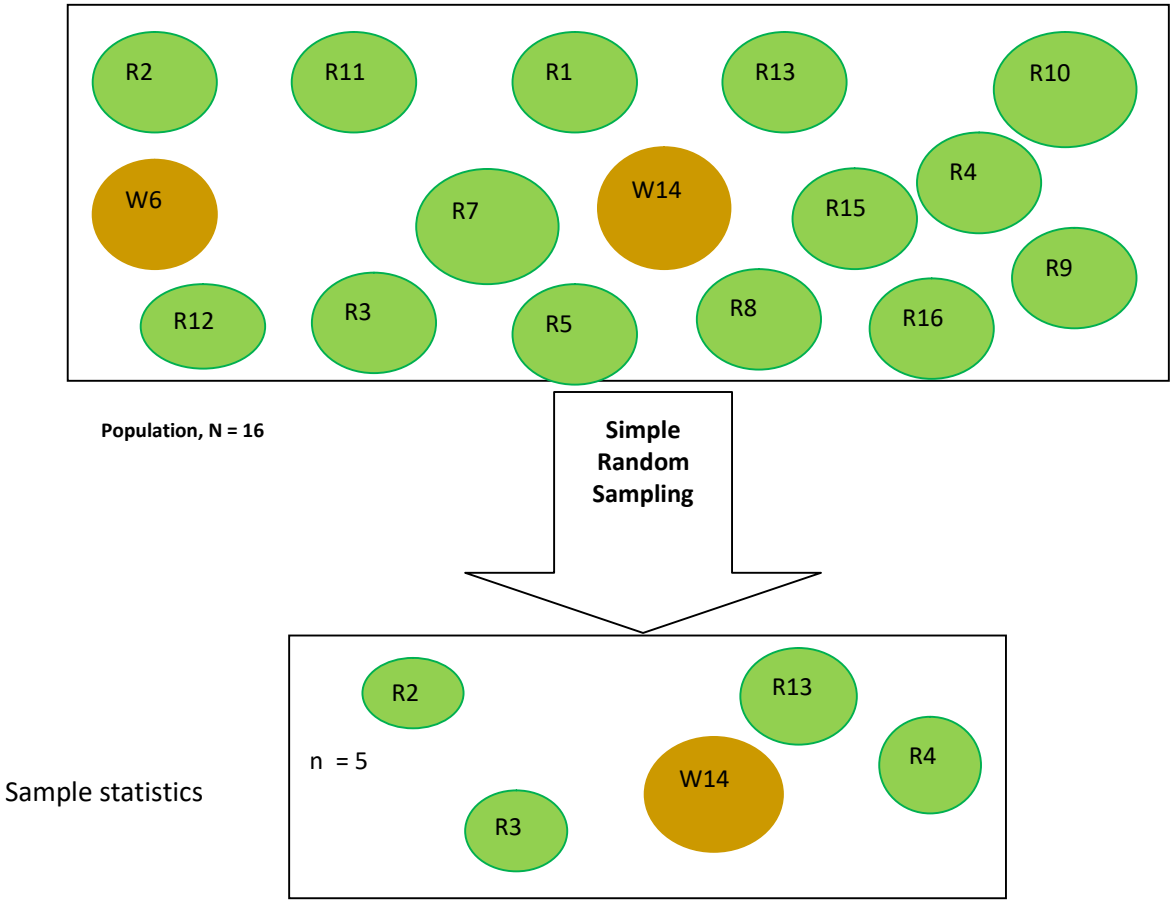


Fig 3.7: Simple Random Sampling for studying rice consumption pattern

Stratified Sampling

When: When: You can divide your population into research-related characteristics.

How: In fact, a stratified sample tries to reproduce the population's statistical characteristics on a smalle r scale.Until sampling, the population is divided into researchrelevant characteristics, such as gender, so cial class, level of education, religion, etc.Instead, within each group or stratum, the population is sampl ed randomly.

Pros: This method attempts to overcome the weaknesses of random sampling by dividing the populatio n into different segments or strata

and selecting entities from each.It means that the sample contains every segment of the population.Stra tified sampling is often used when there is a low incidence of one or more sections in the population co mpared to the other sections.

Cons: Stratified sampling is the sampling form that is the most complex. This sets out expectations that might be difficult to meet and put a heavy strain on your available resources.

Example: A researcher intents to study the reach of different schemes under making of Digital India ina rural district of Maharashtra. If in a chosen rural district of Maharashtra, 35% of the population is college-educated and 65% of the population have not been to college(U), then 35% of the sample is

randomly selected from the college-educated (E) subset of the population and 65% of the sample is randomly selected from the non-college-going population. Maintaining the ratios while selecting a randomized sample is key to stratified sampling for this study. As a first step the given population is divided into two strata's viz Educated and Uneducated strata and in the second step the sample is chosen from each strata at random in the ratio of E to U as 1:4 as shown in Fig.

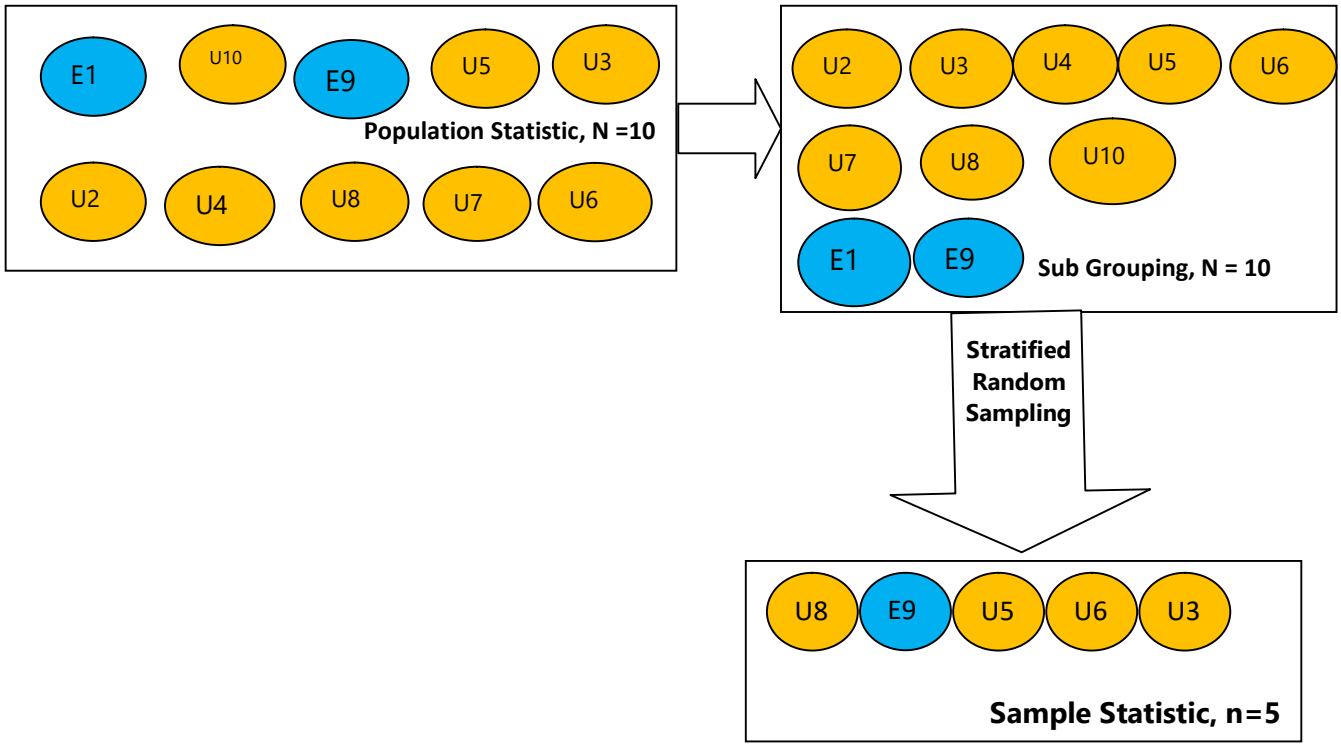


Fig 3.8: Stratified Random Sampling to study the reach of Digital India schemes in a rural district

Cluster Sampling

When: With cluster sampling, every member of the population is assigned to one, and only one, group.
How:A cluster is named for each group A set of clusters is picked using a system of chance (often simple random sampling). Also, individuals are surveyed within the sampled clusters.
Pros: For stratified sampling, the sample includes elements of each stratum.In comparison, the study incl udes elements from sampled clusters only with cluster sampling.Observing clusters of units in a populati on is more economical than randomly selected units distributed across the entire population.

Cons: If there is a biased opinion of the population group chosen as a cluster sample then it is inferred th at the entire population has the same opinion. The other probabilistic methods give less sampling error than cluster sampling.

Example: The prevalence of goitre was highest in Maharashtra (11.9%) and West Bengal (9%) as per the survey conducted by the National Nutrition Monitoring Board (NNMB). If a researcher intends to study the iodine deficiency in a particular district, then, the researcher can start with collection of the following data from the census data of Nashik, a district in Maharashtra, which is chosen for the study.

- Population of the district (6,534,981 as on 2019)
- Number of villages in the district is 73.
- Village wise population from the census records (Table)

From the above listed data, the researcher can identify the sample clusters for the study as shown in Fig which illustrates Cluster Sampling of the villages of Nashik district, Maharashtra to study prevalance of iodine deficiency in people.

Table3.2: Village wise population statistics of Nashik district, Maharashtra

Cluster Id	Village Name	Village Population	Cluster Id	Village Name	Village Population	Cluster Id	Village Name	Village Population
1	Ambebahula	1,640	26	Jakhori	2,699	51	Pimpri Sayyad	10,353
2	Babhaleshwar	1,149	27	Jalalpur	2,940	52	Rahuri	1,190
3	Belatgavhan	2,653	28	Jategaon	1,639	53	Raigadnagar	1,098
4	BelgaonDhaga	2,842	29	Kalavi	698	54	Rajewadi	1,239
5	Bhagur Rural	276	30	Kashyapnagar	879	55	RajurBahula	1,746
6	Chandgiri	1,385	31	Kotamgaon	1,575	56	Sapgaon	1,167
7	Chandshi	2,047	32	Ladachi	3,108	57	Sarul	2,002
8	Dahegaon	1,573	33	Lahvit	6,826	58	Sawargaon	722
9	Dari	2,011	34	Lakhalgaon	4,282	59	Shastrinagar	1,818
10	Devargaon	3,714	35	Lohashingwe	1,533	60	Shevgedarna	1,769

11	Dhondegaon	2,263	36	Madsangavi	4,941	61	Shilapur	2,436
12	Donwade	1,130	37	Mahadeopur	1,708	62	Shinde	8,498
13	Dudgaon	1,380	38	Mahirawani	3,169	63	Shivangaon	1,049
14	Dugaon	3,387	39	Manoli	599	64	Subhashnagar	1,366
15	Ganeshgaon Naik	1,191	40	Matori	2,843	65	Sultanpur	600
16	GaneshgaonTrimbak	1,707	41	Mohagaon	1,115	66	Talegaon Anjaneri	2,503
17	Gangamhalungi	1,834	42	Mungsare	2,680	67	Tiradshet	1,008
18	Gangapadali	960	43	Nagalwadi	1,105	68	Vadgaon	2,865
19	Gangavhare	1,381	44	Naikwadi	1,779	69	Vaishnavnagar	963
20	Gaulane	2,122	45	Nanegaon	3,667	70	Vanjarwadi	2,202
21	Girnare	6,013	46	Odha	3,321	71	Vinchurgavali	1,747
22	Govardhan	5,997	47	Ozarkhede	2,877	72	Wasali	1,342
23	Govindpuri	394	48	Palashe	11,721	73	Yashawantnagar	849
24	Hinganwedhe	1,488	49	Pimpalgaon Garudeshwar	1,501	Data Source: https://data.gov.in/		
25	Indiranagar	1,864	50	Pimplad Nasik	3,810			

To Do Activity

Suppose that the Department of Agriculture wishes to investigate the use of pesticides by farmers in a state. A cluster sample could be taken by identifying the different states in India as clusters. A sample of these states (clusters) would then be chosen at random, so all farmers in those counties selected would be included in the sample. It can be seen here then that it is easier to visit several farmers in the same county than it is to travel to each farm in a random sample to observe the use of pesticides. What sampling method is used in this case study? Explain.

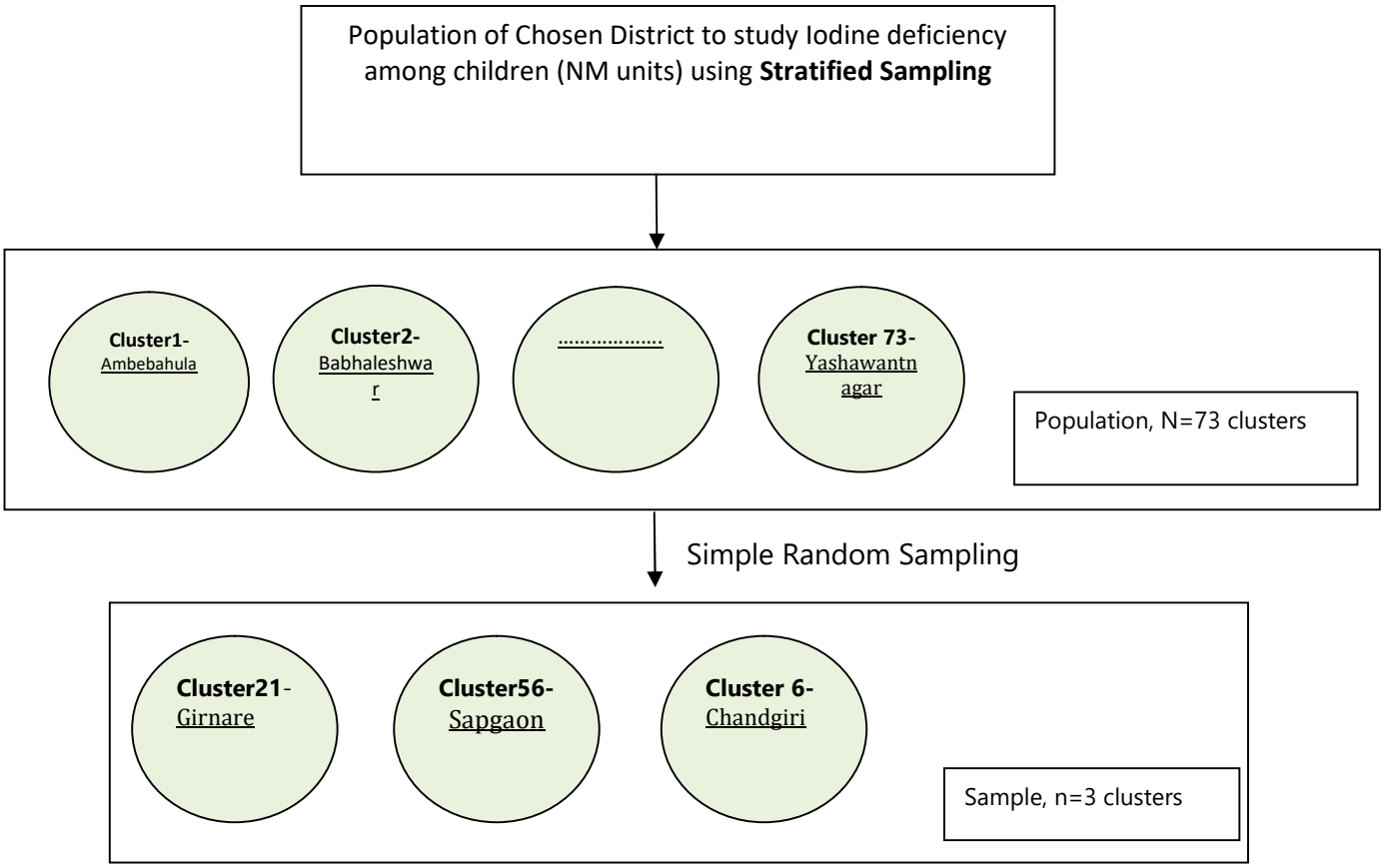


Fig 3.9 Stratified Sampling

When: Your given population is logically homogenous.

How: In a systematic sample, after you decide the sample size, arrange the elements o the population in some order and select terms at regular intervals from the list.

To Do Activity

Suppose a farmer wishes to work out the average milk yield of each cow type in his herd which consists of Ayrshire, Friesian, Galloway and Jersey cows. He could divide up his herd into the four sub-groups and take samples from these. What sampling procedure has he used here?

Pros: The simplicity is the biggest advantage of using systematic sampling. The certainty that the populat ion will be distributed uniformly is another benefit of systematic random sampling.
Cons: The possible weakness of the method that may compromise the randomness of the sample is an inherent periodicity of the list. This can be avoided by randomizing the list of your population entities before proceeding with systematic sampling.

Example: Let us consider the previous example of the study the iodine deficiency in the rural villages of Nashik district of Maharashtra (Refer to Table 3.2). There are 73 villages (population N =73) from which we intend to choose a sample of 3 villages (sample, n = 3) for the study.
The sampling interval is calculated as s = N/n = 73/3 = 25.66
Now, randomly choose a number between 1 and 25 and they choose 11. This number is the equivalent to the 1st village sample (cluster) to survey and the subsequent samples are chosen as given in Table:

Table 3.3: Calculation of Village Numbers using decimal sample intervals

Sample No of Villages (cluster) chosen	Calculation	Total	Village chosen by Systematic random sampling
1	11	11	11
2	11 + 25.66	36.66	37
3	36.66+25.66	62.32	62

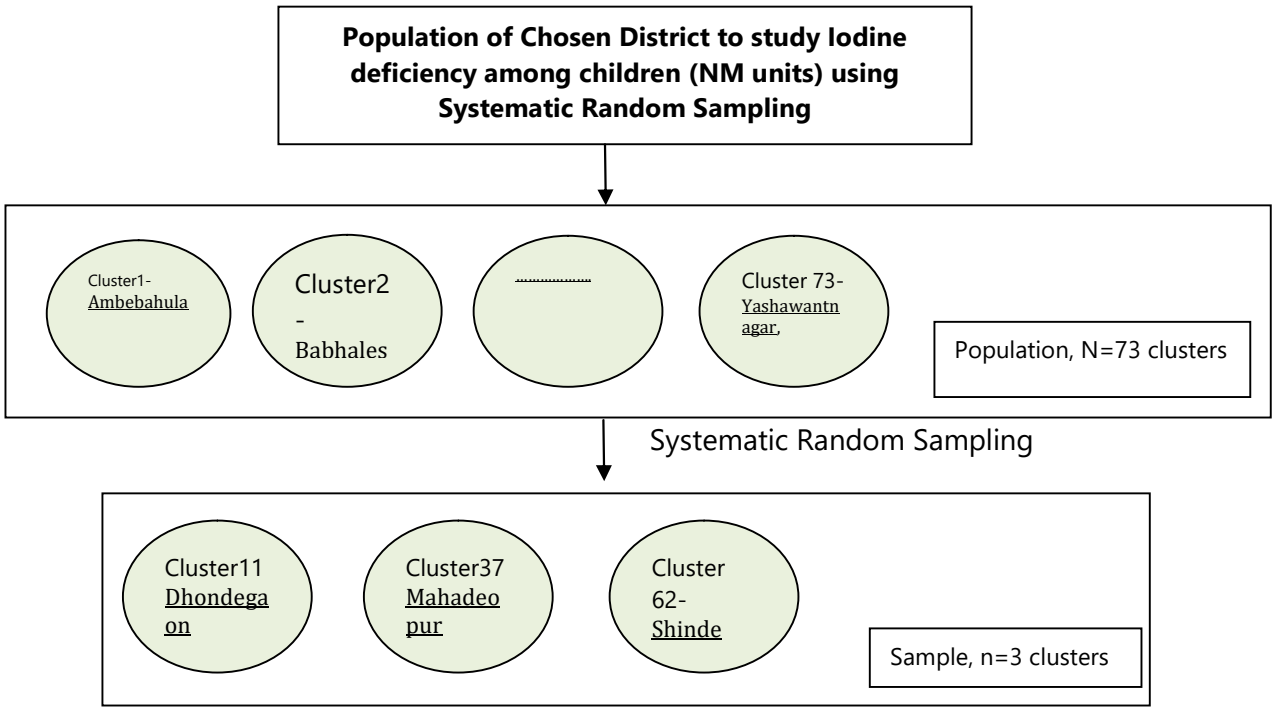


Fig 3.10: Systematic Random Sampling

To Do Activity

If the Government of India wishes to evaluate the number of immigrants living in India, they can divide it into clusters on the basis of states such as Assam, Guwahati, Kolkatta, etc. This way of conducting a survey will be more effective as the results will be organized into states and provides insightful immigration data.

To Do Activity

Analyze characteristics of people belonging to different annual income in a village. Create strata (groups) according to annual family incomes such as Less than Rs.30,000, Rs.30,000-Rs.60,000, Rs.60,000- Rs.1,00,000, etc. and draw conclusions of which income strata have which characteristics. Marketers can analyze which income groups to target for a particular product.

3.5 Metrics for Errors

A sample is a part of a population. A sample may not represent all characteristics of a population may not be present in the sample drawn from the population. Any statistical measure say mean of the sample, may not be the same as the mean of the population from which the sample was drawn. These discrepancies are known as Errors in sampling. Errors in sampling are of two types.

Sampling Errors: They are inherent in the method of sampling. Errors in sampling primarily occur due to the following reasons:

- Faulty selection of the sample: This is due to faulty selection of sampling technique for a particular research.
- Substitution: Sometimes a researcher while collecting the information from a particular sampling unit, substitutes another person and as a result this may lead to some bias as the characteristics possessed by the sampling unit may be different from the substituted unit.
- Faulty demarcation of the sampling units may also lead to the errors.
- Variability of population or the high heterogeneity of the population may also lead to sampling errors.

Non-Sampling Errors or Bias automatically creep in due to human factors which always varies from one investigator to another. Bias may arise in the following ways:

- Due to negligence of researcher
- Due to faulty selection of samples
- Due to improper sample survey
- Due to wrong questionnaire framing.
- Due to wrong response from the respondents
- Due to error in compiling
- Due to applying wrong statistical measure.

Standard Error of a Statistic

The statistical measure of standard deviation may be computed both from the observations of the population and also from the sample. Since, standard deviation is a measure of the average amount of

the variability of all observations of variable from their mean.

When the average amount of the variability of the observations of a population is computed, it is called the standard deviation. But, when the average amount of variability of the observations of a sampling distribution is computed, it is known as Standard Error.

To recall the formula for Variance and Standard deviation for a given population
The sample variance for a sample of n **ungrouped** measurements is equal to the sum of the squared distances from the mean divided by (n - 1). In symbols, using s^2 to represent the variance of population

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

Variance = s^2 =
In calculating s^2, we divide by (n – 1) instead of n to get a better estimate of the population variance.
The sample standard deviation, s, is defined as the positive square root of the sample variance, s^2.

$$s=\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$$

where $\bar{x} = mean$ of population.

To compute Standard Error of sample , given the population is an ungrouped data, we get,

$$SE = \frac{s}{\sqrt{n}}$$

i.e.

$$SE = \frac{\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}}{\sqrt{n}},$$ where $\bar{x} = mean$ of sample and n is the sample size.

The standard error is just the standard deviation divided by the square root of the sample size. Hence, the standard error becomes smaller when the sample size becomes larger because the standard deviatio n is the same.

To do activity

Given that you want to go to a neighbouring fast food restaurant during your lunch hour and asks people who walk into the restaurant if they would be willing to fill out a questionnaire. Which of the following methods of sampling among the following will you use? Why?

(i) Convenience sampling
(ii) Judgemental sampling
(iii) Quota sampling
(iv) Probability sampling

Example: Given the population of all villages of Udhagamandalam, Nilgiris District. Compute the mean population of all villages, the variance and standard deviation. Using Simple random sampling, select a sample of 3, 5 villages and calculate corresponding standard error of each sampling.

