RESEARCH METHODOLOGY AND HEALTHCARE ANALYTICS

MASTER OF BUSINESS ADMINISTRATION (HOSPITAL ADMINISTRATION)

FIRST YEAR, SEMESTER-II, PAPER-IV

DIRECTOR, I/c.

Prof. V. Venkateswarlu

M.A., M.P.S., M.S.W., M.Phil., Ph.D.

CENTRE FOR DISTANCE EDUCATION
ACHARYA NAGARJUNA UNIVERSITY
NAGARJUNA NAGAR 522 510

Ph: 0863-2346222, 2346208
0863- 2346259 (Study Material)
Website www.anucde.info
E-mail: anucdedirector@gmail.com

MBA (HA): Research Methodology and Healthcare Analytics

First Edition : 2025

No. of Copies :

This book is exclusively prepared for the use of students of MASTER OF BUSINESS ADMINISTRATION (Hospital Administration) Centre for Distance Education, Acharya Nagarjuna University and this book is meant for limited circulation only.

Published by:

Prof. V. VENKATESWARLU
Director, I/c
Centre for Distance Education,
Acharya Nagarjuna University

Printed at:

FOREWORD

Since its establishment in 1976, Acharya Nagarjuna University has been forging ahead in the path of progress and dynamism, offering a variety of courses and research contributions. I am extremely happy that by gaining 'A+' grade from the NAAC in the year 2024, Acharya Nagarjuna University is offering educational opportunities at the UG, PG levels apart from research degrees to students from over 221 affiliated colleges spread over the two districts of Guntur and Prakasam.

The University has also started the Centre for Distance Education in 2003-04 with the aim of taking higher education to the door step of all the sectors of the society. The centre will be a great help to those who cannot join in colleges, those who cannot afford the exorbitant fees as regular students, and even to housewives desirous of pursuing higher studies. Acharya Nagarjuna University has started offering B.Sc., B.A., B.B.A., and B.Com courses at the Degree level and M.A., M.Com., M.Sc., M.B.A., and L.L.M., courses at the PG level from the academic year 2003-2004 onwards.

To facilitate easier understanding by students studying through the distance mode, these self-instruction materials have been prepared by eminent and experienced teachers. The lessons have been drafted with great care and expertise in the stipulated time by these teachers. Constructive ideas and scholarly suggestions are welcome from students and teachers involved respectively. Such ideas will be incorporated for the greater efficacy of this distance mode of education. For clarification of doubts and feedback, weekly classes and contact classes will be arranged at the UG and PG levels respectively.

It is my aim that students getting higher education through the Centre for Distance Education should improve their qualification, have better employment opportunities and in turn be part of country's progress. It is my fond desire that in the years to come, the Centre for Distance Education will go from strength to strength in the form of new courses and by catering to larger number of people. My congratulations to all the Directors, Academic Coordinators, Editors and Lesson-writers of the Centre who have helped in these endeavors.

Prof. K. Gangadhara Rao
M.Tech., Ph.D.,
Vice-Chancellor I/c
Acharya Nagarjuna University.

MASTER OF BUSINESS ADMINISTRATION
(HOSPITAL ADMINISTRATION)
Programme Code: 197
PROGRAMME SYLLABUS
1st YEAR – I1nd SEMESTER SYLLABUS
204HA26: RESEARCH METHODOLOGY AND HEALTHCARE ANALYTICS

Unit-I: Research Methodology: An introduction- meaning of research-objectives of research-types of research- research process- Defining the Research Problem – Criteria for a good research problem.

Unit-II: Theory of Sampling: concept of sampling- probability and non-probability sampling techniques- size of the sample- sampling distribution- sampling error- criteria for selecting- sampling procedure- census and sample.

Unit-III: Measurement and Scaling: nature- types of measurement scales- nominal, ordinal, interval and ratio- validity and reliability in scaling- MDS & ranking and rating scale-Likert's Summated scale- Thurstane's Equal Appearing intervals- out line of MDS-Q sort.

Unit-IV: Data collection and Presentation: primary and secondary data- methods of data collection- questionnaire- designing questionnaire, pre testing questionnaire- **Data Presentation**: editing of primary data, classification of data, guidelines for class selection-objectives and types of classification, frequency distribution- charting of the data- bar chart, histograms and two-dimensional graphs, mean, median, S.D, Regression analysis correlations- chi-square.

Unit-V: Healthcare Analytics and Report Writing: introduction of health care analytics-data- utilization of basic data-sources of health statistics - problems in collection of sickness data- measurement of sickness- vital statistics- Report Writing.

Reference Books

1. Luck.J. David. Ranald S. Rubin, Market Research, Prentice Hall of India, New Delhi,1999.
2. G.C. Beri, Marketing Research, Tata MacGrawhill, New Delhi, 1996.
3. Green.E. Paul. Danald S. Tull, Gerald Albaum, Research for Marketing Decisions,Prentice Hall, New Delhi, New Delhi, 1996.
4. Adrian Payne, The Essence of Services Marketing, Prentice Hall, New Delhi, 1996
5. Luck.J. David, Hugh G. Wales, Donald a Taylor, Ronald S. Rubin, Marketing Research, Prentice Hall, 1982.
6. C.R.Kothary" Research Methodology"New Age International Pvt Ltd Publishers; 2nd edition 2009
7. CR Kothari GauravGarg Research Methodology methods and Techniques, New Age International Publishers.
8. **Healthcare Analytics : Foundations And Frontiers 1St Edition by Ross M. Mullner Edward M. Rafalski, T&F/Routledge**
9. **Chandan K Reddy &Charu C Agarwal, Healthcare Data Analytics, CRC Press**

PandianSundara P , Muthulakshmi S , Vijayakumar T. Research Methodology and Applications of SPSS in social science research.
10. Luck.J. David, Hugh G. Wales, Donald a Taylor, Ronald S. Rubin, Marketing Research, Prentice Hall, 1982.

CONTENTS

Lesson 1

INTRODUCTION TO RESEARCH METHODOLOGY

Structure of the Lesson

1. Meaning of research
2. Objectives of research
3. Importance and scope of research in healthcare

1.Introduction

Research in common parlance refers to a search for knowledge. Once can also define research as a scientific and systematic search for pertinent information on a specific topic. In fact, research is an art of scientific investigation. The Advanced Learner's Dictionary of Current English lays down the meaning of research as "a careful investigation or inquiry specially through search for new facts in any branch of knowledge." Redman and Mory define research as a "systematized effort to gain new knowledge." Some people consider research as a movement, a movement from the known to the unknown. It is actually a voyage of discovery. We all possess the vital instinct of inquisitiveness for, when the unknown confronts us, we wonder and our inquisitiveness makes us probe and attain full and fuller understanding of the unknown. This inquisitiveness is the mother of all knowledge and the method, which man employs for obtaining the knowledge of whatever the unknown, can be termed as research.

2. What is Research?

- Research can be in a common parlance
    - ❖ Answers to Questions (why shape of the earth was flat)
    - ❖ Solutions to problems
    - ❖ It is a purposive investigation
    - ❖ It is an organized inquiry
    - ❖ It seeks to find explanations to unexplained phenomenon

- To clarify doubtful propositions
- To correct the misconceived facts
- "Search for a knowledge"
- In a short-term the research means a systematic method of finding the solution to a particular problem
- Scientific and systematic search for pertinent information on a specific topic;
- Research is an art of scientific investigation
- A careful investigation or inquiry specially through search for new facts in any branch of knowledge
- Systematized effort to gain new knowledge
- Some people consider research as a movement from the known to the unknown

3. **Meaning of Research**:
   - The Advanced Learner's Dictionary of current English lays down the meaning of research as **"*A careful investigation or inquiry specially through search for new facts in any branch of knowledge.***
   - **Redman and Mary define** ***research as a systematized effort to gain new knowledge***
   - ***According to Clifford woody research comprises defining and redefining problems, formulating hypothesis or suggested solutions, collecting, organising and evaluating data making deductions and research conclusions: and at last carefully testing the conclusions to determine whether they fit the formulating Hypothesis.***
   - **Webster's Twentieth century dictionary** defines the term Research as a careful, pertinent, systematic diligent enquiry or examination in some field of knowledge undertaken to establish facts or principles.

- ❖ The term Research refers to the **systematic method consisting of stating clearly the problem, formulating a Hypothesis, collecting the facts or data, analysing the data and reaching certain conclusions either in the form of solutions** towards the concern problem.
- ❖ Research is, thus, an original contribution to the existing stock of knowledge making for its an advancement.

4. SCIENCE AND SCINETIFIC RESEARCH

The word science is derived from the Latin word "scientia" which means knowledge. Science refers to a systematic and organised body of knowledge in any area of inquiry that is acquired using "the scientific method"

Science can be grouped into the two broad categories.

1) Natural Science
2) Social Science.

Natural science is the science of naturally occurring objects (or) Phenomena, such objects, matters earth celestial bodies (or) the human body, natural science further classified into Physical Science, Earth science, Life sciences and others.

Physical science consists of disciplines such as Geology.

Life science includes disciplines such as human Biology, Zoology (The science of Animals) and Botany (science of Plants).

Social science is the science of people (or) collections of people such as groups, firms, societies, (or) economics and their individual (or) collective Behaviours.

Social science can be classified into disciplines such as Psychology (the science of human Behaviour), sociology (the science of social group) and Economics (the science of firms, markets and economics) etc.

5. OBJECTIVES OF RESEARCH

The purpose of research is to discover answers to questions through the application of scientific procedures. The main aim of research is to find out the truth which is hidden and which has not been discovered as yet. Though each research study has its own specific

purpose, we may think of research objectives as falling into a number of following broad groupings:

1. To gain familiarity with a phenomenon or to achieve new insights into it (studies with this object in view are termed as *exploratory* or *formulative* research studies);
2. To portray accurately the characteristics of a particular individual, situation or a group (studies with this object in view are known as *descriptive* research studies);
3. To determine the frequency with which something occurs or with which it is associated with something else (studies with this object in view are known as *diagnostic* research studies);
4. To test a hypothesis of a causal relationship between variables (such studies are known as *Hypothesis-testing* research studies).

6. Characteristics of Research

- Research is a systematic and critical investigation
- It is not a mere complication, but a purposive investigation
- It adopts scientific method
- It is a objective and logical, applying possible tests to validate the measuring tools and the conclusions reached
- It is based on observable (or) complex evidence
- It is directed towards finding answers to pertinent questions and solutions to problems

7. Importance and Scope of Research in Healthcare

Introduction: Research in healthcare refers to the **systematic investigation and scientific study** aimed at generating new knowledge, improving patient care, enhancing public health, and strengthening healthcare systems. It involves clinical research, health systems research, epidemiological studies, operational research, and health analytics. Healthcare is a complex field influenced by biological, social, economic, and technological factors, and research provides evidence-based solutions for these challenges.

7.1 Importance of Research in Healthcare

Healthcare research is critical for improving the quality, efficiency, and effectiveness of health services. Its major importance includes:

1. Improves Patient Care and Outcomes

- Research provides **evidence-based guidelines** for diagnosis, treatment, and patient management.
- Helps healthcare professionals choose the **best treatment protocols**.
- Reduces errors and enhances patient safety.
- Leads to discovery of better drugs, procedures, and technologies.

Example: Research on infection control reduced hospital-acquired infections globally.

2. Development of New Treatments and Technologies

- Clinical trials lead to the development of **new medicines, vaccines, diagnostic tools, medical devices, and surgical techniques**.
- Biomedical research helps understand diseases at the genetic, molecular, and biochemical levels.

Example: Rapid development of COVID-19 vaccines was possible due to ongoing medical research.

3. Evidence-Based Decision Making in Healthcare Management

- Research provides data to support **administrative and policy decisions**.
- Helps hospital administrators decide

- Resource allocation
- Staffing models
- Cost containment
- Quality improvement measures

Example: Research showing reduced waiting times with online appointment systems led to digital health reforms.

4. Enhances Public Health and Preventive Care

- Epidemiological research helps identify **causes, risk factors, and patterns** of diseases.
- Supports design of public health interventions like:
  - Immunization programs
  - Screening initiatives
  - Health education campaigns
  - Outbreak response

Example: Research on tobacco hazards shaped anti-smoking legislation.

5. Improves Healthcare Quality and Safety

- Research enables:
  - Continuous quality improvement (CQI)
  - Patient feedback systems
  - Clinical audits
  - Standard treatment protocols
  - Accreditation requirements

Example: WHO's Safe Surgery Checklist is a result of global research improving surgical outcomes.

6. Cost Reduction and Efficient Use of Resources

- Health economics research helps identify **cost-effective treatments**.
- Avoids unnecessary tests and hospitalizations.
- Assists in optimizing supply chain and reducing operational costs.

Example: Research on telemedicine showed reduced hospital visits and cost savings.

7. Supports Policy Making and Health Planning

Governments and institutions rely on research to formulate policies on:

- Disease control
- Health insurance
- National health programs
- Hospital management
- Healthcare financing

Example: National Family Health Survey (NFHS) data guides India's health policy decisions.

8. Helps Understand Patient Behavior and Community Needs

- Research identifies:
  - Patient satisfaction
  - Barriers to healthcare access
  - Cultural beliefs affecting health
  - Community health needs

This helps hospitals plan patient-centered services.

9. Improves Training and Education of Healthcare Professionals

- Medical and paramedical students use research to:
  - Update knowledge
  - Learn modern techniques
  - Understand global best practices

Example: Evidence-based nursing practice improves clinical competencies.

10. Promotes Innovation in Healthcare Systems

- Research encourages hospitals to adopt:
  - Digital health systems
  - AI & analytics
  - Process automation
  - Smart hospital technologies

It transforms healthcare delivery through innovation.

8. Scope of Research in Healthcare

Healthcare research is broad and multidisciplinary. Its scope includes various types, fields, and levels of research.

1. Clinical Research

Focuses on improving patient care through studies on:

- New drugs
- Medical devices
- Treatment protocols
- Surgical methods
- Diagnostic techniques

Includes Clinical Trials (Phase I–IV).

2. Epidemiological Research

Studies patterns and determinants of disease in populations:

- Incidence & prevalence
- Risk factors
- Disease outbreaks
- Surveillance systems

Essential for public health planning.

3. Health Systems & Administrative Research

Covers research on:

- Hospital management
- Quality assurance
- Patient satisfaction
- Human resource planning
- Hospital information systems
- Supply chain management

Used for improving efficiency and service delivery.

4. Community Health & Preventive Medicine Research

Deals with:

- Community-based health problems
- Environmental and occupational health
- Health promotion and education
- Maternal and child health

Supports national health programs.

5. Health Economics Research

Studies costs, effectiveness, and efficiency:

- Cost–benefit analysis
- Cost–effectiveness analysis
- Resource allocation
- Healthcare financing

Important for policymakers.

6. Behavioral & Social Science Research

Examines human behavior related to health:

- Health-seeking behavior
- Patient compliance
- Lifestyle and habits
- Mental health

Helps design patient-centered interventions.

7. Operational Research in Hospitals

Focuses on improving day-to-day hospital operations:

- Reducing waiting times
- Optimizing bed occupancy
- Improving emergency services
- Enhancing workflow efficiency

8. Biomedical Research

Studies biological and physiological mechanisms of diseases:

- Genetics
- Microbiology
- Immunology
- Biotechnology

Supports discovery of cures and treatments.

9. Data Analytics & Health Informatics Research

Rapidly growing with digital transformation:

- Big data in healthcare
- Predictive analytics
- AI in diagnostics
- Electronic health records (EHR) analysis

Improves population health predictions and clinical decisions.

10. Global Health Research

Addresses international health challenges:

- Pandemic control
- Health disparities
- International policies
- Global disease surveillance

9. Conclusion

Research is the backbone of modern healthcare. It ensures that healthcare services are **scientific, efficient, patient-centered, and cost-effective**. With growing challenges like ageing populations, chronic diseases, and pandemics, healthcare research continues to expand in both importance and scope.

Self Assessment Questions

1. **Define Research? Explain its importance in Healthcare Research**
2. **What is research explain its objectives in detail**

****

Lesson -2

TYPES OF RESEARCH & THE RESEARCH PROCESS

Structure of the Lesson

1. Introduction
2. Types of research (basic/applied, qualitative/quantitative, exploratory/descriptive/diagnostic, etc.)
3. Steps in the research process

Introduction: Research in common parlance refers to a search for knowledge. Once can also define research as a scientific and systematic search for pertinent information on a specific topic. In fact, research is an art of scientific investigation. The Advanced Learner's Dictionary of Current English lays down the meaning of research as "a careful investigation or inquiry specially through search for new facts in any branch of knowledge." Redman and Mory define research as a "systematized effort to gain new knowledge." Some people consider research as a movement, a movement from the known to the unknown. It is actually a voyage of discovery. We all possess the vital instinct of inquisitiveness for, when the unknown confronts us, we wonder and our inquisitiveness makes us probe and attain full and fuller understanding of the unknown. This inquisitiveness is the mother of all knowledge and the method, which man employs for obtaining the knowledge of whatever the unknown, can be termed as research

TYPES OF RESERACH

1. ***Descriptive vs. Analytical Research.***
2. ***Applied vs. Fundamental Research.***
3. ***Quantitative vs Qualitative Research.***
4. ***Concept vs Empirical Research.***
5. ***Explorative Research.***
6. ***Historical Research.***

7. ***Decision oriented Research. Also called as conclusive research***
8. ***Longitudinal Research***

Descriptive research: Descriptive research includes surveys and fact finding enquiries of different kinds. **The major purpose of descriptive research is description of the state of affairs as it exists at present. In social science and business research we quite often use the term *Ex post facto research* for Descriptive research studies** The main characteristic of this method is that the researcher has no control over the variables, he can only report what has happened or what is happening.

Business organizations preserving databases of their employees, customers and suppliers keep significant data to conduct descriptive studies utilizing internal information. 'This sort of study is generally admired in business research because of its versatility across disciplines.' across organizations, **descriptive studies create a vast appeal to the managers for planning, monitoring and evaluating.**

Ex: frequency of shopping, preferences of people or similar data

Analytical Research: on the other hand, the researcher has to use facts, or information already available and analyse these two make a critical evaluation of the material.

Ex: stock markets, financial status of a company, using existing healthcare data one can conduct analytical research etc.

Applied *vs* Fundamental Research: Research can either be applied (action) or fundamental research. **Applied (action) Research aims at finding a solution for an immediate problem facing a society or an industrial/business organisation**. Whereas fundamental research is mainly concerned with generalisation and with the formulation of a theory. Gathering knowledge for knowledge's sake is termed pure (or) basic research.

Example: Pure mathematics is an example.

The central aim of applied research is to discover a solution for some **pressing practical problem**, where as basic research is directed towards finding information that

has broad base application and thus adds to the already existing organized body of scientific knowledge.

Quantitative *vs* Qualitative : Quantitative research is based on the measurement of quantity or amount. It is applicable to phenomenon that can be expressed in terms of quantity.

Qualitative research on the other hand, is concerned with qualitative phenomenon, i.e. phenomena relating to or involving quality or kind. For instance, when we are interested in investigating the reasons for human behaviour (i.e., why people think or do certain things,), we quite often talk of motivational research. An important type of qualitative research. This type of research aims at discovering the underlying motives and desires, using in depth interviews for the purpose.

Other techniques of such research are word association tests, sentence completion tests, story completion tests and similar other projective techniques.

Qualitative research is especially important in the behavioural sciences where the aim is to discover the underlying motives of human behaviour. Through such research we can analyse the various factors which motivate people to behave in a particular manner or which make people like or dislike a particular thing. It may be stated, however, that to apply qualitative research in practice is relatively a difficult job and therefore, while doing such research, one should seek guidance from experimental psychologists.

Conceptual *vs* Empirical: **Conceptual research is that related to some abstract idea(s) or theory. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones.** On the other hand, empirical research relies on experience or observation alone, often without due regard for system and theory. ***It is data-based research, coming up with conclusions which are capable of being verified by observation or experiment.*** We can also call it as experimental type of research. In such a research it is necessary to get at facts first-hand, at their source and actively to go about doing certain things to stimulate the production of desired information. In such a research the researcher must first provide himself with a working hypothesis, then works to get enough facts (data) to prove or disprove his hypothesis.

