PREFACE

With its grounding in the "guiding pillars of Access, Equity, Equality, Affordability and Accountability," the New Education Policy (NEP 2020) envisions flexible curricular structures and creative combinations for studies across disciplines. Accordingly, the UGC has revised the CBCS with a new Curriculum and Credit Framework for Undergraduate Programmes (CCFUP) to further empower the flexible choice based credit system with a multidisciplinary approach and multiple/ lateral entry-exit options. It is held that this entire exercise shall leverage the potential of higher education in three-fold ways – learner's personal enlightenment; her/his constructive public engagement; productive social contribution. Cumulatively therefore, all academic endeavours taken up under the NEP 2020 framework are aimed at synergising individual attainments towards the enhancement of our national goals.

In this epochal moment of a paradigmatic transformation in the higher education scenario, the role of an Open University is crucial, not just in terms of improving the Gross Enrolment Ratio (GER) but also in upholding the qualitative parameters. It is time to acknowledge that the implementation of the National Higher Education Qualifications Framework (NHEQF), National Credit Framework (NCrF) and its syncing with the National Skills Qualification Framework (NSQF) are best optimised in the arena of Open and Distance Learning that is truly seamless in its horizons. As one of the largest Open Universities in Eastern India that has been accredited with 'A' grade by NAAC in 2021, has ranked second among Open Universities in the NIRF in 2024, and attained the much required UGC 12B status, Netaji Subhas Open University is committed to both quantity and quality in its mission to spread higher education. It was therefore imperative upon us to embrace NEP 2020, bring in dynamic revisions to our Undergraduate syllabi, and formulate these Self Learning Materials anew. Our new offering is synchronised with the CCFUP in integrating domain specific knowledge with multidisciplinary fields, honing of skills that are relevant to each domain, enhancement of abilities, and of course deep-diving into Indian Knowledge Systems.

Self Learning Materials (SLM's) are the mainstay of Student Support Services (SSS) of an Open University. It is with a futuristic thought that we now offer our learners the choice of print or e-slm's. From our mandate of offering quality higher education in the mother tongue, and from the logistic viewpoint of balancing scholastic needs, we strive to bring out learning materials in Bengali and English. All our faculty members are constantly engaged in this academic exercise that combines subject specific academic research with educational pedagogy. We are privileged in that the expertise of academics across institutions on a national level also comes together to augment our own faculty strength in developing these learning materials. We look forward to proactive feedback from all stakeholders whose participatory zeal in the teaching-learning process based on these study materials will enable us to only get better. On the whole it has been a very challenging task, and I congratulate everyone in the preparation of these SLM's.

I wish the venture all success.

Professor (Dr.) Indrajit Lahiri
Vice-Chancellor

Netaji Subhas Open University

Four Year Undergraduate Degree Programme

Under National Higher Education Qualifications Framework (NHEQF) & Curriculum and Credit Framework for Undergraduate Programmes

Bachelor of Arts (Honours) in Sociology

Course Type : Skill Enhancement Course

Programme Code : SEC

Course Title : Quantitative Research Methods in Social Science

Course Code: NSE-SO-01

First Print : April, 2025

Printed in accordance with the regulations of the University Grant Commission – Distance Education Bureau

Netaji Subhas Open University

Four Year Undergraduate Degree Programme

Under National Higher Education Qualifications Framework (NHEQF) & Curriculum and Credit Framework for Undergraduate Programmes

Bachelor of Arts (Honours) in Sociology

Course Type : Skill Enhancement Course

Programme Code : SEC

Course Title : Quantitative Research Methods in Social Science

Course Code: NSE-SO-01

: Course Writers:

- **Units 1 & 2** : **Maitrayee Adak,** *MA, NET.*
- **Unit 3** : **Dr. Subhrajit Chatterjee**, A*ssistant Professor & HOD, Sociology, Kashipur Michael Madhusudan Mahavidyalaya, Purulia*
- **Units 4 & 5** : **Dr Esha Chatterjee,** *Assistant Professor of Sociology, Faculty of Arts and Humanities, Kalinga University, Raipur*
- **Units 6 & 7** : **Dr. Sathi Naik,** S*ACT – 1, Department of Sociology, Women's Christian College, Kolkata*

- **Units 8-10** : **Dr. Soumitra Sarkar,** *SACT-I, Department of Sociology, Kabi Sukanta Mahavidyalaya, Bhadreswar, Hooghly*
- **Units 11 & 12** : **Dr. Arunima Dey,** *Assistant Professor, Department of Social Work, Centre for Distance and Online Education, Rabindra Bharati University*
- **Unit 13** : **Dr. Reshma Anjum,** *Associate Professor at Vidyasagar School of Social Work, Affiliated to Vidyasagar University*
- **Unit 14** : **Ranajit Saha,** *Ph.D. Scholar and Junior Research Fellow (JRF) at West Bengal State University, Department of Sociology*
- **Unit 15** : **Dr. Tapas Pal,** *Post-Doctoral Fellow, Eudoxia Research University, USA.*

: Course Editor :

Dr. Srabanti Choudhuri
Assistant Professor, Discipline of Sociology, NSOU

: Format Editor :

Dr. Srabanti Choudhuri
Assistant Professor, Discipline of Sociology, NSOU

Notification

Ananya Mitra
Registrar (Add'I Charge)

Netaji Subhas Open University

UG : Skill Enhancement Course (SEC)

Course Title : Quantitative Research Methods in Social Science
Course Code: NSE-SO-01

MODULE - I

Introducing Research in Social Sciences

Unit 1 □ Introduction to Social Research

Structure

1.1 Learning Objectives

- To understand the concept of social research .
- To know the importance of social research.
- To study various definitions of research and know its types.
- To know briefly about research methods and methodology.
- To briefly study different scopes for research.
- To connect the differences and similarities between natural science and social science research.

1.2 Introduction

In every corner of human society, research plays a pivotal role in highlighting profound insights. Research fundamentally deals with discovering something new that adds the

knowledge in the field. Research provides new facts and verifies information as well as questions the unknowable facts. Social research systematically explores social happenings. It replaces the old knowledge by uncovering the new ones. It essentially unfolds the complexities of human life within social structures and cultural changes. Social Science research tries to examine human behavior in every aspect of life. The prime motive of sociological research is to find new evidence on interaction, relationships which shape social groups and social institutions.

Finding the truth that lies hidden and unexplained is the primary goal of the inquiry (Kothari, 2004). According to Kothari (2004), research seeks to gain new viewpoints on it. Research accurately represents the characteristics of a specific person, circumstance, or group (Kothari, 2004). According to Kothari (2004), it assesses variables and their connections to one another. Sociological research not only brings social change but also resist changes which might be harmful for society. Social Research in social science is done through the following three research methods: Qualitative Research, Quantitative Research and Mixed methods. Qualitative Research is a non-numerical research that goes deeper into the meaning of the data collection on a social phenomena. Quantitative Research is a numerical research which deals with the statistical packages and highlights the objective data more than the subjective one. On the other hand, mixed method research forms a triangulation and integrates quantitative and qualitative research to collect data that will lead to a thorough result. As stated by Kothari (2004), a good undertaking for research is always methodical, rational, empirical, and repeatable.

Do you know?

Research is different from investigation. Social research is a systematic study about social issues by understanding human behaviour and their interactions. But investigation is about gathering statements, information, evidence about the suspect.

1.3 Definitions of Research by Different Scholars

The encyclopedic Oxford English Dictionary defines "Social Research as the systematic investigation into the study of materials, sources etc. in order to establish facts and reach new conclusions; an endeavor to discover new or collate old facts etc. by the scientific study of a subject or by a course of critical investigation" (Walliman 2006, p-14).

Did you know?

A research always helps you to investigate on small details of a topic systematically to reach conclusion.

Leedy (1989) defines social research as "a procedure by which we attempt to find systematically, and with the support of demonstrable fact, the answer to a question or the resolution of a problem" (Walliman 2006, p.-14) . Kerlinger (1970) defined social research as "the systematic, controlled, empirical and critical investigation of hypothetical propositions about presumed relations among natural phenomena" (Walliman 2006, p.-14).

1.3.1 Types of Research

- Descriptive versus Analytical: Descriptive research encompasses many forms of fact-finding inquiries and surveys (Kothari, 2004). The primary goal of descriptive research is to describe the current state of circumstances (Kothari, 2004). The term Ex post facto research is used for descriptive research studies (Kothari, 2004).The primary feature of this approach is that the researcher can only describe what has taken place or is going on; he has no control over the variables (Kothari, 2004). On the other hand, in analytical research, the researcher must employ pre-existing facts or information and analyze it in order to critically assess the content (Kothari, 2004).
- Fundamental vs Applied. Another name for basic research is fundamental research (Kothari, 2004). Its goal is to increase our comprehension of several issues that frequently arise in social contexts and how to address them (Kothari, 2004). It is carried out exclusively to improve our basic and broadly useful knowledge (Kothari, 2004). There may be no intended or immediate purpose for this kind of study (Kothari, 2004). It is "Learning for the sake of learning" (Kothari, 2004). When there is a societal issue that needs to be resolved, applied research is conducted (Kothari, 2004). Its main goal is to provide valuable and practical answers to questions (Kothari, 2004). The outcomes are used in practice to address current issues (Kothari, 2004). Normative prescription is involved. Knowledge that can be used right away is the focus of applied research (Kothari, 2004).
- Quantitative vs Qualitative. The process of gathering and evaluating numerical data is known as quantitative research (Kothari, 2004). It can be applied to evaluate causal linkages, make predictions, identify trends and averages, and

extrapolate findings to larger populations (Kothari, 2004). By producing numerical data or data that may be converted into useful statistics, quantitative research is utilized to quantify the issue (Kothari, 2004). In addition to generalizing findings from a broader sample population, it is used to measure attitudes, views, actions, and other defined characteristics (Kothari, 2004). On the contrary, qualitative study focuses on qualitative phenomena, or those that have to do with quality or kind (Kothari, 2004). For example, "motivation research," a significant subset of qualitative research, is frequently discussed when examining the causes of human behavior, or why individuals think or act in particular ways (Kothari, 2004). Through in-depth interviews, this kind of research seeks to uncover the underlying desires and motivations (Kothari, 2004). Qualitative research also includes attitude or opinion research, which is research aimed at determining people's feelings or thoughts on a specific topic or organization (Kothari, 2004).

- Comparing Conceptual and Empirical Research. Conceptual research is concerned with abstract concepts or theories (Kothari, 2004). Philosophers and thinkers typically utilize it to create new ideas or reinterpret pre-existing ones (Kothari, 2004). Conversely, empirical research frequently disregards systems and theory in favor of relying solely on experience or observation (Kothari, 2004). It is data-driven research that generates results that can be confirmed through experimentation or observation (Kothari, 2004).
- Some other types of research. Like longitudinal study or one-time research (Kothari, 2004). The research is conducted throughout multiple time periods in the latter situation,

(Educba, Pedamkar, 2023)

While it is limited to a single time period in the former (Kothari, 2004). Whereas, clinical or diagnostic research determines the fundamental causal relationships, such research use case-study techniques or in-depth methodologies (Kothari, 2004). These studies typically use very tiny samples and highly penetrating data collection tools to delve deeply into the causes of objects or events that interest us (Kothari, 2004). Historical research is the study of previous events or concepts, including the philosophical beliefs of individuals and communities at any distant point in time, using historical sources such as records, artifacts, etc (Kothari, 2004). On the other hand the exploratory research view on formalised research projects are those with a significant framework and particular hypotheses to be evaluated, exploratory research aims to develop hypotheses rather than test them (Kothari, 2004).

1.3.2 Research Methods Vs. Research Methodology

The behavior and tools we employ when conducting research operations, such as making observations, documenting data, and data processing methods, are referred to as Research techniques (Kothari, 2004). The actions and tools involved in choosing and developing a research technique are referred to as Research Methods (Kothari, 2004). A methodical approach to addressing the research topic is known as research methodology (Kothari, 2004). It could be viewed as a science that studies scientific research methods (Kothari, 2004). In it, we examine the many approaches that a researcher typically takes to investigate his research problem and the reasoning behind them (Kothari, 2004). The researcher must be knowledgeable about both the approach and the research methods (Kothari, 2004).

There are several methods of research methods: (i) observation method, (ii) interview method, (iii) through questionnaires, (iv) through schedules, and (v) other methods which include: warranty cards, distributor audits, using mechanical devices, through projective techniques, depth interviews, and content analysis (Kothari, 2004). When using the observation method, the investigator directly observes the situation to get information without consulting the respondent (Kothari, 2004). The interview method of data collection entails presenting oral-verbal information and receiving responses in the form of oral-verbal responses (Kothari, 2004). Personal interviews and, if feasible, telephone interviews can be employed to implement this strategy (Kothari, 2004). A questionnaire is made up of several questions that are typed or printed on a form or series of forms in a specific order (Kothari, 2004). Respondents get the questionnaire via mail, and it is assumed that they will read it, comprehend its questions,

and write their answers in the designated space on the questionnaire (Kothari, 2004). The interviewer or research worker typically fills out the schedule and, if needed, can interpret questions (Kothari, 2004).

1.4 Importance of Social Research

Importance of doing Social research are as follows:

1. Understanding social issues at their most basic level and working toward a speedy and effective solution are the goals of social research.
2. Social research creates new theories and concepts to advance society.
3. Social research expands human knowledge by dispelling outdated notions that are no longer relevant and identifying novel ones.
4. Social research contributes to the betterment of society and the general well-being of people.
5. In addition to providing answers for facts that cannot be explained, social research facilitates the transition from the known to the unknown.

1.5 Scope of Social Research

The scope of social research includes:

1. Social research helps in problem investigation by understanding its causes and consequences it may uphold in future.
2. Research evaluates the social policies and tries to understand the effectiveness of it.
3. Social research tests certain theories and then builds a new one based on the evidence gathered.
4. Social research creates a structural social change in society.
5. Scope for social research is vast and wide and hence it understands human society and its totality.

1.6 Difference and Similarity between Scientific Research and Social Science Research

Dissimilarities and Commonalities, the Intersection of Social Science and Scientific Research. The two main categories of science are the social sciences and the natural sciences (Ogunbemeru, 2009). Astronomy, biology, chemistry, geology, and physics are examples of natural sciences, whereas sociology, psychology, economics, political science, and anthropology are examples of social sciences (Magaji and Sa'eed, 2017). These sciences have given rise to a number of specialized subfields over time, and scholars contend that no scientific discipline can live in total isolation (Reidpath et al., 2011). There are several relationships between natural sciences, social sciences, and other scientific fields, showing that information in one area frequently influences and enhances information in another (Author, 2011). Social science research and scientific research (natural sciences) both adhere to methodical approaches to investigation and have basic scientific traits (Temitola, 2015). They vary, nevertheless, in their applications, approaches, and areas of interest (Magaji and Sa'eed, 2017). To appreciate how different but connected these professions are, it is essential to comprehend these contrasts and commonalities (Magaji and Sa'eed, 2017).

Despite these distinctions, there are a number of important similarities between social science and scientific (natural sciences) research (Temitola, 2015):

1. Reasonability and objectivity Research in the social and scientific sciences aims for objectivity by drawing findings from empirical evidence rather than subjective judgments or assumptions (Magaji and Sa'eed, 2017). Their methodical approach to investigation guarantees that research findings are logically connected to supporting data (Magaji and Sa'eed, 2017).
2. Using Empirical Approaches Empirical methods—knowledge acquired from observations, experiments, or real-world data—are emphasized in both kinds of study (Magaji and Sa'eed, 2017). The social sciences employ surveys, case studies, and interviews to collect empirical data, whereas the natural sciences depend on experiments and physical measurements (Magaji and Sa'eed, 2017).
3. Methodology and Systematic Arrangement Research in both domains must be conducted methodically (Magaji and Sa'eed, 2017). Researchers in the social sciences and scientists in general use systematic approaches like: Developing a theory and creating a study or experiment. Gathering information via surveys, tests, or observations (Magaji and Sa'eed, 2017). Using statistical or qualitative

techniques to analyze the findings (Magaji and Sa'eed, 2017). Making evidence-based decisions and publishing results for review by other researchers and additional confirmation Credibility, dependability, and transparency are guaranteed in both natural and social science research by this methodical procedure (Magaji and Sa'eed, 2017).

4. Testability & Verifiability Verifiability by other researchers is a prerequisite for both scientific and social science research (Magaji and Sa'eed, 2017).This implies that in order to verify the correctness and dependability of the results, they must be examined by colleagues or subject matter experts (Magaji and Sa'eed, 2017). In a similar vein, testability is a key feature of both disciplines (Magaji and Sa'eed, 2017). A research assertion needs to be verifiable by data collecting, experimentation, or observation (Magaji and Sa'eed, 2017). A hypothesis is not considered scientific or social science study if it cannot be tested (Magaji and Sa'eed, 2017).
5. Applying Statistical Analysis and Data for both fields to produce insightful findings, data gathering and analysis are essential (Magaji and Sa'eed, 2017). The social sciences employ both quantitative and qualitative data, including survey results, interviews, and historical documents, whereas the natural sciences rely on quantitative data, such as numerical measurements, chemical compositions, and biological activities (Magaji and Sa'eed, 2017). Both disciplines frequently use statistical methods to interpret data and determine relationships between variables, including regression analysis, hypothesis testing, correlation, and probability models (Magaji and Sa'eed, 2017).
6. Knowledge and Society Contribution to research in the social and scientific sciences both advance human understanding and enhance society (Magaji and Sa'eed, 2017). While social sciences have an impact on education, governance, social development, and policy, natural sciences drive advances in engineering, technology, medicine, and environmental protection (Magaji and Sa'eed, 2017). Research in sociology and psychology, for instance, aids in addressing problems like mental health, crime prevention, and educational reform, while medical research in biology and chemistry results in the development of new medicines and therapies (Magaji and Sa'eed, 2017).
7. Moral Aspects / Ethical considerations: When performing research, both disciplines must abide by ethical standards (Magaji and Sa'eed, 2017). Environmental effects, responsible technological usage, and experiment safety procedures are among the ethical issues in the natural sciences (Magaji and Sa'eed, 2017).

Ethics in the social sciences center on participant safety, informed consent, confidentiality, and objective reporting of results (Magaji and Sa'eed, 2017). For instance, in order to safeguard participants from psychological injury, psychological research involving human beings must adhere to stringent ethical standards (Magaji and Sa'eed, 2017). Clinical trials in medical research must also adhere to ethical guidelines to guarantee that therapies are evaluated safely before being taken to patients (Magaji and Sa'eed, 2017).

1.7 Conclusion

Social research is essential to understanding human civilization since it looks at relationships, institutions, and interactions. It promotes social growth through addressing societal concerns, developing new theories, and assessing policies. Social research guarantees reliable information production through the use of methodical methodologies, supporting both academic research and real-world social developments. The main areas of difference between social science and natural science scientific research include topic matter, methodology, objectivity, and predictability. In order to maintain objectivity and reproducibility, the natural sciences use quantitative, experimental, and strictly regulated methodologies to study physical and biological phenomena. On the other hand, social science research acknowledges subjectivity and cultural factors while utilizing both qualitative and quantitative methods to investigate human behavior and society systems.

1.8 Summary

Social research methodically investigates human civilization, confirming pre-existing information and revealing fresh insights. It uses combined, qualitative, and quantitative approaches to investigate relationships, behaviors, and social systems. Social research effectively addresses new social issues and advances theory development, policy assessment, and societal transformation through methodical investigation. In the end, both research in the social and scientific sciences advances human knowledge, impacting innovation, policy, and societal advancement. By acknowledging their differences and valuing their similarities, researchers may create comprehensive answers to pressing issues that benefit people and communities everywhere.

1.9 Questions

Long Questions :

1. Discuss the importance of social research in understanding human society.
2. Define social research according to different scholars and analyze their perspectives.
3. Examine the scope of social research in modern society.
4. What are the key differences between scientific research in natural sciences and social science research? Explain with examples.
5. Why is predictability important in scientific research? How does it differ in natural sciences and social sciences?
6. Explain the key differences between descriptive and analytical research with examples.

Short Questions:

1. What type of research explores social phenomena in depth?
   Ans: Qualitative
2. Which research method relies heavily on statistical data?
   Ans: Quantitative
3. Who defined research as a "systematic investigation"?
   Ans: Kothari
4. Which research method combines qualitative and quantitative approaches?
   Ans: Mixed
5. What is the main goal of social research?
   Ans: Inquiry
6. Which branch of science studies human behavior and society?
   Ans: Sociology
7. Which type of research is conducted over multiple time periods?
   Ans: Longitudinal.
8. What is another term for fundamental research?
   Ans: Basic.

1.10 References

1. Asghar, J. (2013). Critical Paradigm: A Preamble for Novice Researchers. *Life Science Journal*, *10*(4), 3121-3127.
2. Babbie, E. (2014). *The practice of social research*. CENGAGE Learning Custom Publishing.
3. Bryman, Alan. (2012). *Social Research Methods*. Oxford University Press.
4. Kothari, C.R. (2004). *Research Methodology: Methods and Techniques.* New Age International Limited.
5. Majeed, I. (2019). Understanding Positivism in Social Research: A Research Paradigm of Inductive logic of Inquiry. *International Journal of Research in Social Sciences, 9*(11), 118-125.
6. Magaji, A., and Sa'eed, M. (2017). Science, Social Research and Scientific Approach. *IOSR Journal Of Humanities And Social Science (IOSR-JHSS)*, *22*(10), 15-21.
7. Nudzor, H.P. (2009). A critical commentary on combined methods approach to researching educational and social issues. *Issues in Educational Research*, *19*(2), 114-127.
8. Pervin, N., and Mokhtar, M. (2022). The Interpretivist Research Paradigm: A Subjective Notion of a Social Context. *International Journal Of Academic Research in Progressive Education & Development*, *11*(2).
9. Temilola, O.M. (2015), Elements of Social Scientific Thoughts in Ogundiya, A.S. and Amzat, J. (eds) *Foundation of the Social Sciences. Lagos*: MalthousePress Limited.
10. Walliman, N. (2006). *Social Research Methods*. Sage Publications, pp. 14-28.
11. https://www.educba.com/types-of-research-methodology/

Unit 2 □ Paradigms, Theory and Social Research; Deductive and Inductive Reasoning

Structure

2.1 Learning Objectives

- To acquaint oneself with the paradigms of research.
- To learn about the types of the paradigms of research.
- To study the meaning of theory.
- To know why theory is important in research and vice versa.

2.2 Introduction

To create good research, it's critical to know about the paradigms (frameworks) that guide the construction of the research methodology in order to gain a deeper knowledge of social research, its forms, and its approaches. Knowing the specific subject of the

research and analyzing its results are made easier by paradigms like positivism, interpretivism, and critical theory.

2.3 Paradigms of Research

A paradigm model is a general idea that aids in providing the broad theoretical framework needed by scientists working in a specific topic. There are four elements of the paradigm: Firstly, ontology which focuses on what is the social reality we are talking about? Secondly, Epistemology which emphasizes on the different perspective of knowledge gathered on reality. Thirdly, methodology the research design used to conduct the study. Fourthly, the method is the tools and techniques used in the field. On the other hand, the various paradigms—such as positivism, interpretivism, and critical theory—help in understanding the particular topic of the study and interpreting its conclusions.

2.4 Positivism

Auguste Comte originated the notion of positivism, which Emile Durkheim later expanded upon. Positivism is the belief in an objective, scientifically proven reality. During the development of positivism, Auguste Comte held the view that all true knowledge is derived from sensory experience and can only be expanded upon through additional experimentation and observation (Walliman, 2006). Comte's ideas of Observation, Experimentation and Comparison are the foundation of studying human society and helped in developing positivism. The first paradigm, positivism, is based on realism, which holds that the world exists and can be understood according to what it really is (Walliman, 2006). The goal of positivism is to identify the universal rules governing society and how people behave in it (Walliman, 2006). The positivist paradigm is a logical structure developed by scientists to provide explanations for human behavior (Walliman, 2006) . Positivist framework creates a simplification of reality, particularly through quantitative analysis and mathematical models (Walliman, 2006). In quantitative terms, positivism demands two primary aspects: observation and verification. The following are assumptions related to positivism:

Firstly, It is only through experience that scientific knowledge should be developed (Majeed, 2019). The notion that social phenomena can be established using the principles

of natural sciences, and that these approaches can be directly applied to the social world (Majeed, 2019).

Secondly, Both social science and science have the same topic area (Majeed, 2019). The study of reality that is external to itself is what natural science studies and positivism takes into account (Majeed, 2019).

Thirdly, Durkheim considered social facts to be things or objects, therefore it does not significantly change that the "objects" of social science, namely individuals, are different (Majeed, 2019). The axiological rule that normative statements uphold a strict distinction between values and facts and do not possess the status of knowing (Majeed, 2019).

Fourthly, Technically beneficial knowledge has been pursued more often due to the generation of "instrumental knowledge" by science (Majeed, 2019). The positivist paradigm, which is questioned by many, attempts to compare natural science with social science (Majeed, 2019). The paradigm faces criticism because natural scientific methods may not be appropriate for social science (Majeed, 2019). Since positivism does not dismiss creativity or reflexivity, interpretivism has gained new momentum (Majeed, 2019). According to hermeneutic theory, meaning that isn't reducible to observation alone must be objectified in order for social reality to exist (Majeed, 2019). Interpretivism became possible as a result of all these critiques.

2.5 Interpretivism

The first person to stress interpretative research was Max Weber. It is opposed to positivist research and related to qualitative research. The interpretative paradigm draws attention to the fact that participants' subjective interpretations form the basis of social reality. Interpretive researchers use sense-making and hypothesis testing to interpret social reality as they see it from interactions with others. Numerous academic fields, including semiotics, anthropology, and sociology, are the foundation of interpretive research. The foundation of interpretivism is idealism, which holds that while the world does exist, various people interpret it extremely differently (Walliman, 2006) Interpretivist discover the way various individuals perceive the world in which they exist (Walliman, 2006).The interpretivist approach looks for significant connections and determines how those connections affect behavior (Walliman, 2006). Interpretivist analysis is holistic but very contextual. It is involved in language, signs and every meaning involved in a particular social phenomenon. However, because interpretivism is based on subjectivity, many positivist scholars see it as inappropriate. According to

Pervin and Mokhtar (2022), Researchers must concentrate on the two levels of interpretation: the first is seeing or experiencing phenomena from the subjective viewpoint of social participants, and the second figuring out the significance of the participant's experience to offer a rich narrative or "detailed description" of the phenomenon of interest that can elucidate the participants' actions.

The data collected in interpretative research is through various techniques like, face to face interviews, telephonic interviews, participant observation, focus group and documentation. It collects data and interprets every meaning of data (Pervin and Mokhtar, 2022). The concept of the "hermeneutic circle" states that our comprehension of a text as a whole is dependent upon our comprehension of every component as well as how those sections relate to the text as a whole. It is an interpretative iterative process that keeps on happening (Pervin and Mokhtar, 2022). The main objective is to obtain a deeper comprehension of the ways in which individuals create, alter, and interpret the social reality they inhabit (Pervin and Mokhtar, 2022). When there is no a priori theory or when an a priori theory is insufficient, interpretivist paradigms are often helpful and advantageous for theory development, therefore researchers may choose to utilize them (Pervin and Mokhtar, 2022) . The interpretative paradigm is seen as important since it pertains to this research since it may detect detailed occurrences in life (Pervin and Mokhtar, 2022). Interpretivism faces criticism for its assertion that it provides deeper understandings of social issues than independent scientific facts (Nudzor, 2009). Interpretivism, it is argued, lacks a cohesive theoretical base and is instead fragmented into other "-isms," such as feminism and interactionism (Nudzor, 2009). Interpretivism's critics claim that it is only appropriate for early phases of research and that it only serves to familiarize researchers with contexts before quantitative techniques can thoroughly evaluate data (Nudzor, 2009). As a result, the criticism paints positivist methods as more reliable and suitable for thorough scientific investigation than interpretivism (Nudzor, 2009). Critics claim that because interpretivism emphasizes subjectivity, it is unreliable (Nudzor, 2009). They contend that interpretivists concentrate on understanding how people view the social context (Nudzor, 2009). They argue that positivists record important details, including pauses and overlaps in encounters, to guarantee more accurate and balanced viewpoints in social research, but interpretivists frequently disregard these (Nudzor, 2009). According to this criticism, positivist methods can produce more reliable or consistent results than interpretivism (Nudzor, 2009).

Though interpretive paradigm is important for investigating how individuals perceive and comprehend their social environment is the focus of interpretive sociology. It all

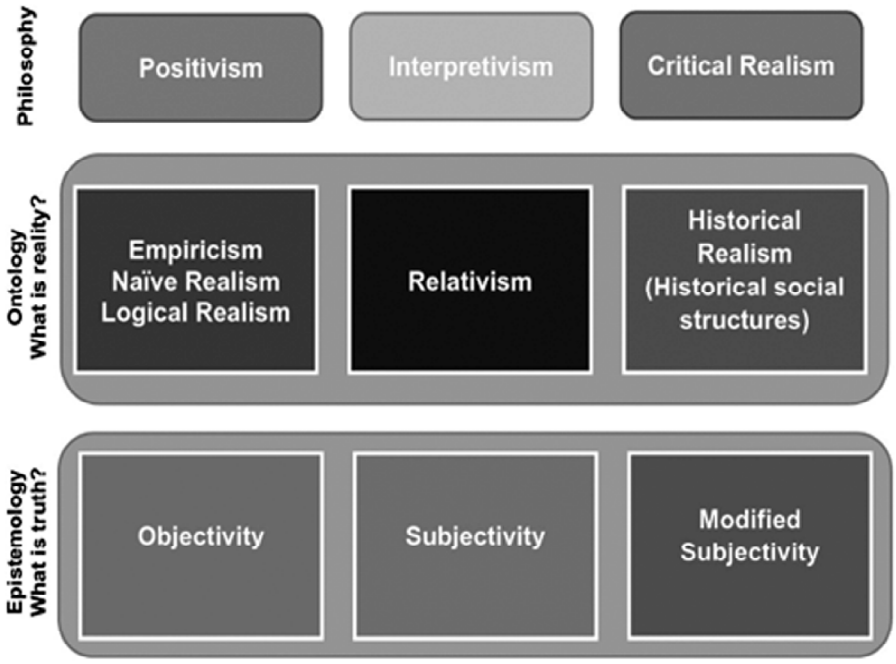
comes down to researching the meanings people ascribe to their experiences in order to fully analyze their viewpoints, behaviors, and identities. Scholars working in this area try to understand these subjective meanings through qualitative methodologies.

2.6 Critical

The three Frankfurt School philosophers, Horkheimer, Adorno, and Marcuse, are recognized for founding the critical theory paradigm (Asghar, 2013). One of the architects of the Frankfurt School and Critical Theory, Horkheimer, defines the term and says that Critical theory aims to liberate human beings from the circumstances that confine them (Asghar, 2013). Critical theory attempts to create a democratic and equitable society by challenging the status quo (Asghar, 2013). The issue of power dynamics in society and the interactions between the various social institutions that form up a social system—such as those related to gender, race, class, education, the economy, and religion—are of particular significance to critical theory (Asghar, 2013). Marx's idea that the economy is the base and all the other aspects of society, such as gender, race, and religion, are the superstructure, had an influence on Horkheimer (Asghar, 2013). They are implying that economic considerations contribute to oppression (Asghar, 2013). The decisions made in critical research methodology enable the continual process of integrating theories and methods (Asghar, 2013).

Any methodology or strategy that could help suggest improvements in the unbalanced social structure can be used by critical theory (Asghar, 2013). They have the option of using mixed, qualitative, or quantitative approaches (Asghar, 2013). The majority of research using a critical theory paradigm deals with conflict and inequality (Asghar, 2013). Studying conflict, inequality, and the resistance they generate is crucial for understanding the dynamics of interpersonal relationships (Asghar, 2013). Rationalism, neoliberalism, post-colonialism, feminism, radicalism, romanticism, and critical race theory all lend themselves to being integrated with and constructed upon critical theory (Asghar, 2013).

Critics argue that by focusing primarily on power dynamics, critical theory oversimplifies complex social concerns and ignores the action of the person. Its emphasis on inequality and conflict may at times result in a deterministic view of society, ignoring variations in the historical and cultural conditions that influence social relationships and institutions. Thus, the Paradigms are the model of certain general assumptions of reality. Whereas, Theory describes a specific fact of the social world in a more concrete way.



(Springer, M.Ali, 2023)

2.7 Definition of Theory

Theories are scientifically proven facts. Theories are used by researchers to understand, explain and make assumptions about the concept. Theory interconnects the idea and formulates knowledge about social phenomena. According to (Goode & Hatt, 2006), Theory means the relationships between facts or to the arrangement of them in some purposeful way.

Theory is the cornerstone of research. Research is based on theories, while theories are dependent on research. Research provides simultaneously an instrument for the advancement of theory and a means of obtaining the evidence necessary to support theory.

2.8 Role of Theory in Research

Theory explains research, assists in fact prediction, and points out new directions. Theories help in understanding both the broad topic and the crucial subject matter

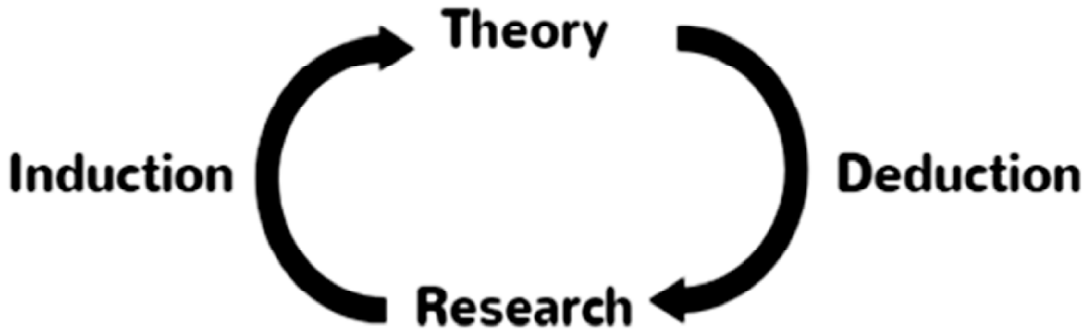
within it, thereby helping in further topic reduction. Theories are the extra knowledge that supports the formulation of the hypothesis that has to be investigated. By analyzing the theories, researchers can determine where there is an analysis gap and develop questions for further investigation.