#	Villages	Population
1	Ebbanad	6,233
2	Hullathi	9,187
3	Kadanad	7,080
4	Kagguchi	10,986
5	Kukkal	5,219
6	Masinagudi	8,783
7	Nanjanad	13,007
8	Thummanatti	9,461
9	Thuneri	7,110
10	Udhagamandalam	8,214

(i) Calculation of Variance, Standard Deviation of the Population (N = 10 villages)

Sl	Villages	Population, x	$x-\bar{x}$	$(x-\bar{x})^2$
1	Ebbanad	6,233	-2,295	5267025
2	Hullathi	9,187	659	434281
3	Kadanad	7,080	-1,448	2096704
4	Kagguchi	10,986	2,458	6041764
5	Kukkal	5,219	-3,309	10949481
6	Masinagudi	8,783	255	65025
7	Nanjanad	13,007	4,479	20061441
8	Thummanatti	9,461	933	870489
9	Thuneri	7,110	-1,418	2010724
10	Udhagamandalam	8,214	-314	98596
		$\bar{x} = mean$ **=8528**		Sum = 47895530

$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$ = 47895530/9 = 5321726, s = $\sqrt{21726}$ = 2306.887

(ii) Using Random sampling, if 3 villages are chosen at random, the sample mean, variance and standard error of the sample is calculated as shown below:

SI	Villages	Population, x	$x-\bar{x}$	$(x-\bar{x})^2$
1	Hullathi	9,187	611	373321
2	Kadanad	7,080	-1,496	2238016
3	Thummanatti	9,461	885	783225
		$\bar{x} = mean$ =8576		Sum =3394562

Standard Error of Sample statistic is given as below:

$$SE = \frac{\sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}}{\sqrt{n}}, \text{ SE} = \frac{\sqrt{3394562/2}}{\sqrt{3}} = 752.17$$

(iii) Using Random sampling, if 5 villages are chosen at random, the sample mean, variance and standard error of the sample is calculated as shown below:

SI	Villages	Population, x	$x-\bar{x}$	$(x-\bar{x})^2$
1	Ebbanad	6,233	-2,650	7022500
2	Kadanad	7,080	-1,803	3250809
3	Kagguchi	10,986	2,103	4422609
4	Nanjanad	13,007	4,124	17007376
5	Thuneri	7,110	-1,773	3143529
		$\bar{x} = mean$ =8883		Sum = 34846823

$$\text{SE} = \frac{\sqrt{34846823/4}}{\sqrt{5}} = 1319.978$$

Example: For the given population of 10 villages, the Standard deviation, s = 2306.887. If you have two data sets with the same standard deviation, where in data set 1 has 3 samples and data set 2 has 5 samples. For each of these calculate the standard error. What is your observation?
Given, for n = 10, s = 2306.887.

(i) For data set 1, n = 3, $= \frac{s}{\sqrt{n}}$, $= \frac{2306.887}{\sqrt{3}}$ = 1331.882

(ii) For data set 2, n = 5, $= \frac{s}{\sqrt{n}}$, $= \frac{2306.887}{\sqrt{5}}$ = 1031.671

Observation: When the sample size gets larger the standard error gets smaller because the standard deviation is the same.

Summary of the Chapter

Summary: The chapter presents a detailed overview of the Probability and Sampling. The concepts essential for understanding the importance of Probability are discussed in detail alongwith appropriate examples and case studies. The chapter also explains the concept of sample space and sample frame and the diffrentiattion between probabilistic and Non-Probabilistic Samples are illustrated using appropriate tools and techniques. The chapter also presents the various methods of sampling as well as basic error calculation metrics.

Model Questions

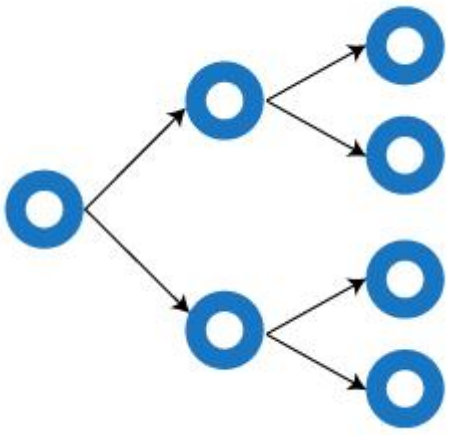
1. A coin is tossed. If it shows head, we draw a ball from a bag consisting of 5 red and 6 green balls; if it shows tail, we roll a die. What is the sample space associated to this experiment?
2. An experiment consists of Rolling a die and then tossing a coin once if the number on the die is odd. If the number on the die is even, the coin is tossed twice. Write the sample space for this experiment.
3. The probability that due to rains the college will close tomorrow is 0.75. What is the probability that the college will remain closed tomorrow?
4. A card is drawn at random from a pack of 52 cards. Find the probability that the card drawn is
    a. a black King
    b. either a black card or a king
    c. black and a king
    d. a Jack Queen or a king
    e. A heart
    f. the ten of clubs
    g. neither a red card nor a queen
    h. the king of Hearts
    i. an ace
5. A bag contains 3 red balls and 5 black balls. A ball is drawn at random from the bad. What is the probability that the ball drawn is :

a. red
b. not red
c. A researcher intends study the most preferred color of females for their clothing. To do this, what sampling technique would you propose and why?

6. An analyst develops the following table of joint probabilities relating the size of firm (measured in terms of the number of employees) and the type of firm.
7. If one is selected at random, find the probability of the following events:
    a. P(A) = The firm employs fewer than 20 employees
    b. P(B) = The firm is in the retail
    c. P(C) = A firm in the construction industry employs between 20 to 99 workers
    d. P(D) = A firm in the retail industry employed more than 20 workers.

Number of Employees	**Industry**		
	Construction	Manufacturing	Retail
Under 20	0.2307	0.0993	0.5009
20 – 99	0.0189	0.0347	0.0876
100 or more	0.0019	0.0147	0.0113

8. After careful examination of the figure given below, mention what sampling method is used for the study?



9. An agro company was in business of processing of basmati Rice, manufacturing of sandal wood oil and producing of castor oil. The Company had started availing credit facilities from the banks under consortium arrangement led by one of the banks. After watching the modus operandi of

the company for one year, the bank had a suspect and had appointed a Head of the Investigation Team for a suspected fraud in a company in a fiscal year. The head was provided with 100,000 journal entries that were entered during the period of suspicion. Without examining all of the transactions in the interest of time to complete the investigation, how do you study the given data and write a report on the investigation. Justify the sampling methodology chosen for the investigation.

a. Sales Vouchers = 70,000
b. Purchase Invoices =12,000
c. Salary Invoices =7,500
d. Purchase of Fixed Assets = 500
e. Others = 10,000

10. You are asked to survey your hometown about the habit of drinking alcohol among the residents. Will you go around the town completely and collect the data or meet a few households around you and ask them about those who drink alcohol? It will not exactly reflect the characteristics of your town, yet the researcher can gain some insights about some major characteristics. Explain what type of sampling technique is used here.
11. A researcher intends to study the price of petrol in 4 major metropolitan cities in India for 6 days as given in table below:

Month			Mumbai	Chennai
January 03, 2020	**75.35**	**7.94**	**80.94**	**78.28**
January 02, 2020	75.25	77.87	80.87	78.20
January 01, 2020	75.14	77.79	80.79	78.12
December 31, 2019	75.14	77.79	80.79	78.12
December 30, 2019	75.04	77.70	80.69	78.02
December 29, 2019	74.88	77.54	80.53	77.85

a. Find the mean price in each city and then state which city has the lower mean.
b. Find the standard deviation of each city's prices
c. Which city has the more consistently priced petrol? Give reasons for your answer.

12. Given the state wise population (classified into Urban and Slum population).
    a. Compute the mean population of all states in India, the variance and standard deviation.
    b. Compute the mean population of all Union territories, the variance and standard deviation.

13. Using Simple random sampling, select a sample of 5, 10,15 states and calculate corresponding standard error of each sampling respectively. What is your observation?

14. Use stratified sampling divide the states of India into North, South, East and West Strata and select 3 states from each strata using random sampling and then calculate the standard error of the urban and slum population respectively. What can you infer?

S. No.	States/UTs	Urban population	Slum population	S. No.	States/UTs	Urban population	Slum population
India	India	377106125	65494604	19	Nagaland	570966	82324
1	Andhra Pradesh	28219075	10186934	20	Odisha	7003656	1560303
2	Arunachal Pradesh	317369	15562	21	Punjab	10399146	1460518
3	Assam	4398542	197266	22	Rajasthan	17048085	2068000
4	Bihar	11758016	1237682	23	Sikkim	153578	31378
5	Chhattisgarh	5937237	1898931	24	Tamil Nadu	34917440	5798459
6	Goa	906814	26247	25	Tripura	961453	139780
7	Gujarat	25745083	1680095	26	Uttar Pradesh	44495063	6239965
8	Haryana	8842103	1662305	27	Uttarakhand	3049338	487741
9	Himachal Pradesh	688552	61312	28	West Bengal	29093002	6418594
10	Jammu & Kashmir	3433242	662062	29	A & N Islands	143488	14172
11	Jharkhand	7933061	372999	30	Chandigarh	1026459	95135
12	Karnataka	23625962	3291434	31	D & N Haveli	160595	NS
13	Kerala	15934926	202048	32	Daman & Diu	182851	NS
14	Madhya Pradesh	20069405	5688993	33	NCT of Delhi	16368899	1785390
15	Maharashtra	50818259	11848423	34	Lakshadweep	50332	NS
16	Manipur	834154	NS	35	Puducherry	852753	144573
17	Meghalaya	595450	57418				
18	Mizoram	571771	78561				

Data Source: https://data.gov.in/

References and Further Reading

1. Ben-Shlomo Y, Brookes S, Hickman M. 2013. Lecture Notes: Epidemiology, Evidence-based Medicine and Public Health (6th ed.), Wiley-Blackwell, Oxford.

2. http://www.stats.gla.ac.uk/steps/glossary/sampling.html

Chapter 4 PRA and PLA

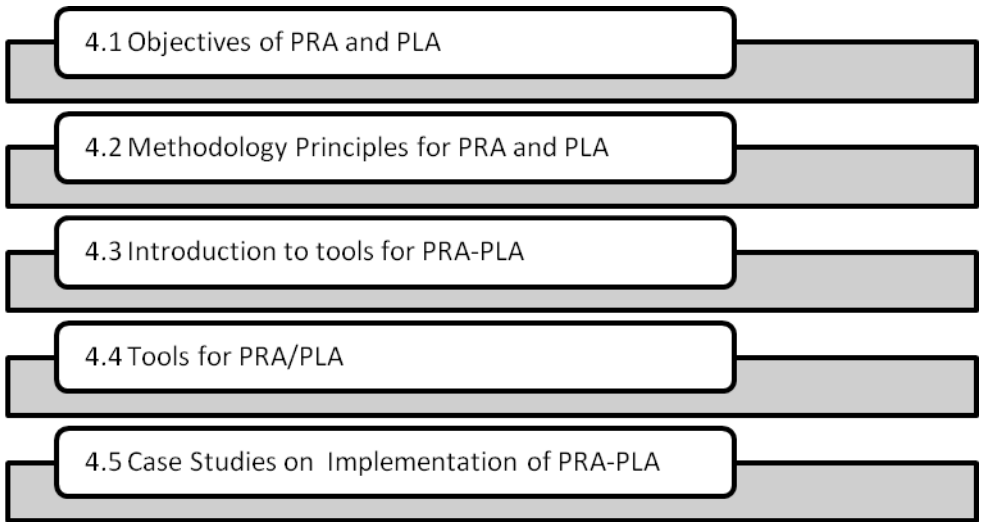
Introduction

Participatory Rural Appraisal (PRA) and Participatory Learning and Action (PLA) are approaches utilized by development practitioners to empower powerful association and arranging with communities. The approaches are supported the philosophy of bottom-up participation and empowerment and recognize that if local people participate within the development processes of designing, implementation and monitoring, they will progressively transform their own lives and surrounding environment.

Objectives of the Chapter

- To familiarise the students with the concept of PRA and PLA
- To discuss the methodology and principles of PRA and PLA
- To introduce various tools used for PRA-PLA
- To present an introduction to Seasonal Calendar, Venn- Diagram and Focused Group Discussions
- To study a case-study on PRA-PLA

Structure of the Chapter



4.1 Objectives of Participatory Rural Appraisal (PRA) and Participatory Learning and Action (PLA)

The Origins and Growth of PRA-PLA

The Brazilian teacher, Paulo Freire, first presented the idea of "conscientization" in his 1968 book, "The Pedagogy of the Oppressed". He alluded to the way toward raising individuals' mindfulness through aggregate self-request and reflection. Freire recommended that by taking a critical look at their own nearby circumstance, individuals would have the option to take activities towards the improvement of their general public. At roughly a similar time, improvement specialists were starting to challenge the traditional top-down, concentrated model of advancement and to advocate for the support of poor, underestimated provincial and urban networks in the improvement procedure.

During the 1970s, improvement professionals presented Rapid Rural Appraisal (RRA), a system for the fast assortment of information inside rustic social orders. RRA enabled advancement specialists to

increase an increasingly complete comprehension of the intricacy of the nearby level circumstance. Participatory Learning and Action began in the southern side of the equator in the 1970s, as a technique for meeting and engaging with village committees in developing nations, particularly in Africa. The principle behind its creation was the same as mentioned above; that poor and exploited individuals can and ought to be empowered to examine their own world, also, to look at their own issues, set their own objectives, and screen their own accomplishments. Basically, it depended on strengthening, for example the promise to helping individuals assume greater responsibility for their lives. To guarantee that individuals were not avoided from cooperation, it built up a number of strategies and methods that abstained from writing as and when possible(due to the fact that literacy rate in villages or rural areas is not that great), depending on the tools of, for example, pictures, physical items and gathering memory.

During the 1980s, PRA rose, followed intently by PLA during the 1990s. Robert Chambers has been a key example of participatory methodologies furthermore, the possibility that the individuals from whom information is customarily gathered could really be perceived as authentic "knower's" and in this way be dynamic members in the exploration procedure. The name PLA further underscores the utilization of the participatory procedure for network learning and activity arranging, yet the reasoning and methods of PRA and PLA are essentially the equivalent.In the last 15 to 20 years the utilization of PLA systems has become progressively well known in the northern side of the equator. It has been received as a way of enabling nearby individuals to evaluate and assess their very own networks, and to distinguish their very own answers for a scope of issues inside nearby networks. Fig. 4.1 illustrates the Growth of PRA/PLA.

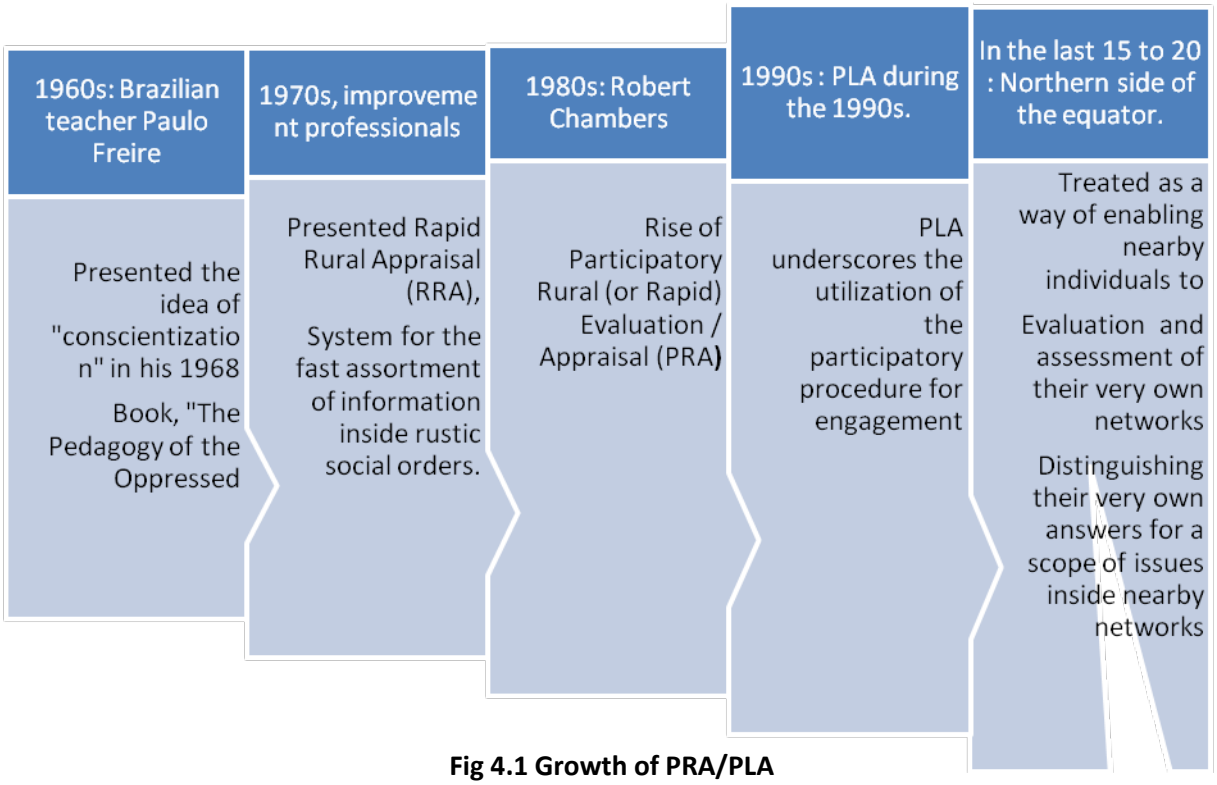


Fig 4.1 Growth of PRA/PLA

Participatory Rural (or Rapid) Evaluation (PRA)& Participatory Learning and Action (PLA)

Participatory Learning and Action (PLA) is a kind of subjective research, which can be utilized to increase an inside and out comprehension of a network or circumstance. It is generally utilized in work including neighborhood networks. PLA is a participatory philosophy, and ought to consistently belled with the full and dynamic cooperation of network individuals. The principle motivation behind PLA is to bolster individuals inside networks to break down their own circumstance, as opposed to have it broke down by outcasts, and to guarantee that any learning is then converted without hesitation.

PLA was initially called Participatory Rural (or Rapid) Evaluation (PRA). It turned out to be mainstream during the 1980s and 1990s, and has since kept its ubiquity with numerous CSOs. PRA was initially intended for use during examinations and needs evaluations in country territories. Nonetheless, it tends to be utilized at any phase of the venture cycle – structure, arranging, observing, survey and assessment – and is currently utilized in urban just as country regions. The name was changed to Participatory Learning and Action (PLA) to mirror its more extensive application, and to accentuate that the procedure is intended to help set moving privately drove activity.

PLA can be portrayed in two extraordinary, reciprocal ways. On one hand, it is a way of thinking and a perspective that underlines inversions in power relations between insiders and outsiders within a community, (for example, scientists, evaluators or then again program organizers). On the other hand, it covers a scope of participatory instruments and approaches that can be utilized to work, design and reflect with and nearby networks.

Major Objective of PRA-PLA

The objective of both PRA and PLA is to stimulate a common taking in process where individuals from outside a network, encourage a procedure with individuals from inside the network to accumulate and break down neighborhood information and figure plans for activity. The methodologies help the essential partners, frequently poor or rustic networks, to assume responsibility for the procedure, to evaluate neighborhood level issues, to discover approaches to distinguish and organize issues and challenges and to draw up handy activity intends to address the difficulties.PLA is situated inside a more extensive field of participatory approaches, which can be portrayed as given in Fig 4.2.

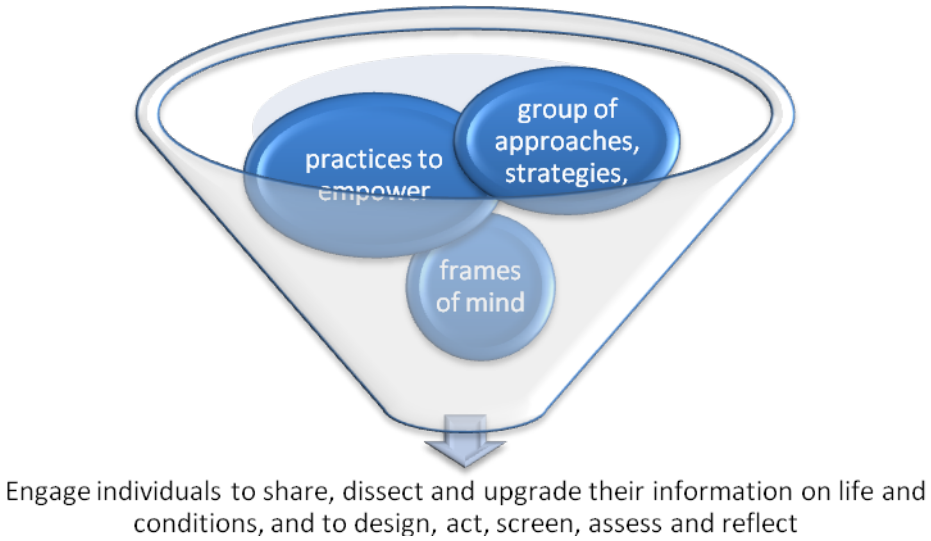


Fig 4.2: Objectives of PRA-PLA

PRA- PLA can be depicted in two extraordinary, integral ways as shown in Fig 4.3. It can either be treated as a philosophy and a perspective that stresses inversions in power relations among communities and outsiders, (for example, specialists, evaluators or program organizers). It may also cover a variety of participatory tools and approaches that can be utilized to work, design and reflect with and close by communities.

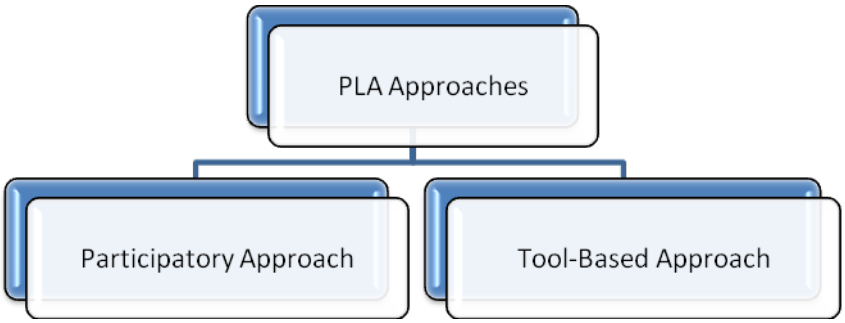


Fig 4.3: Approaches to PRA-PLA

PRA- PLA as a philosophy

In its unadulterated structure, PLA is a philosophy which underscores the requirement for outsiders to find out about circumstances from insiders. This theory looks to turn around power relations between communities and outsiders. It developed somewhat in response to the top-down arranging procedures of the 1960s and 1970s. The essential thought of PLA is that networks are bolstered to break down their very own circumstance, decide about how to best handle their issues, and, subsequently, feel enabled to make a move.

Principles of Participatory PLA

Participatory approaches to deal with learning are dynamic methodologies that urge individuals to have an independent mind. Members effectively add to instructing and adapting, instead of inactively getting data from outside specialists, who might not have neighborhood comprehension of the issues. The methodology urges individuals to share data, gain from one another, and cooperate to take care of basic issues. As individuals become progressively experienced with the methodology, they assume expanding liability for arranging their own learning sessions. They figure out how to cooperate in a gathering. They likewise gain involvement with utilizing the exercises and visual instruments to do their own hands on work.

Members can take what they have realized back to their very own associations and networks, and keep on utilizing facilitator procedures and participatory apparatuses locally. Participatory adapting likewise eventually gives individuals a system of aptitudes that they can use in any circumstance to investigate issues and make a move.

Essence of Using Participatory Approach

Participatory approaches are utilized in circumstances where various individuals must cooperate to determine a typical issue. Great critical thinking requires contribution from an assortment of individuals with numerous sorts of understanding and aptitude. It likewise incorporates everybody who is keen on finding the best arrangement — the partners. Experience shows that when everybody adds to the learning procedure, at that point individuals feel more responsibility for issue and grow progressively fitting answers for their specific circumstance.