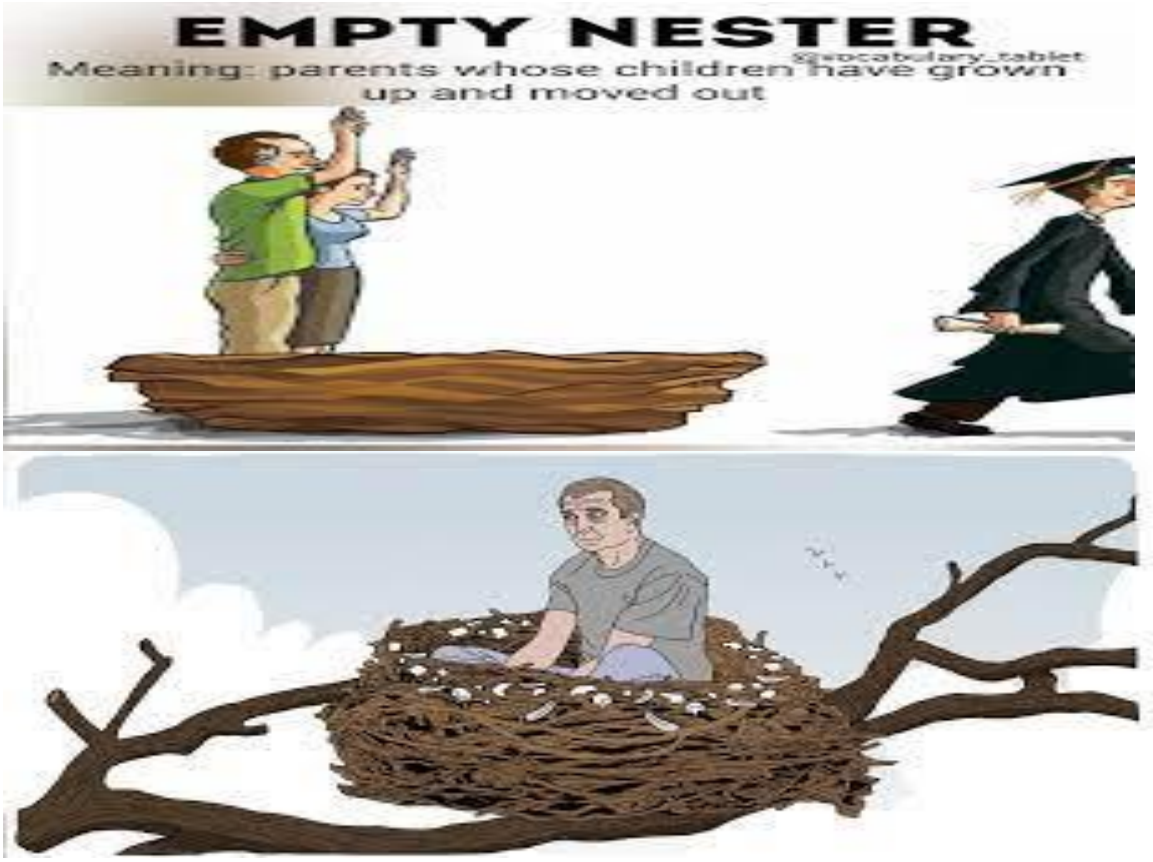
Exploration-Exploratory Research Design **:** When a research is conducted on a new topic about which *nothing is known* or *little is known* , and existing theories do not explain the phenomenon, the purpose of the research is said to be *exploration.*

The researcher conducts the study to get familiarity with the topic and he/she is likely to undertake a full fledged study later.

- **The purpose of the research may be to explore if psycho-social well-being of empty nesters (**a parent whose children have grown up and left home) **affects their longevity leading to early death Or, if a researcher wishes to know about social implications of *live-in-relationship* it has to be exploratory study.**

The purpose of the research may be to explore if *live-in-relationships* affect the social institution of marriage. Or, a study of association between emotional intelligence and *marital satisfaction* would have to be exploratory study.

Exploratory research study often starts with observation, discussions and interviews. Based on the information , interview schedule is prepared and used as a tool of data collection in exploratory research studies. **Exploratory research studies do not have hypotheses.**

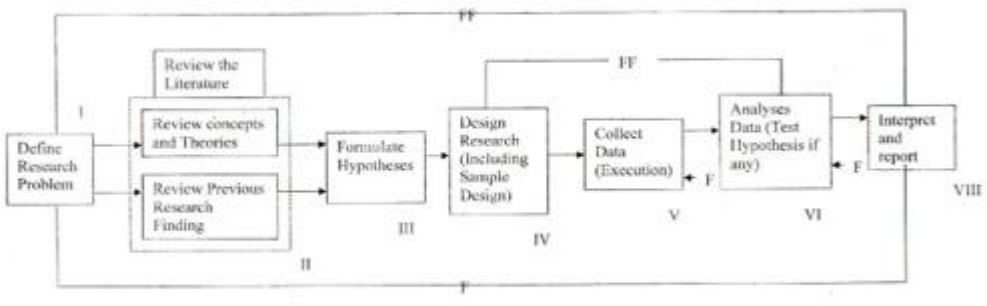


Historical Research: In historical research, past events are studied systematically and chronologically so that a clear description can be given as to the happening of events in past. In historical research scattered recorded facts are collected. Historical research studies also don't have hypothesis.

Decision oriented Research: it is always for the need of a decision maker and the researcher in this case is not free to embark upon research according to his own inclination.

Longitudinal Research: A longitudinal study is a type of correlational research study that **involves looking at variables over an extended period of time**. This research can take place over a period of weeks, months, or even years. In some cases, longitudinal studies can last several decades.

RESEARCH PROCESS IN FLOW CHART



Where F – Feed back (Helps in controlling the sub-system to which it is transmitted)
FF – Feed forward (Serves the vital function of providing criteria for evaluation)

Lesson -3

DEFINING A RESEARCH PROBLEM

Structure of the Lesson

1. Meaning of research problem
2. Process of identifying and defining a research problem
3. Criteria for a good research problem

Research Problem

In a research process, the first and foremost step happens to be that of selecting and properly defining a research problem. Selection of research problem is the first step in every research. No research can undergo by the researcher, unless this first step is properly and scientifically accomplished by the researcher. Usually, it is observed that academic research work undertaken by the students or research scholars are selected on certain parameters i.e., novelty of topic, suitability of topic, convenience in data collection or on option of supervisor. Though, these considerations cannot fulfill the real objective of academic research. On other hand, if any research project is supported by the funding agencies, sometimes research problems are already designed by such agencies and researchers are expected to research on such problems. In all such cases researcher's involvement and his relation with research are not established up to expected level. Concept of problem is basically a psychological aspect which can be caused by either physical facts or mental facts. No research can be performed by the researcher without his psychological involvement, his realization and sensitization towards the research problem. In many instances of academic research it has been observed that, without serious consideration of research problem researcher has undergone complete research

What is research problem?

A research problem, in general, refers to some difficulty which a researcher experiences in the context of either a theoretical or practical situation and wants to obtain a solution for the same. Usually we say that a research problem does exist if the following conditions are met with:

1. There must be an individual or a group or an organization let us call it "*I*"

2. There must be at least two courses of action, say C1 and C2 to be pursued

3. There must be at least two possible outcomes, say O1 and O2

4. The courses of action available must provides some chance of obtaining the objective

Defining the Research Problem

A research problem needs to be identified and clearly defined. Bryman, Alan defined the research problem as a statement, a field of concern, a condition to be made better, a difficulty to be removed, or a disturbing question in scholarly literature, in theory or in practice that points to the need for purposeful comprehension and deliberate investigation. 'The research problem defines the goal of the research in clear terms. Without a clear cut idea of the goal to be achieved, research task would become a worthless exercise. A research like any other human activity is goal directed. 'It is not always possible for a researcher to design his problem in a simple, clear and complete manner. He may often have only a rather general, diffuse, even confused notion of problem. It may even taken an investigator's years of exploration, thought and research before he is in a position explicitly say what queries he has been seeking answers to' The statement of the problem is one of the most important parts of research. The difficulty or impossibility in the satisfactory statement of a research problem does not justify that the researcher ignores the desirability and necessity of stating a research problem. The fundamental principle is that if one desires to find a solution for a problem, one must well understand the problem. It can be stated that a larger portion of the solution to a problem lies in knowing the problem it that the researcher is attempting to do. And the rest lies in knowing what a problem is. A good problem statement should possess certain characteristics. One, the problem is stated in question form. What are the effects on job performance of different types of incentives? Second, the problem states a relation between variables, in this case between incentives and job performance. Thus, a problem is stated in an interrogative form or statement that asks" what relation exists between two or more variables? The answer is what is being sought in the research. The problem statement relates incentives to job performance. There are three features of good problems and problem statements. One, the problem should show a relation between two or more variables. Second, the problem should be clearly stated and evidently express in question form. Questions have the quality of posing problems directly. Third, the problem and the problem statement should be such as may be empirically tested. A problem that holds implications to be tested its stated relations is a scientific problem.

How a Research Problem Originates:

1. Contemporary Interest
2. Own Interest
3. Gaps in the Field
4. Other sources

5. **Criteria for a good research Problem:**

1. Clear and unambiguous
2. Logical
3. Empirical
4. Verifiable
5. Interesting
6. Management

6. **Major problems of Corporate Hospitals**

Time Motion study, patient satisfaction, operations management, Quality, Digitalization, Lack of infrastructure, R&D, High Attrition rate, employee dissatisfaction, skilled administration, scarcity of reputed doctors and nurses, supportive services

Conclusion: The importance of the formulation of research problem cannot be overemphasized. It is not only critical to identify the decision to be made but also to formulate it in such a form that it can lend itself to scientific enquiry. This is a well-integrated, linked and stepwise process beginning by clarifying doubts and getting the research perspective on the basis of discussions with experts. These could be both industry and subject experts. Next the researcher moves to getting the various perspectives of other researcher or theorists on the topic is to conduct a comprehensive examination of the previous studies. In case the research and is intended to be carried out in a particular industry or organization, it is critical to obtain a detailed dossier on the history and current practices of the organization. Some researchers also undertake a brief loosely-structured survey with respondents from the population to be studied to further fine-tune the statement of intent. Now the researcher arrives at a clearly stated research problem that can lend

itself to scientific enquiry. There are some essential elements of a typical research problem including specifying the unit of analysis, clear definition and categorization of the concept or constructs to be studied. At these stages, the researcher should be able to specify what is the causal or independent variable and which the effect on dependent variable understudy is. Also, it is best to acknowledge the effect or presence of any external variables which might have been a contingent effect on the cause and effect of relationship that is to e studied, further classified as moderator, intervening and extraneous variables. It is advisable to the research to construct a model or theoretical framework based on the stepwise conceptualization that the researcher carried out in the process of problem formulation. This is a recommended but not necessarily an essential steps as some studies might be of a nature that the intent is to conduct the study and then arrive at a theory or a model. The research questions having been designed, they need to be broken down and as tasks or objectives to be accomplished for seeking the answers to the research questions. Now, the researcher draws the lay out for the proposed research. The research questions can be broken down into examining questions. If the proposal is for a causal study, the objectives can be formulated into a hypothesis.

Self Assessment Questions

1. **Define research problem?**
2. **Explain criteria for good research problem**

****

Lesson- 4

INTRODUCTION TO SAMPLING

Structure of the Lesson

1. Introduction
2. Concept of sampling
3. Census vs. sample
4. Probability and non-probability sampling techniques

INTRODUCTION

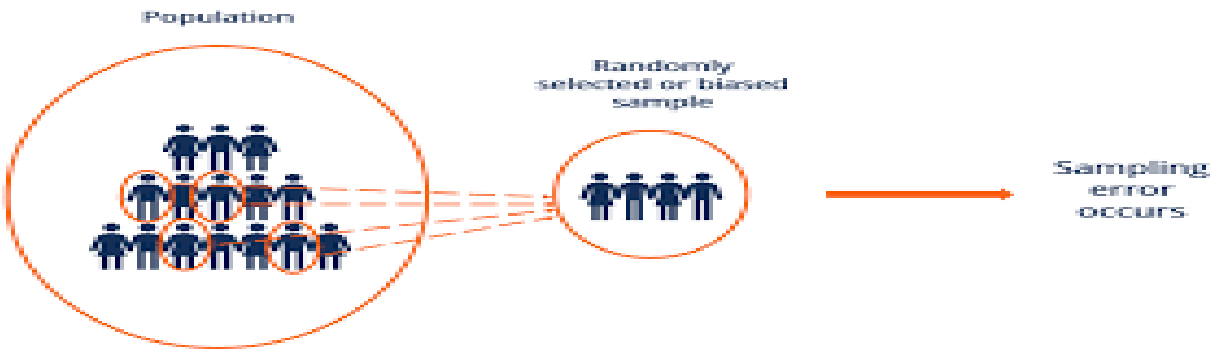
Sampling is an integral part of everybody's life. It is in fact a long established practice. Day in and day out we use sampling methods in carrying out various tasks. For example: a housewife picks up only a handful of grain to test the quality of the material she is going to buy. Similarly companies ask for the samples before they can approve of the quotations for the supply of material. The underlying assumption in these cases is that we expect the whole material is similar or at least it is closely representative of the whole.

sample is a smaller representation of a large whole

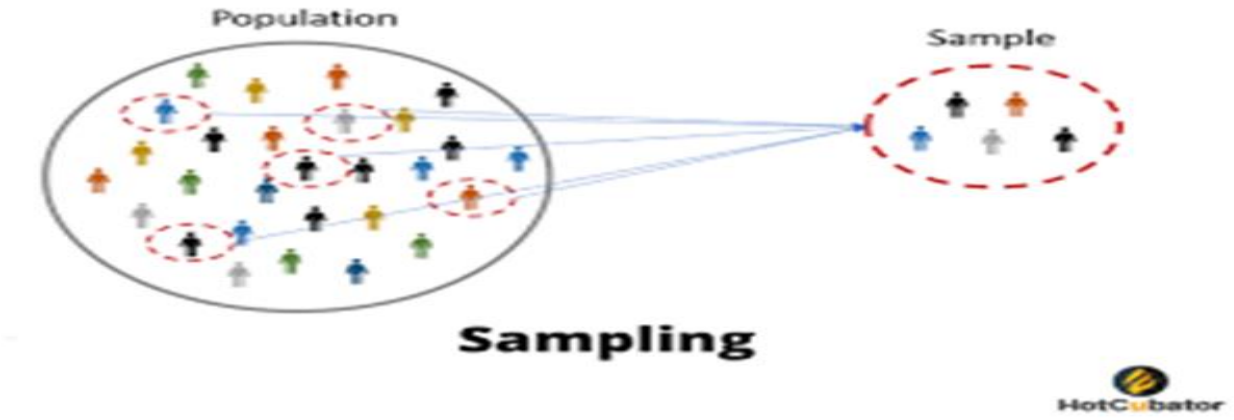
Before discussing about sampling let us need to know what is Universe and Population.

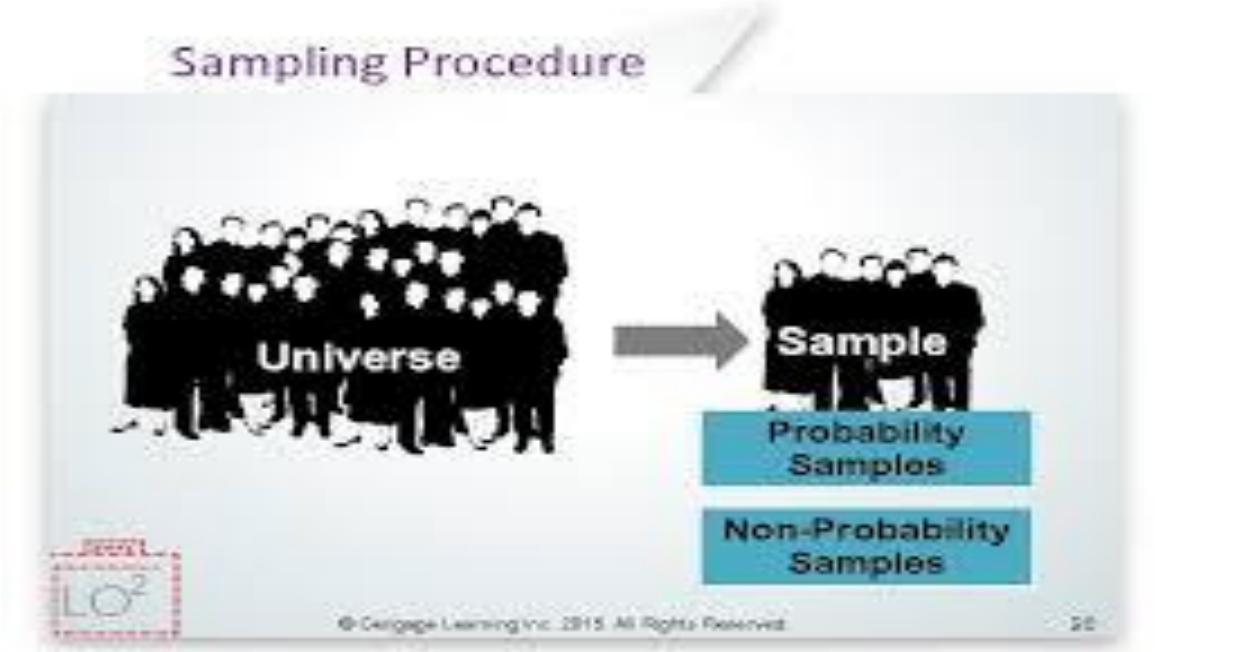
Universe: Universe represents the entire group of units of the study. Thus, the universe could consist of all the persons in the country, or those in a particular geographical location, or a special ethnic or economic group, depending on the purpose and coverage of the study.

A universe could also consist units such as farms, houses or business establishments.




Population
Sample
Associates with glasses for research

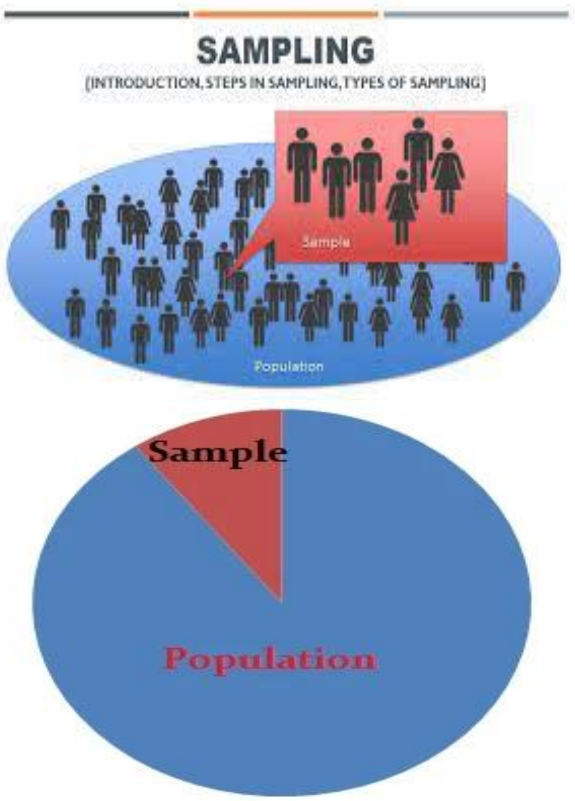

Population
Sample
Sampling
HotCubator


Sampling Procedure
Universe
Sample
Probability Samples
Non-Probability Samples
LO2
© Cengage Learning Inc. 2015. All Rights Reserved.

Sampling is the process of selecting units (e.g., people, organizations) from a universe / population so that by studying the sample we may fairly generalise our results back to the universe / population from which they were chosen.

A sample is a part of the universe / population which is studied in order to make inferences about the whole population

Sampling may be defined as the selection of some part of an aggregate. In other words, it is the process of obtaining information about an entire population by examining only part of it. Or sample is a smaller representation of a large whole



Their selection process (or) technique is called <u>sample design,</u> and the survey conducted on the basis of sample is described as <u>sample survey</u>. Sample should be <u>truly representative</u> of population characteristics without any bias so that it may results in valid and reliable conclusions.

Need of sampling : There are large economic benefits of selecting a sample rather than conducting a census. The cost of taking a census survey may go up to lakhs of rupees

interviewing all 5000 employees of an organization. It can be identified what is to be known by choosing a sample of few hundred. The quality of a study conducted with a sample is usually more than with a population. Research findings provide competent evidence of this opinion. In one study, more than 90 percent of the total survey error was from non-sampling sources, and only 10 percent was error from random sampling. The results of a study from are quicker than from a study of census. The speed of execution reduces the time between recognizing of a requirement for data and the data within reach. When the population of the study is small and the variability high, as well as the components completely different from each other, the census study is more appropriate. If the universe is small and the variability is high, the selected sample may not be representative of the universe. The results drawn from the sample are not accurate as estimates of the population values. When the sample is taken appropriately, however, some sample elements underestimate the parameters and some other overestimates them. Variations in these values act in opposition to each other this counteraction arises in a sample values that is usually near to the population value. For these offsetting effects to happen, however, there must be adequate members in the sample, and they must be drawn in a way that neither underestimation nor overestimation occurs.

Sampling is used in practice for a variety of reasons such as

1. ***Sampling can be save time and money***
2. ***Sampling may enable more accurate measurements for sample study is generally conducted by trained and experienced investigators***
3. ***Sampling remains the only way when population contains infinitely (many members)***
4. ***Sampling remains the only choice when a test involves the destruction of the item under the study.***
5. ***Sampling usually enables to estimate the sampling errors and thus assists in obtaining information concerning some characteristic of the population.***

ABOVE ALL: The point to be kept in mind is, if we can get almost same results by studying a carefully selected small group of people why should we study the large group at all.