Research empirically helps to gather experienced and observable data. After finding an appropriate theory, the researcher finds the gap in the theory and formulates a hypothesis. Based on the hypothesis, researchers build the research questions and in order to gather research questions, researchers many times use empirical methods. The empirical research is restricted to what can be observed and the facts it produces, and that needs to be either verified, rejected, or are neutral to a hypothesis.

For example, functionalist theory proponents Emile Durkheim and Talcott Parsons claimed that society is a complex system composed of different parts, each of which is necessary for society to function. They argued that all parts of society help to maintain order and stability, and that norms and values keep society integrated and functional. Research was conducted on the impact of the education system during the pandemic. While conducting this research, the researcher used the interview method and asked students from different schools how they studied and kept in touch with each other during the stressful pandemic times. Many students answered that during the pandemic, the school followed a routine pattern of classes via video calls. Every student had to stay active, and teachers sometimes conducted instant tests to keep them engaged with their education. This norm of attending classes and giving tests, including oral exams, helped the younger generation to remain functional during one of the most difficult phases in the world.

Theories are developed in response to research, and theories can be revised in response to new information that researchers have discovered while doing their research. In addition to making up for the deficiencies of a single theory, the integration of ideas from other fields of study produces a multifaceted view of a social situation and increases the interdisciplinary, comprehensive and inclusive nature of research.

Hence theory and research both are inextricably intertwined with each other. Both are essential for contributing knowledge to the field of social science. In order to develop theories for research, there is a need for two types of reasoning- Inductive and Deductive reasoning. Both are vital approaches in social science research.



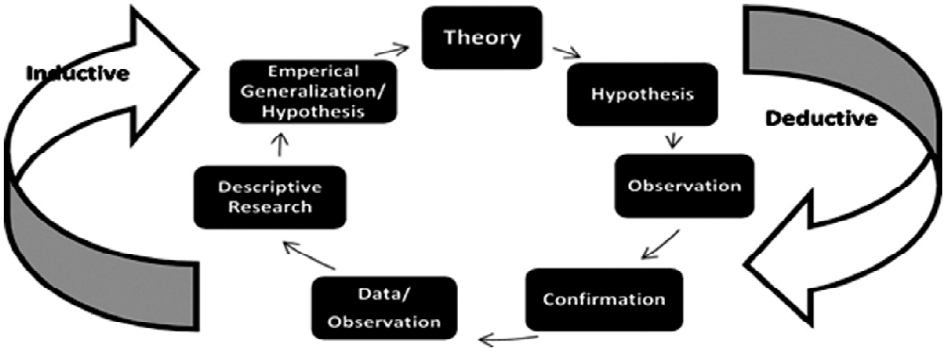
(Mind the graph, Salomao, 2023)

2.9 Inductive Reasoning

Inductive reasoning begins with particular observations and derives broad conclusions from them (Walliman, 2006). This simple illustration will show the way of reasoning: " All swans which have been observed are white in color. Therefore one can conclude that all swans are white " (Walliman 2006, p-16). Perhaps the most well-known theories that those who created them asserted to have originated through inductive reasoning are Darwin's theory of evolution and Mendel's discovery of genetics (Walliman, 2006) It is an empiricist approach (Walliman, 2006). Three stages of inductive reasoning are involved: The researcher first gathers and observes data. Second, the pattern is identified by the researcher. Third, a broad conclusion is drawn by the researcher. Do youngsters who use social media get addicted to it? This could be an example of inductive reasoning. To gather the answer to this question, surveys of teenagers must be conducted in order to determine how much time they spend on phones. Following that, the investigator will attempt to find the trend, specifically the duration of the day during which adolescents use social media. If a study finds that teens use social media for eight to twelve hours a day, it can result in low grades in school, increased isolation, and a drop in physical activity. A broad conclusion that teens are more vulnerable to social media use and that it has adverse effects on their life may be drawn from the research findings and patterns that have been noticed. Inductive reasoning merely sticks to the bottom-up approach.

2.10 Deductive Reasoning

A deduction-based reasoning commences with a few generalizations and uses reason to arrive at a particular conclusion (Walliman, 2006). The simplest kind of this type of argument is a logic, which is composed of a main general premise (statement), a minor, more specific premise, and a logically followed conclusion (Walliman, 2006). Here's an easy instance: "All live mammals breathe. This cow is a live mammal. Therefore, this cow breathes" (Walliman 2006, p-17). The Ancient Greeks were the first to establish deductive thinking (Walliman, 2006). Deductive reasoning operates in the subsequent ways: A theory already exists; one then develops a hypothesis, gathers and analyzes data, and, at the end, the theory can be completely rejected by making observations that contradict its predictions (Walliman, 2006). This illustrates how science is thought to work through trial and error: a theory is examined and proposed when it is rejected, and the theory that holds strongest in logic prevails ahead (Walliman, 2006). Deductive reasoning concentrates on a single solution, almost ensures a correct result. It simplifies the process of making decisions and demonstrates the systematic process that guides the decision.



(Springer, Haque, 2022)

2.11 Conclusion

Social research, which is essential to revealing individual social insights, is based on a variety of viewpoints and concepts. The complex subjective experiences that are essential for a comprehensive knowledge are frequently ignored by positivism's empirical rigor. Despite offering profound contextual insights, interpretivism is

criticized for lacking scientific rigor and generalizability. Though it can be unnecessarily generally biased and predetermined, critical theory effectively draws attention to power disparities. While the interaction of deductive and inductive thinking is essential, each approach has drawbacks as well, such as limited applicability and potential researcher bias. To address each paradigm's shortcomings and maximize on its advantages, an essential integration of different approaches is necessary. This integrated approach advances sociology's theoretical development as well as its practical applications by promoting a more accurate, nuanced, and balanced knowledge of social processes. Since research produces new concepts and technology, it is the bedrock of human development. The world advances and develops as a result of these new concepts' and it contributes to the creation of fresh predictions and objective truths.

2.12 Summary

The chapter intends to give a brief idea on research paradigm: positivism, which stresses in objective, scientific methods; interpretivism, focusing on subjective meanings and qualitative insights; and critical theory, which examines power structures and aims for social equity. The chapter highlights theory, explaining its important role in guiding research and forming assumptions. On the other hand, it delves into inductive reasoning, which derives general conclusions from specific observations. The deductive reasoning, which tries to understand general principles to reach specific conclusions. The integration of these paradigms and reasoning approaches is vital for developing an exact and balanced understanding in social research.

2.13 Questions

Answer in Detail:

1. Compare the paradigms of positivists, interpretivists, and critical theory in social research.
2. Define the concept of theory in the role of research.
3. Differences between inductive and deductive reasoning in the research process.
4. How does research contribute to development and testing of theories?

Short one word answers:

1. What Paradigm focuses on Objective Reality?
   Ans: Positivism.
2. Which research method uses statistical analysis?
   Ans: Quantitative.
3. What is the key feature of interpretivism?
   Ans: Subjectivity.
4. Which research process starts with a theory?
   Ans: Deductive.
5. What type of research focuses on exploring new phenomena?
   Ans: Exploratory.

2.14 References

1. Asghar, J. (2013). Critical Paradigm: A Preamble for Novice Researchers. *Life Science Journal, 10*(4), 3121-3127.
2. Babbie, E. (2014). *The practice of social research*. CENGAGE Learning Custom Publishing.
3. Bryman, Alan. (2012). *Social Research Methods*. Oxford University Press.
4. Kothari, C.R. (2004). *Research Methodology: Methods and Techniques*. New Age International Limited.
5. Majeed, I. (2019). Understanding Positivism in Social Research: A Research Paradigm of Inductive logic of Inquiry. *International Journal of Research in Social Sciences, 9*(11), 118-125.
6. Nudzor, H.P. (2009). A critical commentary on combined methods approach to researching educational and social issues. *Issues in Educational Research, 19*(2), 114-127.
7. Pervin, N., and Mokhtar, M. (2022). The Interpretivist Research Paradigm: A Subjective Notion of a Social Context. *International Journal Of Academic Research* in Progressive Education & Development, 11(2), 419-428.
8. Walliman, N. (2006). Social Research Methods. Sage Publications, pp. 14-28.
9. https://mindthegraph.com/blog/inductive-vs-deductive-research/
10. https://link.springer.com/chapter/10.1007/978-981-19-5441-2_5
    https://link.springer.com/chapter/10.1007/978-3-031-25470-3_2

Unit 3 □ Ethics in Social Research

Structure

3.1 Learning Objectives

- Understand the fundamental concepts of ethics and how they apply specifically to research practices in the social sciences.
- Identify and evaluate the various ethical challenges and dilemmas that may arise in the process of conducting social research.
- Familiarize yourself with the ethical guidelines established by the American Sociological Association (ASA) to ensure responsible conduct in sociological research.
- Explore the ethical standards set by the National Committee for Research Ethics in the Social Sciences and the Humanities (NESH) and their application in research.
- Recognize the crucial role of ethics in maintaining integrity, credibility, and public trust in social research.

- Learn about the specific actions and protocols necessary to uphold and enforce ethical standards throughout the research process.
- Analyze the significance of adhering to ethical codes of conduct in promoting fairness, transparency, and accountability in social research.

3.2 Introduction

Ethics in social research is a foundational aspect that ensures the integrity, respect, and responsibility of scholarly inquiry involving human subjects. It is essential to follow ethical guidelines when doing social research to safeguard participants' rights, welfare, and dignity as it explores the intricacies of human behavior, communities, and relationships. The goal of this lesson on "Ethics in Social Research" is to examine the morally and scientifically good research techniques, guidelines, and ethical considerations that one must follow.

A wide range of concerns are covered by the ethical framework in social research, such as informed consent, confidentiality, and preventing harm. In order to ensure that participants are completely aware of the nature, purpose, and potential dangers of the study, researchers are required to acquire informed permission from them. Participants' autonomy is respected throughout this process, which also gives them the freedom to choose how involved they want to be. Confidentiality is another important aspect where researchers have a responsibility to protect participants' personal information and ensure that data is handled with the highest level of security, which not only protects participants from potential harm but also maintains the integrity of the research process. Furthermore, ethical social research involves considering the potential impacts of the research on participants and their communities. Researchers must be vigilant to avoid exploitation or harm to individuals or groups. This involves sensitivity to issues of power dynamics, cultural differences, and the potential to reinforce stereotypes or prejudices.

Ethical principles in social research are defined by Institutional Review Boards (IRBs), which monitor and ensure compliance with established codes of conduct and ethical standards. These guidelines help researchers navigate complex ethical dilemmas and make informed decisions throughout the research process.

This unit will provide an in-depth examination of these ethical principles, explore case studies that illustrate ethical challenges, and offer practical strategies for implementing ethical practice in social research. By understanding and applying these

principles, researchers can contribute to the advancement of knowledge while maintaining the highest standards of ethical behavior, ultimately promoting trust and credibility in social research.

3.3 Meaning of 'Ethics' and 'Research Ethics' in Social Science

Ethics is a branch of philosophy that deals with questions of morality, focusing on concepts such as right and wrong, good and evil, justice, and virtue. It involves the systematic examination of how people ought to act and what kinds of actions are morally permissible or obligatory. Various scholars have defined and interpreted ethics in different ways, often influenced by their philosophical perspectives. For instance, Aristotle defines ethics as a branch of philosophy concerned with the principles of human conduct. He emphasizes the concept of "virtue ethics," where the moral character of an individual is central to ethical behavior (Ross 1999). In the Collins Dictionary of Sociology, it is stated that 'ethics' refers to the moral code of a person or society (Jary & Jary 2000). According to the Oxford Dictionary of Sociology, 'ethics' is often defined as the concern with what ought to be (Marshall 2004). It can be said that ethics examines the principles and guidelines that govern individuals' behavior, helping them determine what is morally correct in various situations.

Research ethics in social science refers to the moral principles and guidelines that govern the conduct of research, ensuring that it is carried out responsibly and ethically. According to M. Israel & I. Hay (2006), research ethics refers to "the moral principles guiding research, from its inception through to completion and publication of results and beyond, to its potential afterlife in terms of uses made of findings." Israel and Hay emphasize that research ethics is not only about the conduct during the research process but also involves considerations of how research findings are used after publication. This comprehensive approach underscores the importance of ethical considerations throughout the entire research lifecycle. In the words of A. Bryman (2016), "Ethics in research is concerned with the attempt to formulate codes and principles of moral behavior for researchers in relation to the rights of those who become the subject of the research." E. Diener & R. Crandall (1978) mentioned that "ethical research in social sciences is the application of ethical standards, including the consideration of potential harm, informed consent, and the privacy of participants." According to D. B. Resnik (2015), "Research ethics encompasses the application of ethical principles to the design, conduct, and

reporting of research, with an emphasis on honesty, integrity, transparency, and respect for human rights."

Therefore, social science research ethics can be said to be a subset of applied ethics, which includes the application of ethical principles in the development, implementation, and documentation of research projects. In other words, research ethics refers to the application of ethical principles and codes of professional conduct to the collection, evaluation, compilation, and dissemination of research data. Research ethics sets the standards of speech and behavior that determine what is appropriate and inappropriate for research. Other factors that fall under the umbrella of research ethics are respect for others and society, how resources and research products are used, scientific misconduct, and research regulation.

3.4 Ethical Considerations in Social Research

To maintain both the integrity of the study and the safety of participants, social researchers must carefully evaluate a number of significant ethical issues. A few of the main ethical issues are as follows (Baker1998; Bryman 2016):

A) **Whether there is Harm to Participants:** Harm to participants is one of the most important ethical issues in social research. This refers to any physical, psychological, social, or emotional discomfort that a research participant may experience while participating in a research study. Although rare in social research, physical harm can occur in certain studies, such as those involving vulnerable populations or risky environments. Psychological harm includes stress, anxiety, depression, or trauma that may arise from recalling painful memories, discussing sensitive issues, or facing uncomfortable questions. Social harm occurs when participants may be exposed to social consequences such as stigmatization, discrimination, or damage to reputation as a result of their association with research. Participants may also face legal consequences if sensitive information is leaked, such as unlawful conduct or illegal conduct. Participants may experience financial costs associated with participating in the study, such as absenteeism or lost wages.

Examples of Harm in Social Research:

1. **Milgram's Obedience Study**: In this experiment, participants were instructed to administer electric shocks to a confederate (an actor) whenever the confederate answered a question incorrectly. Many participants experienced

significant psychological distress, believing they were harming another person. Although no physical harm occurred, the emotional and psychological toll was significant, raising serious ethical concerns (Milgram 1963).

2. **Stanford Prison Experiment**: Conducted by Philip Zimbardo, this study involved participants assuming roles of guards and prisoners in a simulated prison environment. The experiment escalated quickly, with guards exhibiting abusive behaviors and prisoners suffering extreme stress, anxiety, and emotional trauma. The study was terminated prematurely due to the severe psychological harm experienced by participants (Zimbardo1973).
3. **Tea Room Trade Study**: Humphreys studied homosexual encounters in public restrooms without participants' informed consent. He secretly observed their behavior, recorded their license plates, and later visited them under a different guise to interview them. This study violated privacy, confidentiality, and posed significant risks of social harm and stigma for the participants (Humphreys 1970).

One of the most important components of ethical social research is addressing harm to participants. To balance the need to do no harm with the pursuit of knowledge, researchers must take extensive measures to protect the safety, dignity, and rights of participants. This requires informed consent, confidentiality, and careful weighing of pros and cons, as well as the values of beneficence, and justice.

B) Whether there is a Lack of Informed Consent: Informed consent is a fundamental ethical principle in social research. Informed consent involves providing participants with comprehensive information about the research and ensuring they voluntarily agree to participate without any coercion or undue influence. It is crucial because it respects the autonomy and rights of individuals, allowing them to make informed decisions about their involvement in research. It ensures that participants are fully aware of the research's nature, purpose, procedures, risks, and benefits before they agree to take part. The absence of informed consent, or obtaining it improperly, raises significant ethical concerns, as it can lead to the exploitation, harm, or misunderstanding of participants.

Examples of Lack of Informed Consent in Social Research:

1. **The Tuskegee Syphilis Study (1932-1972):** Conducted by the U.S. Public Health Service, this study aimed to observe the natural progression of untreated syphilis in African American men. Participants were not informed

about the true nature of the study, nor were they given the proper treatment for syphilis, even after penicillin became available. This case is a classic example of lack of informed consent, leading to severe health consequences and ethical outrage (Jones 1993).

2. **The Stanford Prison Experiment (1971):** Conducted by Philip Zimbardo, this psychological study examined the effects of perceived power. Participants were assigned roles as guards or prisoners in a simulated prison environment. Although participants initially consented, they were not fully aware of the psychological risks involved, and many experienced severe emotional distress. The lack of thorough informed consent procedures raised ethical concerns about the study's design and oversight (Zimbardo1973).
3. **Henrietta Lacks and HeLa Cells (1951):** Henrietta Lacks' cancer cells were taken without her knowledge or consent and used extensively in research. While her cells contributed significantly to scientific advances, her lack of consent highlights ethical issues regarding privacy and exploitation, particularly for marginalized individuals (Skloot 2010).

In order to safeguard participants, uphold ethical standards, and promote trust in the study process, informed consent is therefore essential in social research. The autonomy, comprehension and welfare of research participants must always come first when researchers attempt to acquire true informed consent.

C) **Whether there is an Invasion of Privacy:** Invasion of privacy is a central ethical issue in social research. This occurs when a researcher intrudes on an individual's space, thoughts, or life without the individual's express consent, causing potential harm or discomfort. This ethical issue is very important because social research often involves the collection of personal and sensitive data from participants that, if not handled properly, may violate their right to privacy. In qualitative research, like ethnography or in-depth interviews, recording participants without explicit consent can be a serious invasion of privacy. This is particularly problematic in settings where individuals may not expect to be recorded, such as private conversations or personal interactions. Invasion of privacy in social research can take several forms, including:

- **Physical intrusion:** Entering a private space without permission.
- **Publication of Private Facts:** Disclosing personal information without consent.

- **False Light:** Presenting misleading information that harms a person's reputation.
- **Appropriation:** Using someone's likeness or personal data for research without permission.

Examples of Invasion of Privacy in Social Research:

Sociologist Laud Humphreys conducted a study on men who engaged in sexual activities in public restrooms (tearooms) without their knowledge. Humphreys posed as a "watchqueen," an observer who guarded the entrance, to collect data. He recorded license plate numbers and later tracked down the men, interviewing them under false pretenses without revealing the true nature of his study. The study was highly invasive as participants were not informed that they were part of a research study. This intrusion into their private lives and behaviors without consent sparked ethical debates on privacy violations (Humphreys 1970).

The invasion of privacy is a significant ethical concern in social research, and it can lead to a range of problems that affect both participants and the integrity of the research process. Participants may lose control over how their personal information is used, shared, or stored, leading to feelings of vulnerability and exploitation. Invasion of privacy can cause emotional distress, anxiety, or embarrassment, particularly if sensitive or stigmatizing information is revealed. Therefore, in the field of social research, invasion of privacy is a serious ethical problem that must be carefully considered and addressed according to ethical standards. Researchers must give top priority to informed consent, confidentiality, and respecting the boundaries of participants, who must strike a balance between the need to preserve individual rights and the advancement of knowledge. Ethical oversight and continuous rethinking of research methods are essential to resolving privacy issues in social research.

D) Whether Deception is involved: Deception in social research involves deliberately misleading or withholding information from participants about the true nature or purpose of a study. This raises significant ethical concerns because it can violate the principles of informed consent, autonomy, and trust between researchers and participants. One of the core ethical principles in research is informed consent, which requires that participants be fully aware of the nature, purpose, risks, and benefits of the study. Deception can undermine this principle by preventing participants from making fully informed decisions. Deception can also infringe on participants' autonomy, as it involves manipulating their understanding or decision-

making process. This can lead to participants making choices they might not have made if they had been fully informed. Trust between researchers and participants are crucial for ethical research. Deception can damage this trust, not only in the specific study but also in the broader field of social research, as participants may become wary of future research efforts.

Examples of Deception in Social Research:

1. **Milgram's Obedience Study**: One of the most famous examples of deception in social research is Stanley Milgram's experiment on obedience to authority. Participants were led to believe they were administering electric shocks to another person as part of a learning experiment. The true purpose was to study obedience to authority figures. The deception involved both the nature of the task and the true focus of the research. Participants experienced significant stress, raising questions about the ethical justification of the deception used (Milgram 1963).
2. **The Stanford Prison Experiment:** Conducted by Philip Zimbardo, this study involved deception regarding the extent of control and power dynamics within a simulated prison environment. Participants were randomly assigned as guards or prisoners, but they were not fully informed of the psychological risks. The study had severe ethical implications, including emotional and psychological harm to participants, highlighting the dangers of deception in research settings (Zimbardo 1973).

   While deception in social research can provide valuable insights, it must be handled with extreme caution due to its ethical implications. Researchers must weigh the potential benefits against the risks to participants' autonomy, trust, and well-being, ensuring that deception is used only when absolutely necessary and with proper safeguards like debriefing. Ethical guidelines from organizations like the American Psychological Association (APA) and the British Psychological Society (BPS) generally discourage deception unless it is justified by the study's significant scientific value and when no alternatives exist.

E) Plagiarism: Plagiarism is the act of using someone else's work, ideas, words, or research without proper acknowledgment, presenting them as one's own. In social research, plagiarism is considered a severe ethical violation, undermining the integrity and credibility of research. It misrepresents the author's intellectual contribution and disrespects the original creator's effort. This ethical breach can have significant consequences, including academic penalties, loss of professional reputation, and even legal repercussions.

The following are the main categories of plagiarism:

a. **Direct Plagiarism:** According to Purdue University's Online Writing Lab, direct plagiarism is "the word-for-word transcription of a section of someone else's work, without attribution and without quotation marks" (Purdue OWL, 2020).

b. **Self-Plagiarism:** As defined by the American Psychological Association (APA), self-plagiarism is "the practice of presenting one's own previously published work as though it were new" (APA, 2020).

c. **Mosaic Plagiarism:** The Council of Writing Program Administrators (CWPA) describes mosaic plagiarism as "borrowing phrases from a source without using quotation marks, or finding synonyms for the author's language while keeping to the same general structure and meaning of the original" (CWPA, 2019).

d. **Accidental Plagiarism:** According to the University of Oxford, accidental plagiarism is "unintentional failure to acknowledge sources correctly, which may result from lack of understanding or carelessness" (University of Oxford, 2021).

e. **Paraphrasing Plagiarism:** As described by the University of Southern California, paraphrasing plagiarism involves "changing a few words and phrases while retaining the essential content of the source" (University of Southern California, 2021).

Plagiarism raises a variety of ethical issues that impact individuals, organizations, and society as a whole. Plagiarism is fundamentally an act of dishonesty. In professional, intellectual, and artistic contexts, this dishonest behavior destroys trust. Plagiarism violates the intellectual property rights of the original author. Violating these rights not only demeans the work of the original artist but can also lead to legal consequences. This can lead to fines, lawsuits, and loss of respect and credibility. Plagiarism by researchers or students undermines the credibility of academic qualifications, devalues genuine intellectual achievements, and harms the educational process. It can also create undue academic advantages, distorting grades and competitive norms. Because plagiarists rarely research the material they represent themselves in, they often lack a full understanding and knowledge of the subject. Inaccurate or poorly understood information may spread as a result, and the caliber of the work may suffer. In academic and professional environments, it lowers the overall caliber of work.

3.5 Ethical Code of Conducts set by ASA

The American Sociological Association (ASA) is the governing organization for sociologists and their research. The guidelines for conducting sociological research are the ASA Code of Ethics. All social scientists must abide by the ASA Code of Ethics. The Code of Ethics in Sociology is used in the service of the profession to promote the development of sociology as an academic discipline and as a foundational liberal arts profession and to advance sociologists' research toward the betterment of society. Social scientists are encouraged to follow a code of ethics to remain honest and fair in their interactions with clients. Failure to adhere to ethical standards may result in the social scientist's professional qualification being revoked and expose them to lawsuits and litigation.

The American Sociological Association (ASA) first released the code of ethics for society in 1971, and it was updated in 1997. It puts forth certain basic principles, such as:

1. Professional Competence
   - **Guideline:** Sociologists must strive to maintain the highest standards of competence in their work. They should recognize their own limitations and only undertake work for which they are qualified by education, training, or experience.
   - **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 1.01-1.06. ASA.
2. Integrity
   - **Guideline:** Sociologists should be honest, fair, and respectful in all their professional activities. They should avoid false or deceptive statements about their research or professional qualifications.
   - **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 2.01. ASA.
3. Professional and Scientific Responsibility
   - **Guideline:** Sociologists are expected to adhere to the highest professional standards and accept responsibility for their work. They should avoid conflicts of interest and ensure that their research is objective and unbiased.
   - **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 3.01-3.05. ASA.

4. Respect for People's Rights, Dignity, and Diversity
   - **Guideline:** Sociologists should respect the rights, dignity, and worth of all people. This includes recognizing differences in race, ethnicity, gender, age, sexual orientation, disability, and other factors.
   - **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 4.01-4.06. ASA.
5. Social Responsibility
   - **Guideline:** Sociologists have a duty to contribute to the public good, support social justice, and respect human rights. They should ensure that their research does not harm individuals or communities.
   - **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 5.01. ASA.
6. Human Rights and Informed Consent
   - **Guideline:** Researchers must obtain informed consent from all participants, ensuring that they are fully aware of the nature, purpose, and potential risks of the research. Consent must be given freely without coercion.
   - **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 12.01-12.05. ASA. L
7. Confidentiality
   - **Guideline:** Sociologists are obligated to protect the confidentiality of their research participants, ensuring that private information is not disclosed without permission.
   - **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 11.01-11.03. ASA. L
8. Non-Discrimination
   - **Guideline:** Sociologists should not engage in any form of discrimination based on race, ethnicity, gender, sexual orientation, religion, age, or any other characteristic in their professional activities.
   - **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 4.03. ASA.
9. Avoidance of Harm
   - **Guideline:** Sociologists must avoid causing harm to research participants, ensuring that any potential risks are minimized and fully explained to participants before the research begins.

- **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 11.04. ASA.

10. Publication and Authorship

- **Guideline:** Researchers must ensure that all contributors to a research project are properly credited and that findings are reported honestly without fabricating, falsifying, or misrepresenting data.
- **Reference:** American Sociological Association. (2018). *Code of Ethics*, Section 14.01-14.06. ASA.

These ethical guidelines by the ASA emphasize the importance of professional integrity, respect for human rights, and social responsibility in social research. For a more detailed understanding and specific scenarios, refer to the full ASA Code of Ethics available on their official website.

3.6 Ethical Code of Conduct set by NESH

Ethics in social research is essential because it guides researchers to conduct research responsibly, protecting the rights and dignity of participants and ensuring the credibility of the research. Here is a detailed discussion of the importance of ethics in social research:

1. **Protection of Participants' Rights and Welfare**: Ethics ensures that participants are not physically, psychologically, or socially harmed. Researchers must obtain informed consent, ensure confidentiality, and respect participants' privacy. This protects vulnerable groups and preserves the integrity of the research process (Diener & Crandall 1978).
2. **Informed Consent**: Participants must fully understand the nature of research, its purpose, and any risks involved. The informed consent can ensure that participants voluntarily agree without the need to participate in the study (Beauchamp & Childress 2013).
3. **Confidentiality and Privacy**: Protecting participants' data is critical in maintaining trust between researchers and participants. Ethics in social research emphasize safeguarding personal information and ensuring that data is not misused (Babbie 2020).
4. **Avoidance of Deception**: Ethical guidelines help prevent the use of deception in research unless absolutely necessary and justified. When deception is used, participants should be debriefed afterward (Kimmel 2011).

5. **Ensuring research reliability and integrity:** Adherence to ethical standards helps ensure that research results are reliable and trustworthy. It prevents fraudulent activities such as data fabrication, forgery, or plagiarism (Resnik 2020).
6. **Social Responsibility**: Researchers have a duty to conduct studies that contribute positively to society and avoid research that could have harmful social consequences. This includes being mindful of the broader impacts of their findings (Israel & Hay 2006).
7. **Respect for Cultural Sensitivity and Diversity:** Ethical research respects cultural differences and avoids biases that could **influence** research **results.** This is **especially** important in studies involving **different** populations (Punch 2014).
8. **Transparency and Accountability**: Ethics promote transparency in the research process, ensuring that researchers are accountable for their actions and findings. This includes disclosing funding sources and potential conflicts of interest (Shamoo & Resnik 2009).

Ethics in social research are foundational to protecting participants, maintaining trust, and ensuring the validity and societal impact of the research. Researchers must adhere to established ethical guidelines to maintain the integrity of their studies and contribute responsibly to social knowledge.

3.8 Actions to Uphold Ethical Standards in Social Research

Adherence to ethical standards in social research is critical to protecting participants, maintaining the integrity of the research, and ensuring the validity of the findings. Here are the key steps to maintain ethical standards:

1. **Obtain informed consent:** Participants must be fully informed about the purpose, procedures, risks, and benefits of the study. Consent should be obtained voluntarily, without any coercion, and participants should have the right to withdraw at any time.
2. **Ensure confidentiality and anonymity:** Researchers should protect the privacy of participants by keeping data confidential and anonymizing it when appropriate. Data must be stored securely and only accessible to authorized personnel.
3. **Avoid Deception:** Deception should be avoided unless absolutely necessary and justified. If deception is used, participants should be debriefed as soon as possible.

4. **Minimize Harm:** Researchers must ensure that their research does not cause physical, psychological, or social harm to participants. Risks should be identified, minimized, and explained to participants beforehand.
5. **Respect for Participants:** Treat participants with respect and dignity, considering their rights, values, and cultural contexts. Special care should be taken when working with vulnerable populations, such as children, elderly individuals, or those with disabilities.
6. **Ensure Voluntary Participation:** Participation should be voluntary, without any undue influence or pressure from the researcher. Participants should feel free to refuse or discontinue participation at any time without any negative consequences.
7. **Conduct Ethical Review:** All research should undergo an ethical review process by an institutional review board (IRB) or ethics committee to ensure compliance with ethical standards.
8. **Provide Debriefing:** After the study, participants should be provided with information about the research findings and the role they played. This helps to address any misconceptions and offers participants a chance to ask questions.
9. **Avoid Plagiarism and Misconduct:** Researchers must ensure the originality of their work and properly cite sources. Fabrication, falsification, or misrepresentation of data must be strictly avoided.
10. **Transparent Reporting and Dissemination:** Findings should be reported honestly and transparently, regardless of whether the results support the initial hypothesis. Sharing data with the academic community and the public, when appropriate, promotes transparency.
11. **Maintain Objectivity and Integrity:** Researchers should strive to remain objective, avoiding biases or conflicts of interest that may influence the research process or outcomes.
12. **Respect Intellectual Property Rights:** Proper acknowledgment should be given to contributors, and permissions should be obtained for using copyrighted materials.

Implementing these actions ensures that social research is conducted responsibly, ethically, and in a manner that respects the rights and welfare of participants.

3.9 Conclusion

Ethics in social research is the cornerstone of ensuring the integrity, reliability and social value of the research process. It emphasizes respecting the rights of the participants, maintaining confidentiality and ensuring voluntary participation, thereby preserving the dignity and independence of the participants. Ethical guidelines, such as obtaining informed consent and avoiding harm, help researchers navigate complex moral dilemmas that arise during research. These principles not only protect the participants, but also increase the credibility of the research results.

Furthermore, ethics promotes transparency and accountability, encouraging researchers to report data honestly and avoid manipulation or fabrication of results. This adherence to ethical standards fosters public trust, which is essential for the acceptance and use of research findings in policymaking and societal progress. Ethical conduct in research goes beyond legal compliance, as it reflects researchers' broader responsibility to make a positive contribution to society. Ultimately, ethics in social research is not just about following the rules but about respecting the fundamental principles of respect, justice, and responsibility, ensuring that research serves the greater good without compromising the rights or well-being of the individual.

3.10 Summary

Here's a summary based on the topics provided:

Meaning of 'Ethics' and 'Research Ethics' in Social Science:

Ethics refers to moral principles that guide individuals' behaviors, distinguishing right from wrong. In social science, **Research Ethics** involves applying these principles to ensure integrity, honesty, and respect for participants in research processes. It ensures that researchers conduct studies responsibly, safeguarding participants' rights and well-being.

Ethical Considerations in Social Research:

Ethical considerations in social research focus on protecting participants, ensuring voluntary participation, obtaining informed consent, maintaining confidentiality, and avoiding harm. Researchers must navigate sensitive issues, manage conflicts of interest, and handle data responsibly, upholding the dignity and privacy of those involved.

Ethical Code of Conducts set by ASA (American Sociological Association):

The **ASA's Ethical Code of Conduct** outlines principles such as professional competence, integrity, and respect for people's rights, dignity, and diversity. It emphasizes the necessity of confidentiality, informed consent, and honesty in reporting research findings. The code guides sociologists in maintaining high ethical standards throughout their professional work.

Ethical Code of Conduct set by NESH (Norwegian National Research Ethics Committees):

The **NESH Ethical Code of Conduct** stresses respect for individuals, groups, and communities in research. It highlights the importance of transparency, accountability, and the ethical management of data. NESH guidelines emphasize informed consent, confidentiality, and the duty to avoid harm to research participants.

Importance of 'Ethics' in Social Research:

Ethics are crucial in social research to protect participants' rights and ensure that research is conducted with honesty and integrity. Ethical research fosters trust between researchers and participants, enhancing the validity and credibility of findings. It also ensures that the research contributes positively to society without causing harm.

Actions that Ought to Be Taken to Uphold Ethical Standards in Social Research:

To uphold ethical standards, researchers should obtain informed consent, ensure confidentiality, and avoid deception unless justified and minimized. They should adhere to established ethical codes, maintain transparency in data collection and reporting, and prioritize participants' welfare. Continuous ethical training and review by ethics committees are also essential for maintaining high standards in research.