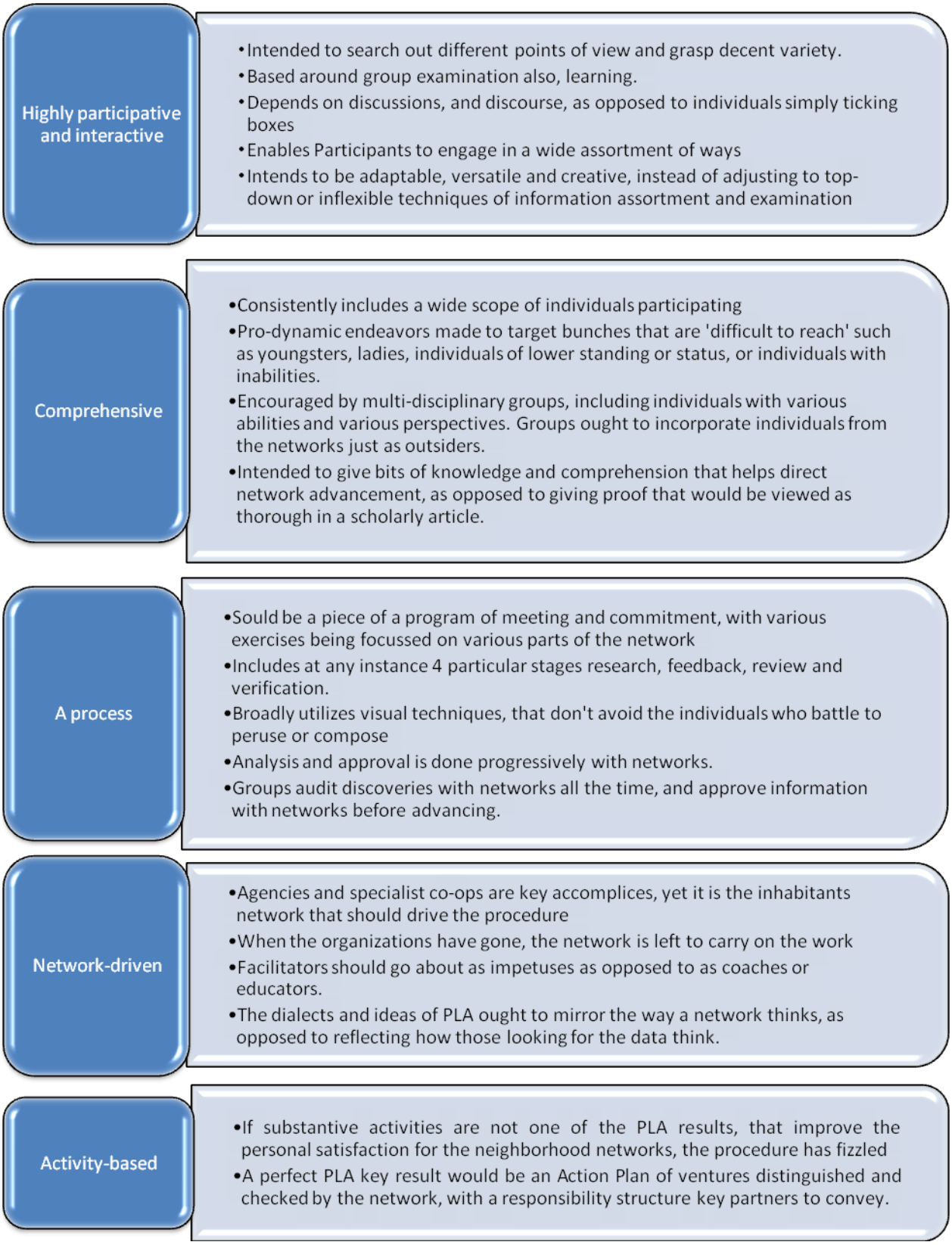


Fig 4.4: Principles of Participatory Approach

The following are the important ground rules of PLA.

- Try not to command and be on a similar level as the members
- Offer everybody the chance to take an interest
- Regard what individuals state
- 'Hand over the pen'
- Pick up from your mistakes
- Try not to surge

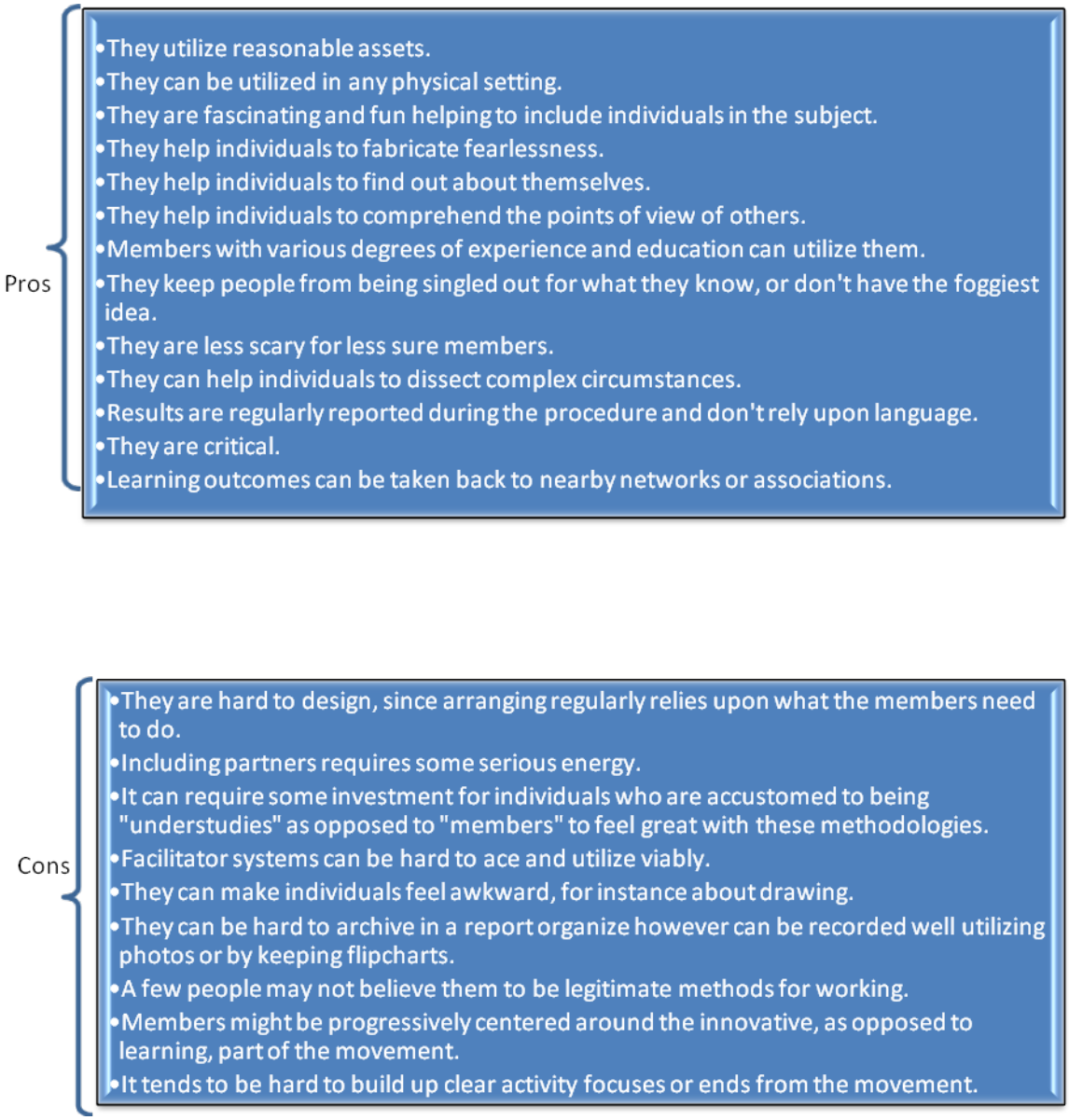


Fig 4.5: Pros and Cons of Participatory Approach to PLA

4.2 Methodology Principles for PRA and PLA

PLA uses a wide extent of systems and techniques, and would subsequently have the option to change in accordance with which bundle is being directed, or to which issue is being tended to. It is an imaginative learning process, which outfits close by people with the aptitudes and conviction to fill in as a comparable partner with associations, expert communities and various accomplices.

It might be isolated from progressively ordinary research strategies on the reason it:

- Is continually a technique, never an event.
- It involves using an extent of instruments rather than just one, as such allowing procedures to be changed in accordance with suit the issue or group.
- Is an instinctive, rather than extractive method.
- Aims to put neighborhood people at the center of the system organizing process

The distinct phases of PLA are outlined in Fig 4.6 and each phase is explained below.

Fig 4.6 Distinct Phases of PLA

Training, Research and Data Collection

In PLA research is done by individuals from the neighborhood network. In numerous PLA extends the procedure starts with the conveyance of PLA preparing to nearby occupants, to furnish them with the information and certainty to attempt the network-based research. The courses can change long from one day to a half year, the more inside and out the preparation the more prominent the degree of interview that the network analysts will have the option to do. A segment of the preparation will constantly be 'at work' with members 'learning by doing', going out into the network and utilizing a scope of PLA strategies, on the overall population, or on explicit pieces of the network. To be sure, this is one of the extraordinary points of interest to PLA.

Analysis and Action/Community Planning

The next stage is to collate, analyse and assemble the data collected into a format that can be presented to stakeholders (as part of the verification stage - see below). This can be done in a variety of ways, but the outcome needs to be a report or draft plan which brings out and highlights the main issues and problems, together with priorities and potential solutions.

Verification

Once the initial findings have been collated and analysed, the resultant report or draft plan needs to be verified. This is the process whereby the main findings are presented and fed back to stakeholders, including key organisations, agencies and the local community. This is to ensure that whatever projects or plans finally emerge, they represent, as accurately as possible, the needs and aspirations of all sectors of the community, and have the support of key partner agencies (e.g. service providers).

Collective Action

Once the verification process has been completed, the final stage will involve the development of a programme of action, or actions, to address the issues that have emerged from the research and Planning stages. Given the philosophy and principles underpinning PLA it is really important that this

action is done with the community, rather than to the community. The inclusive, highly participatory nature of PLA, and the involvement of residents from day 1, through the training) should allow for residents to stay involved as plans and actions get implemented.

Contemplations in Developing the PRA Strategy

PRA is a developing mix of approaches and techniques that encourage the sharing, cross-treatment, investigation, assessment and upgrade of occupation encounters and life conditions among key on-screen characters or partners inside a predetermined setting. Thus,

1) In settling on the instruments to be applied what's more, the power of the PRA preparing, the past information on PRA by the field assessors or facilitators must be evaluated. This assessment will help decide if the starting fragment of the preparation ought to be worried about characterizing PRA, or with examining why PRA is helpful and how the apparatuses could be applied to yield the ideal consequences of the exploration.
2) Learners must turn out to be completely occupied with the commonsense use of the most appropriate PRA tool for the recommended exercise. Be that as it may, PRA is in excess of an assortment of strategies. It is a general methodology, the adequacy of which depends vigorously on the frame of mind of professionals. It will as it were function admirably if specialists tune in to and regard the perspectives on the individuals who live the specific circumstance(s) being considered.
3) The PRA group must be facilitators of a procedure wherein members are catalyzed to become subjects rather than objects of the procedure. In this manner, facilitators of the PRA procedure need to:
4) Accept that there is no imposing business model of intelligence and information.
5) Listen to villagers and regard their information, encounters, assessments also, observations, too be tolerant of their mentalities and customs.
6) Create the feeling helpful to picking up the trust of the villagers furthermore, their ability to share their information and encounters.
7) Be adaptable in the booking of PRA exercises and give openings for critical thinking and learning, in light of the felt needs of villagers and their separate family units and networks.
8) PRA empowers members to participate in introducing their own data, drawing their own decisions and offering their very own investigations and arrangements. In the last investigation, PRA facilitators ought to become repetitive and the members what's more, their constituents enabled to deal with the change procedure viably.
9) When characterizing a fitting PRA device pack, each PRA group ought to decide the devices fitting for use in their individual site(s) or circumstances; having taken full thought of the time, human, innovative and different assets available. Visual systems are a piece of the suite of systems utilized in PRA.
10) The PRA team(s) must guarantee the following so as to ensure the effective execution of the PRA work out, to be specific:
    - Effective Communication
    - Appropriate preparing of members
    - Adequate transportation
    - Participant duty also, co-activity
    - Time the board
    - Group exertion/Team work
    - Proper choice/testing
    - Clear destinations

- Proper arranging
- Well characterized yields/key zones for examination.

11) The exploration region ought to speak to a specific set of villagers or generation framework and must take record of the cultivating network's intrigue and pledge to this procedure.
12) Having decided the arrangement of factors to be explored, a significant test is choosing which participatory approaches and methods would encourage the catch of the base set of information and data required. In this respect, there is the requirement for due industriousness in deciding the way to deal with the preparing and the arrangement of tools that are ideal appropriate for the endorsed examinations.

4.3 Introduction to Tools for PRA-PLA

PRA-PLA consolidates a consistently developing toolkit of participatory and visual strategies with normal meeting systems and is expected to encourage a procedure of aggregate investigation and learning. The methodology can be utilized in distinguishing needs, arranging, checking or assessing undertakings and projects. While a ground-breaking discussion device, it offers the chance to go past negligible interview and advance the dynamic support of networks in the issues and intercessions that shape their lives. This section illustrates all the different tools that can be utilized as Part of a PLA methodology to deal with research and discussion. They have been placed into groups, determined by which phase of the procedure you are at as shown in Figure 4.7.

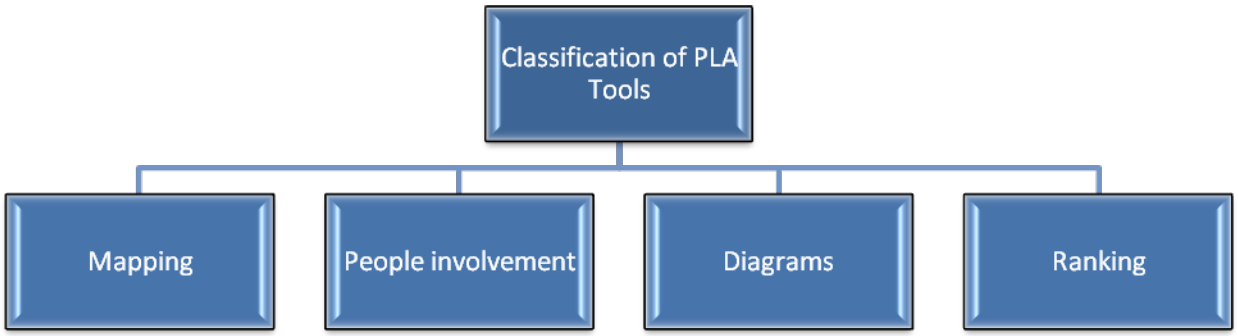


Fig 4.7:Broad Classification of PLA Tools

PRA embraces a series of techniques which are used for using people skills and knowledge to learn about the local conditions, identify the local developmental problems and plan responses accordingly. Different tools are put into PLA umbrella under different classifications as shown in Figure 4.8.

Mapping Tools and Techniques is a drawing of an area, used to describe the area and identify key features from the perspective of the community.

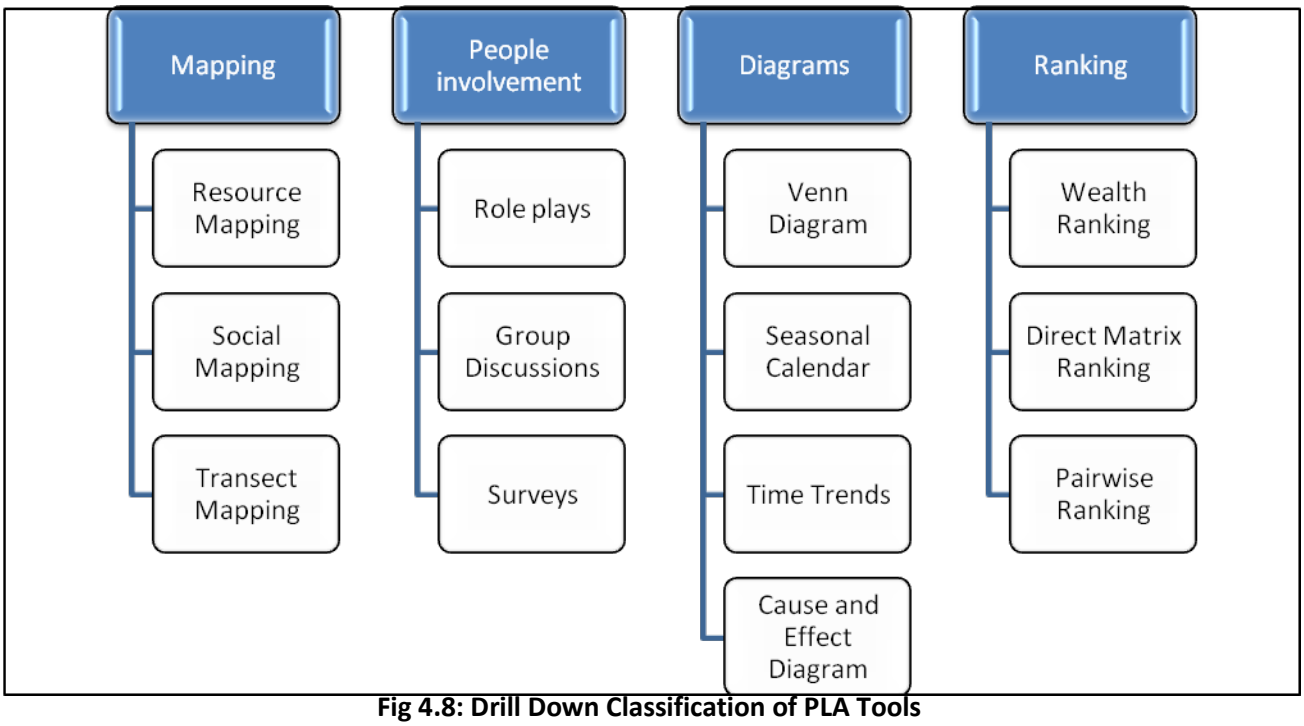


Fig 4.8: Drill Down Classification of PLA Tools

Mapping

Resource mapping is one of the most ordinarily utilized PRA techniques. It centers on the natural assets in the area and portrays land, slopes, streams, fields, vegetation and so forth. An asset guide may cover residence too. Now and again, the differentiation between the resource and social map may get obscured. A resource map in PRA isn't attracted to scale. It is done not by specialists, yet by the neighborhood individuals. The locals are considered to have an inside and out information for the environment where they have made due for quite a while. Thus, the resource map drawn by the neighborhood individuals is viewed as precise and confined. It is imperative to remember, be that as it may, that it mirrors the individuals' recognition instead of exact estimations to scale. Along these lines, a resource map reflects how individuals see their very own region regarding regular asset.

Applications of Resource Mapping

Resource maps have been used for depicting of various aspects related to the natural resource management of a locality including:

- Topography, terrain and slopes
- Forest, vegetation and tree species
- Soil-type, fertility, erosion and depth
- Land and land use, command area, tenure, boundaries and ownership
- Water, water bodies, irrigation sources, rivers and drainage
- Watershed development, various soil and water conservation measures, etc.
- Agricultural developments, cropping pattern, productivity, etc.

Resource maps have been found especially useful because they provide a focused spatial structure for discussion and analysis.

Example: Show the resource map of Tamil Nadu state with respect to its soil type (Fig 4.9)

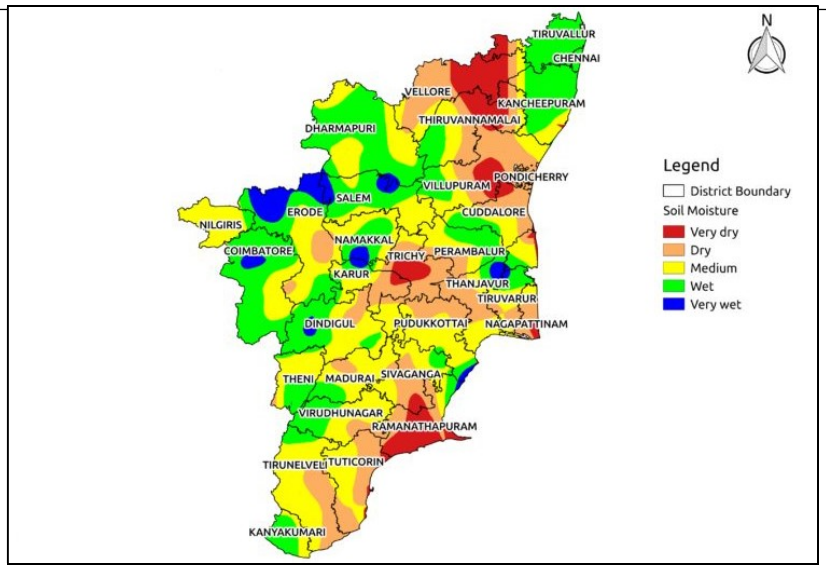


Fig 4.9: Tamil Nadu soil moisture map

To do activity: Resource Mapping

Guntur district of Andhra Pradesh state is geographically classified into the following areas as per the data given below. Try mapping it in the map of Guntur as per the approximate percentage of total geographic area.

SI	Classification of geographic area	Percentage of total geographical area
1	Forest	14.30
2	Cultivable Area	68.86
3	Barren and Uncultivable land	3.40
4	Land put to non-agricultural use and other	13.90
5	Sown area	50.14
6	Irrigated area	56.53
7	Cultivable waster	2.50
8	Permanent pastures and grazing land	1.50
9	Miscellaneous	2.40
10	Other fallow land	3.90
11	Current fallows	8.00
12	Others	0.46

Source:http://irrigationap.cgg.gov.in/wrd/getDistricts

Social Mapping

Social mapping is probably the most mainstream strategy in PRA. These are maps of a village or area which show where groups of people live (for example, rich, poor, literate, illiterate, different ethnic or religious groups etc.). Social map also delineates the normal assets – land, water sources, widely varied vegetation, and so on. A Social Map is a map that is drawn by the occupants and which shows the social structures and organizations found in a zone. It additionally encourages us to find out about social and financial contrasts between the family units.

Indications

- It will be useful to draw a straightforward model.
- During the entire procedure, take care that once someone has given an announcement, you ask the others whether they concur, differ or need to include something.
- The motivation behind the social guide must be obvious to all members, ensure that the members don't have wrong desires. For instance, they may believe that the poor family units will get nourishment gifts, which is totally off-base.
- Make sure that the target of having all family units appeared on the guide will be accomplished.

Applications of Social Mapping

The central component of a social guide is that it is a major assistance in building up an expansive comprehension for the different features of social reality, viz., social stratification, socioeconomics, settlements designs, social framework, and so forth. The various utilizations of social maps include:

- Developing an exhaustive comprehension of the physical and social parts of town life.
- Collecting segment and other required data family – shrewd
- Providing a gathering of exchange in high to disentangle the different parts of public activity
- Serving as an observing and assessing instrument.

Example:A researcher studied his small village and gathered some information about the village as given below. Illustrate the same with the help of social map.

Temple	**2**	**Anganwadi**	**1**
Well	4	Post office	1
Water Tank	1	Primary Health centre	1
Hand pump	186	Railway track	1
School	1	Livestock	1150
Pucca house	204	Roads of village	11
Semi pucca	8		
Kucha	6		

The social map of the surveyed area as per the above data is as shown in Fig 4.10.

Figure 4.10: Social Map

To do activity on Social Mapping

- Visit a nearby village to
  - Find out about the social structures in the Village and the distinctions among the family units by ethnicity, religion and riches.
  - Find out about who is living where.
  - Find out about the social organizations and the various perspectives nearby individuals may have with respect to those foundations.
- Key Considerations:
  - What number of family units are found in the Village and where are they found?
  - Is the quantity of family units developing or contracting?
  - What are the social structures and foundations found in the Village?
  - What ethnic gatherings are found in the Village?
- Material required:
  - If drawing on the ground: delicate ground, sticks and nearby material for images, or
  - If drawing on a paper: big piece of paper, pencils, markers

Transect Maps are constructed by walking through an area with a key informant, using direct observation to note specific features and factors, and talking to people met on the way.