Requirements of a good sample : The sample should possess the following essentials

1. Representativeness : The sample should be a representative of the population in some sense.
2. Homogeneity : There should be no basic difference between the two samples drawn from the same population
3. Adequacy : The number of sampling units in the sample should be adequate.
4. Independence : The sampling units composing the sample should be independent.
5. Similar Regulating Conditions : The regulating conditions should be similar for every sampling unit

Sampling Design: Key Issues

A Sampling Design is a definite plan for obtaining a sample from the sampling frame. It refers to the technique or the procedure the researcher would adopt in selecting some sampling units from which inferences about the population is drawn. Sampling design is determined before any data are collected.

- **Who will be the respondents of study?**
- **How the respondents will be selected?**
- **How many respondents will be studied?**

Attributes of a good sample The decisive test of a sample design is the extent to which a sample is representative of the population. The sample must be valid in terms of accuracy and precision.

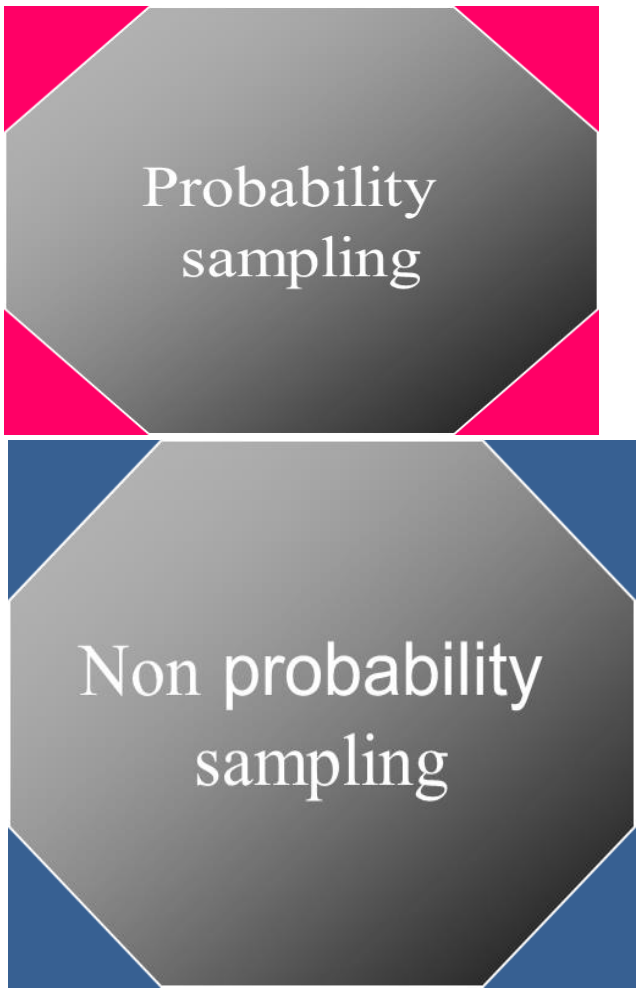
Accuracy Accuracy is the extent to which bias is nonexistent from the sample. In an exact sample the under estimator and the over estimators are adjusted among the members of the sample. From an accurate sample, the systematic variance is absent. The variation in measures some known or unknown effects create resulting in the inclination of scores in one direction more than another. In a study the researcher came to know that in selecting a particular route for his newspaper readership sample, the time of the day, day of the work, and season of the year of the survey strikingly brought down the accuracy and validity of his sample. A classic example of a sample with systematic variance was the presidential election poll in 1936, in which greater than 2 million persons took part. The poll said Alfred Landon would win against Franklin Roosevelt for the presidency of the US. Even the large size of this sample did not counteract its systematic bias. Later evidence exhibited that the poll selected its sample from the middle and upper classes, while Roosevelt's appeal was largely among the much larger working class.

Precision of Estimate The precision of estimate is another criterion of a good sample design. No sample will entirely represent its universe in all respects. The numerical descriptors that give an account of the sample may be expected to differ from those that describe population due to of random fluctuations in the process of sampling. This sampling error indicates the effects of chance in selecting the sample unit remains after all understood sources of systematic variance have been taken into consideration. Sampling error is composed of random fluctuations only, although some systematic variance not known may be contained in when too many or too few sample members possess a particular feature. The standard error of estimate, a measure of standard deviation, gauges the precision; the low value of the standard error of estimate indicates the high degree of precision of estimate of the sample. An appropriate sample design creates a small standard error of estimate. However, all kinds of sample design do not provide estimates of precision, and samples of the same size can generate different amounts of error variance.

Parameters of interest Universe parameters are hard but easily breakable descriptors of interest variables within the universe. Sample statistics are descriptors of the relevant variable computed from sample information. Sample statistics estimate population parameters and constitutes the premise of our inferences regarding the population as a result that they are the simplest estimates of the population. When the variables of interest of the study are measured on interval scale or quantitative relation scale, the sample mean is employed to estimate the population average and also the sample variance is employed to estimate the population variance. If a study focuses on nominally scaled data, the population proportion parameter would be of interest. Proportion measures are essential for nominal data and are extensively applied to measures also. The percentage is the most frequent concentration measures; the variance equivalent is the pq ratio.

SAMPLING METHODS

- **PROBABILITY SAMPLING**
- **NON-PROBABILITY SAMPLING**



Probability

- Simple random sampling
- Systematic random sampling
- Stratified random sampling
- Cluster sampling
- Stage sampling
- Multi-phase sampling

Non probability

- Convenience sampling
- Judgment sampling
- Quota sampling
- Snowball sampling
- Dimensional sampling
- Volunteer sampling
- Theoretical sampling

Probability Sampling :

- Probability sampling methods provide one excellent way of selecting samples that will be quite representative.
- Each element of the population has a known and equal probability of selection in the sample.
- It relies on a random selection of elements.
- It is used in case of 'Finite Population'

Thus Probability sampling is based on the concept of Random selection (or) chance sampling. In the sampling every item of the universe has an equal chance of inclusion in the sample.

Non Probability Sampling:

It is based on the Non Random, it is also called deliberate, Purposive and Judgement sampling. It is used in case **of infinite population**. In non probability sample Random selection of elements is not necessary it relies on personal judgement of the researcher.

- Researchers most often conduct study where the population is not known.
- In such cases it is not possible to specify, for each element of the population, the relative likelihood that it will be included in the sample.
- In such situations we select sample using non-probability sampling.
- In non-probability sample random selection of elements is not necessary.
- It relies on personal judgment of the researcher.
- The researcher can arbitrarily or consciously decide what elements to include in the sample.
- Suppose we wish to study the socio-economic background of People Living with HIV.
- There is no list of People Living with HIV, nor it is possible to create such a list.
- In such situation non-probability sampling procedures are called for.

Simple Random Sampling: This method is the most common and familiar type of probability sample. In this method of sampling that relies on random (or) chance selection. Each number of the population has an equal probability of being included in the sample. Therefore simple random sampling is a method of selecting 'n' units out of the population size 'N' by giving equal opportunity to all units. It is based on the following methods.

A) Lottery Method, B) Tippers Random Numbers

A) ***Lottery Method***: in this method, a ticket, chit or token may be associated with each unit of population then all the tickets are placed in a drum or container, in which a thorough mixing or reshuffling is possible before each draw. Draws are continued until a sample of the required size is obtained.

However the procedure of numbering units on tickets, chits or tokens and selecting the sample becomes a cumbersome process. If the sample size is large, it may be rather difficult to achieve. Human bias and prejudice may also creep in the method

B) Use of random numbers: (various statisticians like tipper, yates, fisher) have prepared tables of random numbers which can be used for selecting random (numbers) sample.

tippet gave 10400 four figure numbers. He selected 41600 digits from the census reports and combined them into fours to give his random numbers which may be used to obtained a random sample.

Illustration: first of all we produce the first thirty sets of tippet's numbers

2952 6641 3992 9792 7979 5911

3170 5624 4167 9525 1545 1396
2370 7483 7203 5356 1300 2693

3408 2769 3563 6107 6913 7691

0560 5246 1112 9025 6008 8126

Suppose we are interested in taking a sample of 10 units from a population of 5000 units bearing a numbers from 3001 to 8000. we shall select 10 such figures, figures from the above random numbers which are not less than 3001 and not greater than 8,000. if we randomly decide to read the table numbers from left to right, starting from the first row itself, we obtain the following numbers. 6641, 3992, 7979, 5911, 3170, 5624, 4167, 7203, 5356 and 7483. the units bearing the above serial numbers would then constitute our required random sample.

Advantages of simple random sampling

- It is very simple technique and serves well to introduce the basic idea of sampling;
- This method is based on the theory of probability and hence it is possible to calculate the sample error;
- Every sample unit has equal opportunity of being considered for selection
- Another advantage in this simple random sampling technique is that accurate, mathematical tests may be applied to judge the randomness of a sampling method.
- It is a fair method of sampling, and if applied appropriately, it helps to reduce any bias involved compared to any other sampling method involved.

- Since it involves a large sample frame, it is usually easy to pick a smaller sample size from the existing larger population.
- The person conducting the research doesn't need to have prior knowledge of the data he/ she is collecting. One can ask a question to gather the researcher need not be a subject expert.
- This sampling method is a fundamental method of collecting the data. You don't need any technical knowledge. You only require essential listening and recording skills.
- Since the population size is vast in this type of sampling method, there is no restriction on the sample size that the researcher needs to create. From a larger population, you can get a small sample quite quickly.
- The data collected through this sampling method is well informed; more the samples better is the quality of the data.

Demerits:

- It is not always possible to have complete universe. It may be infinite even we come across finite universe;
- This sampling method is that it is often statistically inefficient
- If universe is heterogeneous groups of different size, simple random sampling technique is unsuitable

Systematic Random Sampling technique

Systematic sampling is often used instead of random sampling. It relies on arranging the study population according to some ordered scheme and elements are selected at regular intervals (known as the sampling interval) through that ordered list.

Systematic sampling involves a random start and then proceeds with the selection of every nth element from then onwards. It is important that the starting point is not automatically the first in the list, but is instead randomly chosen from within the first to the nth element in the list. As long as the starting point is randomly selected, systematic sampling can be classified as a type of probability sampling. It is easy to implement and the induced stratification can make it efficient, if the variable by which the list is ordered is correlated with the variable of interest. *A simple example would be to select every 10th name from the electoral list. As long as the list does not contain any hidden order, this sampling method is as good as the random sampling method*. Its only advantage over the random sampling technique is simplicity. Systematic sampling is frequently used to select a specified number of records from a computer file.

For example, there are 100,000 elements in the population and a sample of 1,000 is desired. In this case the sampling interval, i, is A random number between 1 and 100 is selected. If,

for example, this number is 23, the sample consists of elements 23, 123, 223, 323, 423, 523, and so on

STRATIFIED RANDOM SAMPLING

Stratified random Sampling: when universe consists of units heterogeneous in character stratified random sampling is generally applied in order to obtain representative sample. In this method universe is divided into sub groups and sample is taken from each sub group. (ex: male, female, manager, doctor, Nurses etc)

If 'N' is taken as the universe, N1,N2, N3 ---------- NK can be its sub group, such that we have following equation.

N = N1 +N2+N3 --------NK

'n' is the sample size then

n = n1+n2+n3+ --------nk

The following three questions are highly relevant in the context of stratified sampling.

a) **How to form strata**

b) **How should items be selected from each stratum**

c) **How to allocate the sample size to each stratum**

Stratified random sampling is a method of probability sampling in which the population is divided into different subgroups and samples are selected from each.

1) Divide the target population into homogeneous subgroups or strata

2) Draw random samples from each stratum

3) Combine the samples from each stratum into a single sample of the target population.

Stratified Random Sampling includes two methods for implementation.

- **Proportionate allocation**
- **Disproportionate allocation**

Proportionate stratified sampling: sampling method in which elements are selected from strata in exact proportion to their representation in the population.

Example: in a study on trade unionism in a unit is 500 worker consisting of females and males.

Then proportionate = 100/500*100=20 therefore female strate is 20 per cent and 400/500*100=80 male strata is 80 per cent.

TABLE 1

DISTRIBUTION OF EMPLOYEES

Employees	Population	Proportion/Percentage of each category of employees
Managers	400	.08/8
Supervisors	600	.12/12
Workers	4000	.80/80
Total	5000	1.00/100

Table 2 Sampling Distribution of Employees (Proportionate)

Employees	Sample Break-up	Proportion/Percentage of each category of employees
Managers	40	.08/8
Supervisors	60	.12/12
Workers	400	.80/80
Total	500	1.00/100

Disproportionate Stratified sampling: Sampling in which elements selected from strata in different proportions from those that appear in the population.

Table 2 SAMPLING DISTRIBUTION OF EMPLOYEES (DIPROPORTIONATE)

Employees	Population Break-up	Sample Break-up	Proportion of employees
Managers	400	200	.33
Supervisors	600	200	.33
Workers	4000	200	.33
Total	5000	600	100

Systematic Sampling

Systematic Sampling: in this method only the first unit is selected with the help of random numbers and the rest get selected automatically according to some predetermined pattern.

suppose N units of the population are numbered from 1 to N in some order.

Let N = nk

where 'n' is the sample size

k is an integer/intervel

Random number less than or equal to 'k' be selected and kth unit thereafter.

For example: we want to select 500 among a list of 10,000 shareholders of a company. N/n=k 10000/500=20.

Then the first number should be 1-20, suppose we selected 15th shareholder, thereafter every 20th shareholder get selected in the sample.

Cluster Sampling : Cluster sampling is a procedure of selection in which the elements for the sample are chosen from the population in groups (or) clusters rather than single. Example: schools, college, factories political subdivisions. (or) if the total area of interest happens to be a big one, a convenient way in which a sample can be taken is to divide the area into a number of smaller non-overlapping areas and then to randomly select a number of these smaller area (usually called clusters). Thus in cluster sampling the total population is divided into a number of relatively small subdivisions which are themselves clusters of still smaller units and then same of these clusters are randomly selected for inclusion in the overall sample.

Multi stage sampling: Sampling in which elements are selected in two or more stages, with the first stage being the random selection of naturally occurring clusters and the last stage being the random selection of elements within clusters.

Multi stage sampling.

Topic : Older Persons in Rural A.P.

No. of Districts : 13

No. of Mandals : 676

No. of Villages : 16158

Universe : About 40 lacs (65+)

Sample : 384

Multi stage Sampling : Procedure

Stage I : Selection of districts (Random)

Stage II : Selection of mandals from selected districts

Stage III : Selection of villages from selected mandals

Stage IV : Selection of Older Persons from selected villages. (Random)

Area Sampling: area sampling is a special form of cluster sampling in which the sample items are clustered on a geographical area basis. In this kind of sampling maps rather than lists or registers serve as basis with in the area, the researcher may select all the members of the area (or) part of the area may be selected. In the area sampling the boundaries of the area must be well defined.

NON-PROBABILITY SAMPLING

It is based on the Non Random, it is also called deliberate, Purposive and Judgement sampling. It is used in case of infinite population. In non probability sample Random selection of elements is not necessary it relies on personal judgement of the researcher.

- **Researchers most often conduct study where the population is not known.**
- **In such cases it is not possible to specify , for each element of the population, the relative likelihood that it will be included in the sample.**
- **In such situations we select sample using non-probability sampling.**
- **In non-probability sample random selection of elements is not necessary.**
- **It relies on personal judgment of the researcher.**

- **The researcher can arbitrarily or consciously decide what elements to include in the sample.**
- **Suppose we wish to study the socio-economic background of People Living with HIV.**
- **There is no list of People Living with HIV, nor it is possible to create such a list.**
- **In such situation non-probability sampling procedures are called for.**
- ***Convenience sampling (or) accidental***
- ***Purposive or judgemental sampling***
- ***Quota sampling***
- ***Snow ball sampling***

Convenience sampling is a variant of non-probability sampling where subjects are chosen due to their **convenience access and nearness to the researcher**. The subjects are chosen just because they are easiest to be approached for the study and the researcher did not think carefully selecting subjects that have the characteristics of the whole population. In all forms of research, it would be appropriate to test the whole population, but in majority of cases, the population is too large to make it impossible for inclusion of every person. This is why most researchers depend on sampling procedures like convenience sampling. Since it is quick, economical, easy, ready availability of the subjects, many researchers hold convenience sampling above other methods.

Purposive or judgemental sampling

- **In purposive or judgmental sampling researchers use their judgment and prior knowledge to select people for the sample who would best serve the purposes of the study.** Purposive sampling can be helpful in situation where the investigator reaches a targeted sample quickly and where proportion of the sample is not the major concern. In this type of sampling, the researcher most probably gets the opinions of his target population, but is also likely to give more weight to subgroups in population that are more readily accessible

For example, if one were conducting a study of the effects of cash transfer to people living with HIV , it might be desirable to choose people for the sample from among those PLWHIV who have been receiving cash (pension).

Quota sampling

- Quota sampling involves dividing a population into various categories and setting quotas on the number of elements to be selected from each category.
- Once the quota is reached , no more elements from that category are put in the sample.

- Quota sampling is like stratified sampling except the selection of elements.
- In quota sampling , the elements are selected depending upon availability of subjects.
- Quotas are normally fixed for demographic variables like , age, sex, castes, religion, nativity, education, etc.

SNOW BALL SAMPLING

- This sampling technique is used when the members of a special population are difficult to locate.
- It thus might be appropriately used to find a sample of drug addicts.
- This sampling technique starts with collecting data from the few members of the target population that one is able to locate, and then asking those individuals to provide information needed to locate other members of that population whom they happen to know.

Summary: The way of selecting for study a fraction of the universe with a view to draw conclusion about the universe or population as a whole is known as sampling. There are large economic benefits of selecting a sample rather than conducting a census. A good sample must have the characteristics of accuracy, precision of estimate and parameters of interest. There are two types of sampling, probability and non-probability sampling. The major types of probably sampling include simple random sampling, stratified sampling and cluster sampling. The simple random sampling denotes that every member of the population has an equal probability of being included in the sample but also renders the selection of every feasible combination of the desired number of cases equally likely In systematic sampling, we randomly enter a stream of elements and sample every k element. It is best when elements are randomly ordered with no cyclic variation. In this sampling, the sample is systematically spread through the population. In stratified random sampling, from each stratum a simple is drawn using a simple random sampling, and such sub-samples are brought together to form the total sample. . In cluster sampling, the sampler first samples from the population certain large groupings or clusters. From these clusters the constituent components are sampled by random procedures Non-probability samples do not possess the virtues of probability sampling and do not make estimate the probability of each element in the population of being selected in the sample. In convenience sampling the subjects are chosen just because they are easiest to be approached for the study and the researcher did not think carefully selecting subjects that have the characteristics of the whole population. The use of judgment and a careful

consideration to attain representative samples by containing presumably typical areas or groups in the sample are the characteristics of purposive sampling. In quota sampling understanding of the strata of the population e.g. sex, race, region and so on, is utilized to choose representative, typical and suitable sample members for predefined research purposes. Snowball sampling begins by identifying individuals who fulfill the criteria for inclusion in the study. Then, they are asked to recommend others whom they may know also meet the criteria.

Self Assessment Questions.

1) **What is sampling why it is important in Research**
2) **Explain probability and non probability sampling techniques**
3) **What is snow ball sampling explain with suitable example**

Lesson- 5

SAMPLING DESIGN & SAMPLING ERRORS

Structure of the Lesson

1. Determining sample size
2. Sampling distribution
3. Sampling error
4. Criteria for selecting a sampling method
5. Sampling procedure

Introduction: Determination of sample size is one of the important decisions to be taken in any sample survey for estimating the population parameters. Too large a sample mean waste of resources and too small a sample will diminish the utility of results. In this paper, the methods for determining the sample size are discussed in cases of estimating the population mean, total and proportion. Determination of sample size for estimating population mean or total is discussed in Section 1. Estimation of population standard deviation is dealt in Section 2. Determination of sample size for estimating population proportion and estimation of population proportion are discussed in Sections 3 and 4 respectively.

DETERMINATION OF SAMPLE SIZE FOR ESTIMATING POPULATION MEAN OR TOTAL

Consider the following hypothetical situation. A Company manufacturing electric bulbs wants to determine the average (mean) life-time (in hours) of its bulbs and requests a statistician to handle the problem. The statistician tells the company that the average life-time of the bulbs can only be estimated on the basis of a sample. He further cautions them that an estimate based on a sample will never (except by rare coincidence) be exactly equal to the population parameter that is being estimated and submits that before organizing the investigation it is necessary to know the degree of precision expected. The company then informs him that they will be satisfied if he can ensure that the difference between the real average and the estimated average is not more than 50 hours.

The statistician again points out that it is not possible to give absolute guarantee that the difference between the real and estimated average will not be more than 50 hours since, while taking a simple random sample, there is always a chance of getting a very unlucky

sample. The company then informs him that they are prepared to take a 5% chance of getting an unlucky sample and will be content to be 95% confident that the difference between the real and estimated averages will not be more than 50 hours.

Now, the statistician is in a position to determine the sample size. Here the population under investigation is the "electric bulbs being manufactured by the company". The variable being measured is the life-time (in hours) of a bulb. The population size 'N' may be infinite or finite. We discuss the determination of sample size 'n' under the following cases.