3.11 Questions

1. What is the definition of 'ethics' in the context of social science?
2. Why is an understanding of ethics crucial for researchers in social science?
3. What are the primary ethical considerations researchers must keep in mind while conducting social research?
4. What are the key principles of the ethical code of conduct set by the American Sociological Association (ASA)?
5. What ethical guidelines are outlined by NESH for social research?
6. Why is maintaining ethics considered essential in social research?

7. How does ethical conduct in research enhance the credibility and validity of social research findings?
8. What are the potential risks of ignoring ethical standards in social research?
9. What actions should researchers take to ensure ethical standards are upheld in their research?

3.12 References

1. Babbie, E. (2020); The Practice of Social Research (15th ed.). Cengage Learning.
2. Baker, Therese L.(1998); Doing Social Research; 3rd Edition, McGraw-Hill College, Boston
3. Beauchamp, T. L., & Childress, J. F. (2013); Principles of Biomedical Ethics (7th ed.). Oxford University Press.New York.
4. Bryman, A. (2016); Social Research Methods (5th ed.). Oxford University Press, New York.
5. Diener, E., & Crandall, R. (1978); Ethics in Social and Behavioral Research. University of Chicago Press, 1427 East 60th Street Chicago, IL 60637 U.S.A.
6. Humphreys, L. (1970); Tearoom Trade: Impersonal Sex in Public Places. Aldine Publishing,New York.
7. Israel, M., & Hay, I. (2006); Research Ethics for Social Scientists: Between Ethical Conduct and Regulatory Compliance. Sage Publications Inc, New Delhi.
8. Jary, David & Jary, Julia (2000); Collins Dictionary of Sociology; Harper Collins Publishers, Glasgo.
9. Jones, J. H. (1993); Bad Blood: The Tuskegee Syphilis Experiment. Free Press, New York City, New York, U.S
10. Kimmel, A. J. (2011); Deception in Social Research: A Conceptual Approach. American Psychological Association.
11. Marshall, Gordon edt(2004); Oxford Dictionary of Sociology;Oxford University Press, Great Clarendon Street, Oxford.
12. Milgram, S. (1963). Behavioral study of obedience. Journal of Abnormal and Social Psychology, 67(4), 371-378.
13. NESH (2016); Guidelines for Research Ethics in the Social Sciences, Humanities, Law and Theology. Oslo: The Norwegian National Research Ethics Committees.
14. Punch, K. F. (2014); Introduction to Social Research: Quantitative and Qualitative Approaches (3rd ed.). Sage Publications Inc, New Delhi.

15. Resnik, D. B. (2015 & 2020); The Ethics of Research with Human Subjects: Protecting People, Advancing Science, Promoting Trust. Springer.
16. Ross, W.D.(1999); The Nicomachean Ethics Of Aristotle ; (Translated by W.D.Ross); Batoche Books, Kitchener.
17. Shamoo, A. E., & Resnik, D. B. (2009); Responsible Conduct of Research (2nd ed.). Oxford University Press, New York.
18. Skloot, R. (2010); The Immortal Life of Henrietta Lacks. Crown Publishers,New York.
19. Zimbardo, P. G. (1973); On the ethics of intervention in human psychological research: With special reference to the Stanford Prison Experiment. Cognition, 2(2), 243-256.

3.13 Suggested Readings

- American Psychological Association (2020); Ethical Principles of Psychologists and Code of Conduct. APA. Retrieved from https://www.apa.org/ethics/code/ On 12 July 2024
- American Sociological Association (1997); Code of Ethics and Policies and Procedures of the ASA Committee on Professional Ethics; American Sociological Association; 1430 K Street NW, Suite 600 ,Washington, DC
- Baker, Therese L.(1998); Doing Social Research; 3rd Edition, McGraw-Hill College, Boston
- Bryman, Alan(2004); Social Research Methods;2nd Edition, Oxford University Press, New York
- Council of Writing Program Administrators. (2019); Defining and Avoiding Plagiarism: The WPA Statement on Best Practices. CWPA. Retrieved from http://wpacouncil.org/positions/WPAplagiarism.pdf On 12 July 2024
- NESH (2024); Guidelines for Research Ethics in the Social Sciences and the Humanities; (5th edition), Retrieved from https://www.forskningsetikk.no/en/guidelines/social-sciences-and-humanities/guidelines-for-research-ethics-in-the-social-sciences-and-the-humanities/ On 14 July, 2024
- University of Oxford. (2021); Plagiarism. University of Oxford. Retrieved from https://www.ox.ac.uk/students/academic/guidance/skills/plagiarism On 12 July 2024
- University of Southern California. (2021); Understanding and Avoiding Plagiarism. USC. Retrieved from https://libguides.usc.edu/writingguide/plagiarism On 14 July 2024

MODULE - II

Fundamentals of Research Design and Measurement

Unit 4 □ Getting Started: Steps in Social Research; Developing a Research Question

Structure

4.1 Learning Objectives

- Understand what is research, its purpose, and how it contributes to knowledge and problem-solving.
- Learn and explore the key steps in the a research process, including literature review, research design, sampling, and data collection.
- Develop skills to identify a research problem, how to choose appropriate methodologies, and structure a study.

- Gain knowledge on techniques for data collection, interpreting analysis and presenting research findings effectively.

4.2 Introduction

Humans have always had an inherent curiosity, driving us to find answers to several questions about why and how things happen. In this regard, research serves as a means to uncover and understand both known and unknown information, thus expanding our knowledge. It is important because it drives innovation, informs decision-making, and helps us better understand our society and the world. The onset of scientific research has greatly transformed society, making it crucial in social sciences as well.

In the context of social sciences, research takes on a unique role by focusing on human behaviour, societies, and social relationships. Social science research seeks to uncover patterns, understand how societies function, and explore the many factors that influence human behaviour. Whether studying the dynamics of families, the impact of education, or the complexities of social inequality, social science research provides valuable insights into the fabric of our lives. In sociology, research is highly significant for understanding complex social structures and phenomena. This unit will introduce you to the steps in social research. Understanding the foundational importance of research will form a background for the more detailed aspects for the detailed aspects.

4.3 What is Research

Research is a planned and organized effort to investigate a specific problem that needs to be solved. It is an addition to our overall understanding of the subject. The word "research" has two parts- '*Re*' which means again and again, and '*Search'* which means to find something. In everyday terminology, research means looking for knowledge. To put in simple words, research is a scientific and systematic way to find relevant information on a particular topic. It is a methodical way of asking and answering questions scientifically.

Some of the popular and comprehensive definitions of research are:

- According to P. V. Young (1966), *"Research is a scientific undertaking which, by means of logical and systematic techniques, aims to: (1) discover new facts or verify and, test old facts, (2) analyse their sequences, interrelationships*

and causal explanations, (3) develop new scientific tools, concepts and theories which would facilitate reliable and valid study 'of human behaviour."

- According to Kerlinger (1973), *"Research is a systematic, controlled, empirical and critical investigation of hypothetical relations among natural phenomena".*
- According to C.C. Crawford (1928): *"Research is simply a systematic and refined technique of thinking, employing specialized tools, instruments & procedures in order to obtain a more adequate solution of a problem than would be possible under ordinary means."*

There are certain characteristics of research can be identified. They are listed below:

- It is a planned and detailed study about a phenomenon.
- It goes much beyond than simply gathering information or facts. Rather, research involves a purposeful inquiry to describe, interpret, and explain a phenomenon.
- It uses the scientific method in the processes involved.
- It is objective and logical, using measurable tools and data in order to validate the conclusions drawn.
- It is grounded in observable experiences or empirical evidence.
- It aims to answer relevant questions and solve problems.
- It focuses on producing generalizations, concepts, or theories.

4.4 What is Social Science Research

Social sciences are not as exactly same as physical sciences because they are concerned with human beings. Social science research uses systematic methods to analyse, explore and understand human life to expand, correct, or verify knowledge about human behaviour and social life. It seeks to explain the unexplained phenomena, clarify doubts, and correct misconceptions about social life. It applies scientific methods in understanding and analysing social life, aiming to verify and improve existing knowledge in a systematic way.

Some of the important definitions of Social Science Research are:

- According to C. A. Moser (1971), *"Social research is a systematized investigation to gain new knowledge about social phenomenon and problems."*
- According to P.V. Young (1966), *"Social research is a scientific undertaking which by means of logical methods, aim to discover new facts or old facts and*

to analyse their sequences, interrelationships, casual explanations and natural laws which govern them."

The main aim of social research is to discover new relationships, new knowledge, and new facts and verify old ones. Human behaviour follows certain values and laws, and social research is aimed at uncovering these laws so that human interactions and behaviour are understood better. Since human behaviour is diverse and complex, it is challenging to find underlying patterns. Social research is the systematic and objective analysis and recording of controlled observations, which can lead to generalizations, principles, or theories that help to predict and possibly control events in society. It seeks to answer and solve social problems.

Hence, it can be said that the objectives of social research are:

- To understand human behaviour better.
- To gain knowledge about social issues, events, problems and phenomena.
- To identify and understand the relationships among social phenomena.
- To discover natural laws that govern social phenomena.
- To standardize concepts like culture, conflict, generation gap, and social distance.
- To generate solutions for social problems.
- To maintain social order, reduce social tension, and correct misconceptions.
- To create plans for social improvement and revival.

There are certain characteristics of social research. They are:

- It is aims towards solving problems, aiming to discover cause-and-effect relationships in social issues.
- It aims to develop generalizations, principles, or theories to help predict future events.
- It depends on observable experiences or empirical evidence.
- It needs precise observations and descriptions, using various qualitative and non-qualitative methods.
- It focusses on collecting new data from primary sources or repurposing existing data.
- It is marked by carefully designed procedures and rigorous analysis.
- It needs expertise. The researchers must be familiar with existing knowledge and previous investigations on the problem.

- It strives to be objective and logical, applying thorough tests to validate the procedures, data, and conclusions.
- It aims to provide answers to unsolved problems.
- It should be conducted in a patient and meticulous way. In the process, the researchers should be prepared to expect setbacks and challenges, since they tackle difficult questions.
- It is recorded and reported in a careful manner, with clear definitions, acknowledging the limitations, detailed procedures, well-documented references, objectively recorded results, and cautious conclusions.
- It is interdisciplinary in nature.
- It sometimes requires courage to pursue.

4.5 Steps in Social Research

The research process consists of a number of connected activities. Each step in the research process is linked to the others, making up the entire research project. This process includes multiple steps, starting from framing of research questions to documenting the findings.

Let us look at the steps involved in Social Science Research:

- Identifying and selecting a Research Problem
- Review of Related Literature
- Research Design
- Determining the Sample
- Carrying out the Research
- Interpretation of the Data
- Presentation a Report

The steps of a research is shown with the flowchart given below:

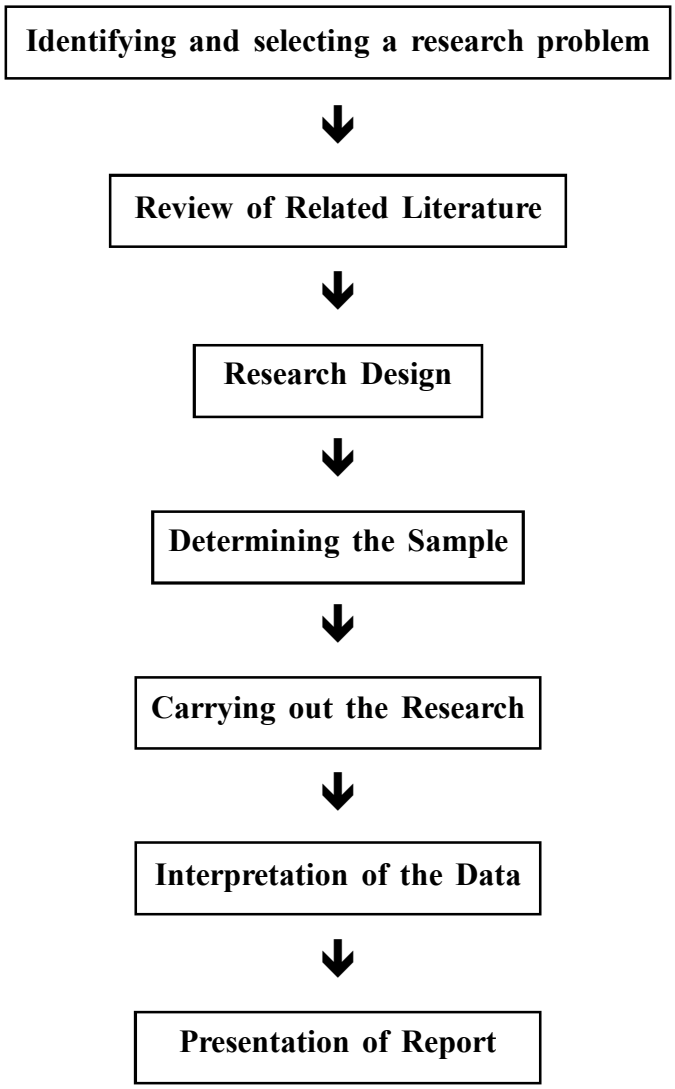


Fig 1.

Now, we shall understand each of these steps in detail with examples.

4.5.1 Identifying and selecting a Research Problem

The first significant step in the research process is to formulate a research problem. In social research, the emphasis is to identify these instead of answering questions that

come up. So, rather than asking, *"What is this?"*, the question posed by the researchers should be *"Why does this exist?"*. Such problem-solving questions do not only explain facts but also provide explanations for why those facts exist in the first place. For example, you want to do a study on the impact on social media on teenagers. Instead of merely asking "what is the impact of social media among teenagers", you can pose a deeper question like: "Why does frequent use of social media lead to increased anxiety and depression among teenagers?" Such a question not only tries to identify the effects of social media on the mental heath of teenagers but also aims to understand the reasons or factors like online peer pressure, cyberbullying, which could explain why social media usage is linked to anxiety and depression. Similarly, if you want to study about the level of gender inequality in workplaces, you can pose specific and sociological research questions like "Why are there fewer women in top positions in corporate environments despite equal qualifications?" Such research problems aim in understanding the underlying causes like gendered biasness or stereotypes, workplace culture, or lack of supportive policies, which prevent women from advancing in their careers despite having the same qualifications as men.

Do you know?

Anthony Giddens (2009), one of the renowned sociologist compares research problems to puzzles. He is of the view that research problems are like puzzles. For him, these problems are not merely about missing information on a topic; rather, they represent gaps in our understanding of that topic.

It is believed that an appropriate research problem should meet three main criteria:

- It should show the relationship between two or more variables.
- It should be framed as a question.
- It should be testable with data from direct observation and experiments.

You can use several sources to develop a research problem or becoming aware of potential issues:

- Reading: By reading books and articles about topics you are interested in, will help you come up with important questions. Also, research papers can suggest you the areas to explore.
- Academic Experience: Class discussions, lectures, seminars, and interaction with fellow students and professors will generate ideas for interesting problems to investigate.

- Day-to-day experience: Life is constantly changing. New experiences spark questions that are worth exploring. Being curious, alert and sensitive to life situations will help you in revealing new research opportunities.
- Field Situations: Undertaking field visits, internships and extension work exposes you to real-world problems which demand investigation.
- Consultation: Discussions with experts, analysts, administrators will help you in finding relevant research problems.
- Brainstorming: Intense debate among a group of people who are interested in a problem also leads to the identification of relevant questions and generating new ideas about the problem.
- Research: Research on one problem often leads to the discovery of newer issues or problems that need to be investigated further.

4.5.2 Review of Related Literature

Once your research problem is identified, the next step involves examining existing literature on the chosen or selected research area. In this phase, the researcher studies the existing and relevant body of literature. Defining review of literature, C. Hart (1998) writes, '*the selection of available documents (both published and unpublished) on the topic, which contain information, ideas, data, and evidence written from a particular standpoint to fulfil certain aims or express certain views on the nature of the topic and how it is to be investigated, and the effective evaluation of these documents in relation to the research being proposed.'* Review of literature sheds light on the area of study and helps to avoid unnecessary repetitions and duplications of research work. It helps in providing the researcher with the existing gaps in the literature.

To summarise, the review of literature has certain goals. They are:

- To find theories, ideas, explanations, or hypotheses that can support in formulating the research problem.
- To prevent doing repetitive studies.
- To generate ideas for hypotheses.
- To identify methods for data collection, procedures for exploring data sources, and statistical techniques suitable for solving the problem.
- To collect and compare data and findings from previous research that can help interpret and analyse results effectively.

In the stage of review of literature, the researcher is expected to review various types of literature, which can include newspapers or magazine articles, journal articles, research reports, book chapters, project reports, working papers, books, conference proceedings, databases or government reports. In this context, the researcher should be careful in choosing the literature for the review. For example, a theoretical literature review will helpful in formulating the practical concepts, definitions and theories to be used in the field of research. Reviewing literature on methodological research will help the researcher to select and decide the specific method for studying the problem. Again, a review of empirical literature on the selected area will help in contextualizing the approach of the researcher. Overall, it can be said that this is a significant step in research as it will help the researcher in developing the arguments of the research questions.

4.5.3 Research Design

After you develop a research problem and review the existing literature, the next vital stage of your research is to prepare a design of the research. The term 'design' refers to drawing an outline and planning a strategy of conducting research. It refers to the overall strategy that the researcher chooses to integrate the different components of the study in a coherent and logical way, thereby, ensuring the effective addressing of the research problem. In the words of Ragin (1994), a research design *"is a plan for collecting and analysing evidence that will make it possible for the investigator to answer whatever questions he or she has posed. The design of an investigation touches almost all aspects of the research, from the minute details of data collection to the selection of the techniques of data analysis.'*

Research design serves as a "blueprint" for empirical research aimed at addressing specific research questions or testing hypotheses (Bryman, 2016). Decisions regarding when, what, when, how much, and how data is collected and analysed are crucial aspects of research design.

Research design outlines three important steps. They are:

- **The data collection process:** It involves deciding how you will gather the information needed for your research. For example, if you want to study how much time students spend on social media, you might decide to collect data through surveys. In that case, you would have to create a questionnaire asking students about their frequency of daily social media usage.
- **Instrument Development Process:** It refers to creating the tools or instruments you will use to collect your data. For example, as in the case of the example

stated above, the "instrument" for understanding social media usage by people, will be the survey -the questionnaire itself. Here, the "instrument" or the questionnaire should be designed carefully so as to ensure that the questions are clear and focussed on the objective the study has. For instance, you could include questions like, "How many hours do you spend on social media each day?*"* with options ranging from "less than 1 hour" to "more than 5 hours."

- **Sampling Process:** Sampling process is about selecting a sub-set of the population from whom you will collect data. For example, if you want to understand the social media habits of your college students, you may decide to take a sample of 100 students from different departments of your college. Then, this selected group of 100 students is your "sample" as it represents the larger population of all students in your college. You will learn more about the sample process in the rest of the units. It will also be discussed in the subsequent section.

4.5.4 Determining the Sample

After designing the research, the next step is to determine the sample of the study. It involves determining the sample from the total universe of study. While defining sample in social research, P. V. Young (1966) writes, 'a *statistical sample is miniature picture or cross section of the entire group or aggregate from which the sample is taken. The entire group from which sample is chosen is known as 'the population', 'universe' or 'supply'.'* So a sample is taken from the total population or universe of study. Each sample has some universe behind it. Here, the universe is understood as the aggregate of all units having certain characteristics (Young, 1966). So, a sample represents the universe. For example, the conceptual category of the universe in a study on 'farm crisis at small and marginal farmer level' refers to all the marginal and small farmers. The sample of the study may be small and marginal farmers located in a particular district or region. Another point to be noted is that, universe can be either finite or infinite. A finite universe has an identifiable number of elements. For instance, the voters of a particular country. Similarly, in the case of infinite universe, the number of elements is indeterminable or cannot be identified. For instance, the number of persons communicated via email in a day. It is something similar to stars in the sky which is difficult to be counted or leaves on a particular tree.

There are two essentials of sampling, that should be kept in mind while determining the sample:

- A sample should accurately represent the entire population. Each sample should reflecting the characteristics proportionately found in the overall population.

The value of a sample is determined by how well it represents the total population. For example, a teaspoon of *rasam* or *sambar*, which a housewife tastes to determine if the entire dish is evenly seasoned, shows a representative sample. If she finds that the salt is inadequate, she will add more salt, stir it thoroughly, and then test it again.

- The sample size should be sufficient to allow for accurate generalizations.

Do you know?

- The sample size, usually shown as the letter **n**, is the number of people or items from whom you gather the necessary information from.
- The sampling design or sampling technique is the method you use to select choose people or items from the larger group.
- The sampling frame is the list of everyone or everything in your population that you want to study. It's a complete list of everyone or everything you want to study. rephrase in simple way.

The choice of sampling design depends on the research design the researcher uses in order to address the research problem. The fundamental assumption of sampling is that a sample should be selected ideally in a way that it represents the entire group or universe. The sampling method can be either a probability or non-probability sample. Probability sampling is based on probability theory. Another alternative name for it is random sampling. It gives each population variable a known non-zero chance of selection. Non-probability sampling, also known as non-random sampling, is not founded on probability theory. This sampling does not give each population variable a chance to be chosen. The advantages of this method of sampling are its

simplicity, ease of use, and low cost.

Both probability and non-probability sampling has certain types. These are shown with the help of the diagram given below in figure 1.2. Each of these types of sampling will be discussed in later units.

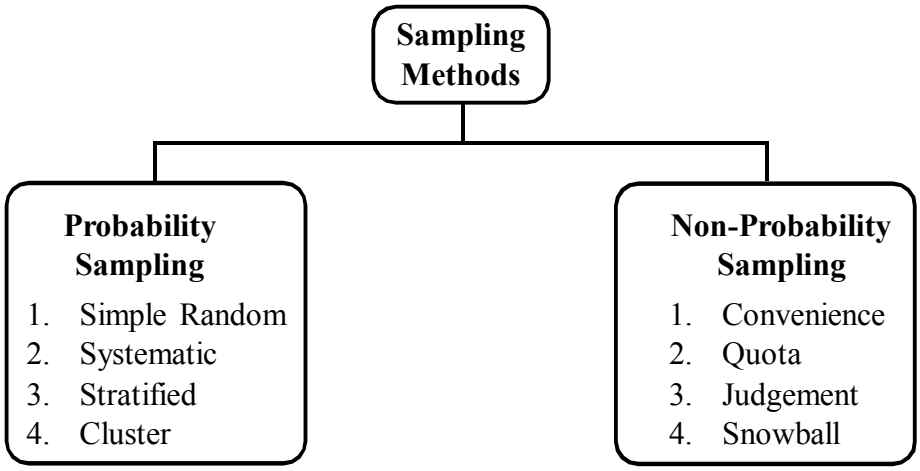


Fig 1.2

4.5.5 Carrying out the Research

After designing the research and determining the sample, the researcher decides the instruments and tools to be used. This selection of instruments such as questionnaires, interview schedules, case studies, group discussions, observations, or narratives should be developed based on the research question and design. A study might use multiple instruments to gather primary data from the field. During data collection, researchers might encounter unforeseen practical difficulties. Also, biasness can sometimes creep into data collection from respondents. For example, during a discussions, a researcher might become too involved in respondents' emotional views, leading to deviations from the research's objectives. Hence, researchers should be cautious to avoid unforeseen difficulties and biased responses during data collection.

4.5.6 Interpretation of the Data

After the research is carried out, the next step is to analyse and interpret the data. Data can be of two types: primary and secondary. Primary data refers to the first hand data gathered by the researcher himself/herself. Secondary data means data collected by someone else earlier. The method of data analysis depends on whether the information collected is quantitative or qualitative. Quantitative research uses numbers and statistics to answer questions about "what" and "how often"; qualitative research uses words and meanings to answer questions about "why" and "how". In this stage, the researcher tries to link the results with the research problem, then test the initial hypothesis, and determine whether the results support or reject it. After the data collection and

analysis, the researcher interprets the data to reach the generalizations. This interpretation helps the researcher to understand the significance of the study and determine whether answers to the research questions have been found.

4.5.7 Presentation of a Report

This is the final step in getting started with a research. The findings can be published in various formats such as a thesis, report, journal article, working paper, occasional paper, or book. It provides a detailed account of the research conducted and justify the research questions addressed. One of the main purposes of documenting the research is to legitimize the research by showing that the results are based on rigorously collected data rather than being arbitrary. When writing the report, the researcher should ensure clarity of thought and language, correct usage of concepts and terminology, clearly mentioning the research problem along with presentation of data and statistics. Sometimes, the findings and reports may sometimes highlight unanswered questions and suggest areas for further research.

Points to Recall:

- Research is logical, objective, and relies on empirical evidence to answer questions and solve problems.
- Social science applies systematic methods to analyse social life, aiming to discover new knowledge, relationships, and generalizations that explain and predict human interactions.
- The research process in social sciences includes identifying a research problem, reviewing related literature, designing the research, determining the sample, conducting the research, interpreting data, and presenting findings. Each step is interconnected and important for a coherent and effective study.

4.6 Conclusion

Research is a systematic and scientific process of inquiry in order to generate new knowledge, verifying existing theories, and solving problems. In the context of social science research, this process is crucial for understanding human behavior, cultural dynamics and social structures. Social science research uses both qualitative and quantitative methods to analyse social issues and contribute to policy-making, development, and academic discourse.

The research process follows a structured step-by-step approach. It starts with identifying and selecting a research problem, which forms the foundation of the study. This is followed by a review of related literature, which helps in understanding existing work, identifying the research gap and refining research objectives. The next step follows a well-defined research design which guides the methodology, ensuring accuracy and reliability. The next step, determining the sample, is very important for collecting data, followed by carrying out the research through data collection methods such as surveys, interviews, or experiments. Once data is gathered, interpretation of the data allows researchers to analyse patterns and draw conclusions. Finally, the research findings are presented through the presentation of a report, which organizes and synthesizes the results for academic, policy, or public engagement purposes.

Therefore, in conclusion, a research process is essential for the systematic exploration of social phenomena. By following the structured steps of research, scholars and practitioners contribute to a deeper understanding of social issues and promote evidence-based decision-making.

4.7 Summary

In this unit, we understood the essential elements of social science research. We explored the steps involved in the research process, including identifying a research problem, reviewing literature, designing a research framework, sampling, conducting research, and interpreting data. Each step has been emphasized for its importance in ensuring that research is methodologically sound and yields reliable results.

4.8 Questions

Answer in very short

1. What do you understand by research?
2. What is the purpose of a literature review in social research?
3. Why is sampling important in research?

Answer in short

1. Define social science research and its significance.
2. What are the key components of a research design?
3. Explain the importance of data interpretation in research.

Answer in detail

1. Describe the steps involved in identifying and selecting a research problem.
2. Explain the way in which a well-structured research report contribute to knowledge building.
3. Discuss the role of ethical considerations in carrying out social science research.

4.9 References

1. Babbie, E. (2016). *The Practice of Social Research.* Cengage Learning.
2. Bryman, A. (2016). *Social Research Methods.* Oxford University Press.
3 Crawford, C. C. (1928). *The Techniques of Research.* The Century Co.
4. Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* Sage Publications.
5. Hart, C. (1998). *Doing a Literature Review: Releasing the Social Science Research Imagination.* SAGE Publications.
6. Jenkins, R. (2014). *Social Identity* (3rd ed.). Routledge.
7. Kerlinger, F. N. (1973). *Foundations of Behavioral Research* (2nd ed.). Holt, Rinehart and Winston.
8. Young, P. V. (1966). *Scientific Social Surveys and Research* (4th ed.). Prentice Hall.

4.10 Suggested Readings

- Kothari, C. R. (2004). *Research Methodology: Methods and Techniques* (2nd Edition). New Age International Publishers.
- Sarantakos, S. (2012). *Social Research* (4th Edition). Palgrave Macmillan.
- Gerring, J. (2001). *Social Science Methodology: A Criterial Framework*. Cambridge University Press.
- Sinha, A. K. (2000). *Research Methodology in Social Sciences*. Commonwealth Publishers.

Unit 5 □ Constructs and Concepts; Conceptualization

Structure

5.1 Learning Objectives

- Understand the difference between concepts and constructs and their importance in academic research.
- Examine the characteristics of concepts and how they form the foundation of theoretical frameworks.
- Analyse the role of constructs in developing measurable variables for research studies.
- Apply concepts and constructs in formulating research questions and hypotheses effectively.

5.2 Introduction

Social Science Research needs a clear understanding of certain components. If these components are not properly understood, then doing a research will be a difficult process. This unit attempts to give an understanding of components like: concepts and constructs that are fundamental to any research project. Concepts are the building

blocks of research—they represent the ideas or phenomena you want to study. Constructs, on the other hand, are more complex ideas that are often composed of multiple concepts. For instance, "social inequality" is a construct that could encompass various concepts like income disparity, educational differences, and access to resources. Understanding these elements is crucial for defining what you are studying and ensuring that your research is focused and clear.

In the end, thus unit looks into the process of conceptualization, which is about refining and clarifying your research focus. By the end of this unit, you will have a clear understanding of how to get started with social research. It will equip with the knowledge to develop research questions, define and work with concepts and constructs, and engage in the process of conceptualization, all of which are critical for conducting meaningful research in the social sciences.

5.3 Concepts and Constructs

In sociology, concepts and constructs are two basic tools that are used to understand and analyse social phenomena. Concepts are basic ideas or categories that we use to make sense of the world. They help us describe and classify different aspects of society. They represent broad ideas that can be observed and discussed in various social contexts. Constructs, on the other hand, are more complex and abstract than concepts. They are ideas that have been developed through theoretical thinking and are often used to explain relationships between concepts.

5.3.1 Concepts

Concepts are fundamental components of the scientific methodology, with the majority of concepts being abstract representations that capture specific facets of reality. While defining concepts, P.V. Young (1966) says *"Each new class of data, isolated from other classes on the other basis of definite characteristics, is given name, a label in short hand concept"*. Hence, a concept serves as a precise definition of a category or set of observations. It functions as a theoretical representation that denotes an entity, an attribute of the entity, or a specific occurrence.

There are some characteristics of a concept. They are :

- Concepts are symbols attached to the meanings we hold.
- They represent only a part of reality.
- Different people may have different concepts of the same thing.
- Concepts represent various degrees of abstraction.

It is believed that a good concept has certain features and fulfil certain criteria:

- It should be clear, definite, and precise.
- It should provide comprehensive and clear information and understanding.
- It should avoid multiple meanings and convey exactly what was intended when they were coined.

There can be certain concepts which derive their meaning from specific theories, and when used in different theories, they may convey entirely different or even opposite meanings. They are called **concepts by postulation.** Let us take the example of the concept of class in social theory. In Marxian terms, class refers to the relationship individuals have to the means of production, such as the bourgeoisie and proletariat; whereas in Weberian sociology, class refers to a group of people with similar levels of wealth, power, and prestige (Ritzer, 2011). The second example is that of social Stratification. While in Marxist theory, social stratification refers to the hierarchical arrangement of individuals or groups based on their access to and control over economic resources (e.g., wealth, income); in functionalist theory, it refers to a system that ensures that the most qualified individuals occupy the most important positions, thereby promoting social stability and efficiency in society.

Similarly, there are also concepts that represent something that is immediately understood, with a meaning that remains constant regardless of who uses them. These are called **concepts by intuition.** For example, racism as a concept is intuitively understood across all cultures as a prejudice or discrimination directed against someone of a different race based on the belief that one's own race is superior. This concept remains largely consistent in sociology despite differing theoretical interpretations and approaches. Again, family as a concept is intuitively understood as a group of individuals related by blood, marriage, or adoption, and it holds a more or less similar meaning across different sociological theories. Both the types of concepts hold equal importance and significance in social science research.

Further, there are largely **two types of concepts** in social research:

- **Concrete Concepts***:* These symbolize material objects that can be seen, touched, and felt, such as books and tables. Examples include: social institutions like family, education system, religion, and government are concrete concepts in sociology. Social movement, such as civil rights movements, feminist movements, or environmental movements, are also examples of concrete concepts in sociology. These are visible and organized collective actions aimed at bringing about social or political change.

- **Abstract Concepts:** These refer to the properties or characteristics of objects, such as weight and height. Power in sociology is an example of this kind of concept. According to Weber, power refers to the ability of individuals or groups to achieve their goals despite opposition from others (Ritzer, 2011). It is an abstract concept because it involves authority, influence, and control that are not physical objects but rather social relations and dynamics. Similarly, identity is an abstract concept in sociology. It is understood as to how individuals perceive themselves and are perceived by others in relation to various social categories such as social class, gender, ethnicity or nationality (Jenkins, 2014). It includes self-concept, social roles, and the sense of belonging to specific groups or communities.

Did you Know?

- The word *concept* comes from the Latin *conceptus*, meaning "something conceived." It refers to an idea formed in the mind to represent reality.
- The study of concepts dates back to ancient philosophers. Plato, the ancient Greek philosopher, introduced the idea of "ideal forms," suggesting that concepts exist as perfect, unchanging entities in the realm of ideas.
- Concepts evolve over time. For example, the concept of *family* has now expanded beyond the traditional nuclear family to include same-sex families, single-parent families, and even cohabiting partners.

5.3.2 Constructs

A concept, as defined earlier, is an idea about some aspect of a phenomenon, such as self esteem or gender or bureaucracy or social classification and so on. Similarly, a construct is a verbal response prompted by objects belonging to the category to which the concept applies. While sometimes concepts like temperature, sound, age, and sex can be directly observed, others such as mental strength, drive, attitude, and motivation cannot be directly observed, as they are presumed to exist within the organism and are termed as constructs.

A construct refers to an abstract idea or concept that is specifically defined for the purpose of study. Constructs are not directly observable; rather are inferred from observable behaviours, attitudes, or phenomena (Ragin, 1994). They are used to explain or describe a theoretical idea or phenomenon and are crucial in developing hypotheses and theories within social sciences. Examples can include constructs like

"social class", "self-esteem," or "attitudes towards climate change", which are defined and operationalized to study their impact in research. Similarly **gender identity** is a construct which refers to an individual's deeply felt sense of their own gender. It may or may not correspond with the sex they were assigned at birth. It is a personal experience of gender that is not directly observable but is understood from behaviours, expressions, and self-identification. Researchers study gender identity in order to understand how individuals perceive and express their gender, and how societal norms and expectations influence these experiences. This construct helps in exploring the different ways people experience and articulate their gender beyond traditional binary notions.