Example of Transect Walk : A research group intends to understand the main land use zones in a village community at Yelagiri, North Arcot district. The following are the observations of the group using direct observations as they walked through the village. Draw a transect map to represent their observations as noted down by the researcher as shown in the Table below:

100m -400m	400m – 800m	800m – 1200m	800m-1000m
Resident area, home gardens, rice fields	Fast growing trees of villagers and indigenous trees	Degraded Forest	Rocky mountain

The transect map for the above observation is as shown in Figure 4.11.

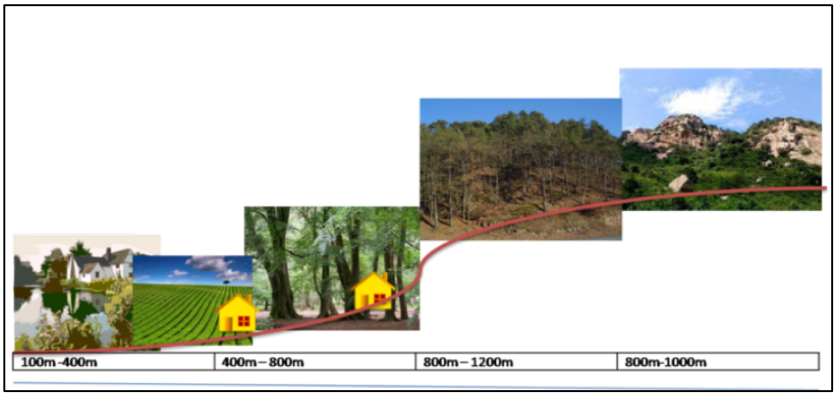


Fig 4.11: Transect map of a village

People Involvement

Role Plays

Different groups (e.g. men, women, young or old people) may role play the same situation from very different perspective. **Example:** A small village has about 40 households and 175 people. They have a health care centre with a doctor and a small school with few teachers. For years, women in the village had to walk to a neighboring village to carry water for household use. The quality of the water was also not good. The Government intends to solve the problem by providing piped supply of water to the village. Due to increase in population in 1 year, this method failed. Later an NGO decided to promote small scale water providers who will bring water to villages in water tanks. They decided to employ village youths for this to ensure regular water supply to the villages to solve water scarcity problem and to promote employment for youth in a small scale. Given the scenario, enact a role play.Role play scenes are shown in Figures 4.12a, b, c and d that will be enacted by the villagers as scenes 1,2,3 and 4 respectively to make the community understand the problem statement better:

Fig 4.12a: Scene1- Depicting water scarcity in village

Fig 4.12b: Scene 2- Solution given to overcome water scarcity in the village by piped supply of water

Fig 4.12c: Scene 3- Failure of piped supply of water to villages

Fig 4.12d: Scene 4- Providing alternative solution- Providing water supply through water tanks

Focused Group Discussion

Focus group discussions (FGDs) are a part of most encounters of participatory research and activity, and maybe the most usually utilized strategy in the participatory toolbox. The name FGD grasps a scope of various techniques, however the common link is that a collection of various sorts of members is shaped, and the grouped individuals are allowed the chance to go into discussion with one another in a sheltered setting. In participatory research, FGD is generally assembled, interceded and recorded by a group of in any event two individuals, including a facilitator and a note-taker.

(FGD) is a trusted method to assemble individuals from comparable foundations or encounters to talk about a particular subject of intrigue. The gathering of members is guided by an arbitrator (or gathering facilitator) who presents points for discourse and encourages the gathering to take an interest in an exuberant and regular talk among themselves.

The quality of FGD depends on enabling the members to concur or differ with one another so it gives an understanding into how a gathering contemplates an issue, about the scope of feeling and thoughts, and the irregularities and variety that exists in a specific network as far as convictions and their encounters and practices. Point by point framework of the FGD is shown in Figure 4.13.

FGDs were carried out with the following groups to gain an understanding of the problems faced by each group. The groups were divided into young people, old age people, farmers, adolescent boys/girls, women self-help group meeting etc.

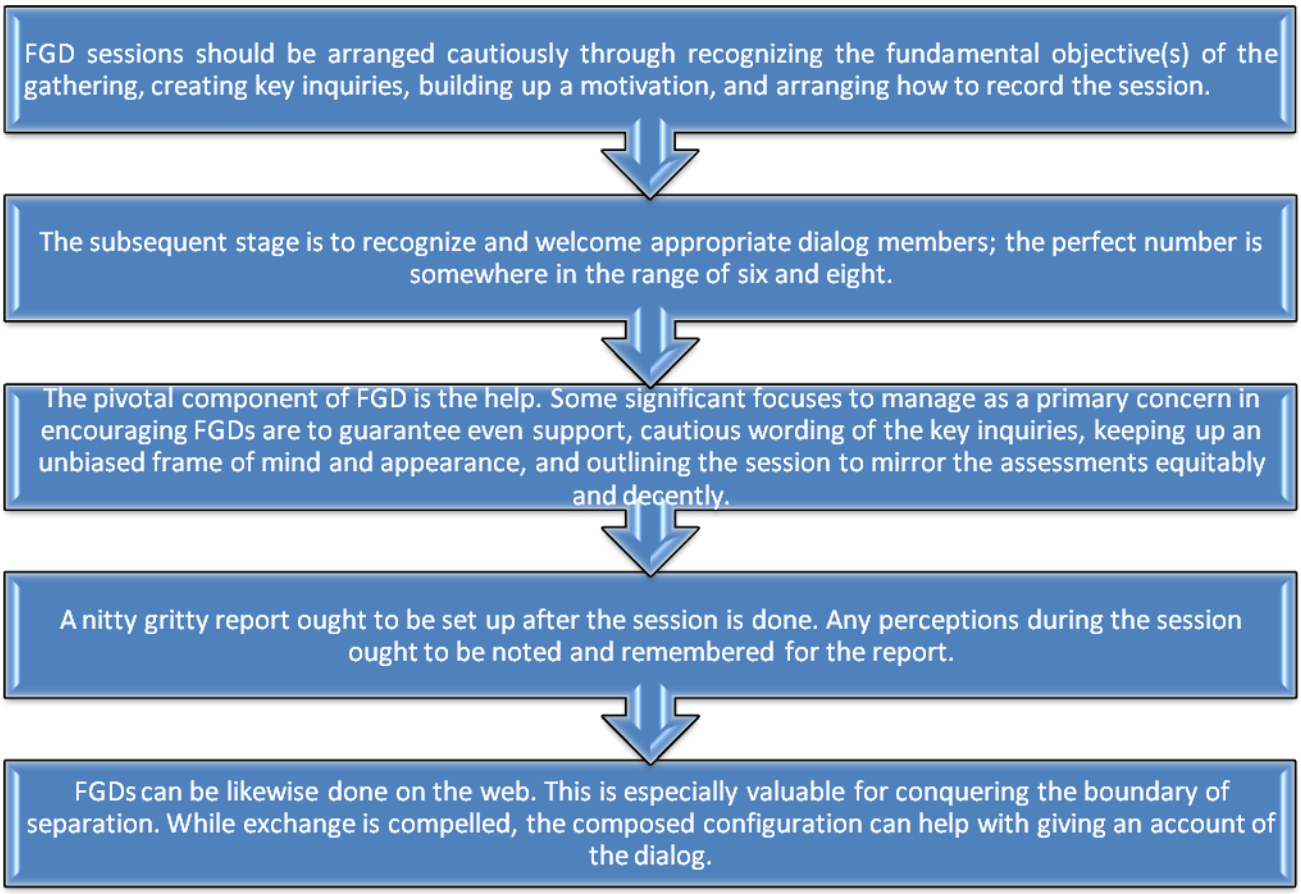


Fig 4.13: Point by point framework of the FGD

Example of Focused Group Discussion

A researcher intends to study the mindset of women in villages and gathers them for a meeting with the following questions. Fig 4.14 illustrates how FGD takes place in a typical rural environment.

(i) Do you buy your own clothes?

(ii) Have you faced any instance of violence against you at home?

(iii) Are you willing to invest your time to learn one of the income generating activity like Piggery farming, ginger cultivation, handicraft and weaving, potatoes cultivation, poultry farming , rajma cultivation etc to earn your own living?

Fig 4.14: Focused group discussion to promote Women empowerment in rural women

At the end of the discussion, they were motivated by the moderator to involve themselves in being part

To-Do-Activity on Focus Group Discussion

Objectives:

of self help groups to earn their own living.

Food and drink are regarded as a source of strength and nourishment. But, usually it is said that "**poor rural economies turned to potatoes for sustenance**". A researcher intends to study if this is true.

Distinguish what network, family unit and individual assets are required to acquire sustenance and family nourishment.

Technique:

Semi-organized group meeting

Choosing Focus Groups:

Plan separate gatherings for people in order to have the option to catch gender based contrasts in recognitions.

Facilitator:

Two PRA-colleagues

Key Areas of Concern:

Sustenance

i. What are the significant medical issues that individuals from the network have looked during the previous year?
ii. In your view, what are the explanations behind these issues and what is generally done to determine the issues?
iii. What assets are required by the network, families and people to turn out to be more successful at keeping wellbeing and nourishment issues from repeating?

Family unit nourishment security

i. What are the nourishments normally eaten in the network during this (dry) season?
ii. Would you be able to rank these nourishments as indicated by their recurrence of utilization?
iii. How does the eating routine change during the other (blustery) season and why?
iv. During the most recent year, what have been the issues in the network, family units and people to get such a sufficient eating routine (to be nourishment secure)?
v. In your view, what were the explanations behind these issues? What did the network and family units do to determine these issues?
vi. How are choices being made inside the family unit as to accomplishing nourishment security or reacting to issues of achieving nourishment security? Who settles on explicit choices? How are assets designated to accomplishing nourishment security? How are assets reallocated in the event of nourishment weakness?
vii. What measures are taken by the network and family units to cater to all nourishment security issues from recurring?
viii. What assets are required by the network, families and people to turn out to be more

Surveys

With an aim to gain an insight into the features, living conditions, difficulties and challenges as well as the available resources in the village, surveys were carried out.

Example: An interviewer wants to survey a remote non-electrified village to understand the financial and economic viability of electrification of that village by the Government. To do so, prepare a survey

template as shown in Fig 4.14 to be done with all households in the village.

Table 4.1: Survey Template for Electrification of a Village

Parameter	Response		Parameter	Response
Name			**Age**	
Occupation			**Education**	
Area Name			**No of Households**	
Major Source of Income			**Assets/ Livestock Processed**	
House Material			**Expenditure on non-food items**	
Sanitation			**Expenditure on food items**	
Expectation for Electrification (Tick the appropriate one)	**Major Lightning Source**	**Use of Electrical Appliances**	**Better Usage of water for irrigation**	**Others (Please Specify**
Willingness to pay for electricity (Tick the appropriate one)	**Extremely Willing**	**Moderately Willing**	**Not willing due to poor Affordability**	

To Do Activity on Creating A Survey Form

Create a survey questionnaire to study the living conditions of any village of your choice. The survey form must collect data about the following:

a. VILLAGE CHARACTERISTICS AND INFRASTRUCTURE
    i. size, caste composition and political structure
    ii. economy and infrastructure
b. ACCESS TO FACILITIES
    i. access to facilities
    ii. access to education
    iii. access to health
c. AGRICULTURE AND FORESTRY
    i. land and irrigation
d. FORESTRY
e. EMPLOYMENT AND MIGRATION
    i. prevailing wages
    ii. migration

4.4 Seasonal Calendar, Venn Diagram and Transect Walk Diagrams

Diagamming plays a major role in participatory research analysis. They make complex processes and

information to be presented in a simple form to make everyone (both literate and illiterate) understand it. Since pictures are much easier to understand than text, diagrammatic representation is considered as one of the most simulating way to represent and analyze.

Venn Diagram

Venn Diagram on Institutions or Chapati Diagram

A Venn Diagram on institutions is a visual strategy for recognizing and speaking to view of key foundations (formal and casual) and people inside and outside a network, their connections, and significance related with the equivalent. It is a tool for depicting institutions, associations, gatherings or people who are significant in the village, just as the resident's perspective on their significance in the network. Furthermore, the chart clarifies who partakes in these gatherings as far as gender and wealth is concerned. The institutional relationship chart likewise demonstrates how close the contact and participation between those associations and gatherings is.

Venn Diagram is additionally prominently known as Chapati Diagram, as the technique utilizes circles of different sizes to speak to organizations or people. The greater the circle, the more significant is the organization/individual and little circle demonstrates little significance. In any case, the separation between circles, speaks to the level of contact between people or organizations. Covering circles demonstrate the level of contact between organization for example a huge cover demonstrates high communication and no cover a far-off relationship.

- It additionally is an instrument which is used to identify the likely entry points for reinforcing or improving connections between key social actors
- It sees how unique network individuals see establishments both regarding investment, basic leadership, availability to and conveyance of benefits inside and outside the network

Venn diagram, one of the regularly utilized strategies for PRA is utilized in institutional setting to deliberate upon:

- Identification of institutions/groups/important persons active in the community and how they interact with each other.
- Identification of the importance or influence of institutions/groups or decision making in the community.
- Importance and elements of different establishments
- Extent of correspondence between associations
- Improving the connections among establishments and associations
- Significant jobs of new associations
- Communities impression of the foundations/people and so on.

Example 1 : Construct a Venn Diagram to illustrate various social gatherings in a particular village where neighborhood individuals of various social gatherings would approach incase of help or help in the outcome of a climatic catastrophe.

Fig 4.15 illustrates the Venn Diagram for distribution of various groups in a village.

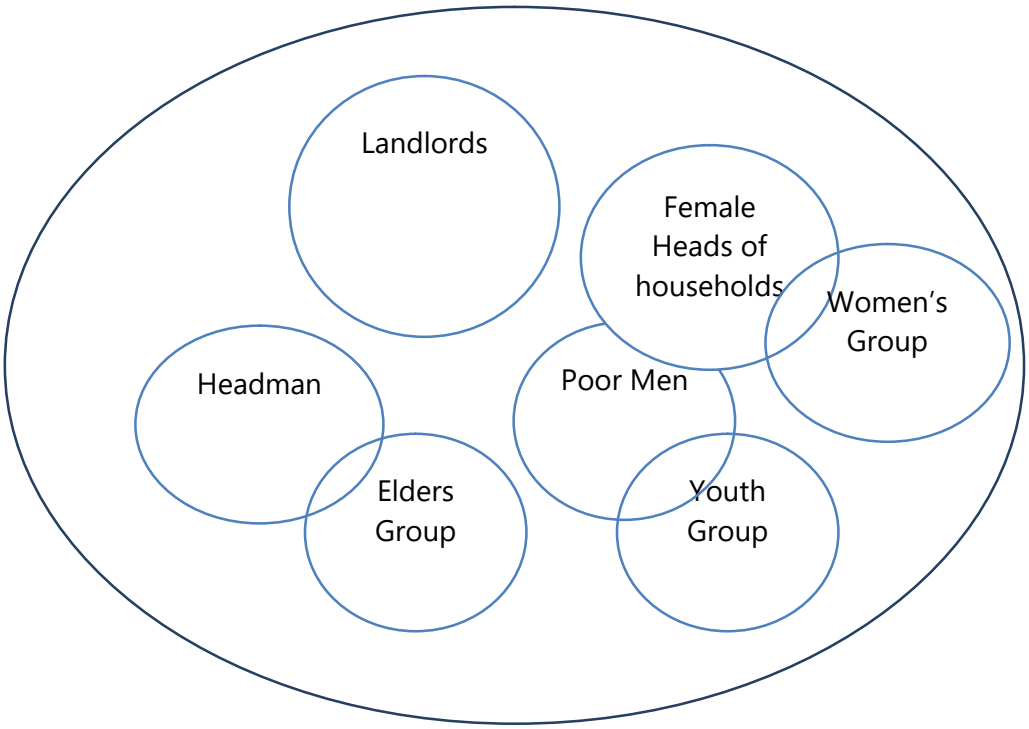


Fig 4.15: Venn Diagram for Distribution of Various Groups in a Village

Example 2: Draw a Venn Diagram to represent the interaction and influence of different communities on the lives of the villagers and explain the steps involved in constructing the same.
Step by step instructions to draw the Venn Diagram is as follows**:**

1) Make sure that you have all material that is required.
2) Ask the members which associations/establishments/groups are found in the village and which different ones from somewhere else are working with them.
3) What sort of methods exists for helping each other among individuals? Which neighborhood groups are composed along ecological issues? Are they political groups? Who settles on significant choices in the village?
4) Ask them to talk about for every association how significant it is for them. Every association/gathering ought to be set apart with the name or image.
5) Ask them to talk about in what direction they profit by the various associations.
6) Ask them to show the level of contact/co-activity among themselves and those establishments by separation between the circles.

Shape of the Venn Diagram	Nature of relationship
to a great extent removed circles	no or little contact or co-activity
hovers near one another	just free contacts exist
contacting circles	some co-activity
covering circles	close co-activity

Finally, Fig 4.16 represents the Venn diagram to represent the interaction and influence of different communities on the lives of the villagers.

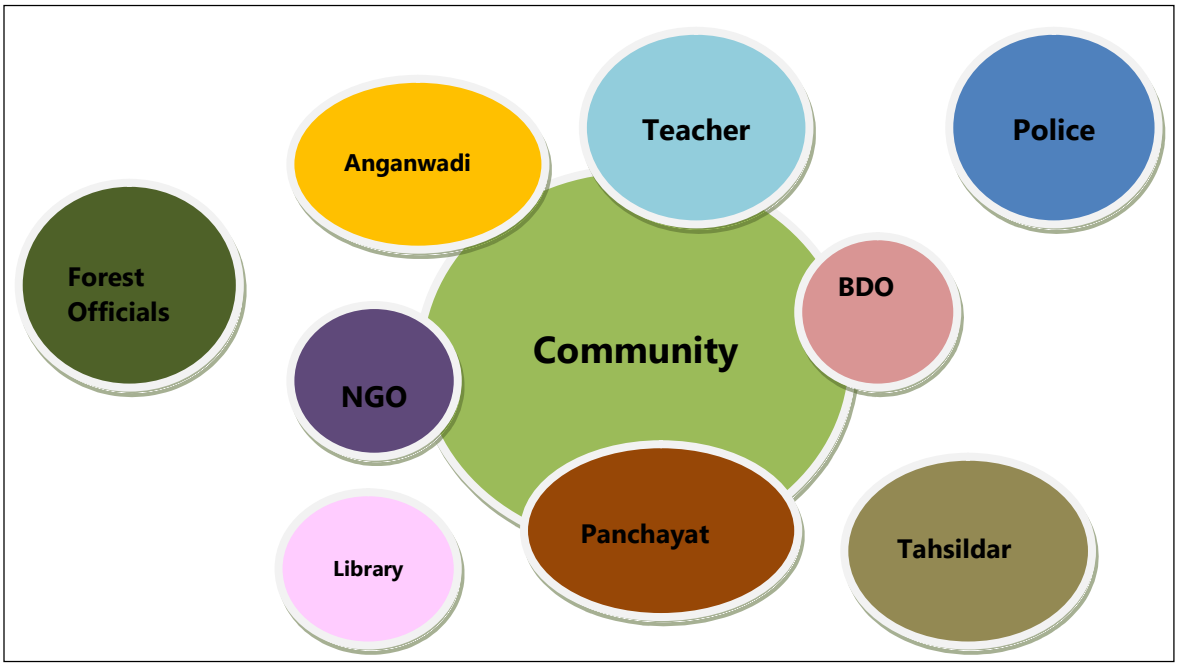


Fig 4.16: Venn Diagram Representing the Interaction of Different Communities on the Lives of the Villagers.

To Do Activity on Venn Diagram

A village is severely hit by draught. Draw a Venn Diagram to represent how the community will interact towards effective draught management with the following objectives:

(i) To recognize outer and inward associations/gatherings/significant people dynamic in the network
(ii) To recognize who partakes in neighborhood associations/organizations by sex and riches
(iii) To discover how the various associations and gatherings identify with one another regarding contact, co-activity, stream of data and arrangement of administrations

Key Questions:

(i) Which associations/foundations/bunches are working in or with the network?
(ii) Which foundations/bunches do the residents view as generally significant, and why?

Material Required:

The notetaker will require the Documentation Sheet for the Venn Diagram, this device sheet, white paper for duplicating the guide.

Seasonal Calendar

A seasonal calendar is a PRA strategy that decides examples of exercises and patterns during the time within a community. Seasonal calendar is one of the famous PRA techniques that has been utilized for worldly examination crosswise over yearly cycles, with months or seasons as the essential unit of examination. It very well may be utilized to discover changes in precipitation dispersion, nourishment

accessibility, horticultural generation, pay and consumptions, medical issues, work request, wood supply for fuel, ailment frequency, movement for business, money crops, livestock and numerous different components that change after some time.
The seasonal calendar can likewise be utilized to gather data on how residents distribute their time just as their work in different exercises inside the village. Seasonal Calendar is additionally called seasonal diagram, seasonal map, seasonal analysis and seasonal activity profile. Seasons are an essential piece of individuals' lives and apply a significant effect upon the work of the nearby individuals, especially in provincial territories. Seasonal outlines have been utilized to investigate what occurs during the year and when. Evaluation and delineation of the size of the different exercises adds to their utility and extravagance.

One of the primary reasons for using these diagrams is to find out about changes in livelihoods throughout the year and to show the seasonal trend of agricultural and non-agricultural workload, nourishment accessibility, human sicknesses, gender-specific salary and consumption, water, rummage, credit and occasions. This PRA tool demonstrates the view of the nearby individuals with respect to seasonal variations from a wide scope of things. A period outline or seasonal calendar is set up by drawing a two-dimensional grid and composing the timespan (for example month, year) on one hub and the distinctive village exercises on the other axis. Villagers are urged to fill in the network of the diagram/schedule by denoting the lattice or by putting stones or different articles on the framework.

Applications of Seasonal Calendar

- Seasonal Diagram recognizes overwhelming remaining task at hand periods, times of relative simplicity, credit crunch, sicknesses, nourishment security, wage accessibility and so on.
- It has end up being helpful in venture arranging, i.e., when to actualize different exercises.
- It has been utilized to recognize periods of stress and to get ready for when mediation is generally required. With a regular outline it is conceivable to recognize and examine the employment design over the year.
- The significant quality of regular examination is that it delineates a scope of things and their sizes, which helps in seeing how these things are identified with and impact each other. These connections can be very uncovering

Example: Table 4.2 represent a Seasonal calendar showing yield of different crops in metric tons per month.

Table 4.2: Seasonal calendar representing yield of crops across a year

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Paddy	xxxx		xxx	xx					xxx		xxxx	xx
Wheat	xxxx		xxx			xxx			xxx			
Jowar		xxxx	xxx		xxxx	xxx		xxxx	xxx		xxxx	
Bajra	xxxx		xxxx					xx	xxxx			xxxx
Ragi		xxx	xx					xxxx	xx			
Pulses	xxx		xxxx						xxxx			

To Do Activity: Seasonal Calendar

Objective:

Construct a seasonal calendar to show the seasonality of agricultural and non-agricultural workload in a village community.

With Whom:

Mixed group for women and men

Time:

2 hours

Key Considerations:

- What are the busiest months of the year?
- When are most agricultural work carried out by women?
- When are most agricultural work carried out by men?
- When is most non-agricultural work carried out by men and women?

Materials needed: Documentation Sheet, the tool sheet, white paper for copying the seasonal calendar.

- If drawing on the ground: soft ground, stones, sticks and other available material to produce symbols, or
- If drawing on a paper: Big sheet of paper, pencils, markers

Timeline or Historical Mapping

Trend and historical diagram: These diagrams were plotted to understand the current trends in the village and how change has impacted the lifestyle of villagers.

Example : A researcher intends to study the lifestyle of rural people from the data source collected by him as follows. Construct a trend diagram and mention any two of your interesting observations.