Case-I: N is infinite

The quantity $\left(1-\frac{n}{N}\right)$ is known as Finite Population Correction (FPC). If N is large relative to n, it can be ignored, i.e., $1-\frac{n}{N}=1$. If μ and σ are the mean and standard deviation of the population then, by central limit theorem, the means of samples of size n will be distributed normally with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ (ignoring the finite population correction). Hence the probability of the mean $\bar{x}$, of a sample of size n, chosen at random, being between $\left(\mu-1.96\frac{\sigma}{\sqrt{n}}\right)$ and $\left(\mu+1.96\frac{\sigma}{\sqrt{n}}\right)$ will be 95% .

i.e. the probability of the maximum difference between μ and $\bar{x}$ being less than $1.96\frac{\sigma}{\sqrt{n}}$ will be 95%. But the company wants to be 95% confident that the difference between the real and estimated averages will not exceed 50 hours.

$$1.96\frac{\sigma}{\sqrt{n}} \le 50$$

$$\sqrt{n} \ge \frac{1.96}{50}\sigma$$

$$n \ge \frac{(1.96)^2\sigma^2}{50^2}$$

If the standard deviation of the population were equal to, say, 200 hours, then

$$n \ge \frac{(1.96)^2(200)^2}{2500} = 61.4656$$

Hence the minimum value of n has to be 62. (Note that the minimum value of n will be 62 and not 61. When n is equal to 61.4656, the confidence level will be 95%. For any value of n less than 61.4656, the confidence level will be less than 95%. Hence the smallest integral

value of n for which the confidence level will be atleast 95%, will be 62). In general, if one is to be C% confident that the error (i.e. the difference between real and estimated means) will not exceed d, then the sample size n should satisfy the relation

$$n \geq \frac{Z^2\sigma^2}{d^2} \qquad \ldots(1.1)$$

where σ is the standard deviation of the population and Z is such that, the area of the standard normal curve between –Z and +Z is $\left(\frac{C}{100}\right)$. (Z will be equal to 1.645, 1.96, 2.57 or 3.0 according as C is equal to 90%, 95%, 995 or 99.75%). C is called the Confidence specification and d is called the Tolerance specification.

Case II: N is finite

When the size of the population is finite and fpc cannot be ignored, the size of sample is given by the formula:

$$Z.\sqrt{1-\left(\frac{n}{N}\right)}.\frac{\sigma}{\sqrt{n}} \leq d$$

i.e. $Z^2\left(1-\frac{n}{N}\right)\frac{\sigma^2}{n} \leq d^2$

i.e. $n \geq \frac{Z^2\sigma^2}{d^2}\left(1-\frac{n}{N}\right)$

i.e $n+\frac{n}{N}\frac{\left(Z^2\sigma^2\right)}{d^2} \geq \frac{\left(Z^2\sigma^2\right)}{d^2}$

let $n_0 = \frac{\left(Z^2\sigma^2\right)}{d^2}$

then $n+\frac{n}{N}\cdot n_0 \geq n_0$

$$n\left(1+\frac{n_0}{N}\right)^n \geq n_0$$

$$n \geq \frac{n_0}{\left(1+\frac{n_0}{N}\right)}$$

That is, if the minimum sample size on ignoring fpc is n_0, then the minimum sample size when fpc is not ignored will be

$$\frac{n_0}{\left(1+\frac{n_0}{N}\right)} \qquad \ldots (1.2)$$

ESTIMATION OF POPULATION STANDARD DEVIATION

The formulae (1.1) and (1.2) depend on population standard deviation σ. So, we can determine n if σ is known. But this may not be so in many practical situations. In such cases we have to first estimate the population variance before determining the minimum sample size required. There are four ways of estimating population variance for sample size determination.

(a) By taking the sample in two stages, the first being a simple random sample of size n from which the values of σ^2 and n can be obtained.

(b) By the results of a pilot survey

(c) Based on previous experience with the same or similar population.

(d) Based on rough estimates about the structure of the population and using some mathematical results.

Method-1

This gives the most reliable estimate of σ^2. But it is not often used since it considerably slows down the progress of the investigation.

First choose a simple random sample of size n_1 and find the variance s_1^2 of this sample. Then the required sample size will be n where

$$n = \frac{z^2 . s_1^2}{d^2}\left(1+\frac{2}{n_1}\right) \qquad \ldots (2.1)$$

Since we have already chosen n_1 items, we have to choose an additional $(n-n_1)$ items for completing the investigation. The formula 2.1 is valid, provided n1 is large enough to neglect terms of the order of $\left(\frac{1}{n_1^2}\right)$.

Method – 2

Sometimes, before launching a survey, in order to study its feasibility and to understand the problem involved, a small pilot survey will be made. If a small, simple random sample is chosen for the pilot survey, then the results of method – 1 can be used for determining the minimum sample size. But often the pilot survey will restricted to a part of the population. Suppose an investigation is undertaken by the statistical department of a life office which has its branches spread over the entire country. During the pilot survey the statistician may like to confine his attention to a few branches to which he can have easy

access or in which he expects to encounter some special problems. The variance of the sample chosen for such a pilot survey usually underestimates the population variance. Allowance has to be made for this fact while taking the sample variance as an estimate of population variance in the formula for the calculation of minimum sample size.

Method – 3

Some times a rough estimate of the population variance can be got from the results of similar surveys undertaken in the past. If suitable past data are available, the earlier value of the population variance (σ^2) may be used after making necessary adjustments for the changes that might have occurred with time. The following rough rule may be followed for making such adjustments.

"If the population mean is known to have increased or decreased by K% since the previous survey, the value of σ^2 can be increased or decreased by (2K)%."

- **Common Misconceptions**
  - **The sample should be a proportion (often 5 or 10 per cent) of the population;**
  - **The sample should total about 500;**
  - **Any increase in the sample size will increase the precision of the sample results.**

- **No such rule-of-thumb method is adequate.**
  - **<u>How large should a sample be</u>?**

1. **The degree of accuracy required.**
2. **The degree of variability or diversity in the population.**
3. **The number of different variables examined simultaneously in data analysis.**

Determination of sample size:

Formula used by Krejcie and Morgan

$$s = \frac{X^2 \, NP \, (1- P)}{d^2 \, (N-1) + X^2 \, P \, (1-P)}$$

s = Sample Size
X^2 = the table value of chi-square for 1 degree of freedom at the desired confidence level (3.841)
N = Population
P = the population proportion (assumed to be .50 since this would provide the maximum sample size).
d = the degree of accuracy expressed as a proportion (.05).

50

- **For example, one may wish to know the sample size required to be representative of the opinions of 9000 high school teachers relative to merit pay increases.**

Calculation

$$s = \frac{3.841 \times 9000 \times .5 \, (1- .5)}{(.05)^2 \, (9000-1) + 3.841 \times .5(1-.5)}$$

= 368.

Determination of sample size using Krejcie and Morgan Table

- **To obtain the required sample size enter the Table at N = 9000.**

The sample size representative of the teachers in this example is 368.

Krejcie and Morgan Sample size table

N	S	N	S	N	S
10	10	460	210	2600	335
20	19	500	217	2800	338
30	28	550	226	3000	341
40	36	600	234	3500	341
50	44	700	248	4000	351
60	52	800	260	4500	354
70	59	900	269	5000	357
80	66	1000	278	6000	361
90	73	1200	291	7000	361
100	80	1300	297	8000	367
120	92	1400	302	9000	368
140	103	1500	306	10000	370
160	113	1600	310	15000	375
180	123	1700	313	20000	377
200	123	1800	317	30000	379
250	152	1900	320	40000	380
300	169	2000	322	50000	381
360	186	2200	327	75000	382
400	196	2400	331	1000000	384

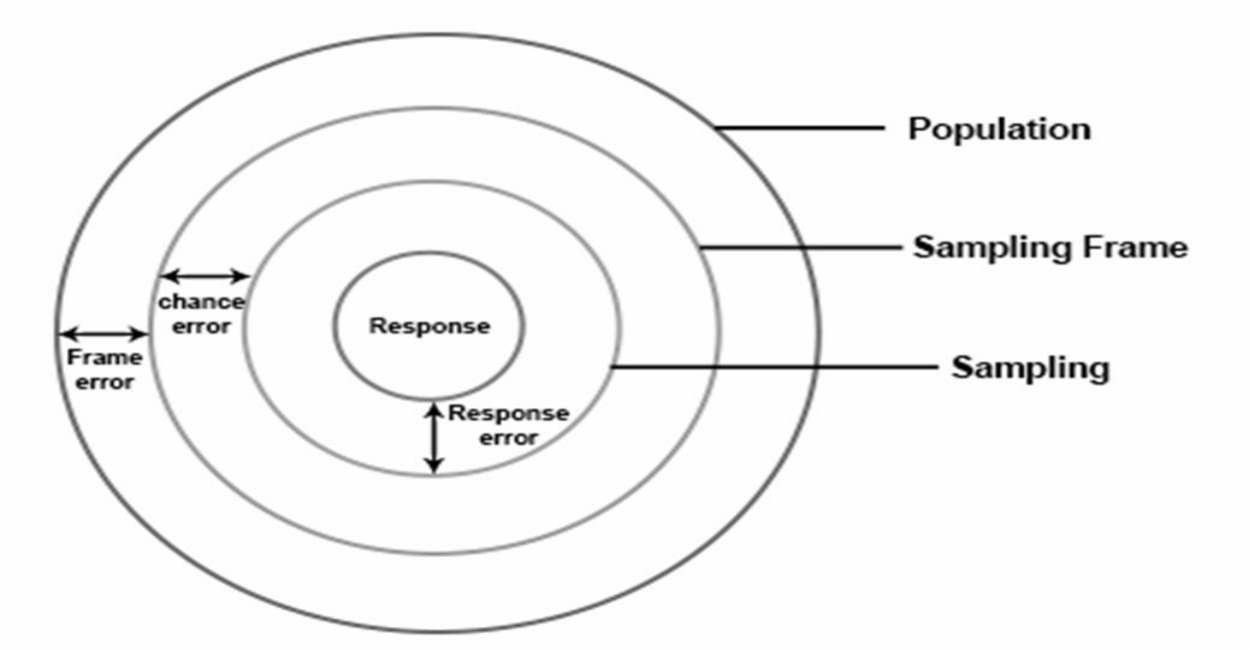
In practice, decision concerning the sample size are more complicated.

1. Researchers have to decide how precise they want their sample results to be, that is how large a standard error they can tolerate.
2. The decision on a sample size is also governed by the way; bivariate or trivariate, the results are to be analysed.
3. If more than one variable is to be studied, a sample that is adequate for one variable may be unsatisfactory for another.
4. In order to use a formula one needs an estimate of population standard deviation (standard error) which leads only to rough estimate of the required sample size.

Sampling Error or Error Variance

Sampling error refers to differences between the sample and the population that exists only because of the observations that happened to be selected for the sample.

Sampling Error: sample surveys do imply the study of a small portion of the population and as such there would naturally be a certain amount of inaccuracy in the information collected. This inaccuracy may be termed as sampling error or error variance. In other words sampling errors are those errors which arise on account of sampling and they generally happen to be random variations in the sample estimates around the true population values.



Sampling error = frame error+ chance error + response error

Types of sampling errors

1) Sample error : errors caused by the act of taking sample.
2) Non sample errors

   a) Non Response error

   b) Response error

- A Non response error occurs when units selected as part of the sampling procedure do not respond in a whole or in a part.
- A response or data error is any systematic bias that occurs during data collection, analysis or interpretation.
    - Respondent errors (sing, lying, forgetting etc)
    - Interviewer bias
    - Recording errors

- Poorly designed errors
- Measurement errors.

SAMPLING PROCEDURE

1. Determining Relevant Population an Parameters
2. Selecting appropriate sampling frame
3. Choosing Between Probability and Non-probability Sampling
4. Selecting the sampling method to be used
5. Determining the Necessary Sample Size
6. Selecting the Sample and Gathering Information
7. Validating the Sample

CENSUS AND SAMPLE

A complete enumeration of all items in the population is known as a census inquiry. It can be presumed that in such an inquiry when all items are covered, no element of chance is left and highest accuracy is obtained.

A census is the procedure of systematically acquiring and recording information about members of a given (total) population, under this method, data is collected for each and every unit, viz. persons, households, field, shop, factory etc. as the case may be of the population (or) universe.

In a sampling technique instead of every unit of universe only a part of the universe is studied and the conclusion is drawn on that basis for the entire universe.

Differences:

CENSUS	SAMPLE
1) Each and every unit of the population is studied.	1) But only few units of the population is studied in a sampling.

2) Census refers to periodic collection of information from the entire population.	2) Sampling is the most convenient method of obtaining data about the population.
3) Census method demands a large amount of finance, time and labour.	3) Less finance, time and labour
4) It is more suitable to use census method if population is heterogeneous in nature.	4) It is more suitable to use sampling method if population is homogeneous in nature.

Self Assessment Questions

1. Write a short note on difference between census and sample
2. Briefly explain about sampling procedure
3. What is sampling error? Explain in detail

Lesson -6

MEASUREMENT & SCALING BASICS

Structure of the lesson

1. Introdution
2. Nature and importance of measurement
3. Types of measurement scales: nominal, ordinal, interval, ratio
4. Validity and reliability in measurement

If you can't measure it you cannot Improve it"…… Peter Drucker

Introduction: In our daily life we are said to measure when we use some yardstick to determine weight, height or some other feature of a physical object, we also measure when we judge how well we like a song, a painting or personalities of our friends. Measurement is a relatively complex and demanding task. The data are collected on quantitative variables like price, income, sales etc., and qualitative variables like knowledge, performance, character etc. The qualitative information must be converted into numerical form for further analysis. This is possible through measurement and scaling techniques. A common feature of survey based research is to have respondent's feelings, attitudes, opinions, etc. in some measurable form. For example, a bank manager may be interested in knowing the opinion of the customers about the services provided by the bank. As a researcher you may be interested in knowing the attitude of the people towards the government announcement of a metro rail in Hyderabad. In this lesson we will discuss the issues related to measurement, different levels of measurement.

As we discussed earlier, the data consist of quantitative variables like price, sales, income, etc. and *qualitative variables like knowledge, performance, character etc. the qualitative information must be converted into numerical form for further analysis.* This is possible through measurement and scaling techniques. A common feature of survey based research is to have respondent's feeling, attitude, opinions etc. in some measurable form.

***Measurement means the process of assigning numbers to object or observations*, the level of measurement being function of the rules under which the numbers are assigned**

Technically speaking, ***measurement is a process of mapping aspects of a domain onto other aspects of a range according to some rule of correspondence***. In measuring we devise some form of scale in the range (in terms of set theory, range may refer to some set) and then transform or map the properties of objects from the domain.

Example: In case we are to find the male to female attendance ratio while conducting a study of persons who attend some show, then we may tabulate those who come to the show according to sex (gender). In terms of set theory, this process is one of mapping the observed physical properties of those coming to show (the domain) on to a sex classification (the range). The rule of correspondence is : if the object in the domain appears to be male, assign to "0" and if female assign as "1", similarly we can record a person's marital status as 1,2,3, or 4 depending on whether the person is single, married, widowed or divorced.

ISSUES IN MEASUREMENT

When a researcher is interested in measuring the attitudes, feelings or opinions of respondents he/she should be clear about the following:

a) What is to be measured?

b) Who is to be measured?

c) The choices available in data collection techniques

The first issue that the researcher must consider is 'what is to be measured'? The definition of the problem, based on our judgments or prior research indicates the concept to be investigated. For example, we may be interested in measuring the performance of a fast food company. We may require a precise definition of the concept on how it will be measured. Also, there may be more than one way that we can measure a particular concept. For example, in measuring the performance of a fast food company we may use a number of measures to indicate the performance of the company. We may use sales volume in terms of value of sales or number of customers or spread of network of the company as measures of

performance. Further, the measurement of concepts requires assigning numbers to the attitudes, feelings or opinions. The key question here is that on what basis we assign the numbers to the concept. For example, the task is to measure the agreement of customers of a fast food company on the opinion of whether the food served by the company is tasty, we create five categories: (1) strongly agree, (2) agree, (3) undecided, (4) disagree, (5) strongly disagree. Then we may measure the response of respondents. Suppose if a respondent states 'disagree' with the statement that 'the food is tasty', the measurement is 4.

The second important issue in measurement is that, who is to be measured? That means who are the people we are interested in. The characteristics of the people such as age, sex, education, income, location, profession, etc. may have a bearing on the choice of measurement. The measurement procedure must be designed keeping in mind the characteristics of the respondents under consideration.

The third issue in measurement is the choice of the data collection techniques. In lesson 7 you will be learning various methods of data collection. Normally, questionnaires are used for measuring attitudes, opinions or feelings.

Objectives of Measurements:

Measurement and scaling techniques helps us to

- To know the concept of measurement and scaling;
- To know the four level of measurement scaling techniques;
- Classify and discuss different scaling techniques, and
- Select an appropriate attitude measurement scale for our research problem.

LEVELS OF MEASUREMENT

The level of measurement refers to the relationship among the values that are assigned to the attributes, feelings or opinions for a variable. For example, the variable 'whether the taste of fast food is good' has a number of attributes, namely, very good, good, neither good

nor bad, bad and very bad. For the purpose of analysing the results of this variable, we may assign the values 1, 2, 3, 4 and 5 to the five attributes respectively. The level of measurement describes the relationship among these five values. Here, we are simply using the numbers as shorter placeholders for the lengthier text terms. We don't mean that higher values mean 'more' of something or lower values mean 'less' of something. We don't assume that 'good' which has a value of 2 is twice of 'very good' which has a value of 1. We don't even assume that 'very good' which is assigned the value '1' has more preference than 'good' which is assigned the value '2'. We simply use the values as a shorter name for the attributes, opinions, or feelings. The assigned values of attributes allow the researcher more scope for further processing of data and statistical analysis.

There are four levels of measurement and it is important to know what level of measurement you are working with as this partly determines the arithmetic and statistical operations you can carry out on them. The four levels of measurement in ascending order of precision are, **nominal, ordinal, interval** and **ratio**. As we go on to describe and give examples of each of these levels of measurement, you'll see that the numbers used to describe nominal data are simply used to classify data whereas the numbers describing interval or ratio measurements are much more precise and represent actual amounts.

In his seminar article entitled "on the theory of scales of Measurement" published in "science" in 1946, psychologist ***stanley smith stevens (1946) defined four generic types of scales.***

1 Nominal Scale

At the first level of measurement, numbers are used to classify data. In fact words or letters would be equally appropriate. For example, labelling men as '1' and women as '2' which is the most common way of labelling gender for data recording purpose does not mean women are 'twice something or other' than men. Nor it suggests that men are somehow 'better' than women. Another example of nominal scale is to classify the respondent's income into three groups: the highest income as group 1. The middle income as group 2, and the low-income as group 3. The nominal scale is often referred to as a categorical scale. The assigned numbers have no arithmetic properties and act only as labels. The only statistical operation that can be performed on nominal scales is a frequency count. We cannot determine an average except mode. Another example is blood groups where the letter A, B, O and AB represent the different classes

2. Ordinal Scale

In ordinal scales, values given to measurements can be ordered. One example is shoe size. Shoes are assigned a number to represent the size, larger numbers mean bigger shoes so unlike the nominal scale that just reflects a category or class, the numbers of an ordinal scale show an ordered relationship between numbered items – we know that a shoe size of 8 is bigger than a shoe size of 4. What you can't say though is that a shoe size of 8 is twice as big as a shoe size of 4. So numbers on an ordinal scale represent a rough and ready ordering of

measurements but the difference or ratios between any two measurements represented along the scale will not be the same. As for the nominal scale, with ordinal scales you can use textual labels instead of numbers to represent the categories. So, for example, a scale for the measurement of patient satisfaction with the care they received in hospital might look like this: | Not satisfied | Fairly satisfied | Satisfied | Very satisfied |

There are many everyday examples of measurements assigned to ordinal scales: social class gradings I, II, III, IV; academic grades A, B, C, D; house numbers 1,3,5…2,4,6.

3. Interval Scale

On an interval scale, measurements are not only classified and ordered therefore having the properties of the two previous scales, but the distances between each interval on the scale are equal right along the scale from the low end to the high end. Two points next to each other on the scale, no matter whether they are high or low, are separated by the same distance. So when you measure temperature in centigrade the distance between 96 and 98°, for example, is the same as between 100 and 102°C. Remember though is that for interval scales, a measurement of 100°C does not mean that the temperature is 10 times hotter than something measuring 10°C even though the value given on the scale IS 10 times as large. That's because there is no absolute zero: the zero is arbitrary. On the centigrade scale, the zero value is taken as the point at which water freezes and the 100°C value when water begins to boil and between these extreme values the scale is divided into a hundred equal divisions. (You may remember calibrating water thermometers at school using this method.)

Temperatures below 0° on the centigrade scale are designated negative numbers. So the arbitrary 0°C does not mean 'no temperature'. But when expressed on the kelvin scale, a ratio scale, a measure of 0 K equivalent to -273°C does indeed mean no temperature!

Other examples of interval measurements are rare, but there's one you will be familiar with. Calendar years are an interval scale. The arbitrary 0 (or 1 depending on your viewpoint) was assigned when Christ was born and time before this is labelled 'BC'.