Did you Know

- While all constructs are concepts, not all concepts are constructs. A concept is a general idea (for example: gender), while a construct is a specifically defined and measurable version of a concept used in research (for example: gender identity, measured through self-perception and social roles).
- The idea of gender as a binary (male/female) is a social construct. Many societies recognize non-binary and third-gender identities, challenging traditional notions of gender.

5.4 Conceptualization

Conceptualization is the process of defining and clarifying the concepts that will be studied or analysed in research. It involves taking an abstract idea and giving it a specific meaning within the context of your research. For example, if you are doing a study on "social inequality" you have to define what exactly you mean by the term—whether it is based on income or education or access to resources. Similarly, if you are studying or discussing "health", conceptualization would involve defining 'health – whether you are referring to physical health or mental health, or overall well-being? Also, does it include factors like diet, exercise, and stress management. So, you ensure that everyone understands exactly what aspect you are focusing on.

Conceptualization in social research involves clearly defining the ideas and terms that researchers want to study. This step is vital because social issues are often complicated, and researchers need a clear grasp of what they are investigating. Conceptualization

involves identifying the key variables, concepts, and constructs that will be studied and developing clear, concise definitions for them.

There are some important aspects of conceptualization. They are:

- **Identifying Concepts:** The research should begin by identifying the abstract concepts or ideas that are central to the research. These could include terms like social class, gender roles, cultural identity, etc.
- **Defining Concepts:** Once the concepts are defined, next step is to define the concepts clearly in a way that is suitable for empirical investigation.
- **Operationalization:** Moving from identifying concepts to conceptualizing and then operationalizing them involves increasing levels of specificity. Research starts with a broad area of interest, specifying the key concepts that are vital, working on defining those concepts, and finally, determining exactly they will be measured. In quantitative research, this final step is known as operationalization. For example, the concept of social anxiety cannot be directly measured. However, it can be operationalized in different ways. For instance, self-rating scores on a social anxiety scale; or a number of recent behavioural incidents of avoidance of crowded places; or intensity of physical anxiety symptoms in social situations.

  Operationalization involves identifying specific indicators which will represent the concepts the researcher aims to study (Babbie, 2016). For example, if you are interested in studying masculinity, indicators like: social roles traditionally associated with men, such as being a breadwinner or a father, can be considered. Such roles can serve as measures of masculinity in the study and the degree to which a man fulfils these roles can be seen as indicators of how masculine he is 'perceived' to be. In quantitative research, conceptualization involves crafting clear, concise definitions for our main concepts.

  It is also important to review existing research and theories to see how other scholars have defined the concepts that we are interested in. This does not mean the researchers have to adopt their definitions, but understanding how these concepts have been defined previously helps us compare and refine our own conceptualizations. It helps in guiding and deciding whether the study wants to challenge existing ideas or build on them for our research. Ultimately, refining our conceptualization helps to craft a specific and clear research question.

Concepts can be operationalized at different levels of measurement:

- **Nominal level:** This helps researchers to categorize data without implying any order or hierarchy. It is useful for identifying and classifying variables like gender, ethnicity, or political affiliation.
- **Ordinal Level:** Here, categories not only differ in quality but most importantly, they have a logical order or rank. For instance, socioeconomic status or educational levels are often measured at this level. Such a ranking enables the researchers to analyse trends or correlations in social hierarchies or preferences
- **Interval Level:** This measurement level assumes equal intervals between values, but lacks a real zero point. Measurements such as temperature in Celsius illustrate the interval data. In social research, interval scales help in statistical operations such as means and standard deviation calculations, and ensure more precise comparisons and analysis of attitudes, perceptions, or scale assessments.
- **Ratio Level:** This is the highest level of measurement with absolute zero points and the same intervals between the values. Variables like: income, age or number of children are usually measured at a ratio level. The ratio data support all statistical operations, including ratio comparison, percentage calculation and complex mathematical operations, and provide detailed information on social phenomena like: economic differences or demographic trend.

Understanding and utilizing the different levels of measurement—nominal, ordinal, interval, and ratio—is beneficial in research for many reasons:

Firstly, it helps in accurate data classification and organization. Each level of measurement helps researchers structure and classify their data effectively. For example, nominal measurement allows for simple categorization of variables without any ranking, making it ideal for basic identification like gender, ethnicity, or political affiliation; whereas ordinal measurement introduces a logical order or rank, which helps in more complex analyses, such as understanding social hierarchies or preferences (e.g., education levels or income categories).

Secondly, it helps in improved data analysis and interpretation. As the measurement levels progress, the potential for advanced data analysis improves. Interval and ratio measurements allow for detailed statistical calculations, such as mean, median, standard deviation, and correlation analysis. Interval scales, despite lacking a true zero point, help in comparison of differences between values (e.g., in attitudes or temperature), while ratio scales, with a true zero,

allow for precise comparisons and mathematical operations, such as calculating income distribution or demographic trends.

Thirdly, it increases research accuracy and validity. By using various measurement levels researchers can apply appropriate statistical tools and tests, enhancing the validity and reliability of their findings. For example, using ordinal data for ranking variables allows for trend analysis, while ratio data gives the most precise insight into economic or demographic differences, contributing to more accurate results.

Points to Recall

- While concepts are general and broad, constructs are more defined and used in empirical research with measurable indicators.
- Conceptualization involves defining and refining concepts/constructs to ensure clarity and consistency in research.
- Well-defined concepts and constructs help in developing theories, designing research, and ensuring valid measurement.
- Conceptualization leads to operationalization, where abstract ideas are turned into specific variables and indicators for data collection.

5.5 Conclusion

In social science research, a clear understanding of concepts, constructs, and conceptualization is essential for conducting meaningful and structured studies. On one hand, concepts are the foundational building blocks of research, providing broad ideas that help categorize and interpret social phenomena; on the other hand, constructs are more abstract and are used to measure theoretical ideas by linking observable behaviours or attitudes with underlying social realities.

The process of conceptualization refines and defines these concepts, ensuring clarity in research focus. This process is further supported by operationalization, which helps to translate abstract concepts into measurable variables, making empirical research possible. Moreover, understanding levels of measurement—nominal, ordinal, interval, and ratio—is very important for accurate data collection and analysis. These levels help in categorizing, ranking, comparing, and statistically analysing data.

5.6 Summary

This unit explores the fundamental components of social science research—concepts, constructs, and conceptualization. Concepts are basic ideas used to categorize and analyse social phenomena, while constructs are abstract ideas inferred from observable behaviours. Conceptualization is the process of defining and clarifying these concepts for research, ensuring precision and consistency. Operationalization further refines concepts into measurable variables. The unit also covers levels of measurement—nominal, ordinal, interval, and ratio—which help in data classification, ranking, and statistical analysis. By understanding these fundamental components, researchers can develop a strong theoretical framework, formulate clear research questions, and conduct systematic investigations that contribute to a deeper understanding of social issues.

5.7 Questions

Answer in very short

1. What is a concept in research?
2. Define a construct.
3. What is conceptualization?

Answer in Short

1. How do concepts and constructs differ in research? Give examples.
2. Why is conceptualization important in social science research?
3. Explain the role of constructs in theory development.

Answer in detail

1. Discuss the process of conceptualization with examples.
2. How do concepts and constructs contribute to the formulation of research hypotheses?
3. Explain how conceptualization helps in defining variables in research studies.

5.8 References

1. Babbie, E. (2016). *The Practice of Social Research.* Cengage Learning.
2. Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* Sage Publications.
3. Jenkins, R. (2014). *Social Identity* (3rd ed.). Routledge.
4. Moser, C. A., & Kalton, G. (1971). *Survey Methods in Social Investigation* (2nd ed.). Heinemann Educational Books.
5. Neuman, W. L. (2013). *Social Research Methods: Qualitative and Quantitative Approaches.* Pearson Education.
6. Ragin, C. C. (1994). *Constructing Social Research: The Unity and Diversity of Method.* Pine Forge Press.
7. Ritzer, G. (2011). *Sociological Theory* (5th ed.). Tata McGraw Hill.
8. Young, P. V. (1966). *Scientific Social Surveys and Research* (4th ed.). Prentice Hall.

5.9 Suggested Readings

- Sartori, G. (1984). *Social Science Concepts: A Systematic Analysis.* SAGE Publications.
- Gerring, J. (2012). *Social Science Methodology: A Unified Framework* (2nd ed.). Cambridge University Press.
- Sinha, A. K. (2000). *Research Methodology in Social Sciences*. Commonwealth Publishers.
- Kothari, C. R. (2004). *Research Methodology: Methods and Techniques* (2nd Edition). New Age International Publishers.

Unit 6 □ Understanding Measurement and Operationalization in Research

Structure

6.1 Learning Objectives

By studying the following unit, the student will be able to:

- Define the concept of measurement in social science and explain its significance in empirical research.
- Describe the process of operationalization and its role in translating abstract concepts into measurable variables.
- Identify key challenges associated with operationalization and discuss strategies to address them.
- Evaluate the importance of measurement accuracy and consistency in maintaining methodological rigor in social science research.
- Apply principles of operationalization to develop measurable indicators for complex social constructs.

6.2 Introduction

Research in various fields relies on accurate data collection and interpretation. To ensure meaningful results, researchers must use precise measurement techniques and properly operationalize abstract concepts. Measurement and operationalization are fundamental to empirical research, as they determine how variables are defined and assessed. This paper explores these concepts, their significance, and their application in different research domains.

6.3 Measurement in Research

Measurement in social science research is essential for translating abstract concepts into quantifiable variables. This process enables researchers to systematically investigate complex social phenomena, test theories, and generate empirical evidence. The precision and accuracy of these measurements hinge on three critical components: operationalization, validity, and reliability. This paper delves into these aspects, highlighting their significance, challenges, and best practices to ensure methodological rigor in social science research. There are three main reasons for the preoccupation with measurement in quantitative research.

- Measurement allows us to delineate *fine differences* between people in terms of the characteristic in question. This is very useful, since, although we can often distinguish between people in terms of extreme categories, finer distinctions are much more difficult to recognize. We can detect clear variations in levels of job satisfaction—people who love their jobs and people who hate their jobs—but small differences are much more difficult to detect.
- Measurement gives us a *consistent device* or yardstick for making such distinctions. A measurement device provides a consistent instrument for gauging differences. This consistency relates to two things: our ability to be consistent over time and our ability to be consistent with other researchers. In other words, a measure should be something that is influenced neither by the timing of its administration nor by the person who administers it. Obviously, saying that the measure is not influenced by timing is not meant to indicate that measurement readings do not change: they are bound to be influenced by the process of social change. What it means is that the measure should generate consistent results, other than those that occur as a result of natural changes.

- Measurement provides the basis for *more precise estimates of the degree of relationship between concepts.* Thus, if we measure both job satisfaction and the things with which it might be related, such as stress-related illness, we will be able to produce more precise estimates of how closely they are related than if we had not proceeded in this way.

In order to provide a measure of a concept, often referred to as an operational definition, a term deriving from the idea of operationalization, it is necessary to have an indicator or indicators that will stand for the concept. There are a number of ways in which indicators can be devised. It may be through a question (or series of questions) that is part of a structured interview schedule or self-completion questionnaire; the question(s) could be concerned with the respondents' report of an attitude (for example, job satisfaction) or their social situation (for example, poverty) or a report of their behaviour (for example, leisure pursuits). Second, it may be through the recording of individuals' behaviour using a structured observation schedule (for example, pupil behaviour in a classroom). It may also be through official statistics, such as the use of Home Office crime statistics to measure criminal behaviour. Lastly, it may be through an examination of mass media content through content analysis—for example, to determine changes in the salience of an issue, such as AIDS, in the mass media.

Indicators, then, can be derived from a wide variety of different sources and methods. Very often the researcher has to consider whether one indicator of a concept will be sufficient. This consideration is frequently a focus for social survey researchers. Rather than have just a single indicator of a concept, the researcher may feel that it may be preferable to ask a number of questions in the course of a structured interview or a self-completion questionnaire that tap a certain concept.

6.4 Levels of Measurement

In the social sciences, four types of scales for measuring a variable have been delineated. Out of the four types, two types are for categorical variables while the other two types are for numerical variables. These scale types or levels of measurement, are useful in helping to classify and catalog variables in a study as well as in designing questions to measure variables. The four basic levels of measurement scales are utilized by social and physical scientists. These include the nominal scale, the ordinal scale, the interval scale and the ratio scale. Each type of scale has unique characteristics and implications for the other statistical procedures that can be used with it.

i. **Nominal Scale** – Nominal scales do not involve highly complex measurement, but rather involve rules for placing individuals or objects into categories. A variable with a nominal level of measurement consists of a set of distinctive categories that imply no specific order. Consider the variable of gender. This variable can take only two forms: male and female. This is no real order between the categories; respondents must simple be one or the other. Therefore, a nominal variable must have at least two categories, but may have as many as needed and these categories must be characterized as having no prescribed order. These categories must namely (i) be homogeneous, (ii) be mutually exclusive and (iii) make no assumption about ordered relationships between categories.

ii. **Ordinal Scale** – This kind of measurement where we move into a higher level of measurement, we encounter variables in which the categories do represent a rank-ordered series of relationships. Variables that have two or more categories with an inherent order among them are measured at an ordinal level of measurement. Thus, the categories in ordinal scales are not only homogeneous and mutually exclusive but they stand in some kind of relation to one another. In addition to having the symmetrical properties of the nominal scale, an ordinal scale is asymmetrical in the sense that certain special relationships may hold between A and B which do not hold for B and A. For example, the relationship greater than (>) is asymmetric in that if A>B, it cannot be true that B>A. transitivity still holds: if A>B and B>C, then A>C. It is these properties which enable us to place A, B, C along a single continuum.

iii. **Interval Scale** – It is readily apparent that an interval scale level of measurement requires the establishment of some sort of physical unit of measurement that can be agreed upon as a common standard and which is replicable i.e. can be applied over and over again with the same results. Length is measured in terms of feet or meters, time in seconds, temperature in degrees of Fahrenheit or centigrade, weight in pounds or grams and income in dollars. On the other hand, there are no such units of intelligence, authoritarianism or prestige which can be agreed upon by all social scientists and which can be assumed to be constant from one situation to the next. Given a unit of measurement, it is possible to say that the difference between two scores is twenty units or that one difference is twice as large as a second. This means that it is possible to add or subtract scores in an analogous manner to the way we can add weights on a balance or subtract 6 inches from a board by sawing it into two. Similarly,

we can add the incomes of husband and wife, whereas it makes no sense to add their IQ scores.

iv. **Ratio Scale** – When it is possible to locate an absolute or non-arbitrary zero point on the scale, we have a somewhat higher level of measurement referred to as a ratio scale. In this case we are able to compare scores by taking their ratios. Therefore, it can be said that, a ratio scale encompasses all the qualities of the earlier forms of scales: it must have more than one category, it must have an implicit order and it must be able to determine the exact distance between the intervals. In addition, however, it must have a true zero point. At the same time it should also be noted that ratio level of measurement can be applied to either continuous or discrete variables.

6.5 Operationalization

Operationalization is the process of defining abstract concepts in measurable terms. This involves identifying specific indicators that can be observed and quantified, allowing researchers to systematically study phenomena that are not directly observable. The term operationalization is often commonly used with the term of conceptualization as both the terms are intimately linked to each other. Conceptualization is the refinement and specification of abstract concepts, and operationalization is the development of specific research procedures (operations) that will result in empirical observations representing those concepts in the real world. For example, operationalizing the concept of "social capital" may involve measuring the number of social interactions, the presence of supportive relationships, and participation in community activities. As with the methods of data collection, social researchers have a variety of choices when operationalizing a concept. But it must be realized that operationalization does not proceed through a systematic checklist.

- **Range of Variation:** In operationalizing any concept, researchers must be clear about the range of variation that interests them. The question is, to what extent are they willing to combine attributes in fairly gross categories? Suppose you are interested in people's attitudes toward expanding the use of nuclear power generators. You would anticipate that some people consider nuclear power the greatest thing since the wheel, whereas other people have absolutely no interest in it. Given that anticipation, it would seem to make sense to ask people how much they favour expanding the use of nuclear energy and to give them answer categories ranging from "Favor it

very much" to "Don't favour it at all." This operationalization, however, conceals half the attitudinal spectrum regarding nuclear energy. Many people have feelings that go beyond simply not favouring it: They are, with greater or lesser degrees of intensity, actively opposed to it. In this instance, there is considerable variation on the left side of zero. Some oppose it a little, some quite a bit, and others a great deal. To measure the full range of variation, then, you'd want to operationalize attitudes toward nuclear energy with a range from favouring it very much, through no feelings one way or the other, to opposing it very much. Finally, decisions on the range of variation should be governed by the expected distribution of attributes among the subjects of the study.

- **Variations between the Extremes:** Degree of precision is a second consideration in operationalizing variables. What it boils down to is how fine you will make distinctions among the various possible attributes composing a given variable. Does it matter for your purposes whether a person is 17 or 18 years old, or could you conduct your inquiry by throwing them together in a group labelled 10 to 19 years old? Don't answer too quickly. If you wanted to study rates of voter registration and participation, you would definitely want to know whether the people you studied were old enough to vote. In general, if you're going to measure age, you must look at the purpose and procedures of your study and decide whether fine or gross differences in age are important to you. In a survey, you will need to make these decisions in order to design an appropriate questionnaire. In the case of in-depth interviews, these decisions will condition the extent to which you probe for details.

6.6 Steps in Operationalization

1. Conceptual Definition: Clearly defining the construct (e.g., intelligence as cognitive ability).
2. Identifying Indicators: Selecting observable elements that represent the construct (e.g., IQ scores for intelligence).
3. Developing Measurement Tools: Creating surveys, tests, or observational methods to assess indicators.
4. Testing and Refinement: Ensuring that the operationalized measures are reliable and valid through pilot studies and statistical analysis.

6.7 Challenges in Operationalization

Operationalization is the process of defining abstract concepts in measurable terms, which is essential for empirical research in social sciences. However, this process presents a few challenges that must be addressed.

- **Abstract Nature of Concepts:** Social science concepts are often abstract and multifaceted, making it difficult to capture all dimensions through measurable indicators.
- **Contextual Variability:** The meaning and manifestation of concepts can vary across different contexts, necessitating context-specific operational definitions.
- **Subjectivity:** The process involves subjective judgment, which can introduce bias and affect the consistency of measurements.

6.8 Application in Different Research Fields

Measurement and operationalization are essential components of empirical research, allowing researchers to transform abstract concepts into quantifiable data. Proper application of these principles enhances the accuracy, reliability, and validity of research findings. Following are a few examples where measurement and operationalization are used.

i. Social Sciences: Operationalizing social constructs like trust and political ideology.
ii. Psychology: Measuring personality traits using standardized tests.
iii. Healthcare: Quantifying pain levels using patient self-reports and physiological indicators.
iv. Business Research: Assessing customer satisfaction through surveys and behavioural analysis.

6.9 Conclusion

Measurement in social science research is a complex and critical process that requires careful consideration of operationalization, validity, and reliability. By addressing the challenges and employing best practices in these areas, researchers can enhance the

precision and accuracy of their measurements, contributing to the credibility and impact of their research. This paper provides valuable insights and practical guidance for researchers seeking to ensure methodological rigor in their studies, ultimately advancing the field of social science and enhancing our understanding of complex social phenomena.

6.10 Summary

Measurement is a cornerstone of social science, providing the means to quantify abstract concepts and phenomena, thereby facilitating empirical investigation and theory testing. The unit discusses operationalization, the process of defining abstract concepts in measurable terms. This involves identifying specific indicators and developing procedures to systematically observe and record data, ensuring that complex social constructs such as intelligence, socioeconomic status, and social capital are translated into quantifiable variables. By emphasizing the importance of careful measurement practices and the challenges associated with operationalization, this paper contributes to the methodological rigor and integrity of social science research. It provides valuable insights and practical guidance for researchers seeking to ensure that their measurements are both accurate and consistent, thereby enhancing the overall quality and impact of their work.

6.11 Questions

Answer the following questions in your own words as far as possible.

1. What role does measurement play in social science research?
2. How does the process of operationalization help in defining abstract concepts?
3. What are some examples of social constructs that require operationalization?
4. What challenges might researchers face when operationalizing abstract concepts?
5. Why is it important for measurements in social science to be accurate and consistent?
6. How does careful measurement contribute to the overall quality and impact of research?
7. What steps can researchers take to ensure methodological rigor in their measurement practices?

6.12 References

1. Babbie, Earl (2014). The Practice of Social Research, 13th Edition, Chapman University.
2. Baker, Therese. L (1999). *Doing Social Research*. 3rd Edition McGraw Hill College.
3. Bryman, Alan (2012). Social Research Methods, 4th Edition, Oxford University Press.
4. Blalock, Hubert M. (1972): Social Statistics (Second Edition), Japan: McGraw-Hill Kogakusha, Ltd.
5. Elifson, Kirk W. (1990): Fundamentals of Social Statistics (Second Edition), Singapore: McGraw-Hill Publishing Company.

Unit 7 □ Ensuring Validity and Reliability in Research

Structure

7.1 Learning Objectives

By studying the following unit, the student will gain knowledge about:

- The concepts of validity and reliability in social science measurement.
- To differentiate between various types of validity, including content validity, construct validity, and criterion validity.
- To explain the significance of ensuring validity in social science research and its impact on research credibility.
- To identify different forms of reliability, such as test-retest reliability, inter-rater reliability, and internal consistency.
- Analyzing the relationship between reliability and validity in the context of social science measurement.

- Evaluating potential challenges in achieving both validity and reliability and propose strategies to enhance measurement accuracy.
- Applying principles of validity and reliability to assess and improve research instruments in social science.

7.2 Introduction

Validity and reliability are two fundamental principles of research that determine the accuracy, consistency, and credibility of findings. Ensuring that a study accurately measures what it intends to and that its results are replicable is crucial for meaningful and impactful research. This paper explores the concepts of validity and reliability, their types, their significance in research, and methods to enhance them.

7.3 Reliability

Reliability refers to the consistency of a measure. A reliable measure produces stable and consistent results over time and across different contexts. Reliability is concerned with the question of whether the results of a study are repeatable. The term is commonly used in relation to the question of whether the measures that are devised for concepts in the social sciences (such as poverty, racial prejudice, deskilling, religious orthodoxy) are consistent. Reliability is particularly at issue in connection with quantitative research. The quantitative researcher is likely to be concerned with the question of whether a measure is stable or not. After all, if we found that IQ tests, which were designed as measures of intelligence, were found to fluctuate, so that people's IQ scores were often wildly different when administered on two or more occasions, we would be concerned about it as a measure. We would consider it an unreliable measure—we could not have faith in its consistency. There are at least three different meanings of the term reliability.

- **Stability:** The most obvious way of testing for the stability of a measure is the *test–retest* method. This involves administering a test or measure on one occasion and then readministering it to the same sample on another occasion—that is:

T1	T2
Obs1	Obs2

  We should expect to find a high correlation between Obs1 and Obs2. Correlation is a measure of the strength of the relationship between two variables.

- **Internal Reliability:** This meaning of reliability applies to multiple-indicator measures. When you have a multiple-item measure in which each respondent's answers to each question are aggregated to form an overall score, the possibility is raised that the indicators do not relate to the same thing; in other words, they lack coherence. We need to be sure that all our designerism indicators are related to each other. If they are not, some of the items may actually be unrelated to designerism and therefore indicative of something else. One way of testing internal reliability is the *split-half* method.

  It is usually expected that a result of 0.80 and above implies an acceptable level of internal reliability. Do not worry if the figures appear somewhat opaque. The meaning of correlation will be explored in much greater detail later on. The chief point to carry away with you at this stage is that the correlation establishes how closely respondents' scores on the two groups of indicators are related.
- **Inter-Observer Consistency:** When a great deal of subjective judgement is involved in such activities as the recording of observations or the translation of data into categories and where more than one 'observer' is involved in such activities, there is the possibility that there is a lack of consistency in their decisions. This can arise in a number of contexts, for example: in content analysis where decisions have to be made about how to categorize media items; when answers to open questions have to be categorized; or in structured observation when observers have to decide how to classify subjects' behaviour.

The above meanings of reliability can be further developed into the various types of reliability.

7.4 Types of Reliability

- **Test-Retest Reliability:** Sometimes it is appropriate to make the same measurement more than once, a technique called the test-retest method. This type of reliability assesses the stability of a measure over time by administering the same measure to the same group at different points in time. If you don't expect the sought-after information to change, then you should expect the same response both times. If answers vary, the measurement method may, to the extent of that variation, be unreliable.

- **Using Established Measures:** Another way to help ensure reliability in getting information from people is to use measures that have proved their reliability in previous research. If you want to measure anomia, for example, you might want to follow Srole's lead. The heavy use of measures, though, does not guarantee their reliability. For example, the Scholastic Assessment Tests (SATs) and the Minnesota Multiphasic Personality Inventory (MMPI) have been accepted as established standards in their respective domains for decades. In recent years, though, they've needed fundamental overhauling to reflect changes in society, eliminating outdated topics and gender bias in wording.
- **Internal Consistency (Split-Half Method):** As a general rule, it's always good to make more than one measurement of any subtle or complex social concept, such as prejudice, alienation, or social class. This procedure lays the groundwork for another check on reliability. The Split-Half Method evaluates the extent to which items within a measure are correlated, indicating they are measuring the same underlying concept. Let's say you have created a questionnaire that contains ten items you believe measure prejudice against women. Using the split-half technique, you would randomly assign those ten items to two sets of five. Each set should provide a good measure of prejudice against women, and the two sets should classify respondents the same way. If the two sets of items classify people differently, you most likely have a problem of reliability in your measure of the variable.

7.5 Challenges in Ensuring Reliability

Ensuring reliability in social science research is crucial for achieving consistent and dependable results. Reliability refers to the consistency of a measure, meaning that the same results should be obtained when the measure is repeated under identical conditions. However, the following challenges can impede the attainment of high reliability.

- **Variability:** Social phenomena can vary over time and across different contexts, affecting the stability of measurements. Human behaviour is also inherently variable and can be influenced by numerous factors, including mood, context, and external circumstances.
- **Subjectivity:** Many social science measures involve subjective judgments, such as ratings, interpretations, and self-reports. Such measures are prone to variability and inconsistency.

- **Contextual Influences:** The context in which data is collected can influence responses. Factors such as the physical environment, social setting, and timing can affect participant behaviour and responses.

By addressing these challenges through careful planning, standardized procedures, and ongoing evaluation, researchers can enhance the reliability of their measurements, leading to more consistent and trustworthy results in social science research.

7.6 Validity

Validity refers to the extent to which a measurement accurately reflects the concept it is intended to measure. Ensuring validity is crucial for the credibility of research findings and the advancement of social science knowledge. Validity is concerned with the integrity of the conclusions that are generated from a piece of research. It is important to be aware of the main types of validity that are typically distinguished. **Measurement validity** applies primarily to quantitative research and to the search for measures of social scientific concepts. Measurement validity is also often referred to as *construct validity*. Essentially, it is to do with the question of whether a measure that is devised of a concept really does reflect the concept that it is supposed to be denoting. It should be appreciated that measurement validity is related to reliability: if a measure of a concept is unstable in that it fluctuates and hence is unreliable, it simply cannot be providing a valid measure of the concept in question. In other words, the assessment of measurement validity presupposes that a measure is reliable. If a measure is unreliable because it does not give a stable reading of the underlying concept, it cannot be valid, because a valid measure reflects the concept it is supposed to be measuring. **Internal validity** relates mainly to the issue of **causality**. Internal validity is concerned with the question of whether a conclusion that incorporates a causal relationship between two or more variables holds water. Internal validity raises the question: can we be sure that national religiosity really does cause variation in religious orientation and that this apparent causal relationship is genuine and not produced by something else? In discussing issues of causality, it is common to refer to the factor that has a causal impact as the **independent variable** and the effect as the **dependent variable.** In the case of Kelley and De Graaf 's research, the 'religious environment of a nation' was an independent variable and 'religious belief' was the dependent variable. Thus, internal validity raises the question: how confident can we be that the independent variable really is at least in part responsible for the variation that has been identified in the

dependent variable? **External validity** is concerned with the question of whether the results of a study can be generalized beyond the specific research context. **Ecological validity** is concerned with the question of whether social scientific findings are applicable to people's every day, natural social settings. This criterion is concerned with the question of whether social research sometimes produces findings that may be technically valid but have little to do with what happens in people's everyday lives. If research findings are ecologically *in*valid, they are in a sense *artefact of the social scientist's arsenal* of data collection and analytic tools. The more the social scientist intervenes in natural settings or creates unnatural ones, such as a laboratory or even a special room to carry out interviews, the more likely it is that findings will be ecologically invalid. The findings deriving from a study using questionnaires may have measurement validity and a reasonable level of internal validity, and they may be externally valid, in the sense that they can be generalized to other samples confronted by the same questionnaire, but the unnaturalness of the fact of having to answer a questionnaire may mean that the findings have limited ecological validity.

7.7 Types of Validity

- **Face Validity:** At the very minimum, a researcher who develops a new measure should establish that it has **face validity**—that is, that the measure apparently reflects the content of the concept in question. Face validity might be established by asking other people whether the measure seems to be getting at the concept that is the focus of attention. In other words, people, possibly those with experience or expertise in a fi eld, might be asked to act as judges to determine whether on the face of it the measure seems to reflect the concept concerned. Face validity is, therefore, an essentially intuitive process.
- **Criterion Validity:** Examines how well one measure predicts an outcome based on another established measure. This can be further divided into concurrent validity (comparison with an existing measure) and predictive validity (ability to predict future outcomes).
- **Concurrent Validity:** The researcher might seek also to gauge the concurrent validity of the measure. Here the researcher employs a criterion on which cases (for example, people) are known to differ and that is relevant to the concept in question. A new measure of job satisfaction can serve as an example. A criterion might be absenteeism, because some people are more often absent from work

(other than through illness) than others. In order to establish the concurrent validity of a measure of job satisfaction, we might see how far people who are satisfied with their jobs are less likely than those who are not satisfied to be absent from work. If a lack of correspondence were found, such as there being no difference in levels of job satisfaction among frequent absentees, doubt might be cast on whether our measure is really addressing job satisfaction.

- **Predictive Validity:** Another possible test for the validity of a new measure is *predictive validity*, whereby the researcher uses a *future* criterion measure, rather than a contemporary one, as in the case of concurrent validity. With predictive validity, the researcher would take future levels of absenteeism as the criterion against which the validity of a new measure of job satisfaction would be examined. The difference from concurrent validity is that a future rather than a simultaneous criterion measure is employed.
- **Content Validity:** Content validity refers to how much a measure covers the range of meanings included within a concept. This involves ensuring that all relevant aspects of the concept are represented in the measurement. For example, a test of mathematical ability cannot be limited to addition but also needs to cover subtraction, multiplication, division, and so forth.
- **Construct Validity:** Some writers advocate that the researcher should also estimate the *construct validity* of a measure. Here, the researcher is encouraged to deduce hypotheses from a theory that is relevant to the concept. For example, drawing upon ideas about the impact of technology on the experience of work, the researcher might anticipate that people who are satisfied with their jobs are less likely to work on routine jobs; those who are not satisfied are more likely to work on routine jobs. Accordingly, we could investigate this theoretical deduction by examining the relationship between job satisfaction and job routine. However, some caution is required in interpreting the absence of a relationship between job satisfaction and job routine in this example. First, either the theory or the deduction that is made from it might be misguided. Second, the measure of job routine could be an invalid measure of that concept.

7.8 Challenges in Ensuring Validity

Ensuring validity in social science research is critical for producing accurate and credible results. Validity refers to the degree to which a measurement accurately

reflects the concept it is intended to measure. However, achieving high validity may be challenging due to a few factors.

- **Complexity of Social Phenomena:** Social phenomena are inherently complex, involving multiple interrelated factors. Capturing the full breadth and depth of these phenomena in a single measure is challenging.
- **Changing Contexts:** The meaning and relevance of concepts can change over time and across different contexts, affecting the validity of measurements.
- **Bias:** Researcher bias, cultural bias, and respondent bias can all impact the validity of measurements.
- **Measurement Error:** Measurement error refers to the difference between the true value and the observed value of a variable. Errors can arise from various sources, including poorly designed instruments, data collection issues, and respondent misunderstandings.

By recognizing and addressing these challenges, researchers can enhance the validity of their measurements, leading to more accurate, reliable, and impactful social science research.

7.9 Interplay Between Validity and Reliability

While a measure must be reliable to be valid, a reliable measure is not necessarily valid. This interplay highlights the importance of ensuring both reliability and validity to achieve accurate and meaningful measurements in social science research. There are, then, a number of different ways of investigating the merit of measures that are devised to represent social scientific concepts. However, the discussion of reliability and validity is potentially misleading, because it would be wrong to think that all new measures of concepts are submitted to the rigours described above. In fact, most typically, measurement is undertaken within a stance that Cicourel (1964) described as ‘measurement by fiat’. By the term ‘fiat’, Cicourel was referring to the notion of ‘decree’. He meant that most measures are simply asserted. Fairly straightforward but minimal steps may be taken to ensure that a measure is reliable and/or valid, such as testing for internal reliability when a multiple indicator measure has been devised and examining face validity. But in many if not the majority of cases in which a concept is measured, no further testing takes place. It should also be borne in mind that, although reliability and validity are analytically

distinguishable, they are related because validity presumes reliability. This means that, if your measure is not reliable, it cannot be valid. This point can be made with respect to each of the three criteria of reliability that have been discussed. If the measure is not stable over time, it simply cannot be providing a valid measure. The measure could not be tapping the concept it is supposed to be related to if the measure fluctuated. If the measure fluctuates, it may be measuring different things on different occasions. If a measure lacks internal reliability, it means that a multiple-indicator measure is actually measuring two or more different things. Therefore, the measure cannot be valid. Finally, if there is a lack of inter-observer consistency, it means that observers cannot agree on the meaning of what they are observing, which in turn means that a valid measure cannot be in operation.