Table 4.3: Trend Map

Year	Cereals and products	Milk and products	Clothing and footwear	Fuel and light	Household goods and services	Health	Transport and communication	Recreation and amusement	Education	General index
2013	118.4	113.8	115.5	112.8	112.1	110.1	109.9	109.2	111.6	115.5
2014	122.9	124.5	122	115.7	117.5	115.1	110.1	113.9	119.5	119.4
2015	125.7	129.4	129	122	123.6	121.4	111.5	119.6	126.2	126.1
2016	132.3	135.1	135.4	126.6	129.2	126.9	116	124.2	133.1	130.4
2017	135.8	141	142	136.6	134.7	133.1	118.5	129	138.5	137.2
2018	137.5	142.1	146.8	142.7	143.2	144.9	123.6	136.8	150.1	140.1

Source: All India Consumer Price Index of rural India, https://data.gov.in/

The trend diagram shown in Figure 4.17 shows some certain interesting facts such as (i) Pulses and

products has seen a drastic change during 2015-16. (ii) Transport and communication index is always at the lower side only.

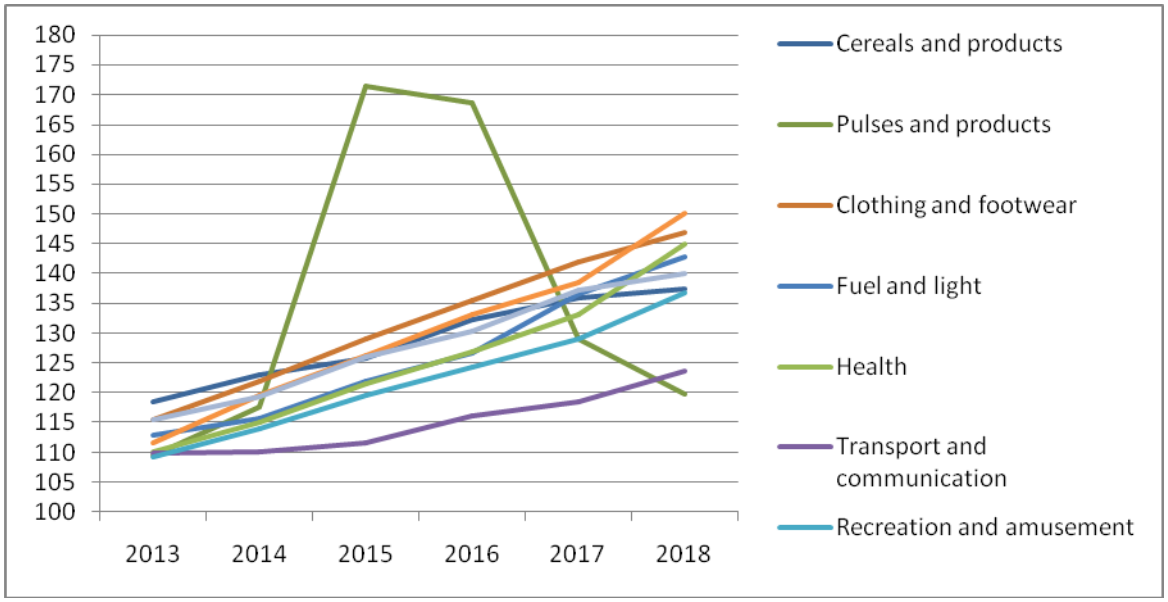


Figure 4.17: Trend diagram of All India Consumer Price Index of rural India

To Do Activity – Timeline Mapping
Table given below illustrates the percentage of villages that are electrified across all states in India. Illustrate the same using a trend graph and record your interesting observations from the graph.

Table 4.4: Percentage of villages electrified across all states in India

State	2008-09	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
A & N Islands	67.07%	67.07%	67.07%	67.66%	67.66%	67.66%	77.78%
Andhra Pradesh	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Arunachal Pradesh	56.82%	56.82%	58.40%	75.51%	75.51%	75.51%	70.25%
Assam	80.79%	78.57%	90.93%	96.15%	96.15%	96.15%	96.84%
Bihar	61.29%	61.29%	77.50%	89.87%	94.18%	94.84%	95.50%
Chandigarh	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Chhattisgarh	96.89%	95.61%	97.13%	97.15%	97.15%	97.37%	97.74%
Dadra & Nagar Haveli	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Daman & Diu	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Delhi	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Goa	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Gujarat	99.72%	99.72%	99.80%	99.80%	99.81%	99.81%	100.00%

Haryana	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Himachal Pradesh	98.22%	98.22%	99.53%	99.83%	99.91%	99.91%	99.70%
Jammu & Kashmir	98.24%	98.24%	98.24%	98.24%	98.24%	98.24%	98.22%
Jharkhand	31.07%	31.07%	88.46%	89.22%	89.22%	89.22%	92.66%
Karnataka	99.92%	99.92%	99.95%	99.95%	99.95%	99.95%	99.88%
Kerala	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Lakshadweep	100.00%	100.00%	100.00%	100.00%	100.00%	75.00%	100.00%
Madhya Pradesh	96.37%	96.38%	97.10%	97.24%	97.59%	97.80%	97.19%
Maharashtra	88.32%	88.32%	99.77%	99.91%	99.91%	99.91%	99.91%
Manipur	85.79%	86.26%	86.26%	86.26%	86.26%	86.26%	86.63%
Meghalaya	59.29%	59.29%	66.45%	76.53%	86.27%	86.27%	80.14%
Mizoram	8.05%	80.51%	87.99%	92.80%	96.19%	95.20%	93.61%
Nagaland	64.40%	64.40%	67.76%	70.11%	70.11%	70.11%	90.79%
Odisha	62.56%	62.56%	76.48%	78.90%	78.90%	81.89%	91.87%
Puducherry	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Punjab	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Rajasthan	69.19%	71.07%	95.05%	96.21%	97.53%	98.17%	90.37%
Sikkim	94.44%	94.44%	94.44%	100.00%	100.00%	100.00%	100.00%
Tamilnadu	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Telangana	57.23%	57.23%	66.32%	71.21%	92.89%	93.01%	96.99%
Tripura	88.27%	88.27%	88.27%	88.27%	88.92%	98.89%	98.67%
Uttar Pradesh	88.27%	88.27%	88.27%	88.27%	88.92%	98.89%	98.67%
Uttarakhand	96.92%	96.52%	98.93%	98.93%	98.93%	98.93%	99.32%
West Bengal	97.34%	99.50%	99.52%	99.67%	99.99%	99.99%	99.99%
India	83.66%	83.87%	92.13%	93.75%	94.41%	96.41%	96.69%

Source: https://niti.gov.in/state-statistics

Flow Diagram

It is a powerful tool that helps to communicate the sequence of actions within a complex system. The most commonly used is a cause and effect diagram.

Steps to build a cause and effect diagram are as follows:

(a) Identify the problem
(b) Put the problem at the center and circle it
(c) Explore the causes of the problem
(d) Discuss and place all causes in correct relationship to the problem
(e) Draw arrows to show the relationship.

Example: To understand the problem of poverty in a village, cause and effect diagram is illustrated as shown in Fig 4.18.

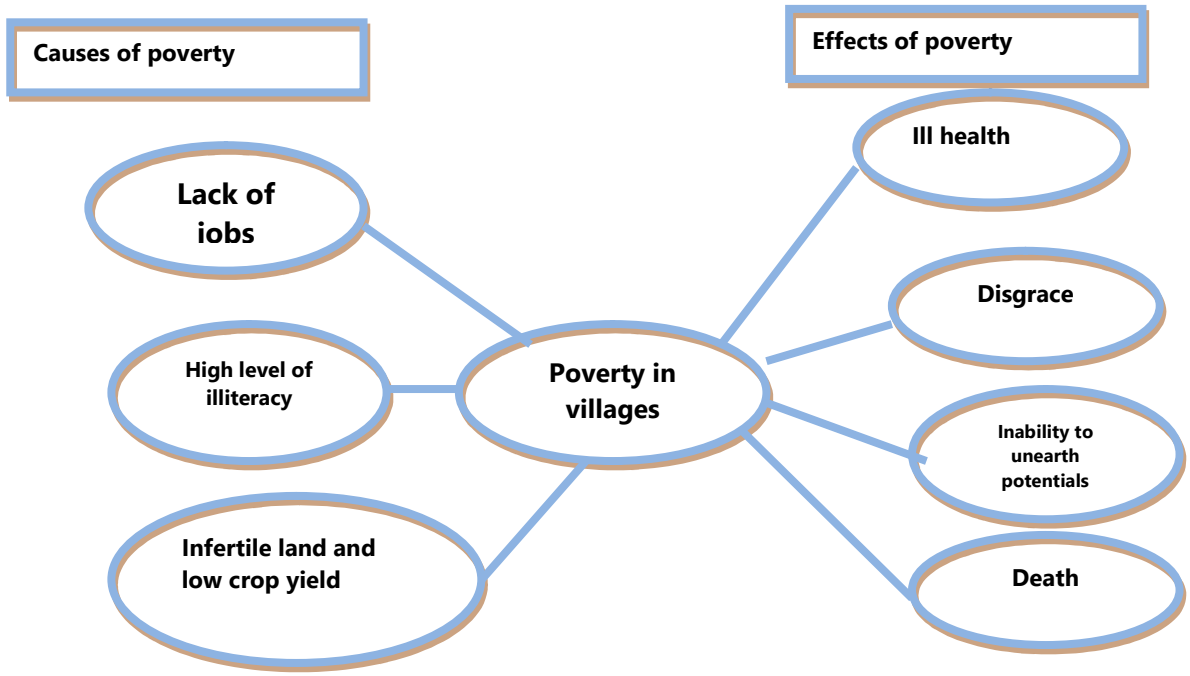


Fig 4.18: Cause and Effect Diagram (for poverty in villages)

Ranking

Ranking is an important **PRA** tool that assess and study the preferences of a process or technology one over another.

Wealth Ranking

Participatory wealth ranking is a technique that captures contrasts in ways of life as seen by the rural community themselves, therefore making it conceivable to pick up knowledge into relative social stratification. Wealth ranking is utilized to find out about riches or prosperity positioning of the residents. The facilitators portray riches positioning instruments and systems in nearby terms to them. Stratification may utilize as a reason for examining families or focusing on poor family units."Wealth" is characterized in every general public utilizing various criteria.

- It is utilized to comprehend the individuals' view of riches and welfare in their own area. Most research organizations utilize Social Map of the town as a reason for test determination, yet in the villages where we have the advantage of having the census data available for the whole village, makes the procedure more tightly and arbitrariness is exact.
- It not simply urges us to recognize the rich and poor groups of the town anyway in like manner gives us a comprehension into what builds up wealth and what involves poor undoubtedly.
- Wealth ranking is done by making a list of all households and ask people in a group/ community to sort them as per their priority.

Example: Government intends to construct an improved irrigation scheme for draught affected villages. The scheme consists of a cemented high storage pond with a capacity of 500 Cu.m, a 1000m long primary canal, and distribution to farms by earthen canals. The scheme intends to study the number of beneficiaries in the villages that it can supply. Create a wealth ranking criteria to categorize the households of the village into three sub-groups like well-off, Middle and poor category. The resultant wealth ranking matrix is shown in Table 4.5

Table 4.5: Resultant Wealth Ranking Matrix

Ranking Criteria	Well-off (Households)	Middle (Households)	Poor (Households)
Livestock holding	≥4 Oxen ≥3 milking cows ≥10 sheep ≥10 goats ≥20 chicken ≥15 donkeys	2-3 Oxen 2 cows 7-9 sheep 1-10 goats 10-20 chicken 10-15 donkeys	No oxen (Use others oxen for ploughing) 1 cow 0-1 sheep 0-1 goat 1-4 chicken No donkey
Land Holding	>20 acres	5-20 acres	<5 acres
Food Security	Food secure all year Surplus income	Food secure all year No surplus income	Food secure for 6 months with a gap of 6 months on an average.
Housing	Separate concrete housing for family and for livestock	A single concrete house with part for family and part for livestock.	A single kutcha house for family and livestock.

To Do Activity: Create a Wealth Ranking Matrix in a nearby village with the following objectives:

Objectives:

Identify and understand local indicators and criteria of wealth and well-being

Map the relative position of households in a community

Methods:

Ranking

Mapping

Selecting Key Informants:

Carry out the exercise with a few key informants who know the community well.

Facilitator:

Two PRA-team members

Key Questions:

What are local perceptions of wealth, well-being and inequality?

What socio-economic groupings are there in the community and who belongs in what group?

Labour	Hire labour occasionally	Don't hire labour	Earn from selling labour

Direct Matrix Ranking

It is a common technique for prioritization. The direct matrix ranking helps to choose a decision when several options are available.

Example: Present a direct matrix ranking to help farmers in cooperatives choose a rice variety to grow by comparing traits of several varieties.

Here, ranking is done based on the number of items chosen for ranking. There are 5 varieties. Hence rank varies from 1 to 5. After ranking the total score for each item is scored and the best alternative is chosen based on the total score.

Note : Best =1; Worst = 5, There are 5 varities. Hence, rank varies from 1 to 5.

As per the Table 4.5 given above, the best option is Variety 1.

Table 4.6: Direct Matrix Ranking

Criteria	Rice Variety				
	Variety 1	Variety 2	Variety 3	Variety 4	Variety 5
Resistance to pests	1	5	4	3	2
Drought resistance	1	3	4	2	5
Length of straw	4	3	5	1	2
Suitability to light soil	1	1	1	2	3
Market price	1	2	1	4	5
Eating quality	1	2	1	3	3
Total Score	9	16	16	15	20

To do activity on Direct Matrix Ranking

From the sample direct rank matrix given below, find the choice of fish to be grown as **fish** breeds for farming in **freshwater** pond.

Parameter	Catla	Mrigal	Rui, Grass Carp,	Silver Carp	Common Carp
Yield	3	2	4	1	5
Demand	1	2	4	5	3
Water Quality resistance	2	1	3	4	5
Market Price	2	1	3	4	3

Note: The ranking given in table purely depicts user's preference.

Pairwise Ranking

Pairwise Ranking, also known as Preference Ranking, is a ranking tool which assigns priorities to the multiple available options. This method uses stones or seeds to rank the preference. For each comparison, a seed will be placed next to the more important problem and as the length of seed grows, the length of line becomes longer which indicates the most severe problem faced by the community.

Example: Perform a pairwise ranking exercise to identify the most important problems in a village. Assume the problems faced by the rural society are as shown in Table 4.7.

Table 4.7: Pairwise Ranking Exercise Conducted at Village.

Sl	Problem	Score (As seeds or stones)	Score
1	Lack of wells	IIIIIIIIII	10
2	Hunger	IIIIIIIIIII	11
3	Broken bridges	IIIIIIIII	9
4	No market for honey	I	1
5	Lack of transport	IIIII	5
6	Dam broken	II	2
7	No hammer mill	IIIIII	6
8	Few oxen for ploughing	IIIIIIII	8
9	Poor cooperation between people	IIIIII	6
10	Theft	-	0
11	Schools not good	IIIIIIIIII	11
12	Roads need repairing	IIIII	5

To Do Activity – Pair-wise Ranking: Identify a village of your choice and perform a pairwise ranking on the probable problems faced may be included as follows:

Sl	Problem
1	Lack of fertiliser and seed
2	Lack of transport
3	Poor roads and bridges
4	Lack of work oxen
5	No consumer shops
6	Lack of clinic
7	Lack of houses
8	Lack of classrooms
9	Lack of market

Finally identify the most important problem faced by the villagers.

4.5 Case Studies on PRA-PLA

Case Study 1: The Case Study described is that of Village Marale in Maharashtra.

Profile of village Marale: Village Marale is located in the lush green and beautiful mountainous area in Taluka-Shahuwadi, DistKolhapur, Maharashtra. This village is located besides the river Barki. It is 65 km away from Kolhapur city (Maharashtra). Two more villages namely Yelewadi and Kante are covered by Marale village group gram panchayat. As per the survey conducted in Marale village by Shivaji University, Kolhapur, Maharashtra (December 2012) the total population of village is 743 (male 388, female 355). Total 65 families are from below poverty line. 222 people are illiterate and main occupation of villagers is farming. For a period of 6 days, the student social workers and their field supervisor carried out the PRA and village micro planning activities. The group applied the following PRA tools:

Mapping: Village map, resource map and health map were prepared by villagers to gain a better understanding of the structure, resources and health related issued in the village.

Transect Walk: This tool was used to understand the farming pattern of this village and various issues.

Seasonal Calendar: To comprehend the adjustments in villager's livelihood status throughout the year and to show the regularity of farming and non-farming workload, nourishment accessibility, human infections, gender-specific income salary and consumption, water, scrounge, credit and holidays this tool was a great facilitator.

Street plays: This tool is used to organize plays on streets and bring to light different issues faced by villagers like child marriage, drug addiction and sanitation.

People's Action plan: To tackle the issues identified through Problem Priority Chart, time bound planning was carried out by some community people who stepped up to the occasion. With the assistance of above procedures particularly; Problem Priority Chart the problems in the town were surveyed.

These issues are as following: Alcoholism, Migration of youth for business, Lack of ladies' investment, School dropout, early relationships, Passive Self-Help Groups, Traditional cultivating strategies, Health issues, Superstitions

Post PRA Intervention

After this PRA, some student social workers were set in this village to manage the issues which were distinguished during PRA with the assistance of dynamic inclusion of the villagers.

Alcoholism: The issue of liquor abuse was one of the most incessant issues of this village. Residents used to spend their cash on drinking liquor and it was influencing their family connections and economy. When all ladies met up and passed an agreement of 'Prohibition on alcohol/liquor' in the town during town gathering. Surveying strategy was utilized and this agreement was passed equitably. Understudy social laborers encouraged and urged town ladies to take care of this issue of liquor abuse by applying social activity strategy.

Women empowerment: Gender sensitive community association practice and Feminist social work are relied upon to be drilled by student social workers. Empowerment is a procedure of progress by which people or groups gain force and capacity to assume responsibility for their lives. It includes access to assets, coming about into expanded investment in basic leadership and dealing power and expanded command over advantages, assets and possess life, expanded self-assurance, confidence and sense of pride, expanded prosperity.

There are five elements of empowerment: political, social, financial, social and well being. National strategy on women accentuate on women empowerment and gender equality. It is a zone to be focused by student social workers while put in villages for field work. First time throughout the entire existence of the village, the womenfolk get-together was organized in village Marale to manage issues of women. The cooperation of women for this get together was wonderful. Politically women met up to examine on the issue of liquor abuse in the village. An agreement of prohibition on alcohol was passed by all residents with the inception of women. Prior to the PRA Activity, the self improvement groups were experiencing number of issues for example the vast majority of the SHGs were shut, absence of mindfulness about examining, wrong advance circulation, nonattendance of records and absence of self control and certainty among ladies to run SHGs and so forth. Women were energized with the assistance of hands on work exercises and were spurred to restart these SHGs and run appropriately. To build women investment different social exercises were done for example rivalry and celebration festivities. Women meetings were conducted to increase their awareness about wellbeing and issues of drop out children.

School Social Work: In this village one elementary school from first to fourth standard is running. Different exercises in the school were completed like; guardians meeting, rivalries, games, melodies, polio rally, home visits to drop out youngsters and spare my interests and so forth.

Youth-Centered Social Work: Youth group meetings were directed to get enrolled and join to Nehru Yuva Kendra. Social projects were sorted out and attempted to organize the network of people. To provide career guidance to the children, special seminars were organized.

Social Work with Minorities: Student social workers ought to apply the counter harsh way to deal with work with deprived and underprivileged communities in the village. The caste system is constantly more grounded in rural than urban regions. In Dalit community group meetings were done to comprehend their issues and granted training about different government programs for minorities, for example, grants and scholarships.

Case Study 2:

This case study intends to answer the following research questions regarding the study of a village called 'Nadukani' in Kerala state. Suggest what PRA/PLA tool would you choose for the same.

1. Identify the physical and topography of the village by interacting with the villagers and open up an interaction with the villagers.
2. Talk to the villagers (both elderly and current population) to study the major events/ changes that has happened in their village to understand the living in the community.
3. Identify the trends of various diseases affecting the community.
4. Understand the variety of treatment methods the villagers are adopted to.
5. Understand the reasons for the women in the village not adopting to family planning techniques, as a result of which their poverty level rises and maternal mortality also increases.

PRA is a means of collecting different kinds of data. To answer each of the above listed question, a different tool must be used.

(i) To identify the physical and topography of the village by interacting with the villagers and open up an interaction with the villagers, a transect map was developed as shown below with an objective to understand the agro-ecological zones in rural areas and also understanding the village, their uses.

Table 4.8: Transect Map in the Nadukani Village.

Features	North East	South West	North East Settlement area
Topography	Ascending	Descending	Slope
Soil	Dry	Dry	Aluminium soil
Trees	Jack fruit, mango	Tamarind, Bamboo	Tamarind, Mango
Crops	Rubber, coconut, coffee, Pepper	Pepper , coffee, coconut	Coffee, coconut
Vegetation	Papaya, plantain	Papaya, plantain	Papaya, bitterguard, plantain

(ii) In order to study the major events/ changes that has happened in a village to understand the living in the community, Timeline maps are generally used. This indicates the major events remembered by the villagers. Elderly villagers can be facilitated to broadly enumerate the important changes with major changes. This information will be useful to understand the development of the village over a period of time. Table 4.8 illustrates the timeline showing different development stages in the village.

Table 4.9: Timeline Showing the Different Development Stages in the Nadukani Village.

Year	Development		Year	Development	
1961	Bullockcart		1990	Television	
1963	Ration shop		1991	Telephone	
1965	School		1992	Primary Health Centre	
1970	Radio		2008	Mobile Phone	
1971	Bus		2013	Internet	
1980	Mosque				

(iii) To understand the majority of diseases that has affected the village, a seasonal map of the most frequent diseases (as reported in the PHC of the village) was mapped as shown in Table 4.9. Also, the age – group of the diseases was recorded to understand the seasonality of the diseases.

Table 4.10: Seasonal Analysis of Five Diseases in the Nadukani village

Disease	Seasonal Analysis												Age group
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	
Eye sore	X	X	X	X				X					ALL
Mumps						X	X	X					CHILDREN
Chicken pox	X	X	X	X	X								ALL
Fever						X	X	X					ALL
Dysentery			X	X	X								ALL

iv) To Understand the variety of treatment methods the villagers are adopted to based on their preference, a preference ranking matrix was created to record their responses (Score = 1 for best and 4 for worst, since 4 methods are compared) and based on the score, the most preferred treatment method can be identified as shown in the Table 4.10.

Table 4.11: Preference Ranking Matrix of Nadukani Village

SI	Criteria	Allopathy	Homeo	Ayurveda	Traditional
1	**Distance to the centre**	4	2	3	1
2	**Treatment cost**	2	1	4	3
3	**Effectiveness**	1	3	4	2
4	**Doctor preference by the villagers**	4	2	3	1
5	**Contagious disease**	1	3	2	4
6	**Serious Illness**	1	2	3	4
7	**Awareness**	1	4	3	2
8	**Speedy recovery**	1	2	3	4
9	**Popularity**	1	3	2	4
10	**Misc Illness**	2	3	1	4
Overall Preference		18 [1st]	25 [2nd]	28 [3rd]	29 [4th]

(v) To understand the reasons for the women in the village not adopting to family planning techniques, as a result of which their poverty level rises and maternal mortality also increases has been studied in the case study by using the impact diagram (Cause and effect diagram) as shown in Figure 4.19.