4. Ratio Scale

Measurements expressed on a ratio scale can have an actual zero. Apart from this difference, ratio scales have the same properties as interval scales. The divisions between the points on the scale have the same distance between them and numbers on the scale are ranked according to size. There are many examples of ratio scale measurements, length, weight, temperature on the kelvin scale, speed and counted values like numbers of people, exam marks – a score of zero really does mean no marks!! Returning to the kelvin scale of temperatures, at the temperature of 0 K the lowest temperature possible, it is so cold that all molecules have stopped moving.

APPROPRIATE DESCRIPTIVE STATISTICS AND GRAPHS

Level of Measurement	**Properties**	**Examples**	**Descriptive statistics**	**Graphs**
Nominal / Categorical	Discrete/ Arbitrary (no order)	Dichotomous • Yes / No • Gender Types / Categories • colour • shape	Frequencies Percentage Mode	Bar Pie
Ordinal / Rank	Ordered categories/ Ranks	Ranking of favourites Academic grades	Frequencies Mode Median Percentiles	Bar Pie Stem & leaf
Interval	Equal distances between values/ Discrete (e.g., Likert scale)/ Metric (e.g., deg. F) Interval scales >5 can usually be treated as ratio	Discrete - Thoughts, behaviours, feelings, etc. on a Likert scale Metric - Deg. C or F	Frequencies (if discrete) Mode (if discrete) Median Mean SD Skewness Kurtosis	Bar (if discrete) Pie (if discrete) Stem & Leaf Boxplot Histogram (if metric)

Ratio	Continuous / Metric / Meaningful 0 allows ratio statements (e.g., A is twice as large as B)	Age Weight VO_2 max Deg. Kelvin	Mean SD Skewness Kurtosis	Histogram Boxplot Stem&Leaf (may need to round leafs)

Validity and Reliability:

RELIABILITY

Measurements must be reliable and objective and the results must be valid. Reliability is the repeatability and the extent to which comparable results are achieved every time a test is repeated. If a muscle test is repeated by one or more therapists who obtain the same grade every time; then the test is reliable. The key to reliability for manual muscle testing is to follow the standard procedures, performing the test in the same way each time and in the same way that other therapists perform it. Reliability is increased if the therapist gives clear instructions to the patient.

Evaluation procedures should exhibit inter-rater and intra-rater reliability:

Inter-rater reliability means that another person who performs the test should arrive at the same results, to an acceptable extent.

Intra-rater reliability means that one person should come out with the same results on every repetition of the test, within acceptable level.

There are three methods that estimate reliability of certain measurement tool; test-retest, parallel-forms and split-half methods. Inter-rater reliability test was done to examine the effects of goniometer size on the reliability of passive shoulder joint measurements. It is concluded that goniometric passive measurement of the range of the shoulder joint motion can be highly reliable when taken by a single therapist (intra-rater reliability), regardless of the size of the goniometer.

VALIDITY

Validity means that a test actually measures what it is supposed to measure. In muscle testing, therapists are testing the strength of a specific muscle. For a muscle test to be valid, the therapist must know the location and function of the muscle being tested and the location and function of surrounding muscles. Validity of assessment means that the therapists evaluate exactly what they are going to do and that the results are correct or true.

1 Face validity

Many of the measurements used in physical therapy clinical practice appear to be based on the assumption of face validity i.e. they are assumed valid as a result of inference. Face validity for a measurement is a lot-like the shine on a new car. It is a nice thing to have but you cannot get very far more than nice. Face validity is the appearance of a justifiable use for a measurement but does not mean there have any data or theory to support the use.

2 Construct validity

Construct validity of a measurement is made through the logic; used existing knowledge. It is a theoretical form of validity which supports the use of a measurement for a specific inference. There is no absolute way of knowing when measurement has construct validity. The manual muscle test was developed with a clear construct in mind. When poliomyelitis was widespread, manual muscle test was developed to characterize the weakness caused by the effect of the virus on anterior horn cells. The construct was simple: fully innervated muscles can generate more tension than partially innervated muscles; while totally denervated muscles can generate no tension.

3 Content validity

Content validity deals with how measurement schemes relate to their constructs, i.e. relating a measuring tool to a general theoretical framework in order to determine whether this tool tied to the theoretical assumption. The classic test of content validity is to ask whether we have chosen an adequate constellation of items to measure from the universe of items, defined by our construct.

4 Criterion-related validity

Construct validity deals with the theoretical basis for a measurement. Content validity deals with implementation of theory. Both construct and content validity are based on theory and can never be directly tested. This is in contrast to criterion-related validity. Criterion-related validity represents the ultimate test of validity. To demonstrate criterion-related validity, a measurement is compared to a criterion to determine whether or not the inference was appropriate. The measurement is tested by comparing it to something else, the criterion.

Conclusion:

There are four levels of measurements: nominal, ordinal, interval, and ratio. These constitute a hierarchy where the lower scale of measurement, nominal, has far fewer statistical applications than those further up this hierarchy of scales. Nominal scales yield data on categories; ordinal scales give sequences; interval scales begin to reveal the magnitude between points on the scale and ratio scales explain both order and the absolute distance between any two points on the scale.

Self Assessment Questions

1) **what is scaling in research why it is important ?**
2) **what is reliability and validity in scaling**
3) **Explain Nominal, Ordinal, Ratio and Interval scales in detail ?**

Lesson- 7

ADVANCED SCALING TECHNIQUES

Structure of the Lesson

1. Introduction
2. Likert's Summated Rating Scale
3. Thurstone's Equal Appearing Intervals
4. Ranking and rating scales
5. Outline of MDS (Multidimensional Scaling)
6. Q-sort technique

OBJECTIVE

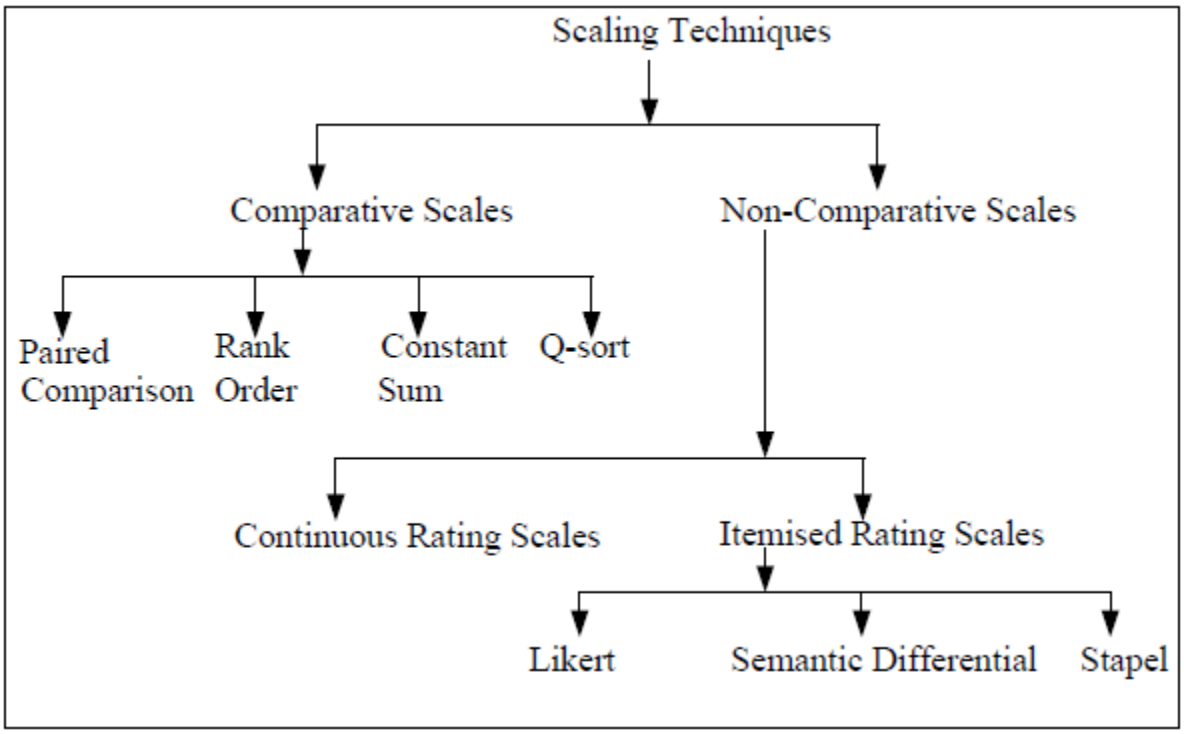
After going through this lesson you should be able to:

- Learn paired comparison, rank order, constant sum and Q-sort scales.

INTRODUCTION

The various types of scaling techniques used in research can be classified into two categories: (a) comparative scales, and (b) Non-comparative scales. In **comparative scaling**, the respondent is asked to compare one object with another. For example, the researcher can ask the respondents whether they prefer brand A or brand B of a detergent. On the other hand, in non-comparative scaling respondents need only evaluate a single object. Their evaluation is independent of the other object which the researcher is studying. Respondents using a non-comparative scale employ whatever rating standard seems appropriate to them. Non-comparative techniques consist of continuous and itemized rating scales. Figure 2.1 shows the classification of these scaling techniques.

Figure 2.1: Scaling Techniques



The comparative scales can further be divided into the following four types of scaling techniques: (a) Paired Comparison Scale, (b) Rank Order Scale, (c) Constant Sum Scale, and (d) Q-sort Scale.

PAIRED COMPARISON SCALE

This is a comparative scaling technique in which a respondent is presented with two objects at a time and asked to select one object (rate between two objects at a time) according to some criterion. The data obtained are ordinal in nature. For example, there are four types of cold drinks - Coke, Pepsi, Sprite, and Limca. The respondents can prefer Pepsi to Coke or Coke to Sprite, etc. In all we can have the following six comparisons.

Coke–Pepsi

Coke–Sprite

Coke–Limca

Pepsi–Sprite

Pepsi–Limca

Sprite–Limca

In general, with n brands we have n(n −1)/2 paired comparisons. The following is the data recording format using the paired comparisons.

Table 2.1

Brand	Coke	Pepsi	Sprite	Limca
Coke	—	Ö		
Pepsi		—		
Sprite	Ö	Ö	—	
Limca	Ö	Ö	Ö	—
No. of times preferred	2	3	1	0

A Ö in a particular box means that the brand in that column was preferred over the brand in the corresponding row. In the above recording, Coke was preferred over Sprite, Coke over Limca, in this case the number of times coke preferred was 2 times. Similarly, Pepsi over Coke, Pepsi over Sprite, Pepsi over Limca, in this case Pepsi was 3 time preferred. Thus, the number of times a brand was preferred is obtained by summing the Ö s in each column.

The following table gives paired comparison of data (assumed) for four brands of cold drinks.

Table 2.2

Brand	Coke	Pepsi	Sprite	Limca
Coke	–	0.90	0.64	0.14
Pepsi	0.10	–	0.32	0.02
Sprite	0.36	0.68	–	0.15
Limca	0.86	0.98	0.85	–

The entries in the boxes represent the proportion of respondents preferring 'column brand' and to 'row' brand. For example, 90% prefer Pepsi to Coke and only 10% prefer Coke to Pepsi, etc. Paired comparison is useful when the number of brands are limited, since it requires direct comparison and overt choice. One of the disadvantages of paired comparison scale is violation of the assumption of transitivity may occur. For example, in our example (Table 2.1) the respondent preferred Coke 2 times, Pepsi 3 times, Sprite 1 time, and Limca 0 times. That means, preference-wise, Pepsi >Coke, Coke >Sprite, and Sprite >Limca. However, the number of times Sprite was preferred should not be that of Coke. In other words, if A>B and B >C then C >A should not be possible. Also, the order in which the objects are presented may bias the results. The number of items/brands for comparison should not be too many. As the number of items increases, the number of comparisons increases geometrically. If the number of comparisons is too large, the respondents may become fatigued and no longer be able to carefully discriminate among them. The other limitation of paired comparison is that this scale has little resemblance to the market situation, which involves selection from multiple alternatives. Also, respondents may prefer one item over certain others, but they may not like it in an absolute sense.

RANK ORDER SCALE

This is another type of comparative scaling technique in which respondents are presented with several items simultaneously and asked to rank them in the order of priority. This is an ordinal scale that describes the favoured and unfavoured objects, but does not reveal the distance between the objects. For example, if you are interested in ranking the preference of some selected brands of cold drinks, you may use the following format for recording the responses.

Table 2.3: Preference of cold drink brands using rank order scaling

Instructions: Rank the following brands of cold drinks in order of preference. Begin by picking out the one brand you like most and assign it a number1. Then find the second most preferred brand and assign it a number 2. Continue this procedure until you have ranked all the brands of cold drinks in order of preference. The least preferred brand should be assigned a rank of 4. Also remember no two brands receive the same rank order.

Format:

Brand	Rank
(a) Coke	3
(b) Pepsi	1
(c) Limca	2
(d) Sprite	4

Like paired comparison, the rank order scale, is also comparative in nature. The resultant data in rank order is ordinal data. This method is more realistic in obtaining the responses and it yields better results when direct comparisons are required between the given objects. The major disadvantage of this technique is that only ordinal data can be generated.

Summatted Scale:

One of the most frequently used methods for assessment of people characteristics especially attitudes in the social science. It has been developed for the measurement of attitudes, beliefs, emotions, feelings perceptions and personality. This scale is introduced by the Rennis Likert's in 1932 article in archive of psychology titled "A technique for the measurement of Attitudes". This idea was expanded by likerts in 1934.

Meaning of summated scale.

1) Scale : a set of numbers, amounts etc. use to measure or compare the level of something.
2) Summated: to add together, total, sum-up.

Example. A patient visited Hospital for healthcare service and he expressed his opinion by using summated scale.

Here 10 statements are given the respondents has to give score to each statement then the respondents needs to sum-up the score then he/she needs to evaluate the level of satisfaction of the respondents.

1. I am satisfied that my Doctor has been taking care of me
2. My doctor explains the Reason(s) for any medical test
3. My doctor explains things in a way that is easy for me to understand
4. I am confident of my Doctor's knowledge and skills, experience
5. My doctor shows respect to what I have to say
6. My doctor listen careful to me
7. My doctor really cares about me as a person
8. My doctor encourage me to talk about all my health concerns
9. My doctor spends enough time with me (as a patient)
10. I would like my doctor to be present in my medical emergency situation.

10x5= 50 – Most favourable response possible

10x3=30 – A neutral attitude

10x1=10 – Most unfavourable attitude

The score for any individual would fall between 10 and 50. if the score appears to be above 30, it shows favourable opinion to the given point of view, a score below 30 unfavourable opinion and a score of exactly 30 would be suggestive neutral attitude.

CONSTANT SUM SCALE

In this scale, the respondents are asked to allocate a constant sum of units such as points, rupees, or chips among a set of stimulus objects with respect to some criterion. For example, you may wish to determine how important the attributes of price, fragrance, packaging, cleaning power, and lather of a detergent are to consumers. Respondents might be asked to divide a constant sum to indicate the relative importance of the attributes using the following format.

Table 2.4: Importance of detergent attributes using a constant sum scale

Instructions: Between attributes of detergent please allocate 100 points among the attributes so that your allocation reflects the relative importance you attach to each attribute. The more points an attribute receives, the more important the attribute is. If an attribute is not at all important, assign it zero points. If an attribute is twice as important as some other attribute, it should receive twice as many points.

Format:

Attribute	Number of Points
(a) Price	50
(b) Fragrance	05
(c) Packaging	10
(d) Cleaning Power	30
(e) Lather	05
Total Points	100

"If an attribute is assigned a higher number of points, it would indicate that the attribute is more important." From the above Table, the price of the detergent is the most important attribute for the consumers followed by cleaning power, packaging. Fragrance and lather are the two attributes that the consumers cared about the least but preferred equally." The advantage of this technique is saving time. However, there are two main disadvantages. The respondents may allocate more or fewer points than those specified. The second problem is rounding off error if too few attributes are used and the use of a large number of attributes may be too taxing on the respondent and cause confusion and fatigue.

Q-SORT SCALE

This is a comparative scale that uses a rank order procedure to sort objects based on similarity with respect to some criterion. The important characteristic of this methodology is that it is more important to make comparisons among different responses of a respondent than the responses between different respondents. Therefore, it is a comparative method of scaling rather than an absolute rating scale. In this method the respondent is given statements in a large number for describing the characteristics of a product or a large number of brands of a product. For example, you may wish to determine the preference from among a large number

of magazines. The following format shown in Table 2.5 may be given to a respondent to obtain the preferences.

Table 2.5: Preference of Magazines Using Q-Sort Scale Procedure

Instructions: The bag given to you contain pictures of 90 magazines. Please choose 10 magazines you 'prefer most', 20 magazines you 'like', 30 magazines to which you are 'neutral (neither like nor dislike)', 20 magazines you 'dislike', and 10 magazines you 'prefer least'. Please list the sorted magazine names in the respective columns of the form provided to you.

Format:

Prefer Most	Like	Neutral	Dislike	Prefer Least
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
_____	_____	_____	_____	_____
(10)	_____	_____	_____	(10)
	_____	_____	_____	
	_____	_____	_____	
	_____	_____	_____	
	_____	_____	_____	
	_____	_____	_____	
	_____	_____	_____	
	_____	_____	_____	
	_____	_____	_____	
	_____	_____	_____	
	(20)	_____	(20)	
		_____		
		_____		
		_____		
		_____		
		_____		
		_____		
		_____		
		_____		
		_____		
		(30)		

Note that the number of responses to be sorted should not be less than 60 or not more than 140. A reasonable range is 60 to 90 responses that result in a normal or quasi-normal distribution. This method is faster and less tedious than paired comparison measures. It also forces the subject to conform to quotas at each point of scale so as to yield a quasi-normal

distribution. The utility of Q-sort in marketing research is to derive clusters of individuals who display similar preferences, thus representing unique market segments.

Conclusion:

The measurement scales, commonly used in marketing research, can be divided into two types; comparative and non-comparative scales. Comparative scales involve the respondent in signaling where there is a difference between two or more firms, brands, services, or other stimuli. The scales under this type are: (a) Paired Comparison, (b) Rank Order, (c) Constant Sum, and (d) Q-sort.

Self Assessment Questions

Lesson -8

DATA COLLECTION METHODS & QUESTIONNAIRE DESIGN

Structure Of The Lesson

After studying the lesson, students will have clear comprehension of the primary data, secondary data, methods of collecting primary data, editing primary data, sources of secondary data and its uses, census and sample experiments, panels and simulation

- **Introduction**
- **Primary data and Secondary data**
- **Methods of collecting primary data**
- **Editing primary data**
- **Sources of secondary data and its uses**
- **Census and sample experiments**
- **Panels**
- **Simulation**
- **Summary**

Introduction: Once the object and scope of inquiry has been laid down, the investigator has to decide about the sources from which the data are to be collected. Here, he has two alternatives:

i. either he may collect the data himself or
ii. he may take the data from published sources

Primary and Secondary Data :

Primary data are those statistics data which are collected for the first time and are original in nature. Primary data are collected originally by the authorities who are required to collect them. The sources from which primary data are collected is called primary source. The method of collection of primary data is also known as the primary method of collection of data. Primary data are collected for the first time by the authorities who require the data

for their own use and treatment. Data collected by field workers, investigators and enumerators are primary data.

Secondary data are those statistical data which are collected and published by one organization and subsequently treated and utilized by other organizations. Secondary data are originally collected and published by the organizations other than the authorities who require them subsequently for their use and treatment. So secondary data of one organization become the primary data of other organizations who first collect and publish them. The source from which secondary data are collected is called secondary source. The method of collection of secondary data is known as the secondary method of collection of data.

Methods of Collecting Primary Data :

i. Direct Personal Investigation (Observation)

Here data are collected by the investigator himself through his personal observations about the behavior of the source. He would not ask direct questions from the source about his likings etc. but would provide him an opportunity of an unmindful selection from various options. Instead of asking from the source whether he prefers ink or a refillpen, the investigator observes his behavior when, of his own, purchases either of them from the market.

Merits :

a) It gives very good results for the intensive and limited enquiries.
b) Because of the direct personal supervision of the investigator it gives maximum degree of accuracy.
c) It is suitable even when the nature of enquiry is confidential
d) It is a most suitable method for collecting data concerning marketing enquiries.

Demerits :

a) The method is quite expensive for detailed and extensive surveys
b) Personal biases will creep into the observations very easily.

ii. Indirect Oral Investigation (Personal Interviews)

This method consists of the collection of data through indirect sources. The investigators, appointed for the purpose, go to the possible sources for recording their statements pertaining to the problem and conclusions are drawn on the basis of the informations supplied by these sources. Generally, a list of concerning questions is prepared and these questions are put before different persons for their replies which are recorded. Such a procedure is adopted by commissions, or enquiry committees etc., appointed by for carrying out some specific statistical investigation.

Merits :

a) The scope of enquiry is extensive but less expensive and time saving.
b) Here the direct contact with the original source is not necessary
c) The investigators are in a position to add questions in doubtful cases. So the data collected, here are sufficient.