7.10 Strategies for Enhancing both Validity and Reliability

Ensuring the accuracy and consistency of measurements in social science research is paramount for producing credible and reliable results. Some key strategies for enhancing both validity and reliability have been discussed below.

- **Pre-Testing and Pilot Studies:** Conducting thorough pre-testing and pilot studies helps identify and rectify potential issues with measurement tools before full-scale data collection. This process can reveal ambiguities, biases, and inconsistencies in the instruments, allowing for adjustments to enhance both validity and reliability.
- **Use of Established Instruments:** Employing measurement instruments that have been validated and demonstrated to be reliable in previous research ensures that the tools are both accurate and consistent. These instruments have undergone rigorous testing and refinement, providing a solid foundation for new studies.
- **Methodological Rigor:** Apply rigorous methodological standards and procedures to ensure the accuracy and consistency of measurements.

By implementing these strategies, researchers can significantly enhance the validity and reliability of their measurements, thereby producing more robust and credible social science research. These practices ensure that the findings accurately reflect the concepts being studied and can be consistently replicated across different studies and contexts.

7.11 Conclusion

Validity and reliability are essential components of high-quality research. While reliability ensures consistency, validity guarantees that research findings are meaningful and applicable. By implementing rigorous methodological practices, researchers can enhance both aspects, leading to more credible and impactful studies.

7.12 Summary

Measurement is a cornerstone of social science, providing the means to quantify abstract concepts and phenomena, thereby facilitating empirical investigation and theory testing. The above unit explores validity, the extent to which a measurement accurately reflects the concept it is intended to measure. Various types of validity are examined, including content validity, which assesses whether a measure covers the full range of the concept; construct validity, which evaluates whether a measure relates to other variables as theoretically expected; and criterion validity, which examines how well one measure predicts an outcome based on another established measure. Ensuring validity is crucial for the credibility of research findings and the advancement of social science knowledge. Reliability, the consistency of a measure, is another focal point of the unit. The discussion encompasses different forms of reliability, such as test-retest reliability, which assesses the stability of a measure over time; inter-rater reliability, which examines the consistency between different observers; and internal consistency, which evaluates the extent to which items within a measure are correlated, indicating they are measuring the same underlying concept. High reliability is essential for the replicability of research and the confidence in the results obtained. By emphasizing the importance of careful measurement practices and the challenges associated with validity, and reliability, this unit contributes to the methodological rigor and integrity of social science research. It provides valuable insights and practical guidance for researchers seeking to ensure that their measurements are both accurate and consistent, thereby enhancing the overall quality and impact of their work.

7.13 Questions

Answer the following in your own words as far as possible.

1. Why is measurement considered a cornerstone of social science research?

2. How do different types of validity (content, construct, and criterion) contribute to the accuracy of a measurement?
3. In what ways can a researcher ensure that their measurement is valid?
4. What are the key differences between validity and reliability in social science research?
5. How do various forms of reliability (test-retest, inter-rater, and internal consistency) help maintain consistency in measurement?
6. Why is high reliability essential for the replicability of research findings?
7. What challenges might researchers face when trying to achieve both validity and reliability?
8. How can researchers improve the methodological rigor of their studies through careful measurement practices?

7.14 References

1. Babbie, Earl (2014). The Practice of Social Research, 13th Edition, Chapman University.
2. Baker, Therese. L (1999). *Doing Social Research*. 3rd Edition McGraw Hill College.
3. Bryman, Alan (2012). Social Research Methods, 4th Edition, Oxford University Press.

MODULE - III

Approaches to Quantitative Research Design

Unit 8 □ Designing Quantitative Research (I): Longitudinal Studies

Structure

8.1 Learning Objectives

- Understanding a longitudinal study, its important characteristics, and different types
- To gather information on the advantages and disadvantages of longitudinal studies
- To explore the technique of application and provide appropriate examples

8.2 Longitudinal Study: Introduction

Longitudinal studies are a kind of study design utilized in which data on the same individuals are routinely collected throughout an interval of time. Depending on the

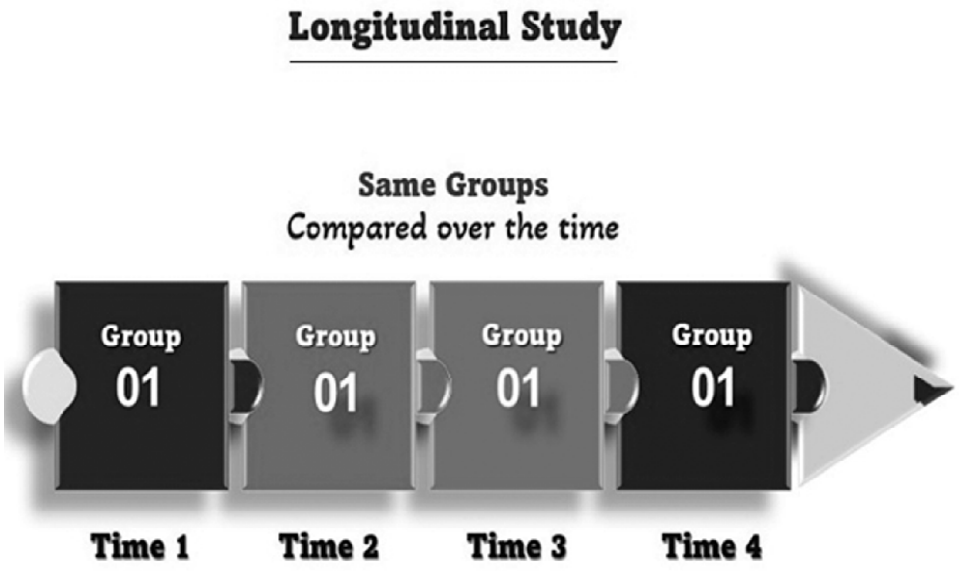
kind of research being done, this might last a few years to several decades. Unlike cross-sectional studies, which examine data from a population at one moment in time, longitudinal studies provide understanding of the dynamics of change and development within the population. In several disciplines, including psychology, sociology, medicine, and epidemiology, these studies are absolutely vital in clarifying long-term processes and results. They may therefore monitor changes in societal views over time, the aging process, or the course of an illness in individuals.

Examining changes and advancements over time inside a community or group via a longitudinal study is a potent research design. Unlike cross-sectional studies, which offer a moment in time view of a population, longitudinal studies follow the same people or groups across several years or sometimes decades. This method lets scientists find otherwise concealed patterns, trends, and causal links. Widely employed in disciplines including epidemiology, psychology, sociology, education, and public health, longitudinal studies provide special insights into the dynamics of human behaviour, health, and social events. The main points, benefits, drawbacks, and uses of longitudinal studies are investigated in this essay in order to underline their importance in both developing knowledge and guiding policy.

An observational research technique known as a longitudinal study is one whereby the same variables are repeatedly observed or measured over a lengthy period. Understanding how people or groups evolve throughout time and finding elements influencing these developments is the main objectives here. To investigate the effects of early childhood education on long-term outcomes, longitudinal research might, for instance, follow the academic performance of elementary school graduates through college. Data collecting at several times points allows researchers to discover causal links, examine patterns, and project results. This temporal component lets researchers see how factors change and interact, therefore offering a dynamic viewpoint not possible with cross-sectional studies. Data from the same people are gathered in longitudinal research at several times. By means of this recurrent measurement, researchers may monitor individual trajectories and spot trends of stability or change. Many times, longitudinal studies center on certain cohorts—that is, people born in the same year or those going through a similar life event. Following these cohorts across time allows researchers to investigate how environmental, social, or historical elements affect results. Longitudinal studies can be retrospective—where past data is examined—or prospective, in which case data is gathered forward in time. While retrospective designs depend on past records or memories, prospective designs are more popular and enable more control over

data collecting.

The capacity of longitudinal research to prove causal links is among their most important sources of leads. Stronger evidence for causality than cross-sectional studies, by tracking changes over time, researchers can ascertain whether an exposure or intervention comes before an outcome. Researching developmental processes—such as cognitive, emotional, or physical growth—longitudinal studies are very useful. To find important times and influencing elements, research can track children's language development from infancy to puberty. These studies are perfect for spotting over time long-term trends and patterns. Longitudinal research on aging, for example, might show how midlife lifestyle decisions impact later year health outcomes. Longitudinal studies reduce the impact of cohort effects—that is, variations resulting from birth in various eras—by tracking the same people or groups throughout time. More accurate comparisons and conclusions follow from this. Often gathering vast amounts of demographic, behavioural, and environmental data, longitudinal studies This richness helps scientists investigate difficult interactions and confusing variables.



Source : Self Creation

By offering a thorough and complete knowledge of social processes, behaviours, and trends over long periods, longitudinal studies are indispensable in sociological study. Unlike cross-sectional studies, which gather data at one point in time, longitudinal studies track changes within individuals, groups, or even society across months, years,

or even decades. With more accuracy, this approach lets sociologists spot trends, establish causal links, and forecast next social events.

8.3 Thinkers Perception

A longitudinal study is a type of research design whereby the same subjects are observed repeatedly over an extended period. Three main categories can help one to classify these studies:

1. One surveys the same people or groups at several times points in panel studies.
2. Focus on particular subgroups, such as birth cohorts, then track them across time.
3. Examining prior patterns and advancements using historical data is known as retrospective studies.

Longitudinal research is a great instrument for sociologists examining social mobility, education, health inequalities, and other dynamic features of society since it helps to expose how and why changes take place.

- **Generational Change and Karl Mannheim**

  Karl Mannheim (1928) maintained that only a longitudinal view would help one properly grasp generational changes in values, attitudes, and behaviours. His efforts on generational awareness showed how historical events impact the perspective of particular groups. Longitudinal research demonstrates that young people's acquired sociopolitical attitudes usually endure and influence long-term society developments.

- **Social Reproduction under Pierre Bourdieu**

  The idea of social reproduction developed by Pierre Bourdieu (1986) stresses the passing on of cultural capital across several generations. Longitudinal studies as those by Annette Lareau (2003) have demonstrated how variations in parenting methods, social networks, and institutional support across time help to sustain educational disparities.

- **The Self-Fulfilling Prophecy by Robert K. Merton**

  Longitudinal investigations have confirmed Robert Merton's 1948 theory of the self-fulfilling prophecy—where personal behaviour is influenced by society expectations. Studies on labelling theory, for example, have shown how early stigmatization—that is, being called a "troublemaker" in a classroom affects later results including criminal activity and career possibilities.

❒ **W.E.B. Du Bois and Racial Inequality**

Early sociological research on racial inequalities by W.E.B. Du Bois set the foundation for current long-term studies on racial inequality. Empirical data on how systematic racism influences economic mobility, health, and educational attainment over generations has come from the National Longitudinal Study of Youth (NLSY) and like studies.

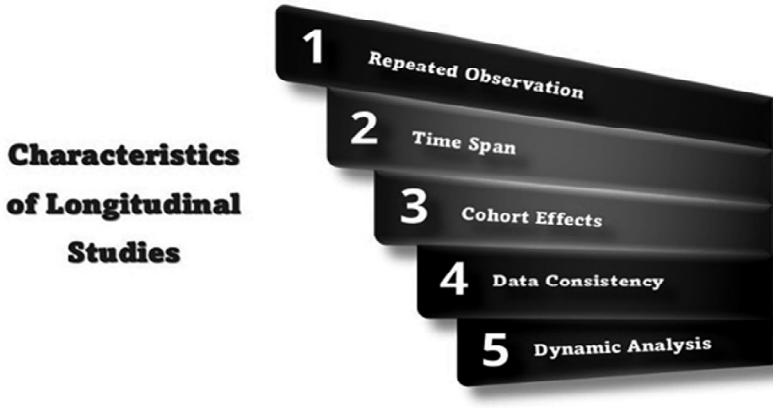
❒ **Social Systems and Talcott Parsons**

According to Talcott Parsons's structural-functionalism theory, accepted institutions help society to remain stable. Longitudinal studies enable sociologists to better grasp how institutions change with society, that is, with regard to family structures, gender roles, and work-life balance.

Longitudinal investigations are crucial since they provide a dynamic viewpoint on societal continuity and change. Tracking people and groups over time helps researchers to find ingrained trends impacting society structures. Theoretically, the contributions of sociologists such as Mannheim, Bourdieu, Merton, Du Bois, and Parsons illustrate the ongoing building upon of the theoretical roots by longitudinal investigations. Notwithstanding the difficulties, these studies remain pillar of sociological research that shapes academic discussions and policies all around.

8.4 Characteristics of Longitudinal Study

Here are the important characteristics of longitudinal study



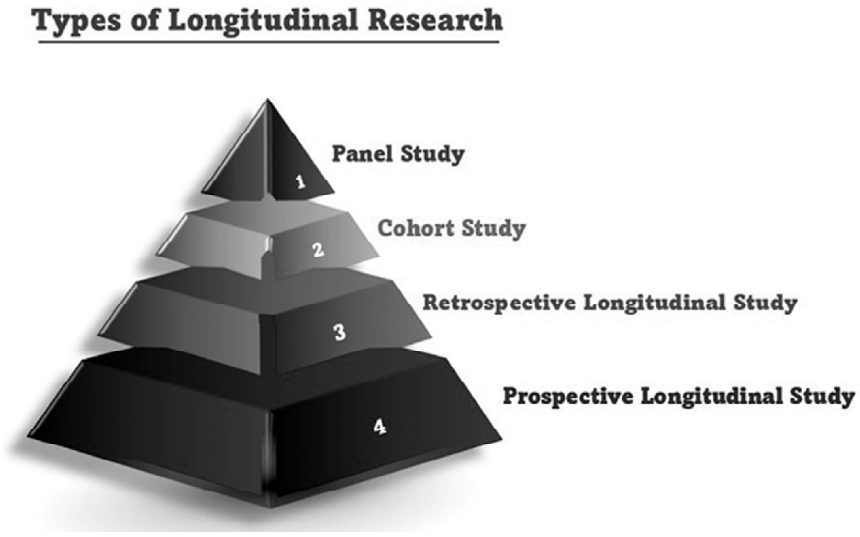
Source : Self Creation

❒ **Repeated Observation:** Longitudinal studies consist of gathering data from

the same people repeatedly over a designated time. They vary from cross-sectional studies, which gather data from many individuals at one moment in time, in that they are multifarious.

- **Time Span:** Longitudinal studies may last anywhere from a few months to many decades; hence their time span varies widely. The kind of the study topic and the phenomenon under examination will determine the time duration.
- **Cohort Effects:** Many times, these studies center on certain cohorts' groups of people with a same trait, like the year of birth. This lets scientists look at how exposures and events change certain populations throughout time.
- **Data Consistency:** Longitudinal studies may provide consistent and similar data over several time periods as the same people are seen frequently.
- **Dynamic Analysis:** Longitudinal studies are perfect for examining processes and trends rather than stationary situations as they let one examine changes and developments throughout time.

8.5 Types of Longitudinal Study



Source : Self Creation

- **Panel Study:** The panel study consists of routinely asking the same people at intervals over an extended time. In panel studies, the same set of subjects is asked often over a period of years. This lets scientists monitor changes on the group as well as the individual levels. One of the main characteristics of the panel study is data collecting from the same sample at many times (Aguinis, 2024). Though they might also be used for qualitative data and unit of analysis, most panel studies are meant for quantitative analysis.
- **Cohort Study:** Tracks a group of people who have a similar trait or experience throughout a designated time in a cohort study. These studies track a certain group of individuals who have a shared trait or experience e.g., born in the same year, attended the same event. Cohort studies may be separated yet further into:
  - **Prospective cohort study:** Prospective cohort studies include participants who are gathered then tracked forward in time.
  - **Retrospective cohort study:** Retrospective cohort studies track from the present back into the past using historical data.

  A cohort study gathers a cohort a group of individuals usually undergoing the same occurrence at a certain moment in time). Cohort investigations are common among medical researchers. Clinical trials might be seen by some as equivalent to cohort studies. Unlike clinical trials in which subjects go through examinations, in cohort studies researchers only see subjects without intervention.
- **Retrospective Longitudinal Study:** Often utilizing existing records and data, the retroactive longitudinal study looks back in time to follow the circumstances of the individuals. A retroactive study makes use of already available data gathered under comparable methods and factors from past research projects. The researcher uses an administrative database, pre-existing medical information, or one-to-one interviews during a retroactive study.
- **Prospective Longitudinal Study:** Data is gathered for prospective longitudinal research after first baseline data collecting. Prospective studies track subjects into the future and gather data at many times after the start of the research. Because they lower the danger of recollection bias and provide greater control over data collecting, they are seen to be more reliable than historical research (Verbeke & Molenberghs, 2000).

8.6 Advantages of Longitudinal Study

- **Temporal Sequence:** Longitudinal studies may clearly show a chronological series of occurrences, therefore guiding the identification of cause-and-effect linkages. Through time, changes allow researchers to deduce how past events or situations could affect present results.
- **Detailed Data:** Rich, comprehensive data on individual changes and advancements these studies give may help one understand trends and patterns that cross-sectional research might overlook.
- **Reducing Recall Bias:** Data collected in real-time or at regular intervals helps longitudinal research reduce the potential of memory bias that is, participant inaccuracy in remembering prior events or experiences.
- **Studying Developmental Trends:** In developmental psychology, where knowledge of how people evolve throughout time is vital, longitudinal research are especially helpful. They may record developmental benchmarks and pinpoint elements affecting development.
- **Control Over Variables:** Longitudinal studies allow one to account for factors that could vary across people as they follow the same participants. This uniformity helps to clearly examine the factors of interest.
- **Identifying Change:** These investigations are well suited to spot over time individual and group level changes.
- **Examining Development and Life Course:** Longitudinal studies are best for investigating developmental processes and life course trajectories.
- **Investigating Rare Events:** Long-term studies allow one to investigate unusual occurrences that could be overlooked in cross-sectional designs.
- **Policy Evaluation:** These studies help to evaluate long-term effects of policies or actions.
- **Greater validation:** Objectives and guidelines have to be developed from the start if long-term research is to be effective. Since this is long-term research, its legitimacy is confirmed ahead of time, so the validity of the findings is really high.
- **Unique data:** Most research projects gather short-term data to ascertain the cause and impact of the underlined topic. Though the data collecting time is different, longitudinal studies use the same ideas. While short-term studies

cannot reveal long-term partnerships, short-term relationships may be under observation in long-term research.

- **Allow identifying trends:** Whether in medicine, psychology, or sociology, the long-term design of a longitudinal research helps one to find patterns and links within the real-time data. One may make significant discoveries and predict future outcomes by using the past data (Hedeker & Gibbons, 2006).
- **Longitudinal surveys are flexible:** Though longitudinal research may be designed to investigate a particular data point, the data gathered can reveal unanticipated trends or meaningful associations in spite of this flexibility of the instrument. This is a long-term study; hence the researchers have flexibility not attainable in other research designs.

8.7 Disadvantages / Limitations of Longitudinal Study

- **Time-Consuming:** Often spanning years or even decades, longitudinal studies need a large time commitment. For researchers looking for fast answers, this prolonged length might be a big disadvantage.
- **Costly:** The length of these research implies they may be costly to run; data collecting, analysis, and participant follow-up call for significant financial resources.
- **Participant Attrition:** Participants may drop out of the research for a variety of reasons—including lack of interest, relocation, or medical problems over time. High attrition rates may compromise the validity of the research and provide biassed conclusions.
- **Changes in Study Situations:** The environment and circumstances under which the research is carried out could change with time and therefore influence the outcomes. The results of the research may be affected by outside elements like policy changes, technical developments, and social transformations.
- **Complexity in Data Analysis:** Examining longitudinal data may be challenging and calls for advanced statistical methods in data analysis. Further adding to the difficulty are handling missing data, guaranteeing data integrity, and deciphering temporal patterns.
- **Changing measurements:** As research cover extended times, measurements may need to be changed, therefore influencing comparability across time.

- **Data Management:** Large, sophisticated datasets produced by longitudinal investigations need for advanced methods of management and analysis.
- **Generalizability:** Cohorts may get less current population representative as they mature.
- **Ethical Considerations:** Long-term participant involvement presents significant ethical difficulties, particularly in sensitive research fields.
- **An Unpredictability Factor:** There is always an unpredictable element; one should consider that the first sample might vanish with time. Longitudinal studies entail the same individuals over an extended period of time, hence what occurs to them outside of data collecting periods will affect the future data collection results.

There are others who could choose to give up helping with the study. Others may not fit the right profile for study (Walls & Schafer, 2006). Should these elements not be incorporated into the first study plan, they may influence the produced results.

8.8 Longitudinal Study: Applications with Examples

- **Medical Science and Health:** Understanding the evolution and spread of illnesses depends critically on longitudinal investigations. For instance, by following hundreds of subjects over decades, the 1948-starting Framingham Heart Study has been crucial in determining risk factors for cardiovascular disease.
- **Psychology:** Longitudinal studies assist researchers in developmental psychology in better understanding of how emotions, cognitive ability, and behaviour change with time. Beginning in 1975, the Minnesota Longitudinal Study of Risk and Adaptation tracks people from infancy into adulthood to provide understanding of how early events affect later development. Research on mental health and behavioural development from infancy to adulthood has come from studies like New Zealand's Dunedin Multidisciplinary Health and Development Study (DMHDS).
- **Education:** Longitudinal studies in education assist to monitor students' development across time, therefore enabling the identification of elements influencing either academic success or failure. Tracking teenage health-related activities into adulthood, the National Longitudinal Study of Adolescent to Adult Health (Add Health) clarifies the long-term consequences of educational

events. These studies enable one to grasp throughout time the effects of educational interventions.

- **Sociology:** Early Childhood Longitudinal Study looks at children's educational experiences and development from early life through elementary school using a longitudinal approach in sociology. Tracking homes across time, the British Household Panel Survey (BHPS) offers statistics on employment, income, health, and social views. Understanding changes in social behaviours and situations is made possible in part by longitudinal research. For instance, the British Cohort Study has shed light on UK physical health, fertility, and unemployment among other issues.
- **Economics:** longitudinal studies allow one to examine over time how circumstances and policies affect people and families. Following households and people, the Panel Study of Income Dynamics (PSID) offers insightful information on poverty transitions, wealth building, and income dynamics.
- **Environmental Studies:** Longitudinal study in environmental studies allows one to track changes in the surroundings on ecosystems and human populations (Babbie, 2020). The Long-Term Ecological Research (LTER) Network, for example, gathers data over long stretches to examine environmental changes and ecological processes.
- **Criminology:** Criminology is the study of elements determining patterns of criminal conduct across the lifetime.
- **Gerontology:** Gerontology is the study of aging processes and elements promoting good aging.

8.9 Data Analysis Techniques in Longitudinal Study

Analysing longitudinal data can be done through these specific statistical methods:

- **Growth Curve Modelling:** This method, known as growth curve modelling, lets one accommodate individual variations in change paths by simulating how a result varies with time.
- **Multilevel Modelling:** Often referred to as hierarchical linear modelling, multilevel modelling addresses the layered structure of longitudinal data—that is, time points nested within persons.
- **Structural Equation Modelling (SEM):** Latent growth curve models and

other complicated interactions between variables over time may be modelled using structural equation modelling (SEM).

- **Time Series Analysis:** Time series techniques allow one to find trends, seasonal patterns, and autocorrelation for data gathered at many times points.
- **Survival Analysis:** Time to event data that is, time until a certain result occurs is analysed using survival analysis.
- **Latent Transition Analysis:** Change in latent class membership throughout time is modelled using latent transition analysis.
- **Fixed and Random Effects Models:** Both fixed and random effects models may assist to explain stable over time unobserved individual variations.

8.10 Moral Issues in Longitudinal Study

Longitudinal studies bring special ethical questions for which researchers have to answer:

- **Informed Consent:** Participants must be advised about the long-term character of the research and their ability to withdraw at any point.
- **Data Privacy and Confidentiality:** Maintaining participants' data over long stretches calls for strong protections in terms of privacy and confidentiality.
- **Reporting of Incidental Findings:** Procedures have to be in place to handle unanticipated health or risk information found during the research.
- **Burden on Participants:** Researchers have to strike a balance between the need for thorough data collecting and the possible load on the subjects.
- **Maintaining Equipoise:** Should early findings in intervention studies point to either definite advantages or disadvantages, ethical questions might develop.
- **Generational Effects:** Studies spanning many generations have to take ethical issues of including offspring of original participants into account.

8.11 Conclusion

A great weapon in the researcher's toolkit, longitudinal studies provide special insights into developmental processes, causal linkages, and long-term trends.

Although their design, implementation, and analysis pose major difficulties, the great wealth of data they provide is priceless in many different disciplines of research. Longitudinal research is probably going to become more and more significant in our knowledge of human development, health, and society as analytical tools and technology keep developing. From the personal to the society level, tracking changes over time offers a depth of knowledge that is hard to reach with other study approaches. Notwithstanding the difficulties, the possibility of longitudinal studies to guide theory, practice, and policy guarantees their ongoing importance on the scientific scene.

For those trying to grasp changes and advancements over time, longitudinal studies are a great weapon. From health and psychology to sociology and economics, their capacity to provide thorough, reliable data makes them indispensable in many disciplines. Notwithstanding their difficulties—time, money, and participant attrition among other things—the insights from longitudinal research may result in major changes in knowledge and practice. Careful design of these experiments and addressing any negative effects will help researchers to keep revealing significant temporal patterns and causal links influencing our perspective of the world.

8.12 Summary

Longitudinal studies are a research design that gathers data from the same individuals over an extended timeframe, spanning from several years to multiple decades. These studies are essential across multiple disciplines, including psychology, sociology, medicine, and epidemiology, to comprehend the dynamics of change and development within populations. They can observe alterations in societal perspectives, the aging process, or the progression of an illness in individuals. Longitudinal studies may be classified as either retrospective or prospective, contingent upon the nature of the data gathered. They offer a dynamic perspective that enables researchers to discern patterns, trends, and causal relationships. Longitudinal studies may be retrospective, analysing historical data, or prospective, concentrating on cohorts. They can establish causal relationships, pinpoint critical periods and influential factors, and monitor developmental processes such as cognitive, emotional, or physical growth. They mitigate the influence of cohort effects by longitudinally monitoring the same individuals or groups, facilitating more precise comparisons and conclusions.

8.13 Questions

1. What is longitudinal study? Write its characteristics.
2. Discuss advantages and disadvantages of longitudinal study.
3. Define thinkers' perceptions regarding longitudinal study.
4. Discuss the types of longitudinal study.
5. Define application and techniques of longitudinal study.

8.14 References

1. Aguinis, H. (2024). *Research Methodology: Best Practices for Rigorous, Credible, and Impactful Research.* Singapore: Saze Publication.
2. Babbie, E. (2020). *The Practice of Social Research.* USA: Cengage.
3. Bryman, A., Clark, T., Foster, L., & Sloan, L. (2022). *Social Research Methods.* UK: Oxford University Press.
4. Dubey, U. K., & Kothari, D. P. (2022). *Research Methodology: Technique and Trends.* London: CRC Press.
5. Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal Data Analysis.* Canada: John Wiley & Sons.
6. Kothari, C. R., & Garg, G. (2019). *Research Methodology: Methods and Techniques.* Delhi: New Age International Publishers.
7. Kumar, R. (2014). *Research Methodology: A Step by Step Guide for Beginners.* London: Sage.
8. McTavish, D. G., & Loether, H. J. (2015). *Social Research: An Evolving Progress.* Chennai: Pearson.
9. Mukherjee, S. P. (2020). *A Guide to Research Methodology.* New York: CRC Press.
10. Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* Belgium: Springer.
11. Walls, T. A., & Schafer, J. L. (2006). *Models for Intensive Longitudinal Data.* New York: Oxford University Press.

8.15 Glossary

- **Longitudinal study:** Longitudinal studies are a kind of study design utilized in which data on the same individuals are routinely collected throughout an interval of time. Depending on the kind of research being done, this might last a few years to several decades. Unlike cross-sectional studies, which examine data from a population at one moment in time, longitudinal studies provide understanding of the dynamics of change and development within the population.
- **Panel Study:** The panel study consists of routinely asking the same people at intervals over an extended time. In panel studies, the same set of subjects is asked often over a period of years. This lets scientists monitor changes on the group as well as the individual levels.
- **Cohort Study:** Tracks a group of people who have a similar trait or experience throughout a designated time in a cohort study. These studies track a certain group of individuals who have a shared trait or experience e.g., born in the same year, attended the same event.
- **Retrospective Longitudinal Study:** Often utilizing existing records and data, the retroactive longitudinal study looks back in time to follow the circumstances of the individuals. A retroactive study makes use of already available data gathered under comparable methods and factors from past research projects. The researcher uses an administrative database, pre-existing medical information, or one-to-one interviews during a retroactive study.

Unit 9 □ Designing Quantitative Research (II): Cross - Sectional Studies

Structure

9.1 Cross Sectional Study: Introduction

- To understand the characteristics and types of cross-sectional studies
- To explore its application in quantitative research
- To find out its advantages and disadvantages
- To understand the difference between a longitudinal study and a cross-sectional study
- To explore the technique of application and provide appropriate examples

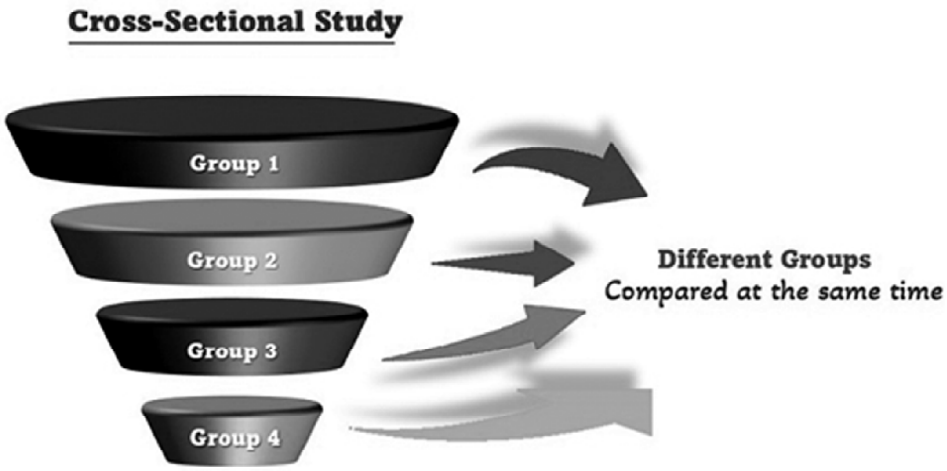
9.2 Cross Sectional Study: Introduction

A cross-sectional study is a kind of observational research in which data from a population, or representative subset, at a designated moment in time is analysed. Cross-sectional studies provide a one moment view of a population, unlike longitudinal studies which track individuals over a time. In several disciplines, including epidemiology, social sciences, psychology, and market research, these studies are extensively used to find prevalence, relationships, and trends among many variables. Understanding the present situation of a population, spotting relationships between variables, and developing ideas for further study depend especially on cross-sectional studies (Bryman, Clark, Foster, & Sloan, 2022). But their single-time-point character makes it impossible to establish causation. The next sections explore the features, forms, benefits, drawbacks, and uses of cross-sectional research by means of comprehensive examples.

From epidemiology to social sciences, psychology to public health, cross-sectional studies are a basic research tool applied in many fields. They offer a moment in time view of a population, therefore illuminating the frequency of diseases, behaviours, or traits within a given group. Cross-sectional studies gather data at a single instant, unlike longitudinal studies, which track individuals over a lengthy period and are thus comparatively rapid and cheap to run. An observational research technique called a cross-sectional study examines data from a population or a representative subset at a given moment. Examining the frequency of an outcome, like an illness, behaviour, or condition, and investigating its relationship with other variables of interest is the main objectives here. A cross-sectional study might look at, for instance, the association between smoking patterns and community respiratory symptom frequency. Simultaneous data collecting on exposure (smoking) and result (respiratory symptoms) allows researchers to see trends and create hypotheses for next studies.

The single time point emphasis of cross-sectional studies defines them. They are not like longitudinal studies, which follow changes across time. Although this design offers a moment, it cannot prove causation or ascertain the chronological order of occurrences. Estimating the frequency of a condition or behaviour within a population makes these studies very helpful. A cross-sectional study might show, for example, that 20% of individuals in a certain area have high blood pressure, therefore guiding public health campaigns. Cross-sectional studies are observational;

hence factors are not under control or intervention by the researchers. Rather, they watch and document data as it arises spontaneously, therefore reducing ethical issues and pragmatic difficulties. From health outcomes and social behaviour to economic trends and educational successes, cross-sectional studies are flexible and can be applied to investigate a broad spectrum of subjects.



Source : Self Creation

Less costly and faster to execute than longitudinal research, cross-sectional investigations do not need for long-term follow-up. For researchers with limited funds or time, this makes them appealing choice. Simple design with one time points data collecting involved. This simplicity makes cross-sectional studies available to researchers from many fields by lowering the complexity of planning and execution. These investigations are quite good in producing hypotheses about relationships between variables. A cross-sectional study might find, for instance, a relationship between low income and poor mental health, which would encourage more investigation into possible causal pathways. Large and varied populations made possible by cross-sectional studies offer a complete picture of a given problem. This large range improves the generalizability of results to the broader public.

9.3 Characteristics of Cross-Sectional Study

- **Single-Time Point Data Collection:** Cross-sectional studies gather data from people at one designated moment in time, single time point data

collection. This offers a moment in the variables of interest but does not follow changes throughout time.