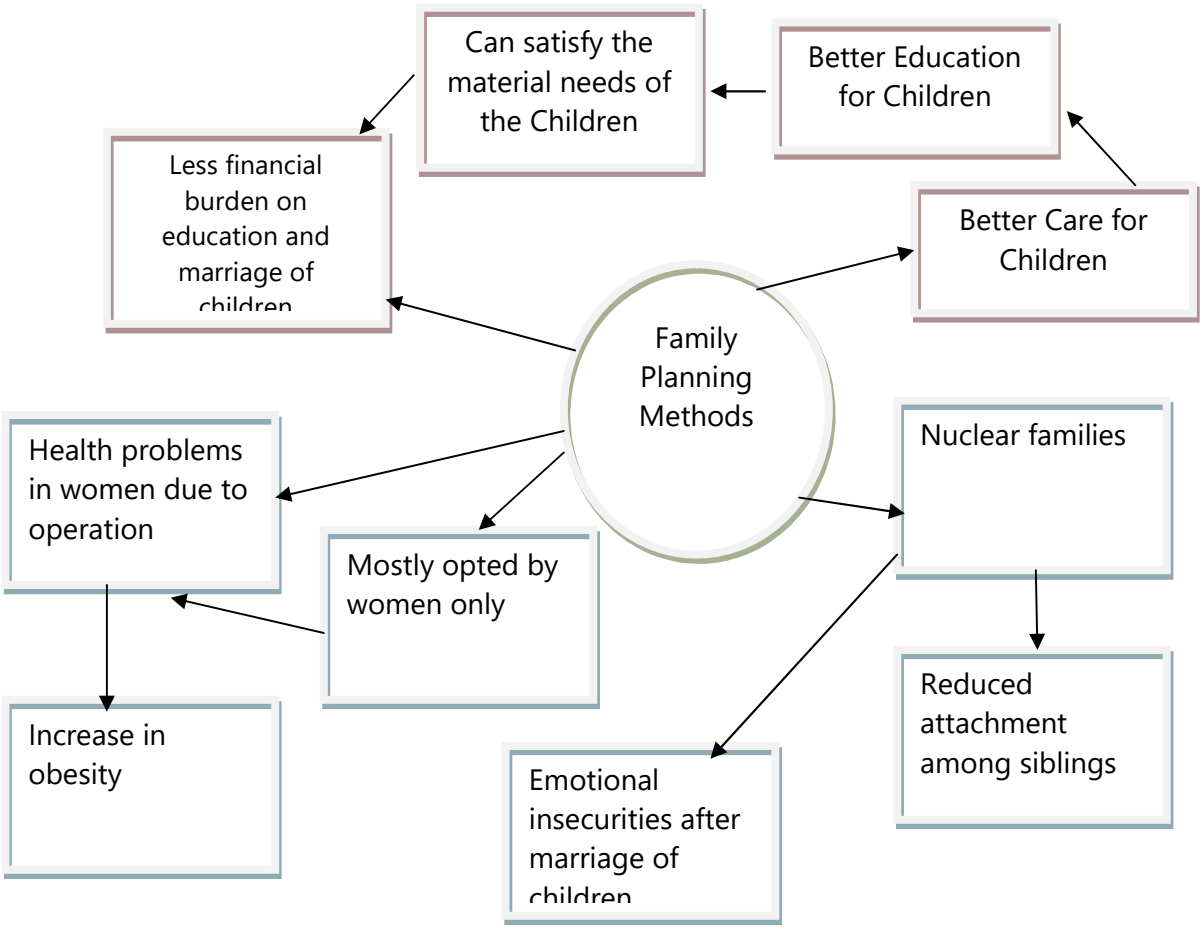


Fig 4.19: Cause and Effect Diagram (for family planning methods)

Summary of the Chapter

Summary

The chapter deals with explaining the concepts of PRA-PLA, the history, growth and present form of the concept. Emphasis was given on the various types of PRA and significance of each type. A detailed description of various kinds of tools used in PRA-PLA was given with focus on to-do-activities for each type of tool as well as applications of the same, The chapter concluded with a case study which depicted the efficacy and essence of this concept in identifying and tackling with problems encountered in the rural areas.

Model Questions

1. What do you understand by PLA-PRA? Throw some light on its genesis and history.
2. Why is the term PLA-PRA replaced by PRA? How many types of classifications are there for the PRA tools?
3. Differentiate between the various approached of PRA. Discuss in detail the participatory approach to PRA.
4. Enlist the basic rules of PRA and elaborate the same on the basis of your understanding of the text.

5. Discuss in detail the four distinct phases of PLA.
6. List the contemplations involved in Developing the PRA Strategy
7. Write Short notes on the following PRA tools:
    a. Resource Mapping
    b. Social Calendar
    c. Wealth Ranking
8. What is the significance of Seasonal Maps? Draw a seasonal map pertaining to the yield of various crops in your area.
9. "Focus Group Discussions are the most basic tool for doing PRA". Justify the statement.
10. What do you understand by a Venn Diagram? What type of problems use this as a tool for visualization? Explain a scenario through Venn/ Chapati Diagram.

To-Do- Activity

1. Find out the most common problem prevalent in your area. Form focus groups and apply PRA tools like Social map, Seasonal map and Venn-diagram to show how these tools can bring out the causes as well as remedies to the problem at hand.
2. Go to some nearby village to get an understanding into the residents' impression of what regular assets are found in the village and how they are utilized.

Identify the following and create a Resource Map.

Identify streets, houses, structures, spans, and so on, water locales and sources; rural terrains (crop assortments and areas); soils, slants, rises; timberland lands; brushing regions; shops, markets; wellbeing facilities, schools, chapels; unique spots (hallowed destinations, graveyards, transport stops, places of worship, and so on).

With whom: Female and male center gatherings

Key Considerations:

a. Which resources are abundantly available?
b. What resources are scarce or rarely available?
c. Does everybody have equivalent access to land?
d. What is the gender distribution of the land ownership?
e. What is the crop cover of the area?
f. What are the main sources of water?
g. What are the main crops grown in the area?
h. Which areas are suitable for what type of crop?
i. Where are the grasslands for grazing animals available ?
j. What sort of advancement exercises do you complete in general network? Where?
k. Which asset do you have the most issue with?

Materials: Sticks, rocks, leaves, sawdust, flour, compost or some other nearby material.

References

1. Chambers, R (1983) Rural Development: Putting the last first. Longman inc., USA, 1983.
2. Chambers, R (2008). Revolutions in Development Inquiry. Institute of Development Studies, 2008, Earthscan, London.
3. Gosling, L and Edwards, M (2003). Toolkits: A practical guide to assessment, monitoring, review and evaluation. Second edition. Save the Children, UK.
4. Mikkelsen, B (1995). Methods for Development Work and Research: A guide for practitioners. London, Sage.
5. Slocum, R; Wichhart, D; Rocheleau, D and Thomas-Slayter, B (eds.) (1995). Power, Process and Participation – Tools for change. London, IT Publications.

Chapter 5 Introduction to Research Report

Introduction

Research report is an orderly presentation of the research investigation done by the researcher by turning the problem into a research question or hypothesis with the intent of the research to answer the question of interest. It requires careful selection of report type. The researcher must ensure that the report adheres to all the characteristics of a good research report and then carefully write the different components of a good research report. Finally, choose appropriate tables and figures to present the research results in the most appropriate form to the intended audience of the report.

Objectives of the Chapter

- To enable the reader to gain an understanding of the types of research report
- To prepare a research plan for a report.
- To enable reader to understand APA style formatting of the research report.
- To take a decision of choosing between tables, figures, and text to present a given data.
- To choose the appropriate graph for a given data/case study.

Structure of the Chapter

5.1 Research Report Structure

Research report is a research document that contains all aspects of the research that has been implemented by the researcher. Be it an inductive research or deductive research, the research report must be systematically presented in the same way in which the research was conducted.

Inductive Research Methods are used to analyze an event that is observed by the researcher. This is used when there is no or very little literature available on the topic because there is no theory to test, but it can never be proven. It consists of three phases of research that are observation, observe a pattern and develop a theory. The figure 5.1 explains the inductive research report with an example which starts from observation, observation of pattern and finally develops a theory. But, this inductive research can be disproved even if one $(n+1)^{th}$ airline flight comes on time after observations of n delayed flights. The complete sequence of research must be explained with clarity in the inductive research report.

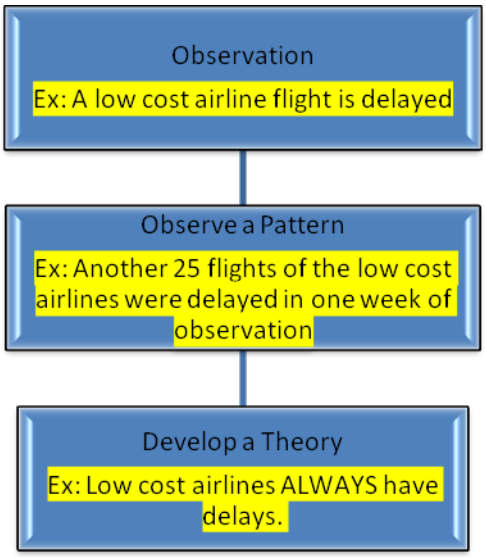


Figure 5.1: Steps of Inductive Research Report

Deductive Research Methods are used to verify the observed event, i.e. start with the results of inductive research. Reasoning deductively means testing these theories, which means deductive research cannot be conducted without theories laid down by inductive research. Figure 5.2 explains the four phases of deductive research with an example and henceforth the report must also have the same sequence of explanations of the work done.

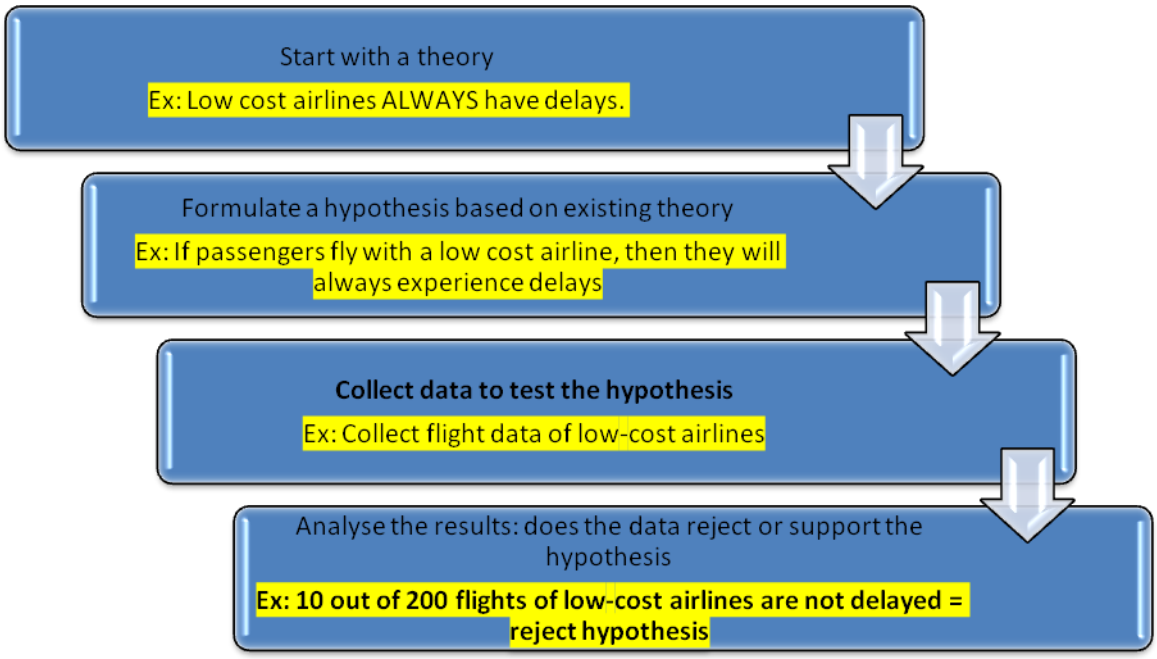


Fig 5.2: Steps of Deductive Research Report

To Do Activity

Study the "impact of classroom environment on the mental health of children" in a neighboring school. Will you conduct an inductive or deductive research to answer the research question?

5.2 Types of Report

The different types of research reports are as presented in Figure 5.3. The functions and examples of each of the mentioned report type is given in Table 5.1.

Fig 5.3: Types of Research Reports

Table 5.1 Types of Reports

Report Type	Functions	Examples
Informational report	Uses facts and information to present the report	A report to present interesting facts about butterfly.
Analytical report	A type of a business report that uses both quantitative and qualitative data to analyze as well as evaluate a complete business process to make data-driven decisions.	How to reduce the patients waiting time in a hospital? This study requires analysis of all required healthcare metrics that would help hospital decision makers to reduce the waiting time in hospitals.
Persuasive report	The report presents a situation, and takes a stand – either in its favor, or against it – to prove to readers whether it is beneficial or harmful for them.	In today's era, most of the youth skip their breakfast. A persuasive report will illustrate the importance of breakfast, especially for students.
Sales report	The sales report or sales analysis report outlines the status of	For a company to actively strengthen its new business a

	sales activities within the company. It shows the trends in sales over time. Sales reports helps a company to identify potential new market opportunities.	report on monthly revenue goals would give a clear picture to improve upon.
Case study report	A case study based research requires you to analyze a business problem, explore the alternative solutions, and propose the most effective solution using supportive evidences. A case study can focus on a single unit of business or entire industry, or program, or a person.	How to consider small customers for bigger profits? A distributor must analyse the profit generated by each customer and customer segment. For this, a case report on customer grouping by size and order profitability will help the distributor to decide which customer segment contributes to steady profits and growth.
Justification/ recommendation report	This report provides different possibilities and suggests ideas to a problem being addressed. Include all relevant data and statistics required. List the possible solutions to solve the problem beginning with the one that is least likely to succeed and ending with your strongest most promising idea. Compare each of the alternative solutions and choose the one that is the best and recommend it with justification of the chosen solution.	A company intends to study the means to stop smoking by employees in the interest of their health and to improve productivity of the organization. Alternative 1: Non-smoking education through lectures and videos for employees willing to stop smoking. Alternative 2: Healthy living events at the workplace for all employees. Alternative 3: Individual professional counseling. After analyzing the pros and cons of each alternative, the researcher would recommend one best option (alternative 3) to the management, so that employees are given a small time during the working hours to take counseling on site to ensure the employees are directly benefitted.

Research report	Study problems scientifically by developing hypotheses, collecting data, analyzing data, and indicating findings or conclusions along with the recommendations.	A company wants to investigate staff attitude to personal mobile phone use in staff meetings. A staff survey (by distributing questionnaires to the staff) on the attitude towards the use of mobile phones in the staff meetings was conducted. The results indicate that majority of the staff use mobile phones. The report concludes that mobile phone usage are disruptive and should be turned off during meetings. The research report has also recommended the banning of mobile phones use during meetings except under unavoidable circumstances.

To Do Activity: Prepare a case study report to analyze the research question : " Does a small business need a website?"

5.3 Components of Research Report

A research report provides a method for scientists to communicate with other scientific community about the results of their research. A standard format shown in Figure 5.4 is used for writing research reports by means of which the author presents the research in an orderly, logical manner. This doesn't necessarily reflect the order in which the research was planned or implemented.

The Title of Research: Important words in the title are capitalized and it must be specific enough to describe the contents of the paper. It should not be complicated that only experts can understand.

Abstract: An abstract or summary gives the preview of what is to come in the report. An abstract should be clear and precise to kindle the interest of the potential readers. It is generally limited to one paragraph about 100-250 words, which summarizes the purpose, methods, results and conclusions of the research report. As a rule of thumb, abstract is usually written after the complete research report is written, so that important sentences of conclusions and recommendations can be taken and included in the abstract. Abstract should be standalone and should not include any literature citations, abbreviations or footnotes. The abstract in general must usually include the following in the given order of presentation:

- Identification of subject population
- Specification of research design
- Data collection methodology
- Summary of results and inferences drawn
- Comparisons (if any) drawn from the results

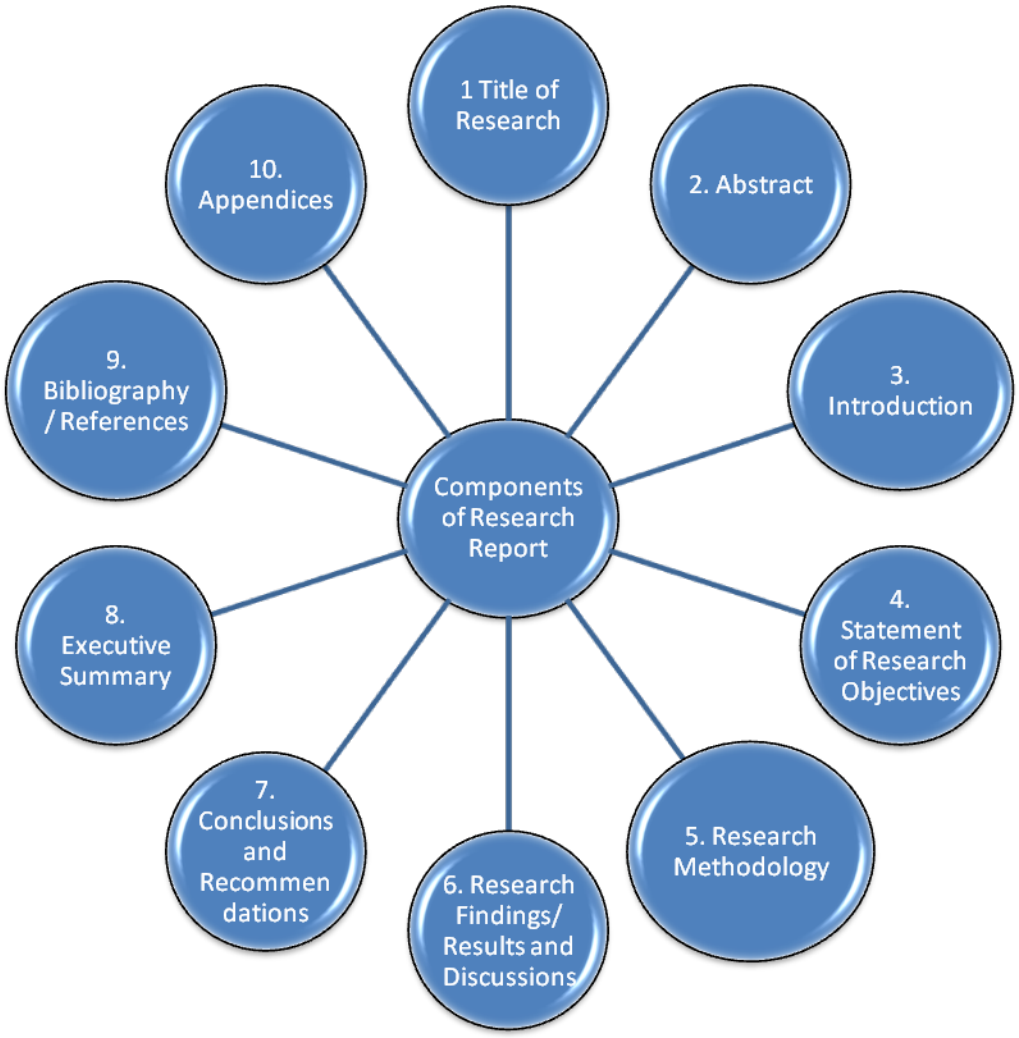


Fig 5.4: Components of Research Report

Introduction: The introduction to research paper is most challenging to write and the length depends on the type of research report. The introduction must include the purpose of research, scope of research, provide context and rationale of work, stating the hypothesis of research. In some reports, the introduction also includes the review of literature and the research gap identified in the study.

Statement of Research Objectives: This section illustrates what is expected to be achieved by the proposed research. The objectives must be defined in simple terms so that even a lay man can understand the intent of the research. It may have an underlying hypothesis to be discussed in the proposed study or it may be a general statement of purpose without the hypothesis.

Examples of research objectives are as given below:

- Objective: To describe what factors does farmer take into account in deciding what crops to grow?
- Objective: To describe the habitat of rural population.
  In the above two examples, the research objectives are explanatory and does not need any hypothesis to be stated.

Given below is a research problem in which the objective is stated with an understanding of the hypothesis (Figure 5.5).

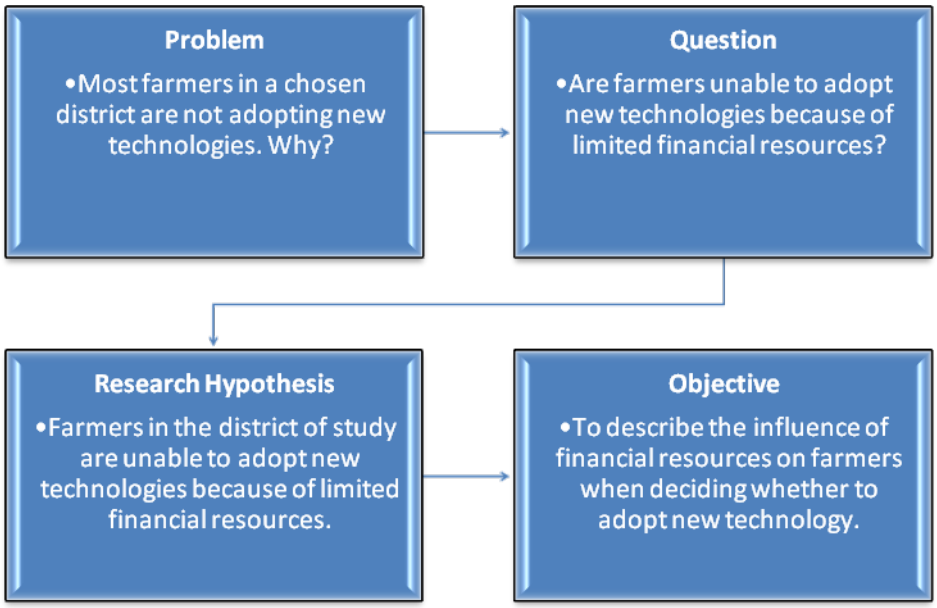


Fig 5.5: Steps in understanding the statement of objective.

Research Methodology: This section details all investigations done. The following are some questions that must be answered in this section.

a. What type(s) of **research design** was used?
b. Why was this type of design used?
c. What **data collection methods** were used? (Primary, secondary; survey, observation) Why were these methods appropriate for this project?
d. What **data collection devices** were used? (Telephone, mail, personal interviewing) Why were these devises appropriate for this project?
e. How was the **population** defined? What were the geographical, age, sex, or other bounds?
f. What **sampling units** were employed?
g. Was a probability or non-probability-**sampling plan** employed? Why?
h. How was the sample actually selected? How large a sample was selected?
i. How was a **list of sampling units** generated? Why was this method used?
j. Were any **difficulties** experienced in contacting designated sample elements? How were these difficulties overcome and was bias introduced in the process? How was the fieldwork done?
k. If questionnaires were used for data collection, was it pre-tested and validated?

Research Findings/ Results and Discussions: Mention all findings and categorize them under each research objective stated earlier in the report. Table and graphs must be used to illustrate the results and appropriately cite them in the text where the explanation is provided. Only important summary tables must be provided in this section, while all the other complex tables detailing the results must be moved to the Appendix section. This section can also be called as “Results and Discussions section”, where the findings are presented as the empirical results of the investigation. Elaboration of the

research findings, also termed as discussion of results, wherein the investigation of the problem objectives proposed in introduction are investigated and reported in terms of the following :

a. Did the investigations accomplish the purpose of research?
b. Did it answer the questions?
c. Did it confirm/ disconfirm the hypothesis
d. Were the results useful? If yes, how? If no, why?
e. Did the investigations lead to new questions?

Conclusions and Recommendations: This section presents some final conclusions about the results of analysis. Relate all important conclusions/findings to the objectives/research questions listed at the beginning of the report. A conclusion for every research objective must be stated clearly. The conclusion should be stated in greater detail than in the executive summary. As part of this, *you should indicate whether the proposed hypotheses were accurate (or not).* If the study does not or could not provide sufficient evidence to make a conclusion about a problem, this should be explicitly stated. The summary must also include recommendations to the probable decision makers.

Executive Summary: This section summarizes the important results and conclusions. This section is usually written at the end of the report.

a. Bibliography: These are the bibliographic reference for each of the works cited in the literature review or in any other section of the report.
b. Appendix: Following is a further list of things, which can be included in the appendix (although it is not limited to these):

    i. Include frequencies and cross tabs.
    ii. Include general information from secondary information.
    iii. Maps used to draw the sample as well as a detailed explanation of the sample design used. Sample size determination.
    iv. Blank questionnaire used in the study.
    v. Summary of all results on a blank questionnaire.

To Do Activity

Prepare a research report as a briefing to a CEO of an international company that is concerned with environmental issues. The CEO has asked for a research report on the role of energy in rural vs urban usage.

You could break down that topic into its parts—the production of food (agriculture), transportation usage, home and other uses—and research those parts in order to present your general perspective and conclusion about how the rural family uses energy. You could bring in information about how families in urban population use energy.