Demerits :

a) Selection of improper witness may yield biased results
b) The biasedness of the informant may colour the facts

iii. Investigation through Local Agencies :

Here investigators do not move for the formal collection of data but the information or correspondents are appointed in the areas under survey to send the required information to the head office. The nature of the appointment of these agents may be regular or adhoc depending upon the nature of the investigation. No doubt, some instructions are issued to these agents but in practice, they send informations according to their own ways and decisions

Merits :

a) This technique is comparatively cheap and easy
b) The data are obtained expeditiously
c) The technique is quite useful when the field of investigation is wide-spread.

Demerits :

a) Degree of accuracy is quite limited

b) The data collected are not that much reliable, because of free hand of the third agency called correspondents or agents.

iv. The Schedule or Questionnaire by Post :

In this method the schedules of questions known as 'Questionnaire' provided with blank spaces for answers, are mailed to the informants with necessary instructions and request that they should return them duly filled in. The questions are so selected that the informant is expected to possess definite as well as accurate knowledge about them.

Merits :

a) This method is comparatively economic and expeditious
b) The informant can fill the questionnaire at his convenience
c) The information furnished, being in the own hand-writing of the informant is more authentic

Demerits :

a) The problem of non-response is rather acute
b) The method can only be used for literate and cautious people
c) The questionnaire cannot be changed during the process of inquiry and so no additional information can be collected

v. Questionnaire through investigators :

In this method investigators are sent to the informants along with the schedules of questions in order to collect the necessary information. The investigators explain the aim and object to such an enquiry to the informants and emphasize upon them to give correct and useful replies.

The questionnaire should have only the simple, straight and non-personal queries contained in it. Also the investigator should be polite, courteous and well versed in social dealing.

Merits :

a) Maximum possible results can be obtained. The investigator can cross –examine for getting the reliable answers.
b) The amount of information is quite large here
c) The field of information is spread over a wide area.

Demerits :

a) This method is quite expensive
b) The process is subject to the condition that the investigator is properly trained

Editing of Primary data :

When the schedules and questionnaires are received from the respondents by post in the case of mail survey or from enumerators in the case of personal interview, the investigator should review them.

After the schedules have been reviewed by the investigator, normally they under go an editing process to prepare them for tabulation. Once the schedules are in the office, the investigator can personally supervise and check them for accuracy. Thus editing is a check on the quality of the interviewing and the response of the respondents.

While editing primary the following points are worth noting.

1. The schedules should be checked to make sure they are complete.
2. It should be found out whether the answers are consistent
3. Accuracy of the data should be ascertained
4. Homogeneity of the data should be determined.

1. **Editing for completeness :** In case the schedule or questionnaire is incomplete, it should be returned for necessary action. But this procedure is costly and time-consuming and should be resorted to only in case where the questions not answered are of great importance. Incase no reply is received, then it is advisable to drop that schedule or questionnaire which is complete.

2. **Editing for Consistency :** The investigator should then find out whether the answers are consistent internally and with the conditions known to exist in the survey area.

For example there may be a discrepancy between the age given and the date of birth. The investigator should try to obtain the correct answers either by exercising his own judgment and discretion.

3. **Editing for Accuracy :** Sometimes answers appear to be highly improbable, if not impossible. There may be significant discrepancies. There may be inaccuracies due to arithmetical errors. Such errors can be easily detected and corrected.

4. **Editing for homogeneity:** The investigator should then ascertain whether the information supplied by the respondents is homogeneous and uniform. For example, as to the question of wages, if some respondents have given daily wage. Others weekly wage and still others monthly wages there is no uniformity in the data and therefore no comparison can be made. The investigator will have to reduce this information to some common base, say weekly wages.

Sources of Secondary data :

There are many sources of published information from which the investigator may make statistical studies. Such sources are discussed below :

i. Government or semi-Government Publications : The publications of the Government or Semi Government agencies such as the statistical Abstract of India published by Central Statistical Organization (CSO) are very good sources as obtaining the secondary data. The statistical material published by the Central or State Governments or by the bodies like Municipal and District boards or the Corporations etc., are quite reliable

ii. Publications of Trade – Associations or Chambers of Commerce : The data may also be obtained from the statistical material published by trade associations, Chambers of Commerce like FICCI, Bank Bodies, Co-operative Societies and Trade Unions etc.

iii. News – Papers and Periodicals : The statistical material on various topics may also be collected from the numerical facts gathered and published in news-papers and periodicals such as : Indian Journal of Economics, Commerce, Capital, Eastern Economist, Economic Times, Indian Finance, State Man's Year Book and 'The Times of India Year Book' etc.

iv. Research Bureaus and Private Academic Organisations : There are several Research Bureaus, Clubs, University Departments, Private Organisations and Scholars which collect statistical material on different topics. These materials are available both in published and unpublished forms. With proper precautions the data from these sources are collected per requisites of the problem

v. Publications of the foreign Governments of the International Bodies : The statistical year book is valued UNO Publication and so is the statistical abstract of United States.

vi. Reports of the various Committees and Commissions appointed by the Government : The observations of the Land Reform Committee, Wanchoo Commission's report on taxation etc., can be the good sources of availing the statistical data.

Uses :

1. It saves money and time.
2. As in the case of Government publications the data collected are reliable, because a separate statistical department is attached to almost all the main Ministries at the Centre and State Levels
3. It covers wide range of areas and departments

Census and Sample Experiments :

Statistical data may be collected by any one of the following ways ;

a) Census / Complete Enumeration : In this system a full enumeration of the population is made and information is collected in respect of all the units of the population. This is also known as Census Survey or Census. The census of the population of the human beings in India is an example of complete enumeration or census survey. A complete enumeration of all the people of India is made at an interval of ten years and information is collected in respect of each and every person of India.

b) Sample Method or Partial Enumeration or sample experiment :

In this system an enumeration of a part of the population or universe is made and information is collected in respect of the Units of the selected part of the population or Universe. This is also known as Sample Survey. The estimates of crop-yield per acre of land

and the control of quality of goods produced on mass production are the examples of partial enumerations or sample survey.

Now the question arises which of these two systems should be adopted in collecting statistical data. There is no hard and fast rule on the choice. The choice between the two depends upon a number of factors; important among which are the following :

1) Nature and Scope of enquiry.
2) Time at the disposal of the investigator
3) The financial provision for the enquiry
4) The degree of accuracy desired in the enquiry
5) The field of enquiry.

The relative merits of the two methods are summarized below :

BASIS	CENSUS METHOD	SAMPLING METHOD
1. Cost (Financial Provisions)	1. It involves more costs	1. It involves less costs
2. Reliability	2. The results are perfectly reliable	2. The results are less reliable
3. Time involved	3. It is highly time consuming	3. It saves time
4. Labour	4. It needs more labour	4. It needs less labour
5. Organisation	5. It needs more organizational skill and large force of investigators	5. It needs less organizational skill and large force of investigators

PANELS:

Panel Data:

Panel data, also called longitudinal data or cross-sectional time series data, are data where multiple cases (people, firms, countries etc) were observed at two or more time periods. An example is the National Longitudinal Survey of Youth, where a nationally representative sample of young people were each surveyed repeatedly over multiple years.

There are two kinds of information in cross-sectional time-series data: the cross-sectional information reflected in the differences between subjects, and the time-series or within-subject information reflected in the changes within subjects over time. Panel data regression techniques allow you to take advantage of these different types of information.

When it is possible to use ordinary multiple regression techniques on panel data, they may not be optimal. The estimates of coefficients derived from regression may be subject to

omitted variable bias -a problem that arises when there is some unknown variable or variables that cannot be controlled for that affect the dependent variable. With panel data, it is possible to control for some types of omitted variables even without observing them, by observing changes in the dependent variable over time. This controls for omitted variables that differ between cases but are constant over time. It is also possible to use panel data to control for omitted variables that vary over time but are constant between cases.

SIMULATION :

It is evident that there are many problems of real life which cannot be represented mathematically due to random nature of the problem, the complexity in problem formulation, or the conflicting ideas needed to properly describe, the problem under study. Under such circumstances simulation is often used when all else fail. This method is often viewed as a 'method of last resort'.

In fact, simulation is the representative model for real situations while visiting some trade-fairs and exhibitions we often find a number of simulated environments therein. For example, a children's cycling park with various signals and crossings in the exhibition is a simulated (represented) model of city-traffic in real system.

Another idea of simulation is involved in flight simulators for training pilots. A computer directs the student's handling of the controls in a simulated aeroplane flight deck. The instruments are then operated by the computer to give the some readings which they would in a real flight. An instructor can intervene with 'Catastrophes' like an engine failure or a bad storm and a television camera is moved over a model of some country side to give the trainee visual feed back of how the air craft is behaving.

Simulation is one of the easiest tools of management science to use, but probably one of the hardest to apply properly and perhaps most difficult from which to draw accurate conclusions. Regardless of the drawbacks, simulation is a useful technique and one which is specially suitable for complicated operations research and systems analysis problems.

Definitions of Simulation:

1. Simulation is a representation of reality through the use of a model or other device which will react in the same manner as reality under a given set of conditions.
2. According to Donald G. Malcohm a simulated model may be defined as one which depicts the working of a large scale system of men, machines, materials and

information operating over a period of time in a simulated environment of the actual real world conditions.

A simulation model mainly consists of two basic phases.

Phase 1 : Data generation : Data generation involves the sample observation of variables and can be carried out with the help of any of the following methods.

i. Random number tables
ii. Mechanical devices

Phase 2 : Book –Keeping : The book – keeping phase of a simulation model deals with updating the system when new events occur. Monitoring recording the system states as and when they change and keeping track of quantities of our interest to compute the measures of effectiveness.

Questionnaire

The success of the questionnaire method of collecting information depends largely on the proper designing of the questionnaire. Designing questionnaire is a highly specialized job and requires a great deal of skill and experience. It is difficult to law down any hard and fast rules to be followed in this connection. Designing of questionnaire is very much an art.

Most of what is known about making questionnaire is based on experience. Neither a basic theory nor even a fully systematized approach to the problem has been developed. The extensive experience of many researchers and organized experiments have led to a considerable understanding of the problem and to a long list of "do's and don't's" rules of thumb. These can help in designing a questionnaire procedure.

Points to be very clear: While developing a questionnaire, the researcher has to be very clear on the following issues:

1. What information will be sought?
2. What type of questionnaire will be required?
3. How that questionnaire will be administered?
4. What the content of the individual question will be?
5. What the form of response of each question will be?

6. How many questions will be used and how the individual questions will be sequenced?
7. Whether the questionnaire shall be structured or unstructured?

10.2 Structured and Unstructured Questionnaires:

A questionnaire can be either structured or unstructured and disguised or undisguised as can be seen from the following:

	Structured	Unstructured
Undisguised	A	B
Disguised	D	C

Structure refers to the degree of standardization imposed on the questionnaire. A highly structured is one in which the questions to be asked and the response permitted are completely predetermined. A highly unstructured questionnaire is one in which questions to be asked are only loosely predetermined, and the respondent is free to respond in his / her own words and in any way he / she sees fit

Disadvantages of Unstructured Techniques:

(1) They are slow and hence, costly administer in the field and to tabulate; and

(2) The data collection process and the interpretation of results are both subjective and hence open to bias. Structured techniques overcome these problems, but they are difficult to use in situations where respondents may hesitate to report their attitudes.

A disguised questionnaire attempts to hide the purpose of the study where as an undisguised questionnaire is one in which the purpose of the research is obvious from the questions posed.

Structured Undisguised questionnaires are the most commonly used type in practice. In such questionnaires the responses as well as the questions are standardized. This is accomplished by employing fixed alternative questions in which the responses of the subject are limited the stated alternatives.

Example: Type of questions regarding people's attitude towards social security.

Do you feel India needs more (or less) social security legislation?

□ Needs more

□ Needs less

□ Neither more nor less

□ No opinion

Structured undisguised questionnaire are simple to administer and easy to tabulate and analyse. The respondent also feels almost no difficulty in replying the question. The question " What is your marital status" is more confusing than is the question. "Are you married, single, widowed or divorced?". The fixed alternative questions are most productive when the possible replies are well known, limited in number and clear cut.

The unstructured undisguised questionnaire is one in which the purpose of the study is not concealed but the response to the question is open ended. Thus consider the question "How do you feel about the need for legislation for more social security measures?". Such measures provide complete freedom to the respondent. However the responses are difficult to tabulate and analyse.

In the unstructured disguised questionnaires, the respondents are not directly told about the purpose of study and the questions are framed in a manner that there is complete freedom for the respondent to answer. The basic philosophy underlying such questionnaires is that the more unstructured and ambiguous a stimulus, the more a subject can and will project his emotions, needs, motivations, attitudes and values, practical difficulties of editing, coding and tabulation of replies impose series limitations on the use of the methods. This method is more often used for exploratory research than for descriptive or casual research.

The structured disguised questionnaires are also not very popularly used in practice. They emerged as an attempt to secure the advantages of disguise in revealing unconscious and hidden motives and attitudes along with the advantages in coding and tabulation common to structured questionnaires. The main advantage of this approach emerges in analysis.

Essentials Of Good Questionnaire/Wording Of Questionnaire

- It should be short and simple questions should arranged in logical sequence;
- Avoiding suggestive answer (smoking is injurious)
- Technical terms should be avoided
- Questions affecting the sentiments of the respondents
- Avoiding ambiguity
- Avoiding questions involving generalization
- Make sure sequence of questions
- Meeting the questionnaire
- Final draft of the questionnaire

Questionnaire Vs Schedule

QUESTIONNAIRE	SCHEDULE
▪ Generally send to through mail ▪ Cheaper method ▪ Non response is high ▪ Incomplete and wrong information is more ▪ Depends on the quality on the questionnaire ▪ Very slow method ▪ No personal contact ▪ Educated only cooperate.	▪ Schedule is filled by the investigator ▪ Costly requires field workers ▪ Non response is low ▪ Depends on honesty of the investigator ▪ Relativity is more collect and complete ▪ Identity of the person is known ▪ Information is collected well in time ▪ Direct personal contact

	▪ Information can be collected from illiterate also.

Summary :

Concepts data, Primary data and Secondary data methods of collecting primary data, sources of secondary data, editing of primary data, census and sample experiments, panels, simulation are discussed.

Self Assessment Questions:

Lesson -9

DATA PRESENTATION & BASIC STATISTICAL TOOLS

Structure of the Lesson

1. Editing and classification of primary data
2. Types and objectives of classification
3. Frequency distribution
4. Charts: bar chart, histogram, two-dimensional graphs
5. Measures of central tendency: mean, median
6. Standard deviation
7. Correlation and regression
8. Chi-square test
9. Summary

Introduction

Primary data are usually contained in schedules or questionnaires. It is necessary to classify and tabulate the information collected so as to bring out the salient features of the data. In secondary data, it may be necessary to rearrange or reclassify them according to the nature of the study. Once data are classified on the basis of some common characteristic, arranged serially and put into the form of frequency distributions and / or tables nothing more need be done because this presentation of the data may give a clear picture of the significance of the information.

Meaning: It stands for grouping of related fact into classes. Classification is the sorting out of a heterogeneous mass data into a number of homogeneous groups and sub groups by their respective characteristics so that the main features and significance of the data are clearly brought out. This grouping is done with respect to some characteristic called a basis of classification.

Objectives of Classification

- ✓ To bring out the unity of attributes out of the diversities persistent in the collected data. Mass of unwieldy data is condensed and arranged in a systematic manner a few classes having common features. This enables easy perception and understanding.
- ✓ To facilitate comparison
- ✓ To give prominence to the important information gathered while dropping out the necessary elements.

- ✓ To bring out the true significance of the characteristics of the data in a suitable manner at a glance.
- ✓ To enable a statistical treatment of the data collected i.e., analysis, interpretation and drafting the required report etc.

Types of Classification:

Data can be classified according to the characteristics that they have. Theses are two types

1. Descriptive 2. Numerical

Corresponding to these characteristics of data, classification is of the following two types

1. By attributes 2. By variables

1. By attributes: In the theory of attributes the objects or individuals are classified according to some property e.g., tall and short, black and white etc. Presence or absence of the attribute chosen, may be counted in individual cases., e.g., literacy and illiteracy, blindness and non blindness. Thus statistics of attributes is a kind of data for which it is not possible to measure the magnitude.

2. By Variable: A quantity which varies from one individual to another is known as a variable or variate. Classification based on numerical characteristics is called classification according to class intervals.

Seriation: Seriation is concerned with the logical listing or arrangement of the data into a particular sequence or order in different classified categories. The statistical series consists of the following types :

1. Chronological, Historical or Time Series
2. Geographical or spatial Series
3. Frequency or Condition Series

1. Chronological, Historical or Time Series: In such a series the basis of the information of the series is time. According time the data are arranged.

Example:

Year	Number of Companies
1968 – 69	27,973
1969 – 70	28,960

1970 -71	30,412
1971-72	32,562

2. Geographical or spatial series or regional series: As the implies place or geographical location is the most important factor in such a series.

Geographical classifications are usually listed in alphabetical order for easy reference. Items may be listed by size to emphasize the important areas.

3. Frequency or Condition series: It is such a series in which the data are arranged with reference to the physical condition such as height, age or any other method of gradation, with respect to their frequency of occurrence at a given time and space.

Example:

Marks	Number of Students
0-10	7
10-20	20
20-30	15
30-40	8

Method of seriation:

The arrangement of the size of items or values can be done in different ways. They are

1. Individual Series 2. Grouped Series

1. Individual Series : In this case the items are listed as they are observed

Example :

Family No :	1	2	3	4	5	6	7	8
No.of Children	1	2	0	3	2	3	1	0

When an item of a given size has been repeated for a number of items, it shall be written as many times as it has appeared. This method can only be adopted when the data are limited. The data given in this form are also called ungrouped data.

Grouped Series: In the case of grouped series, data are grouped into some class –intervals of certain sizes. Each class or group will show the frequency of occurrence of observations.

The grouped series or distribution can take two forms.

1. Discrete series 2. Continuous series

This distinction is based on the nature of the variables. Variables are of two kinds (i). Continuous and (ii) Discrete. A series in respect of a continuous variable is known as 'Continuous Series'. On the other hand, the one in respect of a discrete variable is known as 'discrete series'

1. Discrete series: Variables which can take only particular values are known as discrete variables. The discrete variables are in whole numbers. Thus a variable is said to be discrete when there are gaps between its one value and the next.

Example:

X : 1, 2, 3, 8

X ; denote the variable; and values it takes it.

2. Continuous Series: Variables which can take any numerical value with in certain range are known as continuous variables. Ranges are known as class intervals. Each group will show the frequency of occurrence of observations.

Example:

Class Interval :	0-1	2-3	4-5
Frequency:	11	14	2

Frequency distribution: Number of observations of an attribute or values of a variable arranged according to their magnitudes either individually in the case of both discrete and continuous series.

Construction of Discrete and continuous frequency distributions Discrete frequency distribution:

1. Prepare three columns – one for the variable. One for Tally bars and the third for the frequency corresponding to the size or value of the variable.

2. In the first column, place all possible values of the variable from the lowest to the highest.
3. In the second column, put a bar opposite the particular value to which it relates.
4. In the third column frequency as counted with the help of bars, its placed opposite the value or size of the variable.

Example: Prepare discrete distribution for the following data :

10	20	30	35	40	25	10	15	15	25
35	40	15	25	30	25	20	25	25	30

Solution:

Marks	**Tally bars**	**Number of students frequency**
10	I	1
15	IIII	4
20	II	2
25	IIII I	6
30	~~/~~HI	3
35	II	2
40	II	2
		Total : 20

Formation of continuous frequency distribution:

We divide the observation into groups having ranges known as class intervals. The class – limits are the lowest and the highest values that can be included in the class. The limits of the classes or groups are to be defined with precision. The following are the two methods used for expressing class limits.

1. 'Exclusive' Method
2. 'Inclusive' Method

1. 'Exclusive Method: Under this method upper limit of the first class is excluded from the class and included in the second class. For example

Class Interval	Frequency
40 – 50	15
50 – 60	20
60 – 70	10

Value of the 50 is included in the class 50 – 60 but not 40 -50; in 40 – 50, 50 is excluded.

It may thus be remembered that if classes or groups are given like 40-50, 50 -60 etc it is always presumed that upper limit is exclusive i.e., the last value shall go with the next class.

Any exclusive method of classifications is specially useful for continuous frequency distributions.

2. Inclusive Method: Under this method, the items of the value of both the lower and upper limits are included in the class.

Example:

Class Interval	Frequency
0-10	2
10-20	6
20-30	3

But as far as continuous

Class –Intervals: The difference between the upper and lower limits of a class is called its magnitude or class interval.

Example: 0 -10, 10-20 Class interval is 10

Class frequency: The number of observations falling with in a particular class is called its frequency or class – frequency.

Guide lines for class selection and forming a frequency distribution:

Number of classes: There should be sufficient classes so that each of the collected individual observations are included in some one of them. The number of classes should not ordinarily exceed 20.