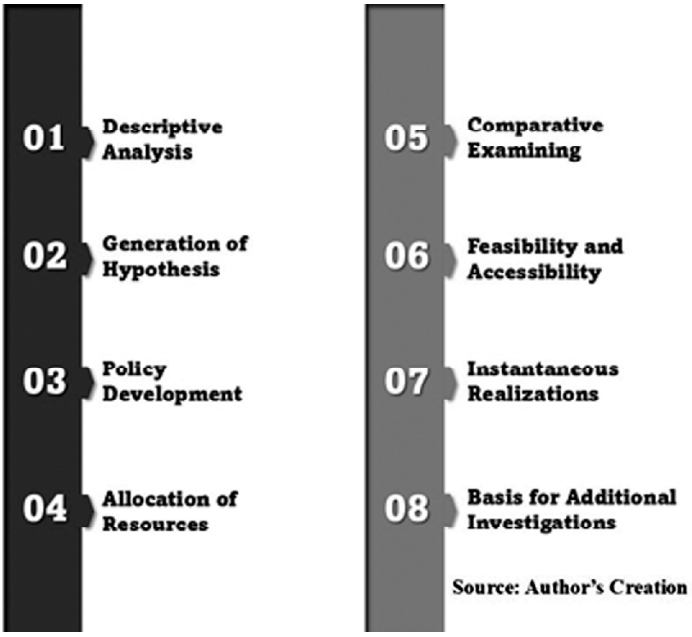
- **Observational Nature:** These investigations track and assess factors without changing the surroundings. They find trends and relationships by depending on current population variances.
- **Descriptive and Analytical Purposes:** Cross-sectional studies may be used for analytical purposes that is, to investigate correlations between variables or for descriptive purposes that is, to estimate the frequency of a condition.
- **Large Sample Sizes:** Cross-sectional studies may need for large, representative samples of the population in order to provide dependable findings. This guarantees that the conclusions are applicable generally.
- **Comparability:** Cross-sectional studies which concurrently examine many groups or subpopulations can compare variables between these groups, therefore stressing both variations and similarities.

9.4 Types of Cross-Sectional Study

- **Descriptive Cross-Sectional Studies:** Descriptive cross-sectional studies seek to characterize, at a given moment in time, the traits or frequency of a condition within a community. For instance, research may project the diabetes prevalence among the people of a city.
- **Analytical Cross-Sectional Studies:** Examining relationships between many variables within the population helps analytical cross-sectional research to go one step farther. Research may examine, for instance, the association between obesity and physical activity levels (Kumar, 2014).
- **Exploratory Cross-Sectional Studies:** Often utilized in uncharted territory of study when little is known, exploratory cross-sectional studies They seek for trends and create hypotheses to guide further research.
- **Comparative Cross-Sectional Studies:** Comparative cross-sectional studies are those which discover variations and parallels by means of comparison between many groups or subpopulations. Research may, for instance, compare smoking rates across many age groups.

Unit 9.5 Importance of Cross-Sectional Study

A basic study design in sociology, cross-sectional studies offer a moment in time view of a population. Understanding social events, spotlights, and guides policy decisions all depend on this approach. Analysing the value of cross-sectional research helps us to understand their contribution in expanding sociological knowledge and handling of social problems. A cross-sectional study is the gathering of single point-of- view data from a population or representative sample. Cross-sectional studies, unlike longitudinal studies—which follow changes across time—focus on the current instant and provide a stationary perspective of social conditions. Descriptive and analytical uses of this design especially help researchers to investigate correlations between variables and create hypotheses for next study.



A. Descriptive Analysis: Describing the features of a population is much enhanced by cross-sectional studies. This approach is used by sociologists to compile information on several social indices including income levels, education, health state, and job rates. Cross-sectional studies support the identification of social trends and inequalities by offering a whole picture of these elements. A cross-sectional study might, for example, show the level of poverty in a given area, therefore stressing the need of focused treatments.

B. Generation of Hypothesis: Cross-sectional research serves mostly as a means of generating hypotheses. Through single point in time analysis of the correlations between variables, researchers might spot possible causal links deserving of more study. A cross-sectional analysis might reveal, for instance, a relationship between income and education level implying that more education results in better-paying employment. More exact study approaches, such longitudinal studies or experiments, allow one to test this theory.

C. Policy Development: Essential data for legislators come from cross-sectional studies. Knowing the present situation of socioeconomic situations helps legislators create interventions targeted at immediate needs. For instance, a cross-sectional study on healthcare access may expose notable obstacles for some demographic groups, which would inspire the creation of laws aiming at enhancing healthcare equity. Cross-sectional data's timely character guarantees that policies are grounded on the most recent facts, hence improving their relevance and efficiency.

D. Allocation of Resources: Cross-sectional research in sociology assist to pinpoint areas of most resource demand. Through the mapping of socioeconomic concerns as unemployment or crime rates, these studies direct the allocation of funds to places most in need. This focused method guarantees effective and powerful interventions, thereby optimizing the advantages for the society.

E. Comparative Examining: Cross-sectional studies help one to compare several groups or areas. Data collecting from many populations allows sociologists to compare social circumstances in different settings, therefore pointing up opportunities for development and best practices. A cross-sectional study might, for instance, show differences between educational results in urban and rural areas that demand focus. These analogues offer insightful analysis that helps create plans to solve social problems.

F. Feasibility and Accessibility: Comparatively to other research approaches, cross-sectional studies are quite easily available and practical. For researchers on a tight budget, they are a sensible option as they call for less time and money than longitudinal investigations. Furthermore, providing versatility in research design, cross-sectional data can be gathered by means of surveys, interviews, and secondary data analysis.

G. Instantaneous Realizations: Cross-sectional studies' timely character lets one have quick understanding of social events. Fast data collecting and analysis by researchers allows them to provide current knowledge on urgent social concerns.

In fast changing societies, where quick data is necessary for understanding and handling developing issues, this immediacy is especially important.

H. Basis for Additional Investigations: Many times, the basis for more investigation is cross-sectional studies. These investigations create questions that can be investigated more thoroughly using other research designs by spotting trends and linkages. A cross-sectional study might find, for instance, a link between social media use and mental health problems, which would encourage more research on the causal pathways behind this relationship.

Sociology depends much on cross-sectional research, which provide a moment in time view of social conditions at a certain location. Their value stems from their capacity to create hypotheses, characterize demographic trends, direct policy decisions, and direct resource allocation. Although they have constraints, their accessibility and viability make them an important instrument for comprehending and resolving societal concerns. Cross-sectional research will remain a vital tool for examining the complexity of human society as sociology develops.

9.6 Thinkers Viewpoints Regarding Cross-Sectional Research

Among the several sociologists in the field, cross-sectional research has attracted interest and controversy. Various academics have presented different opinions on the usefulness, constraints, and uses for cross-sectional studies. Analysing these opinions helps us to better appreciate the function and importance of cross-sectional research in sociological study.

- **Emile Durkheim: Social Reality and Group Consciousness**

  One of the pioneers of sociology, Emile Durkheim underlined the need of researching social facts—patterns of behaviour existing apart from personal activities. Particularly his study on suicide, Durkheim's writings depended on cross-sectional data to spot societal trends and relationships. Cross-sectional research, he thought, may expose a society's collective awareness and offer understanding of the moral and social factors influencing individual action. Since Durkheim considered cross-sectional research as a tool for revealing the fundamental patterns of society, his view of it was essentially favourable. He was aware, meanwhile, of the limits of this approach, especially in proving causality. With an essentially descriptive attitude to cross-sectional study, Durkheim concentrated more on spotting social trends and patterns than on elucidating their origins.

- **Max Weber: Exercises in Interpretive Sociology**

  Another well-known sociologist, Max Weber, developed the idea of Verstehen, or interpretive knowledge, which stresses the need of realizing the subjective interpretations people attach to their activities. Less dependent on quantitative techniques like cross-sectional research and more interpretive in nature, Weber's approach to sociology. Cross-sectional research was seen by Weber as somewhat dubious. Although cross-sectional studies can offer important information on social events, he thought they sometimes missed the subtle meanings and motives driving human action. Weber maintained that while cross-sectional statistics by itself cannot adequately depict the subjective experiences of people, sociology should concentrate on these aspects. Weber agreed, despite his misgivings, that cross-sectional research was useful in offering a general picture of social circumstances. Although he considered it as a helpful tool for developing hypotheses and spotting trends, he underlined the need of supplementary qualitative techniques in reaching a better knowledge of social events.

- **Karl Marx: Social Change and Historical Materialism**

  Rooted in historical materialism—which emphasizes society's material conditions and the dynamics of class struggle— Karl Marx's approach to sociology was Marx's writings mostly focused on comprehending the fundamental economic causes of social change as well as its mechanisms. Marx seemed somewhat dubious of cross-sectional studies. Cross-sectional studies, with their emphasis on a single point in time, he felt frequently missed the historical and dynamic character of social events. Marx maintained that in order to grasp the mechanisms of social change and the development of class relations, sociology should give longitudinal and historical investigations top priority. Marx understood, therefore, that cross-sectional investigations may offer insightful analysis of the present situation of social conditions. Particularly in respect to class and economic level, he considered them as a helpful instrument for spotting discrepancies and injustices. Marx's criticism of cross-sectional studies was a need for a more all-encompassing approach including historical and longitudinal viewpoints rather than a rejection of the technique itself.

- **Pierre Bourdieu: Habitus and Social Capital**

  Work by Pierre Bourdieu concentrated on the ideas of social capital, cultural capital, and habitus—that is, the resources and attitudes people pick up from their social surroundings. Combining qualitative and quantitative techniques,

Bourdieu's theoretical and empirical approach to sociology was both. Bourdieu has a complex view of cross-sectional research. Cross-sectional studies, he realized, were valuable for spotting trends in social inequality and resource allocation. He also underlined, though, the need of knowing the fundamental processes generating these trends, which often call for more in-depth qualitative research. Cross-sectional research should, according to Bourdieu, be supplemented with qualitative techniques to properly depict the intricate interaction between social structures and personal agency. According to him, sociology should seek to expose the covert processes of social reproduction as well as the ways people negotiate their social surroundings.

- **Anthony Giddens: Structured Theory**

  Emphasizing the duality of structure and agency, Anthony Giddens' structural theory holds that social systems are both the medium and the result of human activities. Giddens approach sociology from a combined global and micro standpoint. Given that Giddens sees cross-sectional research as a useful instrument for comprehending the interaction between structure and agency, his view of it is essentially favourable. As well as the ways in which people help to reproduce and change social structures, cross-sectional studies can reveal how personal conduct is shaped by social structures. Giddens clearly acknowledges, nevertheless, that cross-sectional research falls short in portraying the dynamic and processual character of social life. To grasp the continuous interaction between structure and agency, he contends that sociology should take a more all-encompassing stance combining cross-sectional and longitudinal viewpoints.

Reflecting their larger theoretical orientations and methodological preferences, different philosophers in sociology have provided several angles on cross-sectional research. While some, such as Durkheim and Giddens, have underlined the importance of cross-sectional studies in spotting social patterns and structures, others, such Weber and Marx, have been more critical of their shortcomings in catching the subjective and dynamic elements of social life. Bourdieu's complex viewpoint emphasizes the need of combining cross-sectional research with qualitative techniques to reach a deeper knowledge of social events. These different points of view generally underline the need of a thorough and integrated approach to grasp the complexity of human society as well as the complexity and richness of cross-sectional research as a tool for sociological study.

9.7 Advantages of Cross-Sectional Study

- **Cost-Effective:** Generally speaking, cross-sectional research takes less time and money than longitudinal investigations. Their only one cycle of data collecting lowers resource demand.
- **Quick Results:** Data captured at one moment in time allows findings to be received rapidly. This benefits timely policy execution and decision-making.
- **Large Sample Sizes:** Studying big populations or samples helps one to investigate with more dependability and generalizability of the results.
- **Multiple Variables:** Cross-sectional studies may simultaneously look at many factors, therefore offering a complete demographic picture. This enables the detection of trends and connections among variables.
- **Public Health and Policy:** These studies are especially helpful in public health and policymaking as they provide current statistics on health status, habits, and other crucial elements that could guide treatments and policies.

9.8 Disadvantages of Cross-Sectional Study

- **No Causality:** While they may find relationships, cross-sectional studies cannot prove causation. They provide a one-point-of- view, so it is hard to ascertain the direction of connections.
- **Snapshot Limitation:** These studies just provide a moment in time view of the population at one point. Their limited capacity to grasp temporal dynamics results from their incapacity to catch changes or patterns throughout time.
- **Selection Bias:** Should the sample deviate from the population, selection bias might take place and provide erroneous findings. The validity of the research depends critically on a representative sample.
- **Recall Bias:** Cross-sectional studies often depend on self-reported data, which might be prone to memory bias. Participants may not precisely recall or document earlier actions or events.
- **Confounding Variables:** The existence of confusing factors could hinder the outcome interpretation. One might find it difficult to consider all possible confounders without randomization.

9.9 Cross Sectional Study: Application with Examples

- **Epidemiology:** Cross-sectional studies are extensively used in epidemiology to evaluate illness and health disorders' frequency. For the United States, the National Health and Nutrition Examination Survey (NHANES) for example tracks population health and nutritional state using a cross-sectional approach. For instance, cross-sectional research may evaluate the prevalence of hypertension in a community and examine how age, gender, and lifestyle choices interact to affect it.
- **Public Health:** Cross-sectional studies enable public health experts to pinpoint risk variables and create solutions. These studies include current information on health habits, healthcare accessibility, and other factors of health. For instance, research may look at how common smoking is among teenagers and how it relates to socioeconomic level, therefore guiding public health efforts to lower smoking rates.
- **Social Sciences:** Cross-sectional research in social sciences enable one to grasp society patterns and habits. They are used to investigate correlations among social factors like health, income, and education. For instance, a cross-sectional study may look at how job happiness and educational level relate among individuals, therefore offering information for legislators and teachers (Kothari & Garg, 2019).
- **Market Research:** Cross-sectional research help companies to better grasp customer behaviour, preferences, and market trends. These studies guide businesses in choices on customer service, marketing plans, and product development. To evaluate client satisfaction with its goods and services, for instance, a corporation may do a cross-sectional survey pointing out areas needing work and chances for innovation.
- **Psychology:** Cross-sectional studies are used by psychologists to investigate the frequency of mental health disorders and their relationships with other elements such stress, coping strategies, and social support. For instance, research may look at how common anxiety is among college students and how it relates to social support networks and academic achievement, therefore guiding campus mental health programs.
- **Education:** Cross-sectional studies provide educational research new perspectives on student performance, learning outcomes, and educational

inequality. They guide policy choices and assist to pinpoint elements influencing academic achievement. For instance, cross-sectional research assessing the effects of resources, instructional strategies, and socioeconomic elements can compare math competence levels across kids from several educational systems.

- **Environmental Studies:** Studies of Environmentalism In environmental science, cross-sectional studies evaluate how environmental elements affect wellness and health. They assist in determining relationships between health effects and environmental exposures. For instance, research may look at how respiratory health in urban and rural populations relates to air pollution levels, therefore guiding environmental laws and public health campaigns.

9.10 Difference Between Longitudinal Studies and Cross-Sectional Studies

There are multiple differences present between longitudinal and cross-sectional studies

1. Definition

Longitudinal studies are types of research projects in which similar variables are repeatedly observed over extended periods of time. They gather data from each of the same set of respondents at many times points. These studies are especially helpful for tracking changes over time or developmental patterns all over the lifetime.

Conversely, cross-sectional studies are observational research plans wherein data gathered from a population, or representative subset, at a designated moment in time. Examining the interaction of factors of interest, these studies provide a "snapshot" of a population at a given instant.

2. Time Scale

Often spanning years or even decades, longitudinal studies are conducted over a lengthy period. Engage many data collecting sites during the course of the research. Let one see changes and patterns throughout time.

Designed either over a brief period or at one moment in time, cross-sectional studies Share a moment in time view of the population at that given point. Track changes throughout time not within people.

3. Features of Samples

Longitudinal studies follow the same cohort of subjects all through the research. May see over time attrition that is, participation loss. Can examine age-related changes within the same person.

Usually include many people for every age range or topic of research. There is no attrition as data collecting occurs only once. Not able to follow personal developments throughout time.

4. Data Collection

Longitudinal studies gather data from the same individuals often over an extended period. At every period, use many techniques (surveys, interviews, physical measures). Usually call for greater resources and dedication from participants as well as from researchers.

Cross-sectional studies call for gathering information from participants only once. Applied merely at one time point, may use comparable techniques as longitudinal research. Usually call for less resources and less participant dedication.

5. Types of Research Questions

Ideal for examining developmental patterns, causal linkages, and changes throughout time are longitudinal studies. may respond to inquiries such "What factors predict changes in Y?" or "How does X change over time?" fit for looking into long-term consequences of exposures or treatments.

Best for looking at frequency, relationships between variables, and group comparison are cross-sectional studies. Can answer inquiries like "What is the current frequency of X?" or "How does Y vary among different groups?" Helped to create hypotheses for next long-term research.

6. Advantages

Longitudinal studies help to demonstrate temporal sequence of occurrences, therefore supporting the conclusion of causation. Let one examine personal changes throughout time. Share thorough, rich information regarding unique paths. Separate cohort effects from aging influences. Perfect for researching unusual exposures or results. may spot sleeper effects impacts seen long after the first encounter (Dubey & Kothari, 2022).

Usually less costly and time-consuming than longitudinal research, cross-sectional studies can be finished quicker, therefore enabling speedier outcomes.

Not one of attrition problems. Applied for simultaneous research of many effects and exposures. can provide a reasonably accurate picture of a population at any one moment. Perfect for hypothesis generation and descriptive studies.

7. Disadvantages

Time-consuming and costly, longitudinal studies Subject to attrition, which can skew things. May experience panel conditioning that is, answers influenced by repeated testing). Affected by historical events throughout the research period. Demand from researchers and subjects a long-term dedication. Missing data and repetitive measurements make data analysis challenging.

Cross-sectional studies cannot prove temporal links between variables or causation. Subject to cohort effects that is, generational variances rather than age differences across age groups might be explained. One cannot examine changes within persons throughout time. Should the single time point be unusual, the image may be distorted. Easily prone to remember bias while gathering past information.

8. Methodologies for Data Analysis

Often use more intricate statistical techniques in longitudinal studies to include time-dependent factors and repeated measurements. Common approaches include

LONGITUDINAL STUDIES	CROSS-SECTIONAL STUDIES
Mixed effects models.	Descriptive statistics
Growth curve modelling.	Chi-square tests
Time series analysis.	T-tests
Survival analysis.	ANOVA
Latent growth models.	Regression analysis (linear, logistic, etc.)
Generalized estimating equations (GEE).	Factor analysis

9. Application in several domains

Longitudinal Studies can apply in these domains

- **Psychology:** Investigating cognitive aging, personality development, or the long-term consequences of early events
- **Medicine:** Examining risk factors for chronic illnesses, long-term therapeutic results, or disease progression.

- **Sociology:** Investigating changes in attitudes over time, social mobility, or how life events affect behaviour.
- **Economics:** Examining income paths, professional development, or long-term economic patterns helps one in economics.
- **Education:** Examining over time academic performance and how early interventions affect later results.

In cross-sectional studies the important domains are

- **Epidemiology:** It helps to ascertain the frequency of illnesses or risk factors within a community.
- **Market Research:** Comparative customer preferences among several demographic groupings constitute market research.
- **Political Science:** Examining at a given moment political sentiments or voting intentions.
- **Public Health:** Evaluating, across different populations, health practices or access to healthcare facilities.
- **Organizational psychology:** contrasting work satisfaction rates among many divisions or sectors.

10. Ethical Consideration

In longitudinal Studies participants must be continuously informed consented upon. Could find it difficult to preserve participant privacy for long stretches. Have to take long-term participation's effects on individuals into consideration. must have strategies in case of inadvertent discoveries throughout the research.

Usually with one-time data collecting, cross-sectional studies raise less ethical problems (Mukherjee, 2020). Still has to make sure participant informed permission and privacy are maintained. May run ethical problems about the use of private or sensitive data.

11. Techniques of Sampling

Purposive sampling is a common method used in longitudinal studies to choose subjects who might be tracked throughout time. May use stratified sampling to guarantee certain subgroup representation. Has to take possible attrition into account when deciding on the first sample size.

Many times, cross-sectional studies seek for representative sample of the target population. Depending on the study objectives, may use stratified sampling,

random sampling, or cluster sampling. Usually speaking, sample size calculation is easier than in longitudinal research.

12. Cost and Resource Implications

Usually more costly because to long-term participant retention initiatives and recurrent data collecting. Demand consistent money over long times. Long-term project management calls for committed research teams. Maybe for repeated measurements, certain tools or facilities are needed.

Usually less expensive as data collecting is only once. Cross-sectional studies One may finish it with shorter-term money. Usually call for less resources and staff. Possibly better appropriate for small funds exploratory research or pilot projects.

13. Results, Publication and Distribution

Many times, longitudinal studies produce many publications as data builds over time. Before the trial ends, may provide interim analysis or reports. Years after the trial began, final findings could not be accessible. The strength of longitudinal data makes findings very influential.

Usually lead to a single report or publication. Results may be shared comparatively rapidly after data collecting. Might be the foundation for developing hypotheses for further longitudinal studies. Results may not be very good in guiding long-term patterns or causal links.

14. Combining Strategies

Although cross-sectional and longitudinal studies are different ways of inquiry, sometimes researchers blend aspects of both:

- **Sequential Designs:** Studying many cohorts over time helps one to combine cross-sectional and longitudinal aspects in sequential designs.
- **Time-Lag Studies:** Comparing cross-sectional data gathered at many times helps one to deduce changes throughout time.
- **Accelerated Longitudinal Designs:** Follow several cohorts for shorter times, then aggregate data to deduce longer-term patterns using accelerated longitudinal designs.

These hybrid methods seek to minimize some of the respective restrictions of both longitudinal and cross-sectional designs by using their respective strengths.

15. Concerning Policy and Practice

Because they may reveal causal links and long-term patterns, longitudinal studies can have major impact on policy. may provide very convincing proof

over time of the success of policies or actions. May guide long-term strategy and planning in many different spheres.

For swift policy choices, cross-sectional studies which may provide glimpses of present circumstances are invaluable. Useful in pointing out areas of concern or differences that could call for further research or action. Usually used to track systems' or population's present situation.

In general, research in many fields depends critically on both longitudinal and cross-sectional investigations. Though they demand great resources and long-term dedication, longitudinal studies provide special insights into changes throughout time and causal links. Usually faster and less resource-intensive to do, cross-sectional studies provide insightful images of people at certain times. The research objectives, available resources, and pragmatic concerns will determine which of these designs best fits you. Usually, both methods used together provide the most thorough knowledge of difficult events. When designing their studies, researchers have to give great thought to the advantages and drawbacks of every design to guarantee they choose the most suitable approach for their particular objectives.

9.11 Conclusion

In research, cross-sectional studies are a useful instrument as they provide important new perspectives on the frequency of illnesses, habits, and other factors within a community at a given moment. Although they have limits that is, the inability to deduce causation and the existence of confusing variables their efficiency, cost-effectiveness, and practicality make them indispensable in public health, epidemiology, social sciences, and beyond. Understanding the features, forms, benefits, drawbacks, and uses of cross-sectional studies helps researchers to properly plan and analyse these studies to guide public health campaigns, policy choices, and next investigations (McTavish & Loether, 2015). Cross-sectional studies are essential for the identification of health concerns, social behaviours, and other elements influencing populations as shown by many instances; so, they help to enhance public health and well-being.

9.12 Summary

Cross-sectional studies are observational research methods that examine data from a population or representative sample at a particular point in time. They offer a

singular snapshot of a group, in contrast to longitudinal studies that monitor individuals over time. These studies are extensively utilized in disciplines such as epidemiology, social sciences, psychology, and market research to ascertain prevalence, correlations, and trends among variables. Nonetheless, their singular temporal nature precludes the establishment of causation. Cross-sectional studies are fundamental research instruments utilized across several disciplines, including epidemiology, social sciences, psychology, and public health. They provide a snapshot of a population, revealing the prevalence of diseases, behaviours, or characteristics within a specific group. They collect data at a certain moment, in contrast to longitudinal studies, which monitor individuals over an extended duration. Although cross-sectional studies are conducted at a single time point, they can assess the prevalence of a condition or habit within a population, rendering them valuable for public health initiatives. They are observational, mitigating ethical concerns and practical challenges. Cross-sectional studies can be utilized throughout diverse subjects, encompassing health results, social behaviour, economic patterns, and educational achievements.

9.13 Questions

1. What is a cross-sectional study? Identify its characteristics.
2. Explore the advantages and disadvantages of cross-sectional research.
3. What are the differences between longitudinal study and cross-sectional study?

9.14 References

1. Aguinis, H. (2024). *Research Methodology: Best Practices for Rigorous, Credible, and Impactful Research.* Singapore: Saze Publication.
2. Babbie, E. (2020). *The Practice of Social Research.* USA: Cengage.
3. Bryman, A., Clark, T., Foster, L., & Sloan, L. (2022). *Social Research Methods.* UK: Oxford University Press.
4. Dubey, U. K., & Kothari, D. P. (2022). *Research Methodology: Technique and Trends.* London: CRC Press.
5. Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal Data Analysis.* Canada: John Wiley & Sons.

6. Kothari, C. R., & Garg, G. (2019). *Research Methodology: Methods and Techniques*. Delhi: New Age International Publishers.
7. Kumar, R. (2014). *Research Methodology: A Step by Step Guide for Beginners*. London: Sage.
8. McTavish, D. G., & Loether, H. J. (2015). *Social Research: An Evolving Progress*. Chennai: Pearson.
9. Mukherjee, S. P. (2020). *A Guide to Research Methodology*. New York: CRC Press.
10. Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Belgium: Springer.
11. Walls, T. A., & Schafer, J. L. (2006). *Models for Intensive Longitudinal Data*. New York: Oxford University Press.

9.15 Glossary

- **Cross-sectional Study:** A cross-sectional study is a kind of observational research in which data from a population, or representative subset, at a designated moment in time is analysed. Cross-sectional studies provide a one moment view of a population, unlike longitudinal studies which track individuals over a time.
- **Descriptive Cross-Sectional Studies:** Descriptive cross-sectional studies seek to characterize, at a given moment in time, the traits or frequency of a condition within a community. For instance, research may project the diabetes prevalence among the people of a city.
- **Comparative Cross-Sectional Studies:** Comparative cross-sectional studies are those which discover variations and parallels by means of comparison between many groups or subpopulations. Research may, for instance, compare smoking rates across many age groups.

Unit 10 □ Designing Quantitative Research (III): Trend Studies

Structure

10.1 Learning Objectives

- Understanding a trend study, its important characteristics, and different types
- To gather information on the advantages and disadvantages of trend study
- To explore the technique of application and provide appropriate examples

10.2 Trend Study: Introduction

Trend studies are a kind of longitudinal research technique used in analysis of changes over time. Unlike other longitudinal techniques, including panel studies, trend studies concentrate on analysing patterns and trends in a population over certain times rather than following the same people. Examining changes in a population or phenomenon over time is the emphasis of trend studies, a form of longitudinal research methodology.

Trend studies examine several samples from the same population at different intervals, unlike conventional longitudinal studies that follow the same people or groups over several times. This method lets academics spot trends, changes in attitudes, behaviours, or conditions inside a larger framework. In sociology, economics, political science, public health, and marketing, trend studies are extensively applied to grasp society changes, project future developments, and guide decision-making.

A trend study is a type of research whereby data from several population samples at several times are gathered and analysed. Finding and understanding trends—patterns of change or stability—in certain variables or events is the main aim. A trend study might, for instance, look at public opinion on climate change over a decade by polling many groups of people every two years. Data across time allows researchers to ascertain whether opinions are getting more or less favourable and investigate possible causes for these changes. Studies of trends stress population-level rather than individual-level changes. This distinguishes them from panel studies, which follow the same individuals across time. Trend studies employ distinct samples from the same population at every time point, unlike longitudinal studies following the same subjects. This lets academics record more general social changes free from participant attrition's difficulties. Trend studies examine data gathered at several times; hence they are by nature temporal. This helps academics to spot trends in the variables of interest—that is, rises, declines, or oscillations. Trend studies derive findings regarding changes by means of historical comparisons. To evaluate economic patterns, research might, for instance, examine unemployment rates in a nation throughout many years. From social views and economic data to health outcomes and technological uptake, trend studies find application across a broad spectrum of subjects.

Longitudinal research is a kind of study design wherein the same variables are repeatedly observed over either short or extended periods. Trend Analysis is a form of longitudinal study wherein many samples from the same population are gathered at several times to spot patterns and changes.

10.3 Characteristics of Trend Study

Here are the important characteristics of trend study

- **Time Frame:** Trend studies cover many decades or few months depending on their various time periods. This lengthy duration lets researchers see long-term trends and patterns as well as track changes in variables throughout time.

- **Temporal Analysis:** Trend studies are distinguished mostly from others by their emphasis on temporal analysis. Different intervals allow researchers to follow the development or regression of certain variables by gathering data. Understanding change's dynamics depends on this time dimension.
- **Constant Intervals:** In trend studies, data collecting takes place at regular intervals—perhaps yearly, biannually, or otherwise specified. Maintaining the dependability and comparability of the data across time depends on this consistency.
- **Standardised Techniques:** Standardized data collecting techniques used at all time points help to guarantee comparability. This standardizing reduces variability resulting from various data collecting methods, therefore improving the validity of the research.
- **Cohort Studies:** Researchers in cohort trend studies choose a certain group of people with a similar trait—such as birth year—then track them throughout time. This method lets one look at changes within that particular group.
- **Panel Studies:** Panel studies follow the same people across many times points. This approach clarifies individual level changes and facilitates the comprehension of the elements causing such changes.
- **Repeated Cross-sectional Studies:** In repeated cross-sectional studies, at all-time point researchers gather data from many samples of the same population. Understanding population-level patterns and changes calls for this approach.

10.4 Types of Trend Study

Trend studies fit many categories depending on their method and focus:

- **Cohort Trend Studies:** To define Cohort Trend Studies, track a certain cohort such as a group of individuals who experience a particular occurrence in the same time period across time. For example: looking at the career growth of other graduates of that same year.
- **Repeated Cross-Sectional Studies:** Poll multiple samples from the same population at various times across a repeating cross-sectional study. For instance, yearly national surveys on public opinion on climate change.
- **Panel Trend Studies:** Combining elements of panel studies and trend studies, this hybrid technique consists of recurrent evaluations on the same population including new respondents. For a given cohort, for instance, tracking health outcomes across several years.

10.5 Importance of Trend Study

A pillar of study in many different fields, trend studies provide priceless insights on how society, businesses, and attitudes change throughout time. Trend studies offer a macroscopic picture of changes, patterns, and developments that mold our planet by use of data gathered from several samples of a population at several times. In disciplines such sociology, economics, public health, political science, and marketing—where knowledge of long-term changes is essential for strategic planning, policy development, and decision-making—these studies are especially vital.

- **Seeing Shifts and Patterns:** Trend studies mostly help us to find trends and changes in populations or events across time. Regular interval data analysis helps researchers identify trends including changes in particular variables, either increasing or declining. A trend study on educational attainment, for instance, would show a consistent rise in the number of people earning college degrees across many decades. Understanding society development and pointing up areas that need work depend on this knowledge. Trend studies also assist in revealing surprising or newly developing trends. For example, a study on internet use may reveal an unexpected increase in online activity during a given period, which would inspire more research on the causes—technical developments or world events like the COVID-19 epidemic. Trend studies help to lay a basis for more thorough investigation and hypothesis formulation by spotting these trends.
- **Advocating Policy and Decision-Making:** Policy and decision-making on local, national, and worldwide levels depend much on trend studies. Trend data is what policymakers use to create evidence-based treatments and distribute funds efficiently. A trend analysis on urbanization, for instance, can show a fast rise in urban population growth, thereby stressing the need of infrastructure development, cheap housing, and sustainable urban planning. Trend studies are very helpful in public health for tracking the frequency of diseases and health practices. For example, the Global Burden of Disease Study records worldwide morbidity and death trends, therefore offering vital information for international health projects. Rising trends in diseases like obesity or mental health problems allow legislators to prioritize treatments and distribute money to handle these problems. In economics, too, trend analyses of indices including GDP growth, unemployment rates, and income inequality guide governments and businesses in their decisions. Trend research revealing a drop in manufacturing jobs, for instance, could inspire legislators to fund retraining initiatives and assist the shift to a knowledge-based economy.

- **Forecasting Next Developments:** Trend studies help one not only to grasp previous and present changes but also to project future developments. Analysing past data helps scientists spot trends that probably will persist or change going forward. Particularly useful in disciplines such business, technology, and environmental research is this predictive capacity. In marketing, trend analyses of consumer behaviour and preferences enable businesses to predict demand changes and modify their plans. Trend research demonstrating a move toward environmentally friendly items, for instance, could inspire companies to fund sustainable processes and create green product lines. Trend studies on adoption rates and innovation cycles in technology enable one to forecast the course of sectors. Research on the acceptance of electric vehicles, for example, may show a rising trend indicating a future decrease in the demand for fossil fuel-powered vehicles and hence driving investments in renewable energy infrastructure. Trend studies on climate change indicators, such temperature increase and sea level variations, give vital information in environmental science that helps to forecast future effects. These forecasts guide world initiatives to slow down global warming and adjust to its consequences.
- **Adding to Practical and Academic Knowledge:** Trend studies add much to both scholarly and useful information. Academically, they offer researchers an abundance of information to examine and understand, therefore generating fresh ideas, models, and hypotheses. Trend studies on social attitudes, for instance, have helped us to better grasp cultural changes including altering perceptions on gender roles, marriage, and religion. Practically, trend studies give governments, companies, and people useful insights. A trend analysis on workforce demographics, for example, may expose an aging population and require companies to create plans for keeping older workers and handling possible labour shortages.
- **Overcoming Social Issues:** Dealing with society issues and advancing development depend on trend research. Through trend analysis in fields including education, health, and economic development, these studies enable nations to solve disparities and raise standard of living. For instance, a trend analysis of literacy rates could expose differences between urban and rural areas, therefore stressing the need of focused educational efforts. Trend analyses also offer a foundation for assessing the success of policies and treatments. Data pre and post-program comparison allows academics to evaluate program impact and generate suggestions for development. For example, trend research on smoking rates could demonstrate a drop after public health programs were launched, therefore proving their success and directing further initiatives.