Present your research paper with the back up of ideas and information presented by other researchers on the same topic.

The research report should include the following:
- An informative and interesting Title
- An introduction to your subject and question

- A conclusion section reviewing what you discussed throughout the paper
- Bibliography

Answer the following questions after the research report is drafted:

1. What specific idea do you want the prospective readers to think, feel and understand upon reading your research report?
2. What are the strengths of your report and did you find any difficulties to write the research report?

5.4 APA Style Essentials & Citing and Referencing Sources

The Publication Manual of the American Psychological Association (6th ed., 2014) and the APA Style web site (http://www.apastyle.org/) provide a comprehensive reference guide to writing using APA style, organization, and content. APA style essentials provide all members of an academic department the minimal standards for any report or assignment that specifies APA style.

APA Sample Title / Cover Page

a. The first page of the research report/ research paper should be the title page.
b. ***Center*** the title, your name and your affiliation. (A sample title page of the research report which intends to study the "empowerment of rural women through self-help groups" is shown in Figure 5.6.)

Formatting

a. Double-space entire document
b. Left Justify all the text.
c. Margins : Use 1-inch margins throughout (top, bottom, left, right)
d. Use a single font type and font size (12 pt) throughout the report.

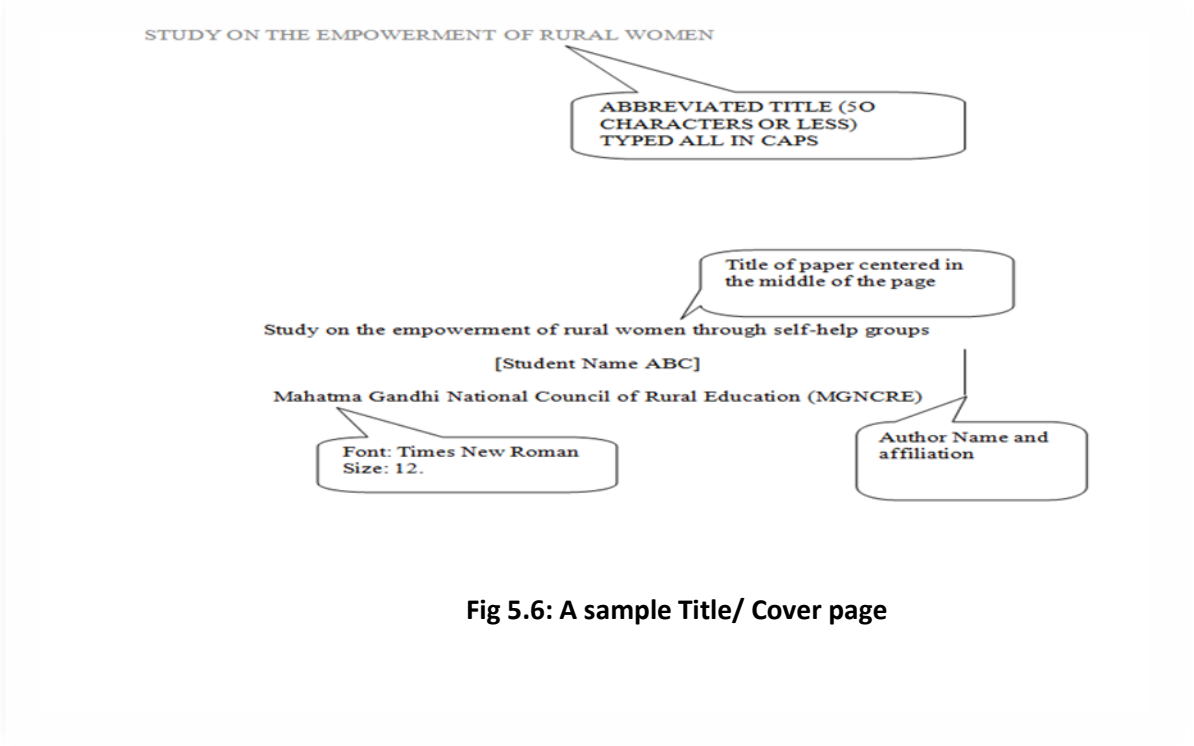


Fig 5.6: A sample Title/ Cover page

Referencing

a. Alphabetize entries by the first word in the citation
b. For sources accessed online, following the citation include the URL: Retrieved from http://url.com
c. If a source has no publication date use the abbreviation n.d. in place of the date. Use n.d. for the in text citation: (Author's Lastname, n.d., p. #)
d. If there is no author, skip it and start with the title of the article, webpage or book. The citation of a text source without an author is mentioned as: ("Article/Book/Page Title," Year, p. #). Abbreviate the title to 1-2 words in your intext citation.

A sample research report in the APA style format is given below:

To do this properly according to APA Style, you will need to include both in text citation and a Reference list. In text citation is a brief parenthetical citation that comes directly after a quote or paraphrased information. The general in text citation format is: (Author's Lastname, Year, p. #).

Here is an example that paraphrases the conclusion of a research study on Rural Entrepreneurship:

Introduction

India is one of among the fast-developing countries with an average GDP growth of around 7% for over last one decade. There has been a good economic progress in the fields of information technology, infrastructure and services. The primary sector (Agriculture, Forestry and Fisheries) which accounts for over 68% of the population (830 million out of 1210 Million, as per 2011 Census) has however been lagging behind with an average annual growth of 4% per annum or less. This is generating a lot of disparity between rural and urban livelihoods (Patil, 2014). Rural economy is in stress because of diminishing returns from agricultural sources. There is, thus an urgent need to accelerate the diversification of rural livelihoods so that rural communities have adequate means and opportunities which are viable and sustainable. Rural Entrepreneurship can address this issue and generate self-employment through pooling of community resources (Shiraleshatti, 2014, p7). The gap between rural and urban development is huge to find a way to fill this gap is still a major issue of developmental process. The only solution to the problem is to develop entrepreneurial activities in the rural sector. The economic, social and educational status of rural India demands skill development for motivating rural people to join business activities along with their routine agricultural activities not only to support agricultural work but to generate additional income for better livelihood. There are a quite good opportunities for Rural Entrepreneurship in areas of harvesting and processing of agricultural, horticultural and fisheries in the rural areas.

References

1. Sangita G. Patil, P.T. Chaudhari (2014), "Problems of Small Scale Industries in India", International journal of Engineering and Management Research, Vol.4, Issue-2, April-2014, ISSN No. 2250-0758; pp. 19-21.
2. A.S. Shiraleshatti;(2014) "Prospects and Problems of MSMEs in India- A study; International Journal of in Multidisciplinary and Academic Research (SSIJMAR); Vol. 1, No. 2, July-August; ISSN 2278-5973; pp 1- 7.
3. www.mgncre.in

5.5 Suggestions for Writing Effective Research Reports

The following are the important characteristics of an effective research report:

a. Focus: an effective report emphasizes the research questions to be investigated.
b. Accuracy: an effective report provides a way to measure how close the sample population is to target population of study.
c. Clarity: a powerful report doesn't confuse the reader.
d. Conciseness: an effective report will be mindful of the readers' time.

Tips to write effective research report are as follows:

a. You will begin writing the report of your study before you have got really done the study:
b. You will begin writing the introductory section of the report as shortly as you have got selected the final approach your study can follow.
c. You will begin writing the results section of the report before you end analyzing the information. However you'll be able to use made-up information to form samples of the tables and figures that may communicate your results. The choices you create in making these tables and figures can assist you decide a way to analyze your information.

Effective Use of Tables and Figures in Writing Research Reports

Table 5.2 helps to decide upon whether to use tables and figures or use plain text to put across key information. Then, follow the best-practice guidelines given in Table 5.2 to ensure that the tables and figures are well-designed.

Table 5.2: How to Choose between Tables, Figures, and Text to Present Data

Use a Table	Use a Figure	Use Text
To show as many numerical values in a precise manner in a small space.	To show trends, patterns and relationships across and between data sets when the study of patterns is more important than the exact data values.	When you don't have extensive data to present.
To compare and contrast data values or characteristics among related items or items with shared characteristics or variables.	Graphs, data plots, maps and pie charts are used to summarize research results	When putting your data into a table would mean creating a table with two or less than two columns/rows.
To show the presence or absence of specific characteristics.	To present visual explanation of a sequence of events, procedures, or physical characteristics.	When the data that you are planning to present is irrelevant to the main study findings.

In places, where mathematical equations are to be written, use the best practices in the context wherever applicable as given in Table 5.3.

Table 5.3:Best Practices in Writing a Research Report involving Mathematical Equations

Context	Best practices to be followed	Examples
Functions and Operators	Functions should be distinguished from ordinary variables	Sin() is a function. S*i*n is a product of three variables s,I,n.
Use of Signs and symbols	Be careful in representing mathematical signs. Use equation editors to represents signs and symbols.	Minus (-) is different from Underscore (_)
Choice of expression of a result as a number or percentage or both?	When numbers are small, represent it as a number only. But, when the numbers are large, use percentage.	138 out of 437 is difficult to process, while 32% is clear. But when the numbers are small, it is better to report the absolute numbers; 2 of 7 is more informative (and easier to visualize) than 29%.
Don't change notation.	Symbols have different meanings in different contexts or situations.	~ This symbol is called a tilde. In symbolic logic (a field of math used a lot in computer sciences) it is used to denote "not." For example, ~A means "not A." In some other contexts, it means "approximately." For example, pi ~= 3.1416.

Finally, a good researcher must be able to effectively communicate the results by means of graphical representations. Hence, knowledge on the types of charts/graphs and when to use them is very important in effective research report writing. Table 5.4 lists down the commonly used graphs and also explains when to use it.

Table 5.4: How to Choose between Different Types of Graphs to Present a Given Data

SI	Graph Type	When is it used	Example of research study
1.	Line Graph	Used to demonstrate trends in data over a period of time.	Trending searches of "Mobile app for farmers" in the last 20 weeks.
2.	Bar Graph	The simplest and most straightforward way to compare various categories is the classic bar graph. The most renowned graph features a series of bars of varying lengths. Categories vs Values are plotted in the two axes of the graphs. The height of each bar is proportionate to the numerical value or the % it represents.	Usage of social media platform in rural regions.
3.	Pie Charts	Pie charts are the simplest and most efficient visual tool for comparing parts of a whole.	A pie chart can quickly and effectively compare various budget allocations, population segments or market-research question responses.
4.	Stacked area charts are frequently used to diagram changes of multiple variables across time.	Stacked area charts are frequently used to diagram changes of multiple variables across time.	Multiple stacked lines can be drawn to track the population changes of various states across time. The area below each line can be colored a different hue to represent the state it signifies, resulting in a graph that clearly represents population trends, while at the same time displaying each state's data in order from least to most populous.
5.	Stacked Bar Graph	When studying groups of people, it's common to compare multiple variables at once. After all, it's enormously more useful to examine category wise ages and gender in addition to total population. A stacked bar graph illustrates trends and proportions in a single graph.	Rather than simply illustrating changes in global population over time with a traditional column bar graph, a stacked bar graph can also represent the tribal makeup of the population during each year and how those proportions have changed during the same period.
6.	Scatter diagrams	Scattergrams, also known as scatter plots, are graphs that show the relationship between	By plotting certain data sets, scientists can discover trends of which they might not otherwise be

		two or more variables. The plots use mathematical coordinates to represent two variables of a data set.	aware. For example, a scatter plot might allow a doctor to plot patients' resting heart rates against their BMI.
7.	Pareto Charts	Sometimes a basic graph doesn't display enough information to draw the necessary conclusion. A Pareto chart combines a bar graph with a line graph to illustrate not only categories' individual values, but also the cumulative total of the entire set.	In a Pareto chart that tracks the type and frequency of food defects, the bars illustrate each type of defects' total occurrences – as reported on one of the charts' axes – while the line charts the cumulative frequency of all categories, from most to least prevalent.
8.	Flowcharts	Following a proper process is most important in many places.	Following the proper process is probably more important in medicine than in any other field. After all, if the surgeon forgets a step, you might very well bleed to death while you sleep. Flow charts are frequently used by hospitals, clinics and other medical facilities to ensure proper procedures are uniformly followed.
9.	Pictograms	In a pictogram, or pictograph, images and symbols are used to illustrate data.	For example, a basic pictogram might use an image of the sun to signify each good weather day in a month and a cloud to symbolize each rainy day.Since images are known to hold more emotional power than raw data, pictograms are often used to present medical data.
10.	Histogram	By definition, a histogram is a special type of vertical bar graph that presents numeric data and its frequency distribution. As its name suggests, the distribution is often illustrated across time, but the data could also be plotted based on any chronological scale, such as temperature, elevation or monetary value.	Plotting the annual average temperature of a region.

Given below are few examples to illustrate the use of different graphs in the context of Indian population study.

Example 1: What graph do you choose to illustrate life expectancy at Birth for male and female in India from the year 1951 – 2006. Why?. What do you infer from the graph?

Life Expectancy at Birth (Years)			
Year	Female	Male	Total
1	2	3	4
1951-61	40.6	41.9	41.3
1961-71	44.7	46.4	45.6
1970-75	49	50.5	49.7
1976-80	52.1	52.5	52.3
1981-85	55.7	55.4	55.4
1986-90	58.1	57.7	57.7
1991-95	60.9	59.7	60.3
1992-96	61.4	60.1	60.7
1993-97	61.8	60.4	61.1
1994-98	62.2	60.6	61.4
1995-99	62.5	60.8	61.7
1996-00	62.7	61	61.9
1997-01	63	61.3	62.2
1998-02	63.3	61.6	62.5
1999-03	63.5	61.8	62.7
2000-04	63.7	62.1	62.9
2001-05	63.9	63.3	63.6
2002-06	64.2	62.6	63.4
Source: Office of the Registrar General, India.			
Notes: Figures for 1951-61 and 1961-71 are based on Census Actuarial Reports and for 1970-75 onwards on the basis of estimates from Sample Registration System.			

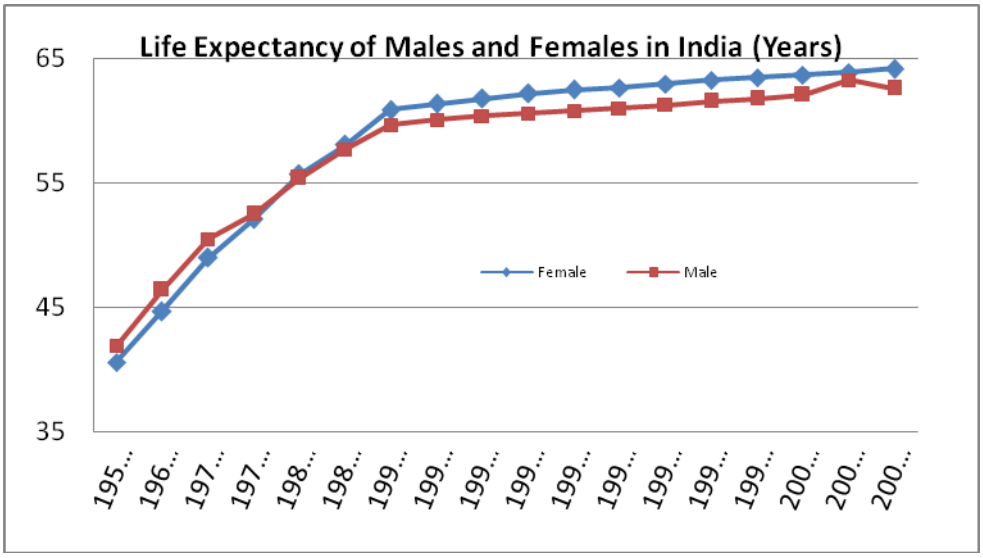


Fig5.7: Line Graph

Observation: Line graph (Figure 5.7) is chosen to illustrate the trend in data over time. It is inferred from the graph that after 1991, the life expectancy of females are higher than the males.

Example 2: What graph will you choose to illustrate the percentage of population of India with respect to the range in age? What do you observe?

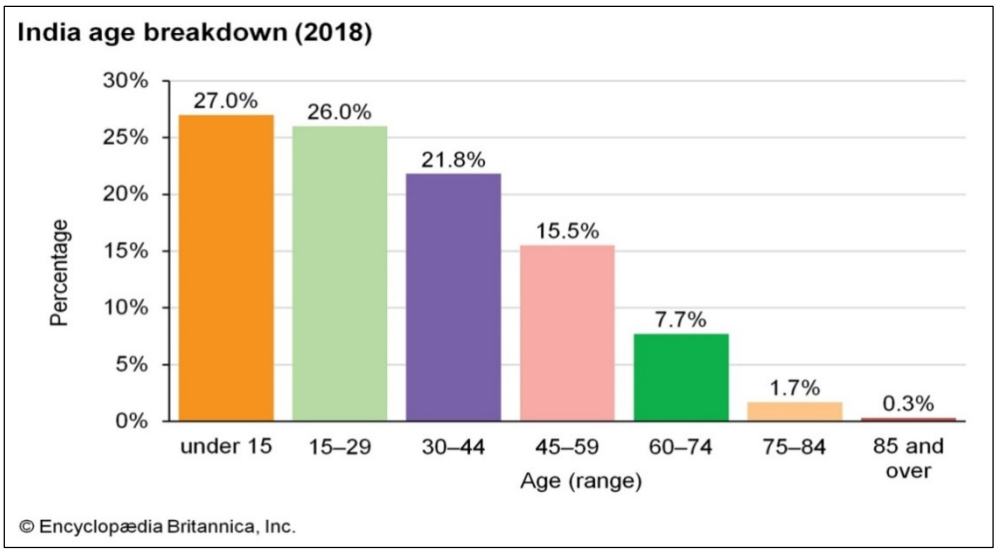


Fig 5.8: Bar Graph

Observation: Bar Graph (Figure 5.8) is used to illustrate the percentage of population of India wrt the varied age groups in one axes and population percentage in the other axes. It is evident from the graph that India has more than 50% of the youth population (less than 30 years of age).

Example 3: The following graph illustrates data of different states regarding population of states in the year 2008. Explain why pie chart (Figure 5.9) is chosen to illustrate this study, given Total population of the given States = 3276000?

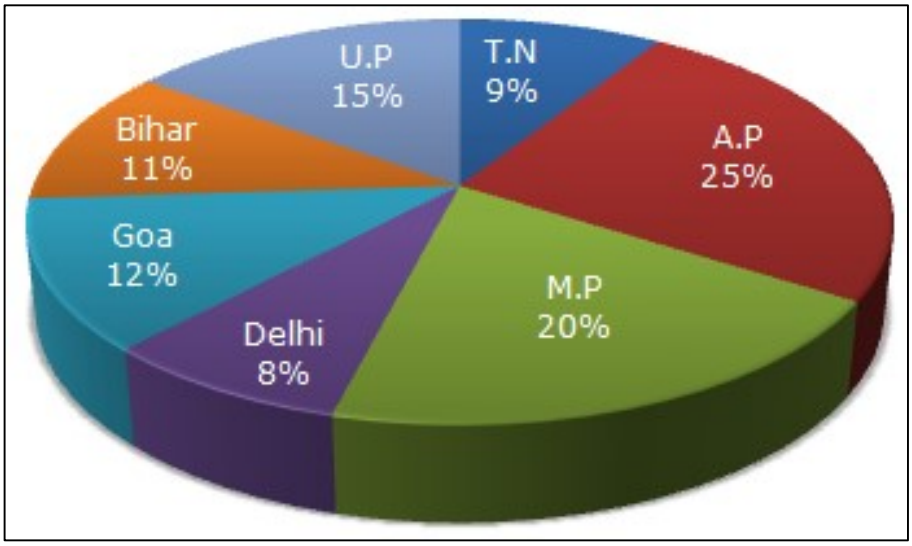


Fig 5.9: Pie Chart

Observation: A pie chart can effectively compare various population segments. Hence, it is used to study the population of different states chosen.

Example 4: Given the budget allocation for various schemes by the Govt of India in 2017-18 and 2018-19 in the table below:

Sectors	2017-18	2018-19
Rural development	1,35,000	138000
Agriculture and allied activities	56589	63836
Subsidy for fertilizer	64974	70080
Subsidy for food	140000	169000

The amount (in Rs is given in crores). What graph do you choose to represent the data? Why?

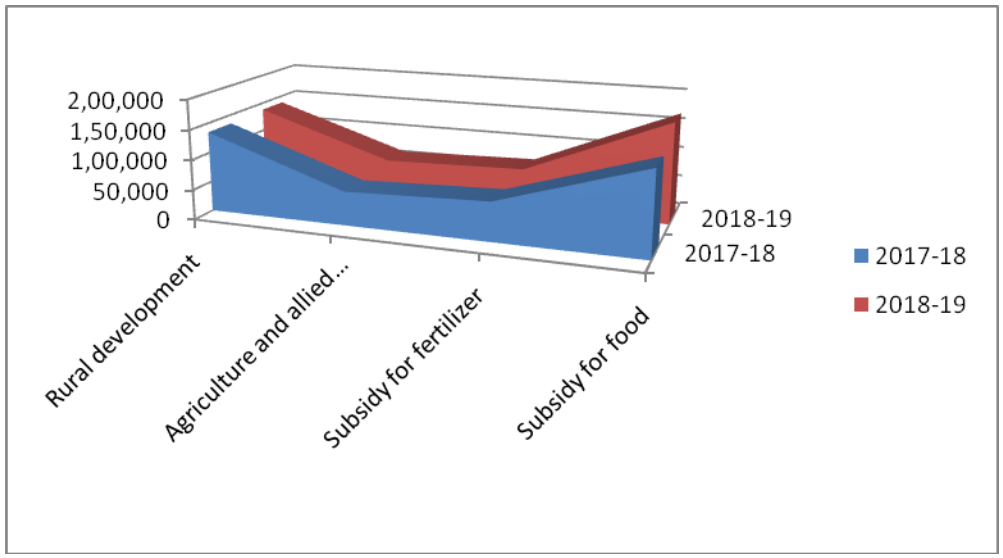


Fig 5.10: Stacked Area Charts

Observation: Stacked area chart (Figure 5.10) is used to represent the given data because they are used to diagram changes of multiple variables across time. The area below each line is colored a different hue to represent the budget head it signifies, resulting in a graph that clearly represents trends, while at the same time displaying each budget head's data.

Example 5: The following are the reason of migration from rural to urban areas of male and female population in a particular year along with the percentage of migration as shown.

Reason for Migration	Employment	Education	Family Moved	Marriage	Business	Others
Female	3.2	1.3	23.7	64.9	0.3	6.7
Male	37.6	6.2	35.6	2.1	2.9	15.7

Which chart will you choose to represent the above data? Justify.

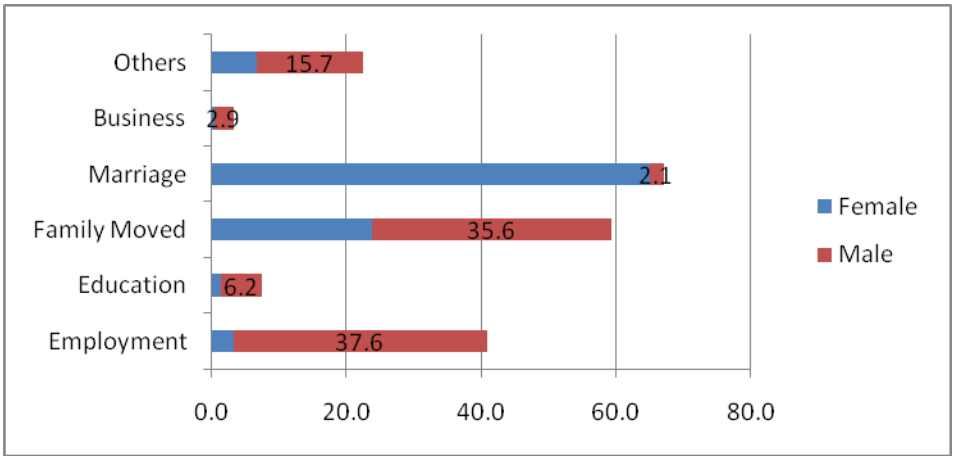


Fig 5.11: Stacked bar graph

Observation: When studying groups of people, it's common to compare multiple variables at once. Hence, a stacked bar (Figure 5.11) graph that combines elements of the traditional bar graph and the pie graph to communicate totals, trends and proportions in a single chart.