The best rule of thumb for determining the number of classes is provided by a formula known as Sturge's rule, which is

K = 1+3.3 log N

$$class\ \text{int}erval = \frac{Range}{Number\ of\ classes(k)}$$

Size of Class Interval: The selection of the class – interval is influenced by the number of items and the range of the variable over which frequencies are found. The choice of class-interval would depend on the following considerations.

i. The class interval should be of equal width as far as possible.

ii. Also open-end classes should be avoided as far as possible.

iii. An interval should be of the size 5 or multiples of 5 or 10 or multiples of 10

iv. Fix the lower and the upper limits in such a way that the observations are evenly distributed over the interval or there may be concentration at the mid points of the classes.

v. Size of class interval

$$\bar{c} = \frac{Highest\ value - Lowest\ value}{1 + 3.3 \log N}$$

vi. It is better to have an integer as the class interval

Cumulative and Relative Frequencies:

In a cumulative frequency distribution, the cumulative frequencies (c.f) are desired by the cumulation (successive adding) of the frequencies of the successive individual class intervals. The cumulative frequency of a given class interval thus represents the total of all the previous class frequencies including the class against which it is written. We can say that the cumulative frequency if it is less than type will represent the total frequency of all classes less than and equal to the class value to which it relates. i.e, if f_1, f_2 --- are frequencies then less than cumulative frequencies are f_1, f_1+f_2, $f_1+f_2+f_3$ and so on.

If it is of a 'more than' type, it will represent the total frequency of classes more than and equal to the class value to which it relates.

Relative frequencies are expressed as a fraction of the total frequency.

A frequency distribution showing the cumulative frequencies against values of the variable systematically arranged in increasing or decreasing order is known as cumulative frequency distribution.

Example: Prepare less than, greater than and relative frequency distributions for the following data.

Values	Frequency
0	6
1	8
2	25
3	31
4	18
5	7
6	4
7	0
8	1

Solution:

Values	Absolute Frequency	Relative Frequency	Less than cumulative frequency	Greater than cumulative frequency
0	6	6/100 = 0.06	6	100
1	8	8/100 = 0.08	6+8=14	100-6=94
2	25	0.25	6+8+25=39	94-8=86
3	31	0.31	70	86-25=61
4	18	0.18	88	30
5	7	0.07	95	12
6	4	0.04	99	5
7	0	0	99	1

8	1	0.01	100	1

Charting of Data

Statistical data can often be presented in chart form or by means of diagrams or graphs. The diagrammatic and graphical method of presentation of data may be employed to render statistical data comprehensible and more specially for those who do not have sufficient time or the inclination to go through a huge mass of data. It makes the unwieldy data readily intelligible and brings to light the salient features of the data at a glance, permitting a visual comparison of the data easier.

General Rules for constructing Diagrams: The following general rules should be observed while constructing diagrams.

1. Title: A short suitable which will carry the general important and purpose of the chart should be put up in the middle at the top.

2. Proportion between width and height: A proper proportion between the height and width of the chart should be maintained so that it gives an attractive look. A ration of 1 (short side) to 1.414 (long side) is recommended.

3. Selection of Scale; The chart should be drawn to proper and accurate scales. The scale showing the values should be in even numbers or in multiples of five or ten. Ex: 2, 4, 6, ……. Or 5, 10, 15, ------ Or 10,20,30, --------- etc. However, where the limits given are 65-69, 70-74, -----etc the class- limits should be taken to be 64.5 – 69.5 and 69.5 – 74.5 etc.

4. Foot-note: Explanatory notes to elucidate the important points should be added at the bottom of the diagram.

5. Index: An index to interpret the symbols, lines, colours etc as well as the scales should be given.

6. Labelling the scales: The vertical and horizontal scales should always be labelled definitely so that each scale represents the units employed on each, for example ‘number of workers in thousands’ etc.

7. Neatness and cleanliness: While drawing a chart, care should be taken that it is neat and clean.

8. Simplicity: The main purpose of diagrammatic representation will be defeated if the diagrams are complex ones which are difficult to understand. It is better to draw two or more than two diagrams for the given data rather than put all the information in one diagram.

One dimensional or bar diagrams:

They are in very common use and enable comparison of simple magnitude of different values or items. They are most frequently used charts in showing the relationship of the parts to the whole.

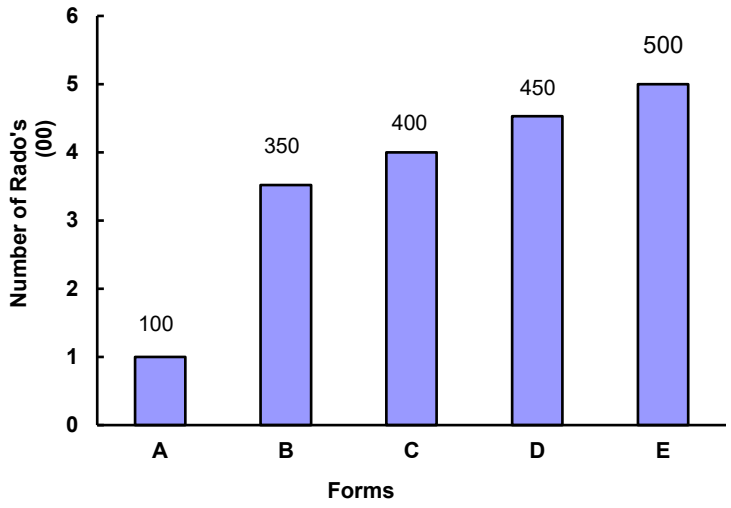
Length to the bars or rectangles is proportional to the magnitude to be represented. The widths of the bars, which are all equal, have no significance and can be taken any convenient size. It is entirely a matter of neatness of presentation. Also when the number of items is large lines may be drawn instead of bars to economise in space. As in such diagrams, only length matters and not the width, they are known as one-dimensional diagrams. The use of colour when possible to add impact is desirable.

Simple bar diagrams: They are very popular but only one variable can be represented by them. For example, a population of a country for a number of decades can be represented with the help of a bar diagram but is not possible to show region-wise or sex-wise distribution.

Example: Draw a bar diagram to represent the following figures relating to the manufacturing of radios by five firms in a city.

Firms :	A	B	C	D	E
Number of radio's :	100	350	400	450	500

Solution : Scale 1 cm= 100 radio's



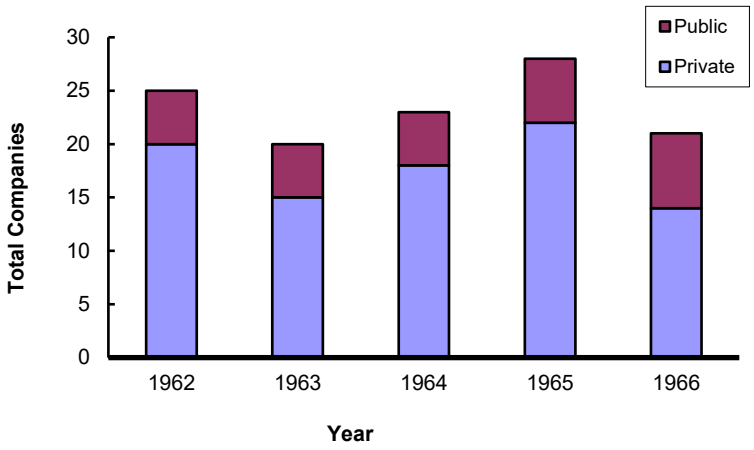
Subdivided or component or composite bar chart

These charts show an aggregate of values and its break up into parts. The bars are drawn proportional in length to the total and divided in the ratios of their components. The components of the bars are hatched differently.

Example: Construct component bar diagram from the following data.

Year	Public Companies	Private Companies	Total
1962	5000	20000	25000
1963	4000	16000	20000
1964	6000	18000	24000
1965	7000	21000	28000
1966	5000	15000	20000

Solution: The data relating to total companies at work are shown below :

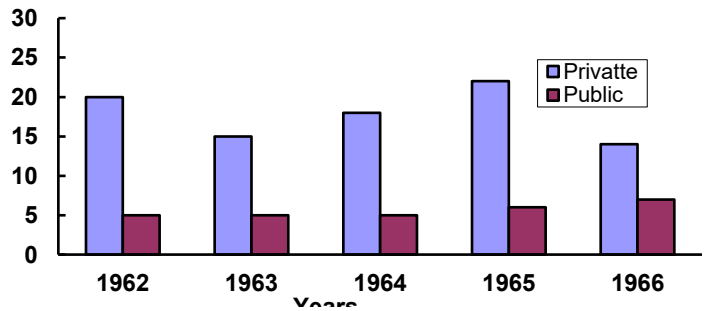


Multiple bar charts (or Compound charts):

Whenever comparisons of more than one variable is to be made at the same time, then multiple bar chart, which groups two or more bar charts together, is made use of. The bars denoting different variables are hatched differently. Comparison of more than one variable at the same time. Such charts also render possible comparison of the same variable over different years. A key to indicate which bars are which is also necessary.

Example: Show the data of above example by multiple bar diagram

Solution:



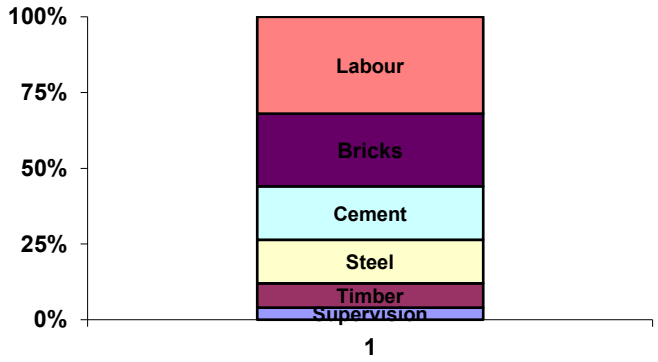
Percentage Bar Diagrams:

A percentage bar chart in which the scale will be a percentage scale and all bars will be of the same weight. They are known as 100% bar charts.

Example: Draw a percentage bar chart for the following data

labour: 25 % Bricks : 15% Cement : 20% Steel : 15% timber: 10% Supervision etc : 15%

Solution: The required diagram is given below:



Deviation bars:

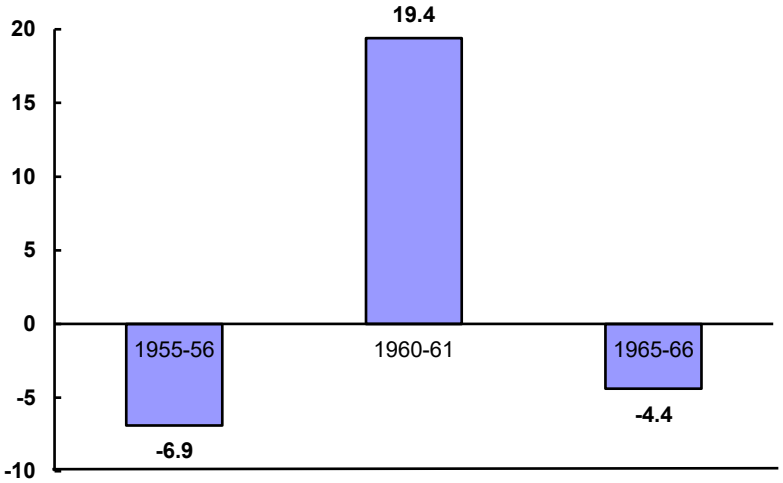
Such bars are employed where net qualities – excess or deficit – are to be represented diagrammatically. They may have both negative and positive values. Positive values are shown above the base line and negative values below the line.

Example:

One Company payments are given below draw the suitable diagram.

Year	**Credit**	**Debit**	**Net (+)**	**(Deviation) (-)**
1955-56	16.2	23.1		6.9
1960-61	46.0	26.6	19.4	
1965-66	47.0	51.4		4.4

Solution: Deviation bar is drawn below:



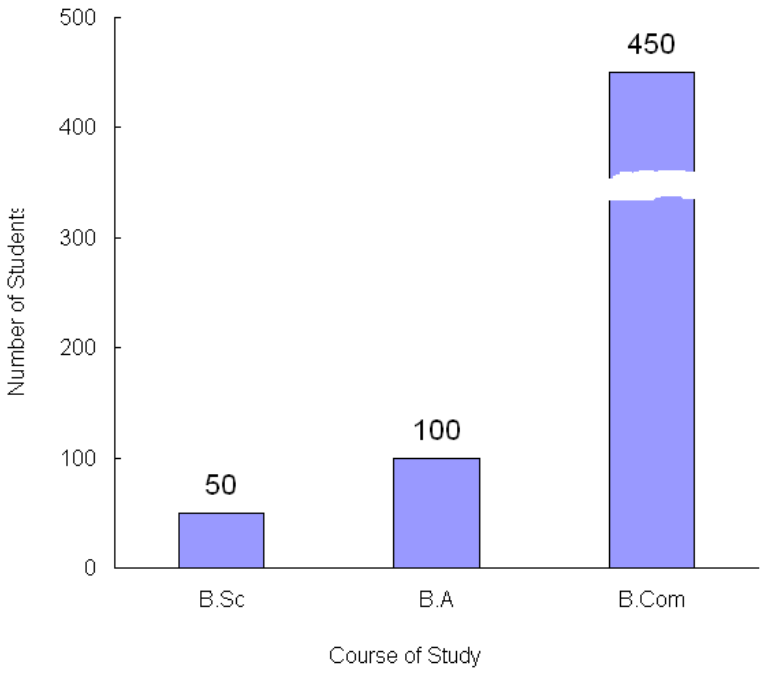
Broken Bars:

Sometimes we may come across certain series in which there may be wide variation in values – some values may be very small and others very large. In such a situation, certain manipulation is necessary. The largest bar / bars may be broken so as to gain space for the smaller bars of the series.

Example: Represent the following data by a suitable diagram

Course of Study	No. of students registered jn a college in 1974-75
B.Com	450
B.A.	100
B.Sc.	50

Solution:



12.4 Two dimensional Diagram (Area Diagrams)

In these diagrams the length as well as the width of the bars is considered. As length and width both are taken into account in such diagrams, they are known as Area diagrams. Following are the three important diagrams under this category.

1. Rectangles 2. Squares and 3. Circles

1. Rectangles: Rectangular diagrams are often used to represent the relative magnitudes of the variable. The area of the rectangles is in proportion to the magnitudes of the values of the variable. These rectangles are often placed side by side with appropriate descriptive titles so that comparison is facilitated.

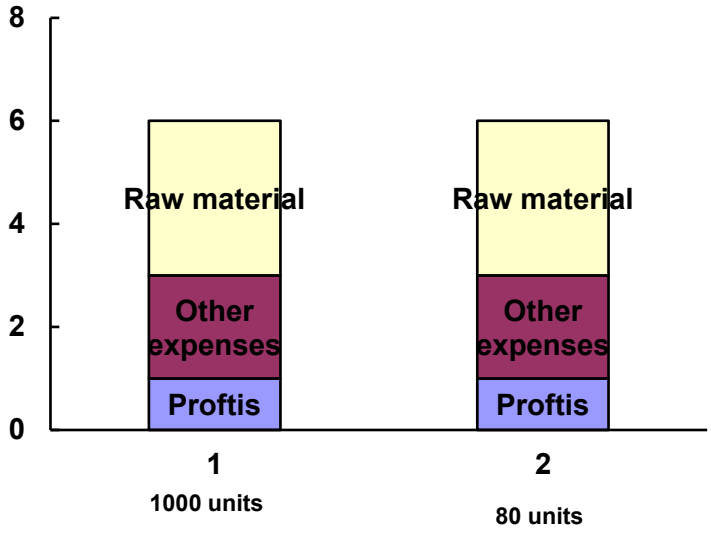
These can be drawn in two ways (i) as the data given (ii) by expressing into percentages.

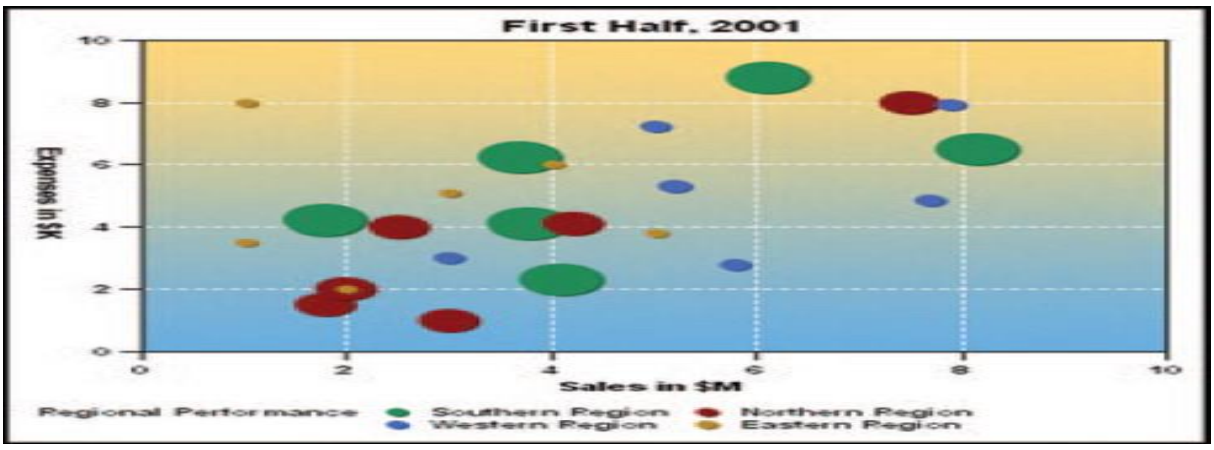
Example: Draw a rectangular diagram to represent the following information.

	Factory A	Factory B
	1000 units	800 units

	Total Rs.	Per unit Rs.	Total Rs.	Per unit Rs.
Value of raw materials	3000	3	2400	3
Other expenses of production	2000	2	1400	1.75
Profits	1000	1	1000	1.25

Solution: The widths of rectangles would be in the ratio 1000 : 800 or 5:4





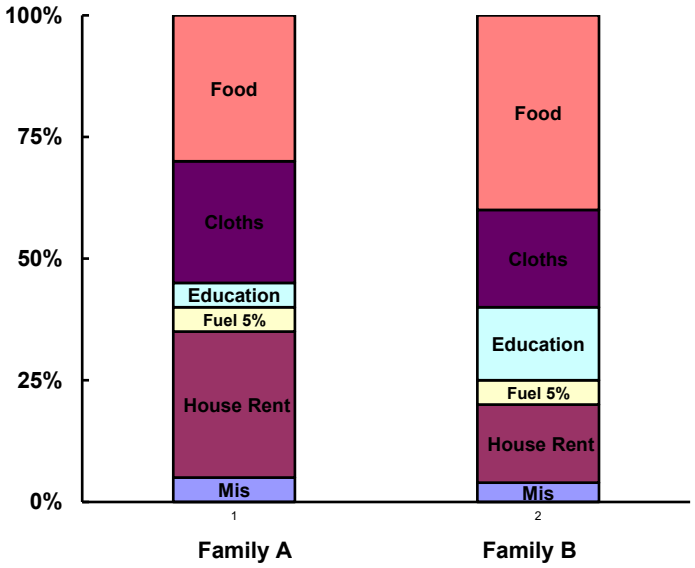
Example: Represent the following data relating to the monthly expenditure of two families A and B by means of a rectangular diagram on a percentage baiss.

Expenditure on	Family A Income Rs. 500	Family B Income Rs. 300
Food	150	120
Clothing	125	60
Education	25	45
Fuel	25	15
House rent	150	48
Miscellaneous	25	12

Solution: converting the given figures into percentages, we get the following table.

Expenditure on	Family A			Family B		
	Rs	%	Cumulative %	Rs.	%	Cumulative %
Food	150	30	30	120	40	40
Clothing	125	25	55	60	20	60
Education	25	5	60	45	15	75
Fuel	25	5	65	15	5	80
House Rent	150	30	95	48	16	96
Miscellaneous	25	5	100	12	4	100

Since the expenditure in family A is Rs. 500 and in family B Rs. 300, the widths of the rectangles would be in the ration of 5:3.



(ii) Squares:

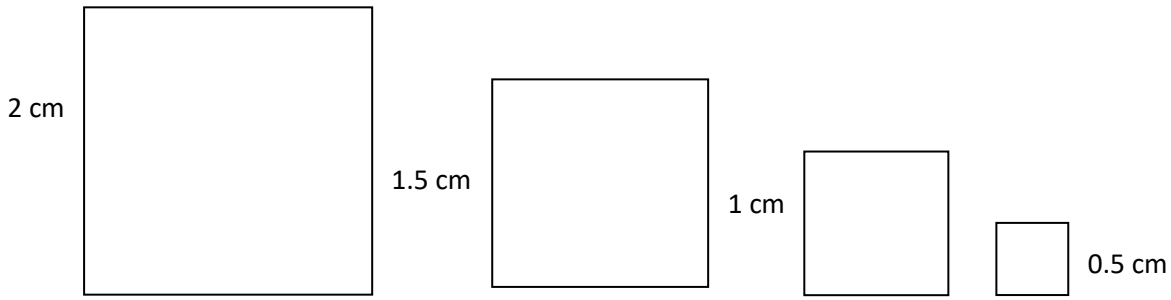
The use of rectangles for diagrammatic representation is not possible where the figures given vary very widely. For instance if we are given income of two families as Rs. 1,440 and 90 the widths of the rectangles would be in the ratio of 16:1. Drawing two rectangles with this ratio would look very odd and unwieldy. Such type of data may easily be represented by the use of square diagrams. In a square, as is known, the length and breadth are equal. On the basis of values of the different series, ratios are first as certained as in the case of rectangles. Then a suitable scale is selected to draw the squares.