10.6 Views of Different Thinkers Regarding Trend Study

As a research technique, trend studies have attracted a lot of interest among academics from several fields. Praised for their potential to expose trends, guide policies, and forecast future developments, these studies—which examine changes in populations or events over time by looking at many samples at several points—have been appreciated. Still, intellectuals from many disciplines have presented several angles on the value, drawbacks, and moral connotations of trend research. Trend studies are increasingly utilized in sociology to investigate changes in social attitudes, practices, and institutions. Though their work precedes the formalization of trend studies, thinkers such as Émile Durkheim and Max Weber have set the foundation for comprehending society changes. Durkheim's emphasis on social truths and communal consciousness fits the macro-level study trend analysis offers. A trend analysis on religious affiliation, for example, would show a drop in conventional religious behaviours, which would imply more general society secularization—a phenomena Durkheim would read as a change in collective values.

Trend studies have helped modern sociologists like Robert Putnam examine changes in social capital. In his foundational book Bowling Alone, Putnam explores the drop in American communal involvement over numerous years. Putnam provides insights on the decline in social cohesiveness by spotting a tendency toward individuality and lower civic participation by means of data analysis spanning several times. His writings emphasize the need of trend studies in comprehending long-term changes in society. Trend studies have long been appreciated by economists in their analysis of economic data and guide of policy. To solve problems including unemployment and inflation, John Maynard Keynes, for instance, underlined the need of knowing economic patterns. Key statistics for Keynesian economic policies meant to stabilize nations come from trend studies on GDP growth, income inequality, and employment rates.

Renowned monetarist economist Milton Friedman also advocated controlled money supply as a technique of inflation control depending on trend data. His examination of past economic patterns produced significant policy recommendations that helped to define late 20th-century monetary policy. Friedman's writings emphasize how important trend investigations are to verifying economic theories and direction of guidance in decision-making. More lately, trend studies have been employed by analysts such as Thomas Piketty to investigate income disparity and wealth. In Capital in the Twenty-

First Century, Piketty examines millennia of economic data to spot patterns in wealth concentration. His results, which expose an ongoing trend toward inequality in capitalist nations, have spurred worldwide discussions on social fairness and economic policy. Piketty's work shows how trend studies could highlight structural problems and guide attempts at their resolution.

Trend studies are crucial in public health for tracking illness prevalence, health practices, and intervention effects. Early types of trend analysis were utilized by thinkers such as John Snow, sometimes regarded as the father of modern epidemiology, to spot trends in disease outbreaks. His research on cholera in 19th-century London, which plotted cases over time, set the stage for applying trend data to grasp and address public health emergencies. Trend studies have helped modern public health experts such as Sir Michael Marmot investigate health disparities. Marmot's research on the social determinants of health uses trend and longitudinal data to demonstrate over time how elements like income, education, and housing affect health outcomes. His results guide strategies meant to lower health inequalities and raise population health. Tracking the spread of infectious diseases such COVID-19 and HIV/AIDS has also been much aided by trend studies. Public health officials can distribute funds, carry out initiatives, and assess their success by examining patterns in infection rates. The efforts of intellectuals such as Anthony Fauci highlight the need of trend research in addressing newly arising health hazards.

Trend studies in political science help to examine changes in political behaviour, public opinion, and government structure. To explain political realignments, intellectuals such as V.O. Key Jr. have underlined the need of grasping election trends. Trend studies on voting behaviour, for instance, might highlight changes in party loyalty or the start of fresh political movements. More lately, researchers like Ronald Inglehart have investigated political and cultural developments using trend studies. Trend data from polls such as the World Values Survey supports Inglehart's post-materialism theory, which holds that as cultures get more rich societies move from materialist to post-materialist values. His works show how deeply ingrained changes in political culture and objectives could be revealed by trend research. Analysis of the emergence of populism and polarization in modern politics has also drawn on trend research. Analysing patterns in voter opinion and political discourse helps scientists to pinpoint elements causing these phenomena and evaluate their consequences for democracy. The writings of intellectuals such as Pippa Norris emphasizes the need of trend studies in comprehending the difficulties of contemporary political systems.

Although trend studies are generally accepted as useful, some analysts have ethical and methodological questions. Critics of the power dynamics inherent in data collecting

and analysis include Michel Foucault and others known as critical theorists. Based on Foucault's writings on surveillance and control, trend studies—especially those carried out by governments or businesses—may be used to track and influence populations. From this viewpoint, ethical issues in the design and application of trend studies become even more important. Feminist academics such as Donna Haraway have also criticised the objectivity sometimes claimed by trend studies, contending that the perspective of the researcher shapes all research. Haraway's idea of "situated knowledges" forces academics to examine the social context in which evidence is gathered and analysed and own their prejudices.

10.7 Advantages of Trend Study

Trend studies have some important benefits

Enhanced Safety

- **Accident Prevention:** One of the benefits of tread research is improvement of vehicle safety. Through tire tread pattern analysis, professionals may spot wear and tear likely to cause blowouts or lack of traction. Preventing traffic accidents depends on this knowledge, hence it is very vital. Regular tread examination, for example, may assist to promptly replace worn-out tires, therefore lowering the danger of skids, hydroplaning, or tire failure.
- **Improved Design:** Tread studies help to create safer tires by means of better design. Manufacturers may create tires with higher grip, longer life, and more resilience to wear and tear by knowing how treads deteriorate over time and in varied environments. Safer driving conditions for customers follow from this ongoing advancement in tire design.

Forensic Uses

- **Accident Reconstruction:** Tread studies are very important in forensic science for accident reconstruction. Examining tire imprints at an accident or crime scene helps professionals ascertain a vehicle's speed, direction, and activities before to a collision. In court proceedings, this material may be rather important as it helps to prove responsibility or clear innocent persons.
- **Criminal Investigations:** Beyond mishaps, tread analysis is very helpful in criminal investigations. For example, distinctive tire patterns might connect a suspect's car to a crime scene, therefore offering vital evidence in instances of hit-and-run events or other vehicle crimes.

Environmental Impact Evaluation

- **Pollution Monitoring:** Tread research have environmental ramifications as well. Tire wear and tear adds to micro plastics, a source of pollution that could harm ecosystems. Analysing tread patterns and tire degradation rate helps researchers to project the impact of tire wear to environmental pollution and create plans of action to reduce it.
- **Sustainable Practices:** Tread study findings may inspire the creation of more environmentally friendly tires. For instance, knowing which materials and designs cause less wear would help to produce tires with a less environmental impact, therefore supporting sustainable practices in the automobile sector.

Economic Benefits

- **Cost Savings:** For customers, consistent tread analysis may result in really large cost reductions. Early identification of unequal wear patterns helps vehicle owners to solve underlying problems such misalignment or poor inflation, therefore increasing the life of their tires and lowering the frequency of replacements.
- **Industry Innovation:** Tread studies may inspire producers. Understanding the limits and strengths of present tread designs helps businesses to make investments in research and development to produce outstanding products. This improves their market situation as well as stimulates industry-wide technical development and rivalry.

Performance Enhancement

- **Tailored Solutions:** Tread studies help to create tires specifically for certain circumstances and needs. For example, the tread patterns of tires meant for long-distance trucks, off-road vehicles, or high-performance sports cars each reflect their respective application situations. This customizing guarantees that cars run at their optimum in different environments.
- **Fuel Efficiency:** Correct tread pattern might also improve fuel economy. Tires with ideal tread patterns lower rolling resistance, so engines do not have to work as hard to move the car. Better fuel efficiency resulting from this energy-saving helps the environment as well as customers.

10.8 Disadvantages of Tread Study

Important disadvantages of trend studies are

High Costs

- **Equipment and Technology:** Doing thorough tread research calls for advanced tools and technologies. Costly investments include high-resolution imaging systems, specialist tire wear testing equipment, and sophisticated data processing software. For individual researchers or small businesses, these costs might be exorbitant.
- **Research and Development:** Additionally, costly is the process of doing tread studies and using the results to guide tire manufacture and design. Research & development must be significantly invested in order to evaluate new materials and designs, therefore influencing the total cost of the finished product.

Time-Consuming

- **Detailed Analysis:** Tread studies, which include painstaking examination of tire wear over long periods can be time-consuming. Accurate and dependable data requires this thorough approach, yet it might slow down research and development's speed.
- **Long-Term Testing:** Long-term testing is crucial to grasp the wear patterns and long-term performance of tires. This may be a long procedure as it entails tracking tires throughout hundreds of kilometres and under many driving circumstances.

Complexity

- **Varied Conditions:** Tires experience a broad spectrum of situations, including varying road surfaces, weather, and driving techniques. This diversity complicates the research and makes it difficult to separate certain elements influencing tread wear.
- **Data Interpretation:** High degree of knowledge is needed to understand the tread study results. Variations in wear patterns may be subtle and impacted by many elements, which calls for qualified analysts to get correct results.

Predictive Accuracy Limitations

- **Unpredictable Factors:** Though tread studies provide insightful information, absolute tire performance prediction remains difficult. Tire wear may be influenced in ways that are difficult to forecast by unanticipated elements such road trash, driver behaviour, and abrupt weather changes.

- **Generalization Issues:** The results of tread investigations could not be always applicable throughout many areas and driving circumstances. A tire that performs well in one setting may not deliver the same performance in another, therefore restricting the general relevance of the study.

Environmental and Ethical Issues

- **Resource Intensive:** Running tread studies may be resource-intensive, requiring the use of several tires and thorough testing. This begs ethical questions about waste creation and resource use, particularly if the investigations do not result in appreciable progress or development.
- **Microplastic Pollution**: Although tread studies assist to reduce environmental effects, the very testing and tire wear-down process adds to microplastic contamination. This dilemma emphasizes the importance of balancing environmental stewardship with research advantages.

10.9 Trend Study: Application

Trend studies have considerable use in many different sectors. Some of their uses are shown below in a comprehensive perspective:

- **Public Health:** Track changes in public health initiatives, personal behaviour, and disease frequency. Monitoring nutritional condition and health among the American population throughout time, the National Health and Nutrition Examination Survey (NHANES)
- **Sociology:** Examining societal changes in public opinion, social customs, and demographic trends in sociology Using the General Social Survey (GSS), analysing society's developments in attitudes, behaviours, and traits.
- **Economics:** Examining consumer behaviour, employment patterns, and economic statistics helps one understand economics. Monthly data on labour market activities includes employment, unemployment, and CPS (current population survey) wages.
- **Education:** Examining policy impacts, educational outcomes, and school performance. Rates American academic performance and educational development, the National Assessment of Educational Progress (NAEP)
- **Business and Marketing:** Knowing consumer trends, market dynamics, and product life span. Nielsen ratings track, for example, variations in television watching throughout time.

- **Environmental Studies:** Environmental studies include public awareness campaigns, conservation efforts, and change monitoring of the surroundings. Environmental issues questions in the European Social Survey help to track public attitudes.

10.10 Conclusion

From improving vehicle safety and supporting forensic investigations to inspiring innovation and supporting sustainability, tread studies have several advantages. They do, however, also present major difficulties like high prices, time commitment, complexity, and environmental problems. Maximizing the benefits of tread research depends on striking a balance between these advantages and disadvantages thus reducing their negative effects. Within the field of car safety, tread studies are very vital. By guaranteeing prompt tire replacements, they not only serve to minimize mishaps but also propel ongoing innovations in tire design. Safer and more dependable tires produced by this iterative process help to ensure road safety generally. Tread analysis gives forensic investigators a great weapon for justice in reconstructing events and connecting suspects to crime sites.

From an environmental standpoint, tread studies show how tire wear contributes to pollution and support the creation of more ecologically friendly tire substitutes. Although the manufacturing process itself may add to pollution, the long-term advantages of producing tires with a less environmental impact exceed these issues. Economically, tread studies will help customers as well as producers. Extended tire life and better fuel economy help consumers save money; manufacturers may use tread analysis information to develop and improve their product lines. These advantages, however, have a price. Particularly for smaller organizations, the great costs related to research, technology, and equipment might be a challenge. The time-consuming character of thorough research and long-term testing might slow down the rate of invention. Analysing several situations and interpreting data also calls for a great degree of knowledge.

Predictive accuracy is still difficult because many erratic elements influence tire performance. Additionally difficult is generalizing results from several areas and underlying driving circumstances. Moreover, the resource-intensive character of tread studies generates ethical and environmental questions that call for a careful strategy to study and development. Tread studies are all things considered a two-edged blade. Their notable benefits in safety, forensic science, environmental effect,

financial gain, and performance enhancement span Still, the related expenses, time commitment, complexity, predictive restrictions, and environmental issues create rather difficult problems. Finding a balance that leverages tread studies' advantages while addressing and reducing their negative effects would help one to fully appreciate their possibilities.

10.11 Summary

Trend studies are a longitudinal research method that investigates temporal changes by analysing patterns and trends within a population across designated timeframes. They contrast with panel studies, which track the same individuals or groups throughout multiple time points. Trend studies examine changes at the population level by utilizing diverse samples from the same population across various time intervals. This facilitates the discovery of patterns in pertinent variables, such as public sentiment regarding climate change. They are temporal, facilitating the study of changes via historical comparisons. Trend studies are extensively utilized across many disciplines, such as sociology, economics, political science, public health, and marketing, to comprehend societal transformations, forecast future trends, and inform decision-making.

10.12 Questions

1. What is trend study? Define characteristics and types of it.
2. Discuss advantages and disadvantages of trend study.
3. What are the differences between trend study and cross-sectional study?

10.13 References

1. Aguinis, H. (2024). *Research Methodology: Best Practices for Rigorous, Credible, and Impactful Research.* Singapore: Saze Publication.
2. Babbie, E. (2020). *The Practice of Social Research.* USA: Cengage.
3. Bryman, A., Clark, T., Foster, L., & Sloan, L. (2022). *Social Research Methods.* UK: Oxford University Press.
4. Dubey, U. K., & Kothari, D. P. (2022). *Research Methodology: Technique and Trends.* London: CRC Press.

5. Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal Data Analysis.* Canada: John Wiley & Sons.
6. Kothari, C. R., & Garg, G. (2019). *Research Methodology: Methods and Techniques.* Delhi: New Age International Publishers.
7. Kumar, R. (2014). *Research Methodology: A Step by Step Guide for Beginners.* London: Sage.
8. McTavish, D. G., & Loether, H. J. (2015). *Social Research: An Evolving Progress.* Chennai: Pearson.
9. Mukherjee, S. P. (2020). *A Guide to Research Methodology.* New York: CRC Press.
10. Verbeke, G., & Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data.* Belgium: Springer.
11. Walls, T. A., & Schafer, J. L. (2006). *Models for Intensive Longitudinal Data.* New York: Oxford University Press.

10.14 Glossary

- **Trend Study:** Trend studies are a kind of longitudinal research technique used in analysis of changes over time. Unlike other longitudinal techniques, including panel studies, trend studies concentrate on analysing patterns and trends in a population over certain times rather than following the same people. Examining changes in a population or phenomenon over time is the emphasis of trend studies, a form of longitudinal research methodology.
- **Cohort Trend Studies:** To define Cohort Trend Studies, track a certain cohort such as a group of individuals who experience a particular occurrence in the same time period across time. For example: looking at the career growth of other graduates of that same year.
- **Repeated Cross-Sectional Studies:** Poll multiple samples from the same population at various times across a repeating cross-sectional study. For instance, yearly national surveys on public opinion on climate change.
- **Panel Trend Studies:** Combining elements of panel studies and trend studies, this hybrid technique consists of recurrent evaluations on the same population including new respondents. For a given cohort, for instance, tracking health outcomes across several years.

MODULE - IV
Methods of Quantitative Data Collection in Social Research

Unit 11 □ Collecting Quantitative Data: Units of Analysis

Structure

11.1 Learning Objectives

The learning objectives of this unit include:

- Explaining the unit of analysis in sociological research.
- Understanding its significance in research design.
- Identifying common units
- Analyzing factors influencing the selection of the unit of analysis.
- Understanding how the unit affects data collection and analysis methods.
- Recognizing the challenges and limitations of different units.

- Formulating research questions with appropriate units.
- Design studies with clear units of analysis.

By achieving these objectives, learners will gain a comprehensive understanding of the unit of analysis in sociological research.

11.2 Introduction

In sociological research, the unit of analysis refers to the primary entity being studied or observed. It is the main element or focus of the study, such as individuals, groups, organizations, social artifacts, communities, or events. By clearly defining the unit of analysis, sociologists can ensure that their research is focused, coherent, and capable of addressing their objectives effectively. Understanding the unit of analysis helps in drawing meaningful conclusions about social phenomena and contributes to the overall rigor and validity of the study. Thus, in research methodology, the unit of analysis refers to the "what" or "who" being studied in a research project, essentially defining the specific element that the researcher is collecting data about and drawing conclusions from; it's the primary focus of the investigation

For example, if a researcher is interested in understanding job satisfaction, here the unit of analysis is the individuals. The researcher might survey employees to gather data on their personal job satisfaction levels. Each employee's response represents an individual data point. This focus on individuals allows the researcher to analyze how various personal factors, such as age, gender, and tenure, affect job satisfaction. By clearly defining individuals as the unit of analysis, the study can yield insights into the specific experiences and perceptions of employees.

11.2.1 Overview of Common Units of Analysis

Units of analysis are the people or things whose characteristics social researchers observe, describe, and explain. Typically, the unit of analysis in social research is the individual person, but it may also be a social group, a formal organization, a social interaction, a social artifact, or other phenomena such as lifestyles or social interactions (Babbie, 2016).

1. Individuals: The individual unit of analysis focuses on the behaviour, attitudes, and experiences of single persons. For example: Researching job satisfaction among employees by surveying individuals to understand personal factors influencing their satisfaction.

Research Question: What factors influence job satisfaction among employees?

Unit of Analysis: Individuals.

Explanation: A study surveys employees across various industries to assess their levels of job satisfaction and identify factors such as age, gender, tenure, and workplace environment that contribute to their satisfaction.

2. Groups: This unit analyzes the characteristics or dynamics of groups, such as families, peer groups, or social classes. For example: Studying how peer groups affect academic performance in high school students by examining different peer groups and their impact on members' grades.

   Research Question: How do peer groups affect academic performance in high school students?

   Unit of Analysis: Peer groups.

   Explanation: A researcher examines different peer groups within a high school to analyze how group dynamics, support systems, and shared study habits influence academic performance.

3. Organizations: The organizational unit of analysis looks at formal institutions and organizations like schools, corporations, or political parties. For e**xample:** Investigating how organizational culture influences employee retention rates by comparing different companies and their management practices.

   Research Question: What is the impact of organizational culture on employee retention rates?

   Unit of Analysis: Organizations.

   Explanation: A study investigates various companies to compare retention rates and analyze how different organizational cultures, management practices, and employee benefits influence the likelihood of employees staying with the company.

4. Social Artifacts: This unit focuses on products of human behaviour, such as books, buildings, or media content. For example: Analyzing gender roles in children's literature by examining how male and female characters are depicted in various children's books.

   Research Question: How are gender roles portrayed in children's literature?

   Unit of Analysis: Children's books (social artifacts).

   Explanation: A content analysis of children's books examines the representation of gender roles, identifying patterns, themes, and stereotypes in how male and female characters are depicted.

5. Communities: The community unit of analysis examines geographic or social communities, such as neighbourhoods or online communities. For e**xample:** Assessing the impact of urban redevelopment on community cohesion by studying different urban neighbourhoods and changes in social ties and engagement.

   Research Question: What is the effect of urban redevelopment on community cohesion?

   Unit of Analysis: Urban neighbourhoods (communities).

   Explanation: A researcher studies different urban neighbourhoods undergoing redevelopment to assess how these changes impact social ties, community engagement, and the overall sense of community among residents.

6. Events: This unit studies specific events or occurrences, such as protests, natural disasters, or elections. For example: Researching the psychological impact of natural disasters by investigating mental health outcomes of individuals affected by events like hurricanes or earthquakes.

   Research Question: How do political protests influence public opinion and policy change?

   Unit of Analysis: Political protests (events).

   Explanation: A study analyzes various political protests to understand their impact on public opinion, media coverage, and subsequent policy changes, examining factors such as protest size, duration, and the issues being addressed.

   These examples illustrate how different units of analysis shape research design and provide unique insights into various social phenomena.

11.3 Importance of the Unit of Analysis in Sociological Research

It is important to understand the unit of analysis for research to be more specific and concise. There are several other factors in that influence the understanding. According to Bryman (2012), those factors are as following:

1. The unit of analysis determines the type of data that needs to be collected and the methods to be used.
2. It influences how research questions are formulated and what aspects of social phenomena are examined.

3. Different units of analysis may require different analytical techniques and approaches.
4. Clearly defining the unit of analysis helps in interpreting the results and understanding their implications for the broader context.

So, specifying the unit of analysis, assists the researchers in formulating the design studies that are focused, coherent, and capable of addressing the research objectives effectively.

11.4 Choosing the Right Unit of Analysis: Factors to consider when selecting a unit of analysis and how the research question influences the choice

It is important to consider few important factors when selecting a unit of analysis. The following explanation helps to clarify the different factors in choosing the right unit of analysis in brief:

1. The research question should directly inform the choice of the unit of analysis. It is necessary to determine what the researcher is primarily trying to understand or explain. For example, if the research question is "How do organizational cultures affect employee satisfaction?" the unit of analysis should be organizations since the focus is on comparing different organizational cultures.
2. The scope of the study, including its breadth and depth, can influence the unit of analysis. The researcher needs to decide whether the study aims to explore broad trends or deep, specific insights. For example, in a study examining the impact of social media on political engagement, if the scope is broad, the unit of analysis might be individuals across different demographics. For a more focused scope, the unit could be specific online communities.
3. The availability and accessibility of data can be a determining factor. It is important for the researcher to ensure that there is sufficient data for the chosen unit of analysis. For example: If studying the effects of educational programs, selecting schools as the unit of analysis might be practical if data on school performance and programs is readily available.
4. The researcher needs to determine whether the study requires a micro, meso, or macro level of analysis. The unit of analysis should align with the intended level of analysis. For example: For a micro-level analysis of social interactions, the unit

of analysis could be individual interactions in specific settings. For a macro-level analysis of economic inequality, the unit might be countries or regions.

5. The chosen research methods can influence the unit of analysis. The researcher needs to select a unit that is compatible with the methods he plans to utilize. For example: For qualitative research using interviews, individuals or groups might be appropriate units of analysis. For quantitative research using surveys, larger units such as communities or organizations might be more suitable.
6. The theoretical framework guiding the study can help in selecting the unit of analysis. The researcher needs to ensure the unit aligns with the theoretical perspective. For example: In a study grounded in social network theory, the unit of analysis might be social networks or groups, as the theory focuses on relationships and interactions within networks.
7. The researcher needs to be aware of potential biases and limitations associated with different units of analysis. He must select a unit that minimizes these issues. For example: If studying the impact of media on public opinion, choosing social artifacts (e.g., news articles) as the unit of analysis can help avoid biases associated with self-reported data from individuals.

Thus, choosing the right unit of analysis is critical for the success of a sociological study. By carefully considering factors such as the research question, scope of study, data availability, level of analysis, research methods, theoretical framework, and potential biases, researchers can ensure that their study is well-designed and capable of yielding meaningful insights.

Also, the research question plays a pivotal role in determining the appropriate unit of analysis for a sociological study.

1. Defines the Scope of Investigation: The research question outlines what the study aims to investigate, which in turn dictates the scope of the analysis. If the research question asks, "How does social media use influence political participation among young adults?", the unit of analysis is likely to be individuals (young adults), as the focus is on their behaviours and experiences.
2. Identifies the Main Focus: The research question identifies the primary focus or entity to be studied, which directly influences the choice of the unit of analysis. If the question is, "What factors contribute to the success of community-based environmental initiatives?", the unit of analysis might be communities or community groups involved in such initiatives.

3. Determines the Level of Analysis: The level of analysis (micro, meso, or macro) is influenced by the research question, guiding the selection of the unit of analysis. For a question like, "How do family dynamics affect adolescent mental health?", the level of analysis is micro, and the unit of analysis could be family units or individual family members.
4. Shapes the Methodological Approach: The research question determines the methodological approach, including data collection and analysis methods, which are aligned with the unit of analysis. A question like, "How do corporate policies impact employee job satisfaction?", suggests a mixed-methods approach, with the unit of analysis being organizations for policy analysis and individuals for measuring job satisfaction.
5. Ensures Coherence and Relevance: The research question ensures that the chosen unit of analysis is coherent and relevant to the study's objectives, leading to meaningful and applicable findings. If the question is, "What role do local governments play in disaster preparedness?", the unit of analysis should be local governments, as this aligns with the study's aim to evaluate their roles and actions.
6. Guides Data Collection and Analysis: The research question guides the data collection and analysis processes, ensuring that the chosen unit of analysis is practical and feasible for the study. For a question like, "How do media representations of gender affect public perceptions?", the unit of analysis could be media artifacts (e.g., TV shows, movies, news articles), focusing on the content and its impact on the audience.

By carefully aligning the unit of analysis with the research question, researchers can design studies that are focused, coherent, and capable of addressing the study's objectives effectively. This alignment ensures that the study yields meaningful insights and contributes to a deeper understanding of the social phenomena under investigation.

11.5 Implications of the Unit of Analysis

The implications of the unit of analysis can be understood from the understanding of how the unit of analysis affects data collection methods, it's impact on data analysis techniques and the potential challenges and limitations.

11.5.1 Unit of Analysis Affects Data Collection Methods

The chosen unit of analysis directly influences the data collection methods used in a study. Different units of analysis require specific approaches to gather relevant data. The following points need to be kept in mind while planning for data collection methods-

- Individuals: Surveys, interviews, observations, and experiments are commonly used to collect data from individual participants.
- Groups: Focus groups, group interviews, and participant observation are suitable for studying group dynamics and interactions.
- Organizations: Document analysis, organizational records, and case studies help gather data from organizations.
- Social Artifacts: Content analysis, archival research, and textual analysis are used to examine social artifacts like books, media, and art.
- Communities: Surveys, ethnographic studies, and community mapping are effective for collecting data from geographic or social communities.
- Events: Event history analysis, narrative analysis, and media reports are used to study specific events and their impact.

11.5.2 Impact on Data Analysis Techniques

The unit of analysis also affects the choice of data analysis techniques. Each unit requires specific methods to interpret and analyze the data accurately. The following points need to be kept in mind while planning for data analysis techniques-

- Individuals: Statistical analysis, thematic analysis, and regression analysis are commonly used to analyze individual-level data.
- Groups: Social network analysis, group comparison, and interaction analysis help understand group-level phenomena.
- Organizations: Case study analysis, organizational analysis, and comparative analysis are suitable for studying organizations.
- Social Artifacts: Content analysis, discourse analysis, and semiotic analysis are used to analyze social artifacts.
- Communities: Spatial analysis, community profiling, and demographic analysis are effective for examining community-level data.
- Events: Time-series analysis, event history analysis, and impact assessment are used to analyze event-related data.

11.5.3 Identifying the potential challenges and limitations:

Each unit of analysis presents unique challenges and limitations that researchers need to consider.

Few of the challenges may be as following-

- Ensuring representative samples, avoiding response biases, and maintaining confidentiality.
- Managing group dynamics, ensuring diverse group representation, and dealing with groupthink.
- Accessing organizational records, obtaining consent from multiple stakeholders, and maintaining objectivity.
- Selecting relevant artifacts, ensuring content validity, and interpreting subjective meanings.
- Engaging community members, dealing with diverse perspectives, and managing logistical issues.
- Obtaining accurate event data, managing temporal aspects, and dealing with retrospective biases.

Few of the limitations may be as following-

- At individual-level data may not capture broader social dynamics.
- At group-level data may overlook individual differences.
- At organizational-level data may not reflect individual experiences.
- Artifact analysis may not capture the context or intent behind the creation. Community-level data may generalize individual experiences.
- Event analysis may not account for long-term effects or broader contexts.

By understanding how the unit of analysis affects data collection methods, data analysis techniques, and potential challenges and limitations, researchers can design more effective and comprehensive sociological studies.

11.6 Conclusion

In conclusion, the unit of analysis is a cornerstone of sociological research that underpins the entire research process. By clearly defining and selecting the appropriate unit of analysis, researchers can ensure that their studies are well-structured, relevant, and capable of yielding valuable insights into social phenomena.

Whether the focus is on individuals, groups, organizations, social artifacts, communities, or events, the choice of the unit of analysis shapes the research design, data collection, and analysis, ultimately contributing to the rigor and validity of the study. As sociologists continue to explore the complexities of social life, a thorough understanding of the unit of analysis will remain essential for producing meaningful and impactful research.

11.7 Summary

The unit of analysis is a fundamental concept in sociological research, referring to the primary entity being studied or observed. Understanding the unit of analysis is crucial for guiding data collection methods, shaping research questions, influencing data analysis techniques, and ensuring coherent and meaningful findings. Common units of analysis include individuals, groups, organizations, social artifacts, communities, and events. Each unit provides unique insights into social phenomena and requires specific approaches for effective study. Factors such as the research question, scope of study, data availability, level of analysis, research methods, theoretical framework, and potential biases play a critical role in selecting the appropriate unit of analysis. By carefully considering these factors, researchers can design studies that are focused, comprehensive, and capable of addressing their research objectives effectively.

11.8 Questions

1. What is the unit of analysis in sociological research?
2. Why is the unit of analysis important in designing sociological studies?
3. What are the common units of analysis in sociological research?
4. Provide an example of a sociological study with individuals as the unit of analysis.
5. What factors should be considered when selecting a unit of analysis?
6. How does the research question influence the choice of the unit of analysis?
7. Explain how data availability impacts the selection of the unit of analysis.
8. How does the unit of analysis affect data collection methods?
9. Discuss the impact of the unit of analysis on data analysis techniques.
10. What are the potential challenges and limitations associated with different units of analysis?

11. If you were to study the effects of a natural disaster on mental health, what unit of analysis would be most appropriate? Explain your reasoning.
12. In researching the role of local governments in disaster preparedness, how would the choice of unit of analysis shape your study's design?

11.9 References

1. Babbie, E. (2016). *The Practice of Social Research* (14th ed.). Cengage Learning.
2. Bryman, A. (2012). *Social Research Methods* (4th ed.). Oxford University Press.

11.10 Suggested Readings

- Neuman, W. L. (2014). *Social Research Methods: Qualitative and Quantitative Approaches* (7th ed.). Pearson.
- Singleton, R. A., & Straits, B. C. (2018). *Approaches to Social Research* (6th ed.). Oxford University Press.
- Trochim, W. M. K., & Donnelly, J. P. (2008). *The Research Methods Knowledge Base* (3rd ed.). Cengage Learning.
- Pressbooks. (n.d.). *Units of Analysis and Units of Observation*. In *Research Methods for the Social Sciences: An Introduction*. Retrieved from Pressbooks.
- McKeever, R. (n.d.). *Units of Analysis*. In *The Craft of Sociological Research*. Retrieved from Pressbooks.

Unit 12 □ Collecting Quantitative Data: Sampling

Structure

12.1 Learning Objectives

The learning objectives of this unit include:

- To identify the different Sampling Methods and distinguish between various sampling techniques.
- To familiarize the different terms related to sampling.
- To understand the importance of Sampling

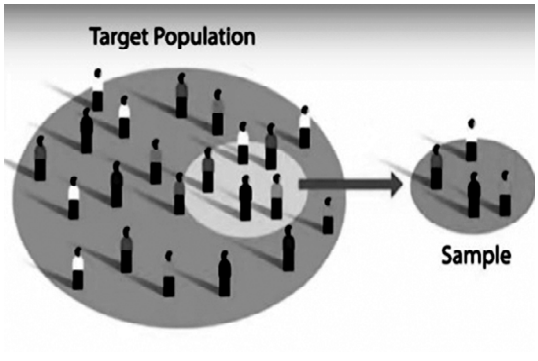
12.2 Introduction

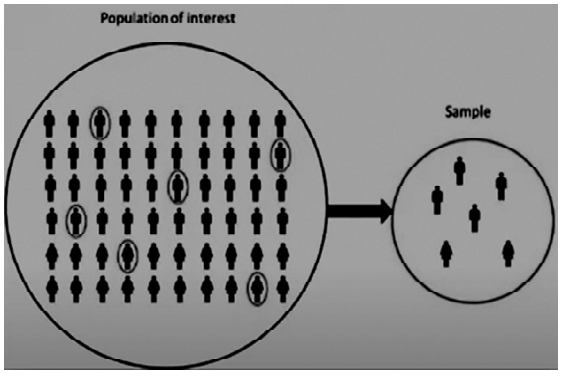
Social research attempts to study a larger population and aims to bring out a conclusion based on the findings. These findings are generalized to the larger population. The process of forming a subset by selecting individuals, cases, or groups from a larger population is known as sampling (Kani Khan & Reza, 2022).

Human minds have been curious. They try to investigate and understand various aspects of our daily lives. To do so, one identifies the area of interest and thus the population of interest. For example- the population of interest may be the women of a city or the persons with disability of a district, and it becomes challenging to cover each of them, thus we form a smaller representative group, the sample.

In social research, good research intends to get more distinct information from a lot of people. This becomes a challenge if the target group is too big or there is a time limit in collecting data. In these cases, to obtain rich data, a smaller portion of the population is considered to provide information. This allows the researcher to follow the sampling technique in social research (Au, 2018).

Sampling is considered as a scientific method or a process depending on the context in which it is used. Sampling is considered as a technique that is used in research to choose a number of units or individuals from a larger population. By using this technique, researchers can generalize about the population as a whole without needing to gather information from each individual. On the other hand, sampling is called as a process as it involves several steps, like, selecting the population, determining a sampling frame, deciding on a sampling technique, and gathering the sample. This process guarantees the validity and reliability of the data gathered as well as the sample's representation of the population.





12.3 Familiarizing with few Terminologies

Before an understanding of different sampling techniques, it is important to familiarize few related terminologies that would facilitate better understanding of the whole topic.

12.3.1 Population, which is denoted by 'N' means the entire group of people, objects, or items to be studied and from where inferences/ conclusions need to be drawn. It varies in size-could be a big group to a smaller one (a country, city, school). Population usually answers for 'who' and 'what' the researcher wants to understand and makes inferences at the end of the study. For example, if the researcher aims to understand the mobile or cell phone use pattern among adolescents of Kolkata, then the population would include all the adolescents of Kolkata.