Example 6: The Human Capital Index (HCI) database provides data at the country level for each of the components of the Human Capital Index as well as for the overall index, disaggregated by gender. The table below shows the HCI of South Asian countries. What chart do you select to plot the given data? Why?

Country Name	WB Code	Region	Income Group	Human Capital Index
Afghanistan	AFG	South Asia	Low income	0.389178
Bangladesh	BGD	South Asia	Lower middle income	0.478669
India	IND	South Asia	Lower middle income	0.440069
Nepal	NPL	South Asia	Low income	0.49014
Pakistan	PAK	South Asia	Lower middle income	0.388698
Sri Lanka	LKA	South Asia	Lower middle income	0.583909

Source: http://databank.worldbank.org/data/source/human-capital-index.

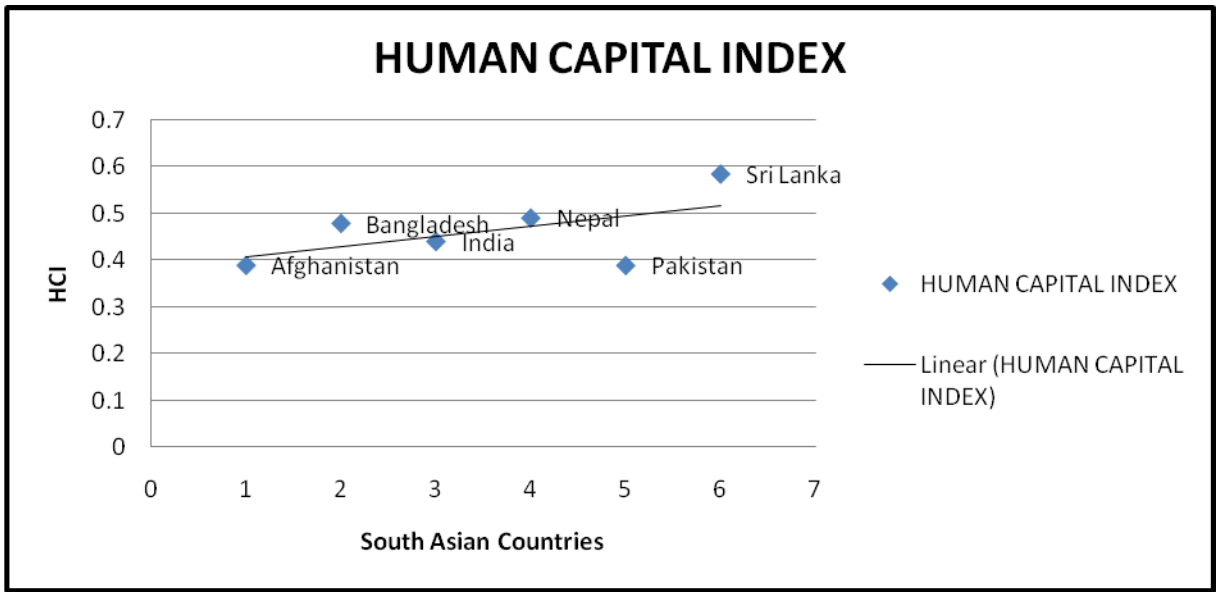


Fig 5.12: Scattergrams

Observation: Scattergrams, also known as scatter plots (Fig 5.12), are graphs that show the relationship between two or more variables. By plotting certain data sets, scientists can discover trends of which they might not otherwise be aware.

Example 7 : A survey was taken in a village to study the reasons for not using several mobile apps to help farmers make use of technology for farming tips. The following table gives the summary of count of the reasons reported by the farmers. What chart do you use to depict this data? Why?

User Complaints	Count
Bad UI Design	2
Lack of technology awareness	78
Frequent crashes	7
Low quality	21
Many bugs	5
Over priced	12
Poor tech support	56
Slow performance	9
Unresolved issues	2
Unresponsiveness	1

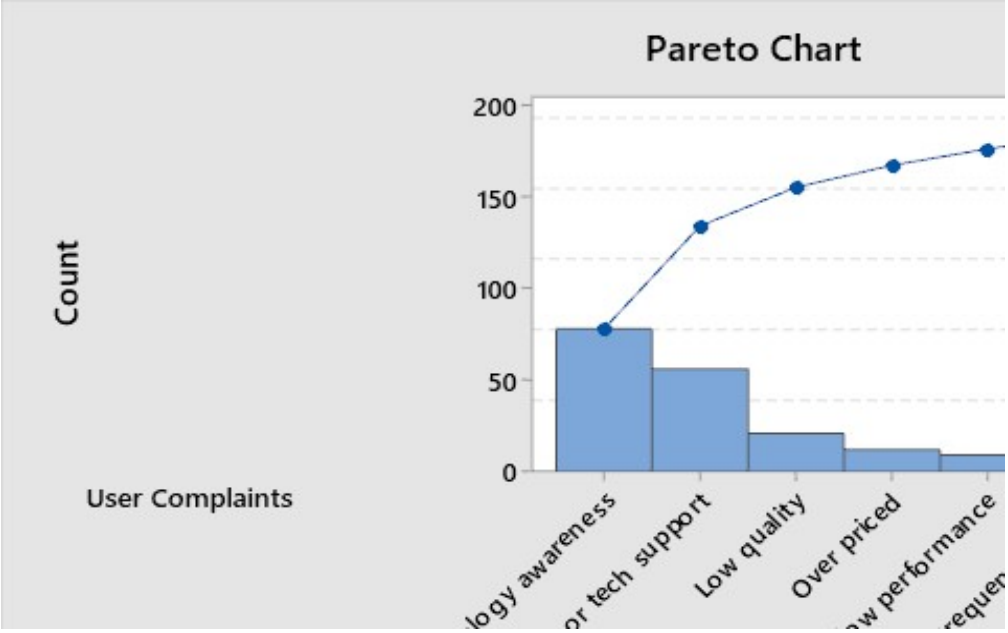


Figure 5.13: Pareto Chart

Observation: In a Pareto chart (figure 5.13) that tracks the type of issue faced by the farmer and the frequency of issues, the bars illustrate each type of issue' total occurrences – as reported on one of the charts' axes – while the line charts the cumulative frequency of all categories, from most to least prevalent. The result is a graph that clearly reflects the most common issue faced is 'lack of technology awareness' and what percentage of the whole each represents.

Example 8: Write a report to illustrate the journey of preparing white rice from paddy.

Observation: The Figure 5.14 given below illustrates the journey of preparing white rice from paddy using a flow chart.

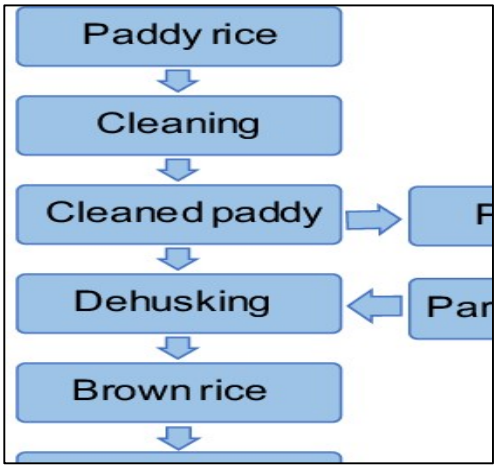


Fig 5.14: Flow chart

To do activity:
After understanding to draw a flowchart, prepare one to illustrate the journey of wheat to bread at

home.

Example 9: A researcher intends to represent the percentage of Indian population that are vegetarians and non-vegetarians across all states of India. What graph will you use? Justify.

Observation: Images are known to hold more emotional power than raw data. Hence, pictograms (Figure 5.15) is used to present such data.

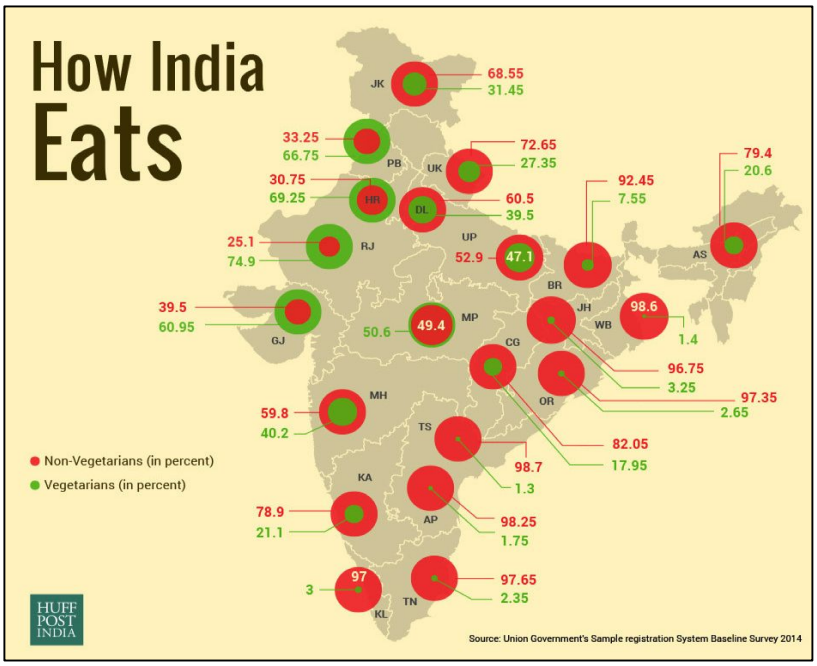


Fig 5.15: Pictogram

Example 10: Non-agricultural jobs in rural areas include jobs like beedi making, sweeping, cleaning, tractor driving, weaving, loading/unloading. The average hourly wages for the listed jobs and the number of workers for each wage category is as given below:

Hourly Wages (Rs)	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
No of workers	8	16	30	45	62	32	15	6

What chart will you choose to show the distribution of hourly wages of non agricultural jobs in a village? Justify.

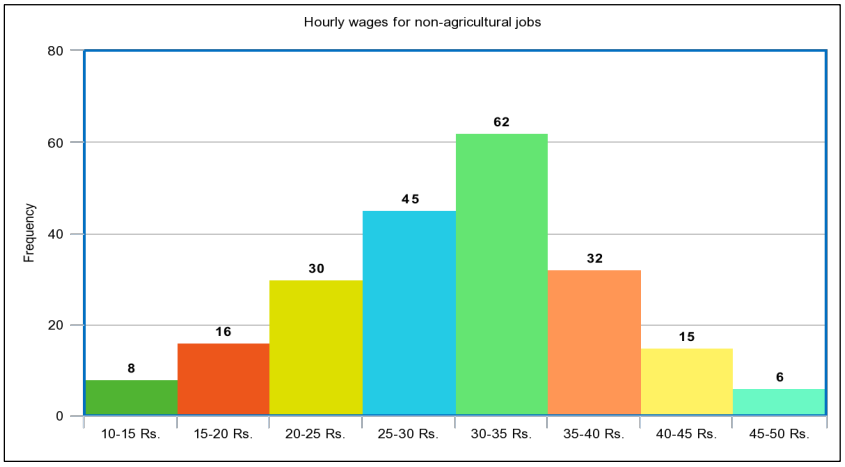


Fig 5.16: Histogram

Observation: As its name suggests, the distribution is often illustrated across time, but the data could also be plotted based on any chronological scale, such temperature, elevation or monetary value as shown in Figure 5.19.

To Do Activity

Institutional Delivery

The family planning programme (now RCH, 1997) since its inception in 1951 has forayed into contributing towards betterment of health of both mother and child along with its initial family planning services. Overall, 47% deliveries took place at a health facility in India during 2007-08. 46.5% couples are protected by at least one of the methods of family planning.

As per National Family Health Survey-III, 2005-06, 99.6% of ever married females in the urban areas are aware of any family planning method as compared to 99.1% in the rural areas. Female sterilization is the most widely known family planning method (known to about 98.4% of married females) followed by male sterilization.

During 2005-06, about 46% of the eligible couples were effectively protected by some method or the other, sterilization being the most widely used method protecting about 29% of eligible couples.

75% pregnant women have received at least one ANC and 51% pregnant women have received 3 ANC visits during 2005-06 in India. **Represent the given data in a graphical representation.**

Summary of the Chapter

Summary:Writing research report is considered as one of the most difficult tasks researchers have to do. Often it takes long days to compile a good report of the research findings. This chapter has given a comprehensive guidelines to make the researchers prepare their research reports in a well-organized, readable, consistent and in a most acceptable and understandable manner to the intended audience.

Model Questions

1. Describe the inductive approach to research, and provide examples of inductive research.

2. Describe the deductive approach to research, and provide examples of deductive research.

3. Explain with example, the different types of research reports.

4. Explain in detail the components of a research report.

5. Explain APA Style formatting.

6. How to choose between tables, figures, and text to present data.

7. Explain some best practises to represent mathematical equations.

8. Given the Daily Average Intake of Energy and Proteins of Rural Population: 2004-06 of Kerala, Tamil Nadu, Karnataka, Andhra Pradesh, Maharashtra, Gujarat, Madhya Pradesh, Orissa and West Bengal. Represent the given data in a suitable graphical representation. Justify the chosen graphical representation.

Age (years)	Sex	Activity	Energy (kcals/day)		Protein (gram/day)	
			Intake	RDI	Intake	RDI
1	2	3	4	5	6	7
1-3	Boy & Girl		719	1240	20	22
4-6	Boy & Girl		1020	1690	29	30
7-9	Boy & Girl		1230	1950	34	45
10-12	Boy		1423	2190	39	54
	Girl		1389	1970	38	57
13-15	Boy		1645	2450	45	70
	Girl		1566	2060	42	65
16-17	Boy		1913	2640	53	78
	Girl		1630	2060	44	63
>= 18	Male	Sedentary	2000	2425	55	60
		Moderate	2184	2875	59	60
>= 18	Female (NPNL)	Sedentary	1738	1875	47	50
		Moderate	1808	2225	49	50

Source: Based on National Nutrition Monitoring Bureau (NNMB) 'Diet and Nutritional Status of Rural Population' Technical Report No. 24 conducted during 2004-06 in rural areas of Kerala, Tamil Nadu, Karnataka, Andhra Pradesh, Maharashtra, Gujarat, Madhya Pradesh, Orissa and West Bengal.

RDI: Recommended Dietary Intakes.

NPNL: Non-Pregnant Non-Lactating.

9. Given the 2004-05 Unemployment rates in the States and Union Territories of India in the table given below. Choose an appropriate graph to represent the same. Justify the chosen method.

Unemployment Rates in States and Union Territories: 2004-05				
State /Union Territory	Rural		Urban	
	Female	Male	Female	Male
1	2	3	4	5
Andhra Pradesh	0.4	1	3.8	3.6
Arunachal Pradesh	0.6	1.1	2.8	1.1
Assam	3.1	2.4	9.1	6.9
Bihar	0.2	1.8	4.1	6.7
Chhattisgarh	0.3	0.8	2.4	3.8
Goa	15.7	9.1	11.8	7.6
Gujarat	0.2	0.8	2.9	2.3
Haryana	1	2.8	7.5	3.2
Himachal Pradesh	2	1.6	10.1	1.7
Jammu & Kashmir	1.3	1.7	10.9	3.7
Jharkhand	0.1	2	2.3	7.5
Karnataka	0.8	0.7	5.7	1.9
Kerala	20.1	5.1	33.4	6.2
Madhya Pradesh	0.1	0.7	1.6	3.1
Maharashtra	0.3	1.5	4.1	3.5
Manipur	0.7	1.4	6.3	5.2
Meghalaya	0.5	0.1	3.5	3.5
Mizoram	0.1	0.5	2.6	1.6
Nagaland	1.4	2.2	7.2	4.6
Orissa	8.3	3.1	26.6	9
Punjab	4.9	3.3	14	2.9
Rajasthan	0.1	1.2	2.9	2.8
Sikkim	1.5	2.8	4.3	3.6
Tamil Nadu	1.1	1.2	4.8	2.9
Tripura	32	9.6	56.8	16.6
Uttarakhand	0.4	2	10.2	4.2
Uttar Pradesh	0.3	0.7	2.5	3.5
West Bengal	3.3	2.2	8.4	5.6
Andaman & Nicobar Islands	12.3	3.7	17.2	6.5
Chandigarh	4.8	2.5	7.5	3.1
Dadra & Nagar Haveli	3.6	3.1	9.1	1.3
Daman & Diu	0	0.4	3.3	2.8
Delhi	0	2	6.4	4.6
Lakshadweep	57.1	0.9	51.5	11.1
Puducherry	3.2	9.4	19.5	4.1
India	1.8	1.6	6.9	3.8
Source: National Sample Suvrey Office, 61st Round (July 2004-June 2005).				
Note: Figures relate to usual status of individuals. The figures represent size of unemployment as percentage of labour force.				

10. Study the table given below:

(i) Plot the given data into a graph. What type of graph would you choose? Justify.

(ii) What are your observations from the graph.

Mean Age at Marriage (in years)	
Year	Female
1	2
1951	15.4
1961	16.1
1971	17.1
1981	17.9
1992*	19.5
1993*	19.6
1994*	19.4
1995*	19.4
1996*	19.4
1997*	19.5
1998	19.5
1999	19.6
2000	19.8
2001	19.9
2002	20.0
2003	20.1
2005	20.2
2006	20.5
2007	20.6
2008	20.7
2009	20.7
Sources:	
1. Population of India : ESCAP Country Monograph No. 10 and Female Age at Marriage; Census of India : Occasional Paper No. 7 of 1988, Office of the Registrar General, India.	
2. Sample Registration System, Office of the Registrar General, India .	
Note: Figures for 1951, 1961, 1971 and 1981 aresingulate mean age at marriage based on population census data. 1992 onwards figures are the mean age at effective marriage based on Sample Registration System.	
***Excludes Jammu and Kashmir.**	

References

1. Ackoff, R. L. (1961), The Design of Social Science Research, Chicago: University of Chicago Press.
2. APA Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting Standards for Research in Psychology: Why Do We Need Them? What Might They Be? American Psychologist, 63(9), 839-851.
3. Caro, S. (2009). How to Publish Your PhD. London: Sage.Kothari, C.R.(1978), Quantitative Techniques, New Delhi, Vikas Publishing House Pvt. Ltd.

Editors' Profile

Dr W G Prasanna Kumar

Dr. W. G. Prasanna Kumar, PhD in Education with basic degree in Social Work and Master's Degrees in Sociology, Public Administration and Political Science has professional education in Environmental Economics, Public Relations, Communication and Training and Development. Presently Chairman, Mahatma Gandhi National Council of Rural Education (MGNCRE) under the Ministry of Human Resource Development, in Government of India strives to promote resilient rural India through Higher Education interventions. The national initiative of reviving Mahatma Gandhi's ideas of NaiTalim, spearheaded by Dr. W G Prasanna Kumar, has met unprecedented success at both national and state levels. The primary objective of this initiative is to promote Gandhiji's ideas on Experiential Learning, NaiTalim, Work Education and Community Engagement, and mainstreaming them in School Education and Teacher Education Curriculum & Pedagogy. As Professor and Head Centre for Climate Education and Disaster Management in Dr MCR HRD Institute, conducted several capacity building and action research programmes in climate education, disaster management and crowd management. He has handled many regional, national and international environmental education programmes and events including UN CoP11 to Convention on Biological Diversity and Media Information Management on Environmental Issues.

He was Director in National Green Corps in the State Government for over 11 years and Senior Social Scientist in State Pollution Control Board for 6 years. Conducted various curriculum and non- curriculum related training programmes in environmental education. He was a Resource Person for AP Judicial Academy, AP Police Academy, AP Forest Academy, EPTRI, Commissonerate of Higher Education and Intermediate Education, State Council for Educational Research and Training and National Council for Educational Research and Training New Delhi, CCRT, Bharathiya Vidyapeet University Pune, CPR Environmental Education Centre Chennai and Centre for Environment Education Ahmedabad. Dr W G Prasanna Kumar was trained in Community Consultation for Developmental Projects in EPA Victoria Australia in 1997 trained as State Chief Information Officer by IIM Ahmedabad and MCRHRDI Government of Andhra Pradesh in 2004 and trained in Environmental Education and Waste Management Technique by JICA, Japan in 2011.

He was awarded Best State Nodal Officer of National Green Corps Award from Centre for Science and Environment, New Delhi, 2008, Jal Mithra Award from Earthwatch Institute of India and Water Aid New Delhi, 2014 and Certificate of Commendation for the services in UN Conference of Parties to Convention for Biodiversity conducted at Hyderabad from 1-20 October 2012 by the Government of Andhra Pradesh 2012.

Dr K N Rekha

Dr K N Rekha, is a PhD Graduate from IIT Madras. She has 14 years of experience in training and education Industry. She works at Mahatma Gandhi National Council of Rural Education (MGNCRE), Hyderabad as Senior Faculty. She is involved in curriculum development on Rural Management and Waste Management. Prior to this, she worked as a researcher at Indian School of Business, Hyderabad, a short stint at Centre for Organisation Development (COD), Hyderabad. She has co-authored a book on "Introduction to Mentoring", written book chapters, peer reviewed research papers, book reviews, Case studies, and caselets in the area of HR/OB. She also presented papers in various national and international conferences. Her research areas include Mentoring, Leadership, Change Management, and Coaching. She was also invited as a guest speaker at prominent institutions like IIT Hyderabad.

AUTHORS' PROFILE

Dr.S.Prasanna Devi is Professor and Head of the Department of Computer Science and Engineering at SRM Institute of Science and Technology, Vadapalani Campus, Chennai. She completed her BE (2004), ME (2007) and PhD (2011) in Computer Science and Engineering from the College of Engineering, Anna University, Chennai. She completed her Post Doctoral research (2015) from the prestigious Indian Institute of Science, Bangalore. She is the recipient of Tamil Nadu Young Scientist Fellowship Award in the year 2015 from the Tamil Nadu State Council for Science and Technology. Dr.S.Prasanna Devi has an overall teaching and research experience of more than 12 Years. She is also the Honorary Secretary of ORSI – Operations Research Society of India, Chennai Chapter. She has 30+ Scopus indexed journal publications as on December 2019 to her credit. She has extensively travelled around India and places abroad like Florida, Paris, Switzerland, Sweden and Germany for participating in many International conferences. Her research interests include artificial intelligence, data analytics, game theory, optimization algorithms and machine learning.

Dr.Meenakshi S Arya is Professor of Computer Science at SRM University in India. She has a diverse academic and professional experience background and has served in many engineering colleges across the length and breadth of the country in various capacities. Dr Arya has done her B.Tech. in Computer Science and Technology from the prestigious SNDT Women's University Mumbai and is a topper in M.Tech. from College of Engineering, Pune. She has done her Ph.D. in the field of Biometric Watermarking and has an array of publications to her credit. Dr Arya has served as Head of the Department of Department of Computer Science and Engineering at G H Raisoni College of Engineeing, Nagpur, Department of Information Technology at Vidyalankar Institute of Technology, Mumbai, Department of Computer Science and Engineering at SRIEIT, Goa and Department of Computer Science and Engineering at BFCET, Bathinda. Her research interests are in Data Science & Analytics and Digital Image Processing. She has published papers in reputed Journals and conferences and has many book chapters to her credit. She is a senior IEEE member and fellow of IETE. Her google scholar citations are more than 100 with h-index as 6 and i-10 index as 5. She serves on the editorial and review board of a multitude of journals related to Computer Science. She has 2 patents and 1 copyright to her credit.

Mahatma Gandhi National Council of Rural Education (MGNCRE)

Department of Higher Education

Ministry of Human Resource Development, Government of India

सत्यमेव जयते

Where there is Rural Wellbeing
there is Universal Prosperity

सत्यमेव जयते

Mahatma Gandhi National Council of Rural Education
Department of Higher Education
Ministry of Education, Government of India

Where there is Rural Wellbeing
there is Universal Prosperity

040 - 2321 2120

admin@mgncre.in
www.mgncre.in

#5-10-174, Shakkar Bhavan, Fateh Maidan Lane
Band Colony, Basheer Bagh,
Hyderabad-500004