Example: Represent the following figures by square diagrams 144, 81, 36, 9

Solution:

Item	**Square root**	**Side of the square in cms.**
144	12	2
81	9	1.5
36	6	1
9	3	1/2

Each figure of the square root has been divided by 6 and the side of the square obtained.



2 cm 1.5 cm 1 cm 0.5 cm

Circles:

Circles may be drawn to represent its area equivalent to the values of a variate. As in the case of rectangular diagrams, both the total and the component parts can be shown. The radii of the circles are proportional to the square roots of magnitudes of the various series. As in the case of square diagrams, square roots are necessary in circle diagrams. But here in circles, the square roots are used for determining the radii of the circles.

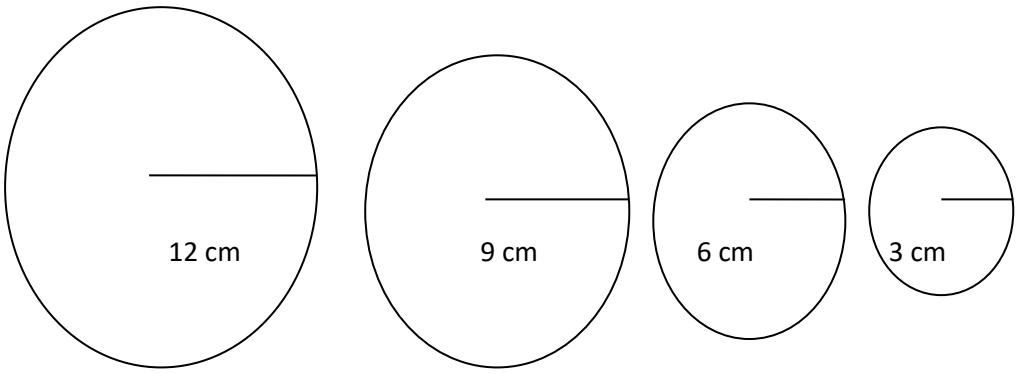
Example: Represent the following figures by circle diagrams 144, 81, 36, 9

Solution: We shall find out square roots of these figures which we will use as radii of different circles or squares.

x : 144 81 36 9

$\sqrt{x}$: 12 9 6 3

Four circle with different radii



Pie Charts or circular charts:

Like the component bar charts, they are used when relationship of the parts is shown to one another and to the total. However, unlike bar charts, in the case of pie-charts, the length of bars are not compared but areas of different segments of the circles are compared.

Example : construct a circular graph from the following information derived from the profit and loss account of companies

Item	Amount spent %
Supplies	46.5
Employees	31.0
Taxes	11.0
Depreciation	3.5
Shareholders	5.0
Miscellaneous	3.0
Total	100.00

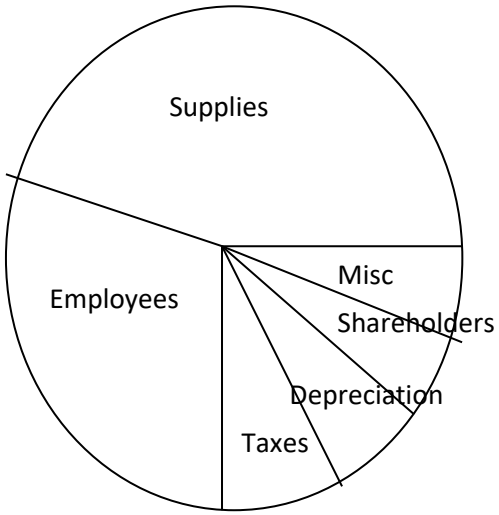
Solution: Obtain the angles i.e,

$\frac{46.5}{100} X 3600 = 167.4^0$; $\frac{31.0}{100} X 360 = 111.6^0$

$\frac{46.5}{100} X 360 = 39.60^0$; $\frac{3.5}{100} X 360 = 12.6^0$

$\frac{5}{100} X 360 = 98.0^0$; $\frac{3.0}{100} X 360 = 10.8^0$

Now draw a circle and radius as shown below.



Limitations:

1. From the point of view of the statistician, they are not of much help in analyzing deata.

2. They do not provide quantitative information, though they give a qualitative appreciation of a

set of data.

Histogram:

A Histogram is a set of rectangles with bases along the intervals between class boundaries and with areas proportional to the frequencies in the corresponding classes. If the class intervals are equal, the heights of the rectangles are also proportional to the frequencies. If the class intervals are unequal the rectangles have unequal bases and their heights are adjusted accordingly. It is area here, and not height, that represents frequency.

Note: In histogram the rectangles are all adjacent, since the bases cover the intervals between class boundaries, not class limits. In a bar diagram, on the other hand the spacing and width of the bars are arbitrary and it is only the height that count.

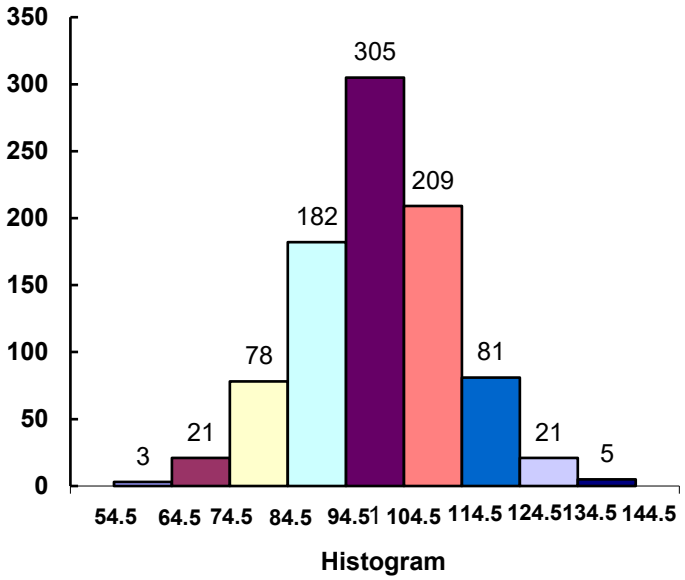
Example: Draw Histogram for the following data.

Class interval	**Frequency**
55 - 64	3
65 - 74	21

75 - 84	78
85 – 94	182
95 – 104	305
105 – 114	209
115 – 124	81
125 – 134	21
135 – 144	5

Solutions: Given data is inclusive class intervals that can be recalculated to draw histogram is as follows

Class interval	**Frequency**
54.5 – 64.5	3
64.5 – 74.5	21
74.5 – 84.5	78
84.5 – 94.5	182
94.5 – 104.5	305
104.5 – 114.5	209
114.5 – 124.5	81
124.5 – 134.5	21
134.5 – 144.5	5

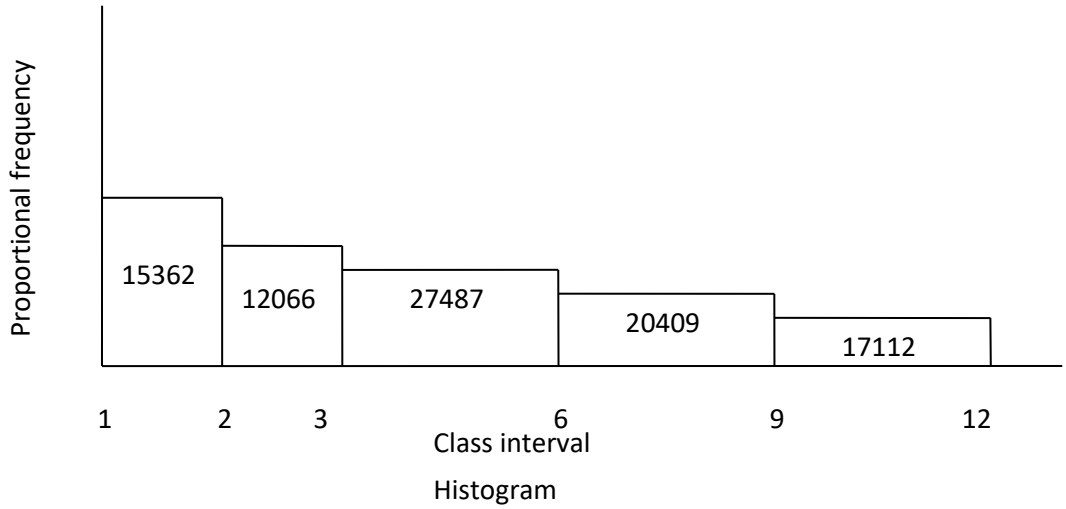


Example: Draw the histogram for the following data

Class interval	frequencies
1 – 2	15,362
2 – 3	12,066
3 – 6	27,487
6 – 9	20,409
9 – 12	17,112

Solution: The revised table will be shown below

Class interval	actual frequencies	proportional frequencies
1 – 2	15,362	15,362
2 – 3	12,066	12,066
3 – 6	27,487	27,487/3 =9162
6 – 9	20,409	20,409/3 =6803
9 – 12	17,112	17,112/3 =5704



Histogram

Advantages of Histograms

- Visual Clarity: Histograms provide a clear visual summary of data distribution.
- Identification of Patterns: They help identify patterns, such as skewness, modality, and outliers.
- Data Analysis: Useful in statistical analysis for understanding data behavior.

Limitations of Histograms

- Loss of Detail: Aggregating data into bins can result in a loss of individual data points.
- Sensitivity to Bin Size: The appearance of the histogram can change significantly with different bin sizes.
- Not Suitable for Categorical Data: Histograms are meant for numerical data and not suitable for categorical data.

Histograms are powerful tools for exploratory data analysis, allowing you to quickly assess the distribution and characteristics of your data.

Rules for Drawing Diagram

Like the preparation of tables, the drawing of diagram has also certain rules to be observed.

1. Selection of a proper scale (selection of a scale is a matter of personal judgement, it should be appropriate to the situation);
2. The vertical and horizontal axes should be properly designed;
3. The diagram should be drawn with the aid of geometrical instruments and neatness should have top priority;
4. The heading should be written on the top of these diagram in bold letter;
5. Distinguish different colours.

Summary :

Concepts of classification, classification of data, types of classification. Construction of discrete and continuous frequency distributions, class selection, cumulative and relative frequency.

Self Assessment Questions

Lesson -10

HEALTHCARE ANALYTICS & REPORT WRITING

Structure of the Lesson

1. Introduction to healthcare analytics
2. Types and utilization of healthcare data
3. Sources of health statistics
4. Problems in collecting sickness data
5. Measurement of sickness
6. Vital statistics
7. Basics of report writing

Introduction

In Units I to IV we have learnt that information is essential to assess scientifically the performance of any organization. Hospital Administration is no exception in this regard. Through lessons 1 to 19 a number of methodologies are important to collect classifying, analyze and interpret any data. With reference to Hospital Administration also these techniques can be adopted provided the data relating hospitals and its associated wings are available. In this lesson and next lesson our objective is to describe the origin and utilization of health and hospital statistics.

A hospital being a mega organization either in public sector or corporate world its administration needs a special focus also parallelly handling the managerial and medical activities. Any hospital worth mentioning its affective functioning it should maintain a perfect documentation that provides information on health and hospital statistics.

Basic Data: it talks about basic data or information about healthcare institution and its infrastructure and about patients, types of problems etc. it includes.

- Hospital Beds

- No. of Doctors
- No. of Nurses
- No. of Supporting & Administrative staff
- No. of In Patients
- No. of out patients
- Different Departments includes clinical and Non clinical
- Types of Diseases
- Bed Occupation Ratio
- No. of Deliveries during a year
- Mortality rate
- Infant mortality
- No. of Deaths
- Infrastructure
- No. of Surgeries etc.
- Medical errors
- MLC cases
- Violation of statutory rules
- Etc.

Utilization of the Basic Data

A hospital is supposed to maintain a medical record system that provides the data source. The documented records of a hospital would provide the required information. The purpose of a medical record is to provide information as and when the data is required. The basic principles of a medical documentation mechanism are clear understanding of the organization of structure, medical staff, hospital administration, medical record librarian. Among these the medical record librarian is the custodian of medical data. He has to monitor development, analysis and technical evaluation of clinical records, preservation of records, development of secondary records. The various formats of preserving data in medical records

is dual records, unit numbering, modification of unit numbering, serial numbering, annual numbering, grouping of digits, terminal digit filing, colour coded folders, straight numerical filing, departments attendance, hospital attendance, the three department system namely out patients, inpatients and the casualty departments. All the data available in this record can be effectively utilized for rating and better administration of the organization with the help of a statistician.

Sources of Health Statistics

Health Statistics primarily comes under a service sector provided by hospitals, primary health centers. Hence health statistics can be obtained from these two sources generally. Either of these organizations shall have well defined functions and systems of their own. In the case of hospitals usually six stages are involved for its functioning with respect to an admitted inpatient. These are (1) Admission (2) Diagnosis (3) Treatment (4) Inspection (5) Control (6) Discharge. Apart from this a hospital will have three components namely, (1) Medical staff (2) Testing Laboratories (3) Support and Administrative Services for general maintenance. The main functions of a hospital are (1) Patient Care (2) Environmental System (3) Social System (5) Cultural Relationships (5) Physician Training (6) Community Health Care (7) Growth and Prosperity of the institution and staff. Other possible data sources are Current population survey, medicare current beneficially survey, national health care survey, and national hospital discharge survey.

Cure is a result of professional and technical quality. Care is an outcome experience of the function of quality of the system. A hospital has to do internal marketing required most by the medical staff. In all these cases the records maintained by the hospital would provide the required data. Each hospital has to device its own proformas of documents that are to be filled and maintained at each and every phase of activity that takes place in a sequential order in the hospital. The moment a patient enters for medical service he or she has to give various particulars to the hospital staff at various stages. The information so collected shall be documented by the respective personal of the hospital. It forms a data base for health statistics. Some of the proformas that form the data base are

- Patient history and physical reports
- Clinic/Office notes

- Operation reports
- Consultation notes
- Discharge summary, Psychiatric reviews
- Lab reports
- X-ray reports
- Pathology reports
- Other investigations

All these proformas should be invariably available with the hospital documentation staff. The information if necessary can be had from these sources of hospital statistics.

From the point of view of a consumer preference of a hospital for medical services is generally decided on the basis of published performance indicator of the hospital computation of these indicators is based on the service data of the hospital. In any way one should require health/hospital statistics. Accordingly collection of hospital statistics is essential in the decision making process. It goes without saying that a data can be obtained from its source of availability. The above are general sources of hospital statistics.

Other sources

CBHI : Central bureau of Health intelligence

NHP : National Health Policy

WHO : World Health organization

ICMR : Indian council of Medical Research

SRS : Statistical report

NSSO : National sample survey organization

NSDUH: National survey on Drug use Health

NHI's : National Health Interviews

CDC : Center for Disease Control and Prevention

Director of Health and family welfare

Census of India

National commission on population

National institution for transforming India

Director of Economics & statistics of respective state Government

Household surveys, Vital records and Administrative reports

Problems in Collection of Sickness Data

The very word sickness some physical inability by the individual with respect to some activity. If the extend of sickness requires a consultancy and cure the help of a medical professional is essential. The information on the types of sickness their professional is essential. The information on the types of sickness their intensity process of cure, consultancy offered, etc., will form sickness data. Like any data collection, sickness data collection also will have its associated problems. The data will be usually as recorded by the diagnostics equipments and procedures. The extent of sickness is known only through an examination and what the clinical test shows. The real extent to which the patient is suffering may not be exactly known different from what is shown by the test. The truth of the data lies with the efficiency of the test which is always questionable because of possible non calibration. For any empirical study if the data on sickness is needed there is no guarantee that the incharge personal of the data source would provide the data. They will have a usual fear that the data given by them would put them to troubles by their employers when they are misused. Therefore a researcher in hospital administration should exhibit on undoubted impression from him that the data would not be misused.

On the other hand in the case of data on out patient which are available only during the time of their registration the researcher would be disappointed for not being listened to by the patients. To over come such situations the researcher may have to interact with the attendants of the patients and get the information in a soft way rather than with a fighting tendency. More over if the data is needed for some past period non availability of records for

that particular period is another bottle neck. In such situations proxy on the missing data variables can be used to generate the missing observations into the data. All these come the missing observations into the data. All these come under sampling problems of sickness data collection. Even if it is assume that the sampling errors such as identifying graphs and charts for the data calculation of statistical constants, interpretation of results all of which are inter dependent. In spite of these problems a researcher should take up a research study in hospital administration as a sincere issue and has to continue his research progress.

Measurement of Sickness

Sickness is a symptom or an experience of an individual in his body. A physician with his technical expertise conforms the intensity of the sickness through a clinical examination process by making the patient undergoing various test procedures. The reports of the clinical examination generally contained the observed parametric values of the patient relevant to a particular sickness as referred by the physician. The parametric values would indicate the level of sickness present in the patient's body and the theoretical ranges to be normally permissible for a non sick individual. Thus the level of sickness indicated in a clinical trail can be defined as measure of sickness and the observation is the corresponding measurement of sickness. In any typical clinical examination a prescribed proforma would be followed to measure the sickness of the patient. The following are some such proforma that would be useful in the measurement of sickness.

SURNAME (capitals)	FORENAME(S)	X-RAY Dept.	X-Ray No.
Mr.			
Mrs.			
Miss.			
Reg. No. / Address	Date of Birth	HOSPITAL	
		Report Required By:	Ward/Dept.
		Consultant:	
Tel. No.		G.P.:	

Previous	X Rays	Examinations Required
	Date	
	Hospital	
10 DAY RULE	**Ignore / Observe ***	APPOINTMENT DATE
DATE of L.M.P. ……		
* delete as required		
Clinical History and Diagnosis		
Signature	Date	Radiographer's Initials
JB – 83866		
		Code ABSP

Fig: Specimen of an X-ray request form

SURNAME (Capitals) Forename(s)	> > G.P.>
	> Other (specify)
Address	Sex Ward CLINICIAN/G.P Signature
Unit No. Date of Birth	
	Date collected Time collected Lab. No.
CLINICAL SUMMARY	INVESTIGATION REQUIRED
FOR LABORATORY USE ONLY	
Date received Date of report Signature Haematology Department	MISCELLANEOUS
A10150 JB-82690	HAEMATOLOGY

Fig: Request form for investigation by Pathology Laboratory

SURNAME (Capitals) Forename(s) Address Unit No. Date of Birth	> > G.P.> > Other (specify) Sex Ward CLINICIAN/G.P Signature Date collected Time collected Lab. No.
CLINICAL SUMMARY	BLOOD REQUIRED FOR: Put 3 a) Group and retain serum (maximum 5 days) b) Urgent transfusion TIME ……. c) Transfusion DATE……. (Bllod/packed cells)……UNITS……
To be completed by clinician Blood group if known ABO …Rh….Antibodies YES/NO Previous transfusion YES/NO Date……. All previous pregnancies YES/NO Dates……. ……. Miscarriages. Early YES/NO Late YES/NO Neonatal jaundice YES/NO	1) 10 ML of clotted blood FULLY labelled with patients forenames. Surname, date of birth. Hospital number (or home address) must Accompany this request. 2) A minimum of 24 HRS notice is required for all routine grouping and cross matching. 3) Blood will be kept for 24 hours ONLY after indicated time of transfusion.
FOR LABORATORY USE ONLY Group ABO Rhesus A typical antibodies Date Signature	

BLOOD CROSSMATCHED **METHODS OF COMPATIBILITY TESTING**

Date	Test No.	Blood Unit No.	ABO	Rh	Sal 20^{0}C	Sal 37^{0}C	Enz	ALB	Coombs	Comments	Fate	Signature

Date received	Date of report	Signature Haematology Department	MISCELLANEOUS
A10150 JB-82690			HAEMATOLOGY

Fig: Request form for investigation by Pathology Laboratory

Phase Use a Ball-point pen	SURNAME Forename(s) Reg. No/Address Date of birth	> > G.P.> > Other (specify) Sex Ward / CLINICIAN / Signature Date collected / Time collected / Lab No.		
BLOOD SAMPLES COPY TICK PRIORITY TEST(S) TO COVER SMALL SAMPLES OR MACHINE FAILURE	CLINICAL SUMMARY including ALL DRUGS please			
	SMA PROFILE Prior arrangement must be made for Urgent Requests			
	Electrolytes Urea> Calcium Phosphate> Creatinine> Proteins>			
A 107SO xxxxxxx (Oct 78) JB - 86832	Date received	Date of report	Signature Pathology Service	CLINICAL CHEMISTRY/SMA

- **Vital Records:** Vital records provides most detailed data on Health conditions because they record the causes of death and the circumstances of birth.
    - It also records the estimated birth rate, death rate and natural growth rate in the country.
    - Estimated birth rate, declined from 25.8 in 2000 to 20.4 in 2016, while the death rate declined from 8.5 to 6.4 per 1000 population over the same period.

Conclusion:

Self Assessment Questions