12.3.2 The population of interest shares specific characteristics in alignment with the research question from where data is to be collected. For example, if the researcher is interested in understanding educational challenges faced by the university students of Kolkata during a pandemic, then the population of interest of the researcher might be the university students of different levels (may be undergraduate or postgraduate) of Kolkata within a specific time frame.

12.3.3 Target population is closely related to population of interest. It also shares specific characteristics in alignment with the study where the experiment or intervention is to be conducted. For example, if the researcher emphasis in understanding the impact of Sukanya Samriddhi Yojana, then the researcher's target population could be the parents of the girls of a specific area.

12.3.4 Study population is common in experimental studies. This is again a specific subset of individuals or units where the researcher directly observes, collects data from, and analyze in his study. It is a subset of population which is smaller than the target population, narrowing the characteristics of the population. It is the group that is considered in statistical analysis. For example, if the researcher conducts a survey on smartphone usage among college students in Kolkata, his study population would be those college students with some inclusion criteria.

12.3.5 Sampling frame is the actual collection of units. It is the list from which units are drawn for the sample. It is also an accessible section of the target population.

For example, if the researcher intends to study certain behaviour patterns of cancer patients, the sampling frame could be formed from the hospital patient registry.

While planning for data collection and statistical analysis, a survey may be conducted by either of two methods, namely, the census method or the sampling method.

- The **census method** deals with the investigation of the entire population. Here data are collected for every unit of the universe. Thus, provides more accurate and exact information as no unit is left out.
- **The sample method** deals with the investigation of specific units selected based on certain criteria. Thus, it is the representation of the larger population aiming to save resources.

Table 12.1: A comparative understanding between census and sample

Parameters	***Census***	***Sample***
Purpose	**-A**ims in collecting data from every individual or unit within a population. -provides a complete and exhaustive snapshot of the entire group.	-Involves selecting a subset (sample) from the population. -Researchers use this sample to estimate population characteristics without surveying everyone.
Utilization	-For small populations (e.g., a small village or a specific organization), a census is feasible and accurate. -In cases where precise data is crucial (e.g., national censuses for policymaking or resource allocation).	-For large populations (e.g., global surveys, market research), sampling is practical.
Advantages	-Provides accurate information about the entire population.- No sampling error; as it covers everyone.	-Saves time, effort, and costs compared to a full census.
Challenges	- Conducting a census can be time-consuming, costly, and logistically challenging.-Collecting detailed data from everyone may raise privacy issues.	-There is always some degree of uncertainty due to sampling variability.-Generalizability-results apply only to the sampled group, not the entire population.

**Source: Author*

12.4 Importance of Sampling in Social Research

In social research, sampling has its importance as to ensure the accuracy and the reliability of the findings, such as-

- **Generalizability:** The results of a sample can be extrapolated or generalized across the entire population where the sample is representative of the population. This is essential to draw valid inferences and predictions.
- **Accuracy and Precision:** Determining the characteristics of a population with precision and accuracy can be achieved by utilizing a meticulously selected sample. Findings are more reliable when sampling methods are used appropriately, thereby minimizing biases and errors.
- **Feasibility:** Due to resource constraints, like, financial, time, etc, studying the entire population becomes unrealistic and even impossible. Utilizing an appropriate sampling method assists researchers to be more efficient and effective in conducting the study.
- **Ethical Considerations:** Sampling can alleviate the workload on participants and prevent superfluous data gathering from the entire community, which can have ethical advantages.
- **Resource Management:** Sampling facilitates the efficient utilization of resources. By narrowing their focus to a smaller group, researchers can optimize the allocation of their resources, thereby guaranteeing comprehensive and rigorous study.
- **Sampling Design** is the systematic approach or strategy employed to choose a representative subset from a larger population. It describes the techniques and strategies for determining which people or units to include in the sample and through which criteria to choose them. Reliable and valid conclusions can be drawn from a sample that accurately represents the population attributable to a well-constructed sampling design.

12.5 Sampling Design and Process

The process of developing a sample involves several steps, such as defining and identifying the population of interest, choosing a sampling frame, defining the sampling technique, calculating the sample size, and finally carrying out the sampling plan.

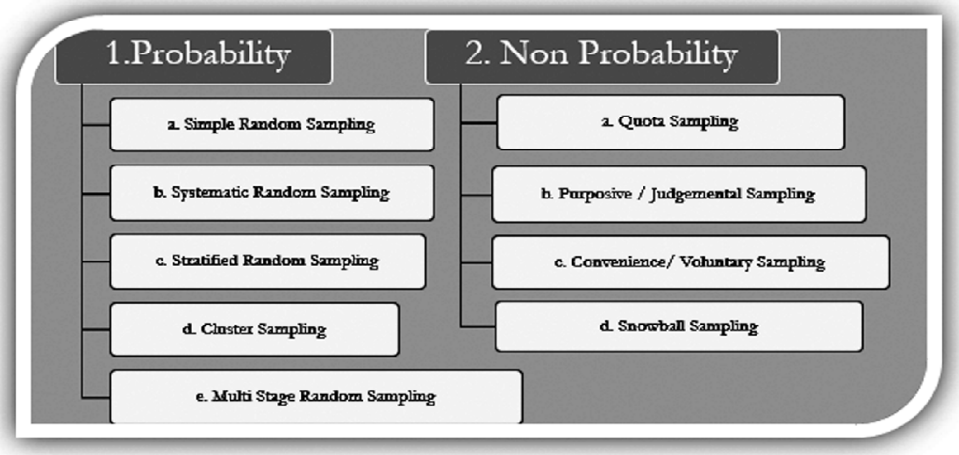
12.6 Types of Sampling Methods

It is crucial to keep in mind that while generating a sample, it must be impartial, representative, and capable of addressing a range of potential difficulties. Sampling biases must therefore be properly managed using a scientific method. Sampling bias, which is typically the outcome of an inadequate sampling design, includes the propensity to favour the selection of participants with specific traits. The most prevalent bias is the non-response, which occurs when certain participants have little or no possibility of being included in the sample.

Sampling techniques are used in both quantitative as well as in qualitative research. The need for sampling is significant as it saves resources, like, time and money, and is most effective when the size of the population is huge. On the other hand, it also has its limitations where a biased sample leads to inaccurate interpretation and personal biases of the researcher lead to erroneous interpretation.

To handle different challenges, two sampling methods, namely, random or probability sampling method, and the non-random or non-probability methods are utilized in social research. In social research, the type of sampling method that is utilized is based on several factors, like, research objectives, availability and access to sampling frame, and the study design. The flow chart (chart no.-12.1) represents the main two types of sampling methods along with its subtypes in a nutshell.

Chart No.-12.1: Sampling Methods



**Source: Author*

12.6.1. In **probability or random sampling method,** participants or units are randomly chosen where all participants have equal chances of being selected. Constructing a sampling frame first, then selecting a sample from the sampling frame utilizing a computer programme that generates random numbers, is a common method that researchers can use to conduct random sampling (Zikmund, et al., 2014). The term probability is associated with the term randomization and examples of randomization are when individuals flip a coin or they throw a dice. For a given degree of sampling error, probability or random sampling offers the most freedom from bias but may also reflect the most expensive sample in terms of time and effort (Brown, 1947). The probability or random sampling methods consists of simple random sampling, stratified random sampling, cluster sampling and systematic sampling.

12.6.2. In **non-probability or non-random sampling** method, participants or units are chosen not randomly but deliberately. Researchers use subjective judgment or specific considerations (such as convenience, availability, or expertise) to select participants using non-random method. Case study research design and qualitative research typically coincide with non-probability sampling. Regarding the last type, case studies usually concentrate on small samples and are meant to investigate a real-life occurrence rather than draw statistical conclusions in respect to the larger population (Yin, 2014). Although strong rationale is required for the inclusion of some circumstances or individuals rather than others, a sample of participants or cases does not need to be representative, or random. The non-probability or non-random sampling methods includes convenience sampling, quota sampling, purposive sampling, and snowball sampling.

Here table 12.2 presents a comparative picture of both probability and non-probability sampling.

Table 12.2: Comparison of the primary types of Sampling

Comparing Parameters	***Probability Sampling***	***Non-probability Sampling***
Selection Method	Through random methods.	Through non-random methods.
Utilization	Mainly used in quantitative research.	Used in both quantitative and qualitative research.
Statistical Validity	Allows to use statistical techniques, like, margins of error, confidence intervals etc to validate the findings.	Limits the use of most of the statistical techniques.

Resource Use	More resource-intensive (e.g., cost, time) compared to non-probability methods.	Less resource-intensive (e.g., cost, time) compared to probability methods
Challenges in Access	Access to complete list for a larger population is a challenge.	Accessible, quicker and cost effective in compared to probability sampling.

**Source: Author*

1. Random Sampling

The random or probability sampling is further categorized as following-

1a. Simple Random Sampling

This is the best probability sampling technique that helps save time and resources and is a reliable method of obtaining information where every single member of a population is chosen randomly & merely by chance. Here every individual has the same probability of being chosen and is more suitable when the population is homogeneous. For example, the researcher wants to select a simple random sample of 100 students from a school. Here, he needs to assign a number to every student in the school database from 1 to 500 where the school has 500 existing students. Then he may use a random number generator to select a sample of 100 numbers.

Types of Simple Random Sampling: This method is of two types-

- **Simple random sampling with replacement**: Every unit of the population has an equal chance of being chosen for the sample with this method, and once a unit is chosen, they remain in the population. It follows that a single unit may be chosen more than once.
- **Simple random sampling without replacement**: In this method, each unit of the population also has an equal chance of being chosen for the sample, but once a unit is chosen, they are removed from the population. This means that each unit can only be chosen once.

Advantage: The advantages of simple random sampling include representativeness of the universe, free from personal bias and prejudices, simple to use method and assists in assessing sampling error.

Disadvantage: If the units are widely dispersed, the sample becomes unrepresentative and it is not applicable when the units are heterogeneous.

1b. Systematic Sampling

Here the participants are chosen at regular intervals that requires the selection of a starting point for the sample and sample size. This process is repeated at regular intervals and has predefined range, hence its least time-consuming. For example, suppose the names of 500 students of a school are sorted in the reverse alphabetical order. To select a sample in a systematic sampling method, the researcher must choose some 50 students by randomly selecting a starting number, say 5. From number 5 onwards, the researcher will select every 5th person from the sorted list. Finally, the researcher can end up with a sample of some students.

Advantage: The advantages of using systematic random sampling includes its simple and convenient method, it is less time consuming and can be used in infinite population.

Disadvantage: Since systematic sampling is a quasi-random sampling and the sample may not be a representative sample.

1c. Stratified Sampling

Here population is divided into smaller groups or strata with no overlapping but representative of the entire population. While sampling, these groups can be organized, and then draw a sample from each group separately. Samples are taken from each group or stratum based on the ratio of the subgroup's size to the total population. If the population is heterogeneous, the researcher must go in for stratified random sampling. Its purpose is to ensure that every stratum is adequately represented. For example, if the researcher is trying to get the opinion on online classes, the researcher would form different strata of students, like, middle school students, high school students, college students, and university students. Here the researcher forms the strata based on the objectives of the study to get a wholesome picture of the problem. Here the strata are homogenous and share common characteristics. So, the researcher will select individuals from each stratum to make the sample representative and cover every type of student.

Advantage: This sampling method ensures to study of all the sub-populations separately. Here an optimum size of the sample can be determined with a given cost, precision, and reliability. It is a more precise sample and is a representation of subgroups in the population. Also, the biases are reduced and are more precise.

Disadvantage: There is a possibility of faulty stratification and hence the accuracy may be lost. The proportionate stratification requires accurate information on the proportion of population in each stratum.

1d. Clustered Sampling

Here researchers divide the entire population into sections or clusters that represent a population. Clusters are identified and are included in a sample based on demographic parameters like age, sex, location, etc. This makes the sampling very simple for a survey creator to derive effective inferences from the feedback. For example, an educational institution has ten branches across the country with almost the same number of students. Each branch of the educational institutions can be treated as a cluster and these clusters are naturally formed and not by the researcher. If the researcher wants to collect some data regarding facilities and other things, the researcher cannot travel to every unit to collect the required data. Hence, he can use random sampling to select three or four branches as clusters.

Advantage: This type of sampling is of low cost, a simple process where the complete list of sampling units within a population is not required, it can estimate characteristics of both cluster & population and ultimately, this method requires fewer resources.

Disadvantage: The main problem with this type of sampling method is cluster members tend to be alike and each stage in cluster sampling introduces more stages, more errors and ultimately, it is not as accurate as simple random sampling.

1e. Multistage Sampling

This type of sampling method is a complex form of cluster sampling. Here the sampling is carried out in stages and uses variety of probability sampling methods. It is actually a combination of stratified and/or cluster and/or simple random sampling. For example, the NFHS (National Family Health Survey) data is collected by multistage sampling. In the urban area, the wards (primary sampling units) are selected through PPS (Probability Proportional to Size), then Census Enumeration Block (Secondary sampling units) are selected again through PPS. Then simple random sampling is used within secondary cluster to choose HHs (ultimate sampling units) from each CEBs. Here sampling frames available at higher stages but not for ultimate sampling units.

Advantage: It is cost efficient, less time consuming and less resources are required.

Disadvantage: It is a difficult and complex method of sampling that involves errors when we consider the primary stages and ultimately, it is a subjective technique of sampling.

2. **Non-Probability Sampling:** Non-probability sampling method where units chosen not randomly but deliberately.

2a. Quota Sampling

It is a non-probability sampling method where researchers create a sample involving individuals that represent a population, researchers choose individuals according to specific traits or qualities and the researchers decide and create that can be generalized to the entire population. It is based on the researcher's knowledge of the population. For example, a quota sample is best for obtaining the big picture about opinions, concerns, situations, and ways of thinking-particularly when an individual researcher needs quick answers. An example may be as follows. Let's assume that the researcher needs to know about the career goals of a particular university student. More precisely, the differences in the career goals among freshers, juniors, and seniors are to be examined. Suppose the concerned university has 10,000 students enrolled. This can be assumed as the population. Now, the researcher must divide our population of 10,000 students into categories such as freshers, juniors, and seniors. Suppose we find that there are 3000 freshers (30%), 2500 junior students (25%), and 2000 senior students (20%). Our sample must have these proportions. It means that if the sample is 1000 students, then the researcher must consider 300 freshers, 250 juniors, and 200 seniors. Lastly, he may start collecting samples from these students based on our proportion. Quota sampling saves money, saves time, it is convenient for fulfilling specific research purposes and it exactly represent the population of interest in terms of researchers' preferences/ interest of the study.

2b. Purposive / Judgemental Sampling

If your research requires specific information from a particular subset of your population of interest, then purposive sampling is the way to go. Also, if you're dealing with a small population of interest, purposive

sampling can help you have a representative sample for your research. It is elective, or subjective sampling which is a form of non-probability sampling where researchers rely on their own judgment when choosing members of the population to participate in their surveys. It enables researchers to squeeze a lot of information out of the data that they have collected, it is extremely time and cost-effective and here the sampling method includes situation where the respondents are directly approachable. But it is prone to researchers' bias. The researchers use their personal judgement to select participants and units of measurement, thus challenging to convince an audience regarding generalization of the findings.

2c. Convenience Sampling

Here the samples are drawn at the convenience of the investigator. The investigator picks up cases which are easily available units keeping the objectives in mind for the study. For example, in convenience sampling, the researcher picks or uses student volunteers known to the researcher, thus researcher sends the questionnaire to the student volunteers and they would act as the sample units. This method is useful for pilot study; here the sampling uses the results that are easily available and the processes of picking people in the most convenient and faster way to immediately get their reactions to a certain hot and controversial topic and ultimately, here minimum time is needed and minimum cost incurs.

2d. Snowball Sampling

The snowball sampling method is appropriate for small specialized populations and is useful in studies involving respondents rare to find. In studies where participants or respondents are hard to locate such as sex workers, drug abusers, etc. This method takes more time, it is most likely not representative and the members of the population, who are little known, disliked, or whose opinions conflict with the respondents, have a low probability of being included.

To sum up, it is important to understand the strengths and weaknesses associated with each respective sampling technique (Refer table 12.3).

Table 12.3: Strengths And Weaknesses of Sampling Techniques

Techniques	***Strength***	***Weakness***
Convenience sampling	Least expensive, least time consuming, most convenient	Selection bias, sample not representative, not recommended by descriptive or casual research.
Judgmental sampling	Low-cost, convenient, not time consuming, ideal for exploratory research design	Does not allow generalization, subjective
Quota sampling	Sample can be controlled for certain characteristics	Selection bias, no assurance
Snowball sampling	Can estimate rare characteristics	Time-consuming
Simple random sampling	Easily understood, results projectable	Difficult to construct sampling frame, expensive, lower precision, no assurance of representativeness
Systematic sampling	Can increase representativeness, easier to implement than simple random sampling, sampling frame not always necessary	Can decrease representativeness
Stratified sampling	Includes all important subpopulation, precision	Difficult to select relevant stratification variables, not feasible to stratify on many variables, expensive
Cluster sampling	Easy to implement, cost-effective	Imprecise, difficult to compute an interpret results

**Source: Malhotra & Birks, 2006*

12.7 Sample for Qualitative Research

When choosing a sample in qualitative studies, careful consideration of the choice of individual or case or community is very important. In that case, it is necessary to justify why the case is suitable for answering research questions. The justifications

may include whether the case study reveals an unusual or neglected aspect or may compare several similar or different cases.

The sample size in qualitative research should be congruent with the objective of the investigation. For example, research that attempts to develop an in-depth understanding of a phenomenon may necessitate a smaller number of participants in comparison to studies that intend to apply their findings to a larger population. Data saturation is a fundamental concept in qualitative research, indicating the point at which no additional information or themes are identified from the data. This frequently influences the choice of the appropriate sample size (Vasileiou, et al., 2018).

12.8 Challenges of Sampling in Social Research

Needless to mention that sampling in social research face different challenges at different levels. Addressing these challenges equip and emphasis the preparedness of the researcher to handle them-

- **Sampling bias and representativeness**: It is vital to reminisce that improperly chosen or biased sampling leads to erroneous or divergent inferences. Consequently, ensuring the selection of a sample that accurately portrays the entire population of interest is of the highest priority for a good researcher, hence it is expected that the sample captures the diversity and features of the population. The research results are skewed when the sample is not representative of the population. Non-random selection or the overrepresentation of specific respondents can serve as the cause (Bhattacherjee, 2019).
- **Ambiguity and Sensitivity**: Along with a range of issues, social research covers a sensitive issue, like sex workers, drug abusers etc. Researchers must work through uncertainty and consider possible effects on respondents' security and personal lives. Research results may potentially be impacted by respondents' dishonesty (Paul, 2023).
- **Non-Response**: In some cases, using questionnaire or interview schedule while conducting survey, selected respondents may not respond or participate. This becomes problematic leading to biased results (Nikolopoulou, 2022).
- **Problem in identifying Sampling Frame**: A report by Survey Point Team (2024) mentions about sampling frame error where missing individuals from the sampling frame or including the units that should not be there, skews the research outcome by 20 percent.

- **Challenge in accessing population:** There are certain respondents who are hard to identify or reach, or may be reluctant to self-identify, for example, drug abuser, sex violence offenders, homeless individuals etc (Raifman, et al., 2022).
- **Ethical Considerations**: While dealing with a variety of issues and cover different respondents, a researcher must ensure the ethical considerations by respecting human rights and privacy (Russell, 2020).

12.9 Overcoming the Challenges

To handle these issues, engaging in meticulous planning, employing suitable sample approaches, and implementing scientific tactics to mitigate bias and errors is necessary.

- Use of random sampling techniques where each member of the population has an equal chance of being chosen to increase representativeness.
- Relying on incentives, personalization in conversing, and reminders to boost response rates.
- Maintaining an accurate representation of the population by regularly updating and verifying the sample frame.
- Based on population variability, estimated effect sizes, and confidence levels, calculate the appropriate sample size using statistical techniques.
- Use digital tools and online surveys to automate data collection and cut expenses by leveraging technology.

12.10 Sampling Error vs. Non-Sampling Error

During the execution of the sample techniques, errors can occur at any point in time. Hence, it is necessary to have an in-depth understanding of what these errors are for and how those might be addressed. Whereas non-sampling error covers other kinds of mistake during data collecting, processing, or analysis, sampling error relates to differences between a sample and the whole population (Assael & Keon, Spring 1982).

12.10.1 Sampling Error

Sampling error makes a sample unrepresentative of its population and comprises differences between sample and population that are due solely to the participants that have been selected. The purpose of sample is to study the population characteristics.

The sample size is not equal to population size except in the case complete enumeration (refer figure 12.1). Therefore, the statistical measurements, like mean of the sample and mean of the population differ. It occurs because no sample can perfectly represent the entire population. For example, if the researcher measures the average height of a sample of population, it will not exactly match the average height of the entire population. Thus, it is challenging to quantify precisely since researcher do not measure data for the entire population. Good sampling practices, like, choosing a proper sample size, random selection in choosing sample etc help in reducing sampling error. The sampling error may be positive or negative or zero.

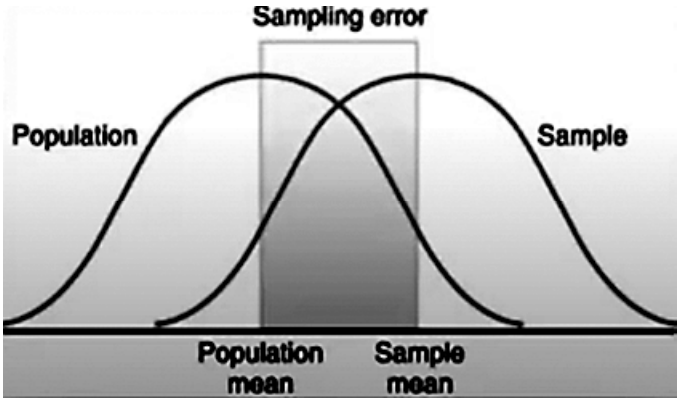


Figure 12.1: Understanding sampling error

**Source: Author*

12.10.2 Non-Sampling Errors

Non-sampling error arises due to various causes right from the beginning stage when the survey is planned and designed to the final stage where the data are processed and analyzed. This issue is more serious than the sampling errors because a sampling error can be minimized by taking a large sample. It is difficult to minimize non-sampling errors, even if a large sample is taken. The sources of non-sampling errors include data entry mistakes, non-response bias, faulty measurement instruments, or errors in data analysis. Non-sampling errors can distort research findings and affect the validity of results. Rigors in developing study design, handling the data carefully and applying quality control measures help in minimizing non-sampling error.

12.11 Sample Size Determination

Bolarinwa (2020) provides essential insights into sample size estimation emphasizing on few parameters (refer table 12.4). This understanding of sample size principles empowers researchers to design vigorous studies.

Table 12.4: Parameters to consider while determining sample size

Parameters	***Specifications***
Purpose of Sample Size Estimation	Accurate sample size determination ensures reliable research outcomes.It balances statistical power, precision, and practical constraints.
Common Study Designs	Different designs (cross-sectional, cohort, case-control) require specific approaches to sample size calculation.Considering the research question, outcome measures, and study duration are important.
Formulae and Guidelines	There are various formulae for various research design and sampling techniques (means, proportions, correlations).It emphasizes adjusting for attrition and clustering effects.
Software and Online Tools	Researchers can use sample size calculators and statistical software.These tools simplify the process and enhance validity.

**Source: Bolarinwa, 2020*

12.12 Conclusion

The success of social research relies mostly on knowledge of and proper application of sample methods. The accuracy, reliability, and wider relevance of the research results directly depend on the sampling technique chosen. Researchers can choose the most effective sampling technique for their studies by carefully weighing the study objectives, demographic characteristics, practical restrictions, and by means of rigorous effort to avoid mistakes and biases. Moreover, reaching statistically significant results without overstretching resources depends on determining the proper sample size. By means of careful preparation and execution, researchers may guarantee that their investigations provide insightful analysis that results in appropriate policy, decision-making, and population inference generation.

12.13 Summary

Thus, this unit deals with sampling in research that involves selecting a subset of individuals from a larger population to make inferences about the entire group. Two main types of sampling methods, i.e, probability sampling (where each member of the population has a known chance of selection, e.g., random, stratified, cluster sampling), and non-probability sampling (where not all members have an equal chance, e.g., convenience, purposive, snowball sampling). The choice of method depends on the research goals, population characteristics, and available resources. Proper sample size determination is essential to ensure accurate results, while minimizing sampling bias and sampling error is crucial for representativeness. Effective sampling helps researchers draw valid conclusions efficiently, avoiding the need to survey the entire population.

12.14 Questions

1. What is sampling? What is it important in research?
2. How does sampling differ from census?
3. Explain the difference between probability and non-probability sampling methods.
4. What are the key features of probability sampling, and when should it be used?
5. Discuss briefly the process of simple random sampling.
6. How does stratified sampling differ from cluster sampling, and when is each method appropriate?
7. What are the advantages and disadvantages of using convenience sampling?
8. In what types of research would snowball sampling be particularly useful?
9. What factors should be considered when selecting a sampling method for a study?
11. How can researchers ensure that their sample is representative of the population?
12. Under what circumstances would a researcher use judgmental or purposive sampling?
13. What is sampling bias, and how can it affect research outcomes?
14. What are some common sources of bias in non-probability sampling?

12.15 References

1. Assael, H., & Keon, J. (Spring 1982). Nonsampling vs. Sampling Errors in Survey Research. *Journal of Marketing, 46*(2), 114-123. doi:10.2307/3203346
2. Au, A. (2018). Sociology and Science: The Making of a Social Scientific Method. *The American Sociologist, 49*, 98–115. doi:10.1007/s12108-017-9348-y
3. Babbie, E. R. (2020). *The Basics of Social Research 15th Edition.* Boston MA: Cengage AU.
4. Bhattacherjee, A. (2019). Sampling. In S. Rowling (Ed.), *Social Science Reseach: Principles, Methods and Practices (Revised Edition)*. Toowoomba, Australia: Pressbooks. doi:https://doi.org/https://doi.org/10.26192/q
5. Bolarinwa, O. A. (2020). Sample size estimation for health and social science researchers: The principles and considerations for different study designs. *Nigerian Postgraduate Medical Journal, 27*(2), 67. doi:10.4103/npmj.npmj_19_20
6. Brown, G. H. (1947). A Comparison of Sampling Methods. *The Jounal of Marketing, 11*(4), 331- 337. doi:10.1177/00222429470110
7. Creswell, J. W., & Creswell, D. J. (2018). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches 5th Edition.* New York: SAGE Publications, Inc.
8. Kani Khan, K., & Reza, M. (2022). Social Research: Definitions, Types, Nature, and Characteristics. In M. R. Islam, N. A. Khan, & R. Baikady (Eds.), *Principles of Social Research Methodology* (pp. 29-41). Singapore: Springer. doi:10.1007/978-981-19-5441-2_3
9. Kothari, C. (2004). *Research Methodology Mthods & Techniques.* New Delhi: New Age International (P) Ltd., Publishers.
10. Kumar, R. (2011). *Resesarch Methodology- a step-by-step guide for beginners.* New Delhi: Sage.
11. Malhotra, N. K., & Birks, D. F. (2006). *Marketing Research: An Applied Approach.* Harlow: FT/Prentice Hall.
12. Nikolopoulou, K. (2022 , November 9). What Is Nonresponse Bias?| Definition & Example. Retrieved July 22, 2024, from https://www.scribbr.com/research-bias/nonresponse-bias/
13. Olsen, W. K. (2004). Triangulation in Social Research: Qualitative and Quantitative

Methods Can Really Be Mixed. In M. Haralambos, & M. Holborn (Eds.), *Developments in Sociology* (Vol. 20). Causeway Press Ltd.

14. Paul, A. (2023). Conclusion: Challenges of Social Research—A Way Forward. In N. Uddin, & A. Paul (Eds.), *The Palgrave Handbook of Social Fieldwork.* Cham: Palgrave Macmillan. doi:10.1007/978-3-031-13615-3_25
15. Raifman, S., DeVost, M. A., Digitale, J. C., Chen, Y.-H., & Morris, M. D. (2022). Respondent-Driven Sampling: a Sampling Method for Hard-to-Reach Populations and Beyond. *Current Epidemiology Reports , 9*, 38-47. doi:10.1007/s40471-022-00287-8
16. Russell, J. M. (2020). *Sampling and Data.* OpenIntro: Pressbooks.
17. Survey Point Team. (2024, May 21). Effective Strategies for Sampling Frame Design. Noida, UP.
18. Vasileiou, K., Barnett, J., Thorpe, S., & Young, T. (2018). Characterising and justifying sample size sufficiency in interview-based studies: systematic analysis of qualitative health research over a 15-year period. *BMC Medical Research Methodology, 18*, 148. doi:10.1186/s12874-018-0594-7
19. Yin, R. K. (2014). *Case study research design and methods (5th ed.).* Oaks, CA: Sage.
20. Zikmund, W. G., D'Alessandro, S., Winzar, H., Lowe, B., & Babin, B. (2014). *Marketing Research.* Sydney: Cengage Learning.

12.16 Suggested Readings

- Bryman, A. (2012). *Social Research Methods* (4th ed.). Oxford University Press.
- Neuman, W. L. (2014). *Social Research Methods: Qualitative and Quantitative Approaches* (7th ed.). Pearson.

Unit 13 □ Surveys Methods in Social Research

Structure

13.1 Learning Objectives

- To learn about the methods of collecting data under survey method.
- To acquaint students with the concept of questionnaires and interviews.
- To develop the questionnaires and schedule and understand the advantages and disadvantages of it.
- To get an idea of differences between a questionnaire and a schedule.
- To understand the use of probing techniques in collection of information.

13.2 Introduction

Data collection is one of the most significant steps in conducting a research. It is the process of gathering information in a systematic way that enables one to answer stated research questions, test hypotheses, and assess the outcomes. It is a very

challenging work which needs proper planning and determination to complete it effectively.

Data collection begins with deciding on what kind of data is required for a particular study followed by the choice of a sample from the universe. Then the instruments to collect the data from the selected sample need to be finalized.

The focus of quantitative research method is on measuring and testing numerical data. This approach is good for getting information from a large number of people in a short span of time.

13.3 Classification of Data

Data are commonly classified into two categories:

A. PRIMARY DATA

B. SECONDARY DATA

A. Primary Data

The primary data are those which have been collected from first hand information, so they are original in nature. The research based on primary data is more reliable and authentic. These are in raw form and collecting this is pretty costly both in terms of time and money. There are various sources from which primary data can be collected. They are

i) Experiments
ii) Survey
iii) Questionnaire
iv) Interview
v) Observations.
vi) Focused Group Discussions
vii) Case studies

Primary data are the raw and original ones collected directly from the source. This allows the researcher to get answers of specific research questions. This helps gain exclusive information and perspectives about the research. These are considered to be more accurate than secondary data.

B. Secondary Data

Secondary data's are collected from a source which has already been published in any form. While doing review of literature in research we make use of secondary

data. Some of the common sources of secondary data are books, journals, records, newspapers, articles on internet etc.

These data are initially collected for different purpose but can be utilized as a basis for further research. These data are readily obtainable and easily accessible. These are less time consuming and cost effective. It is vital to cautiously assess and authenticate secondary data before using it for any research studies.

In this Chapter we will focus on one of the important methods of quantitative data collection from primary sources which is Survey.

13.4 Surveys

Survey is the most commonly used method in social science research. It means collecting information about a group of people by asking them questions and analyzing the results. Survey allows researcher to gather insights linked to a particular area of interest. It aims to collect information, attitudes or behaviours from selected respondents, often representing a larger population. An effective survey can be conducted by keeping in mind the following steps:

i. **Decide on the target population for survey**: Before conducting the survey, the research questions should be clear and based on that the target population has to be identified. The survey should focus to get such results which can be generalized to the whole population. A sample has to be drawn from the total population which should be a representative sample.

ii. **Choose on the type of survey**: There are two main types of survey - Questionnaire and Interview. Which method to select for data collection depends on the nature of the study, unit of enquiry, size of sample, educational level of respondents, type and depth of information to be collected and availability of trained and skilled manpower.

iii. **Plan the survey questions and design**: It is very important to design the survey questions carefully. The types, content, phrasing and ordering of questions should be taken into consideration. The questions should be clear, precise and arranged in a logical order. The types of questions whether it will be close ended or open ended or a combination of both has to be decided. The content of the questions should be such that it can ensure the validity and reliability of results.

iv. **Conduct the survey**: Plan the survey to be conducted by keeping in mind the where, when, how and with whom will the survey are conducted. Based

on the apt research design the survey will be conducted by the selected method.

v. **Analyze the responses**: The collected data has to be processed by editing, classification and coding, Tabulation and Graphical representation. This analysis can be done statistical software like SPSS.

vi. **Write down the findings**: The survey results have to be compiled in a report format.

The survey method helps in faster data collection from a larger size of population. The data collected through this method is easy to analyze and commonly used in social science research.

13.5 Types of Surveys- Questionnaire & Interview

There are two major types of surveys:

A. QUESTIONNAIRE

B. INTERVIEW

The type of survey to be used for a particular research study depends on the sample size, location, and focus of the research.

A. Questionnaire

A questionnaire, as described by Sreejesh (2014), is basically a structured series of questions offered to respondents during an interview, with proper instructions. This instrument holds considerable flexibility across diverse research domains, encompassing survey and experimental design, providing the researchers with a planned structure to gather insights methodically for data collection and analysis.

The questionnaire is a set of questions having both open ended and close ended questions. Open ended questions give an option to the respondents to answer in their own way whereas close ended question have an option to select answers from the given list of choices. This is used to gather data from respondents about their attitudes, experiences, or opinions. It is used to collect data both in the qualitative and quantitative way. Questionnaires are mostly self administered which is sent through mail, online or in pen paper format. These are easy to administer over a large population and is very cost effective. The main drawback is that it can be utilised only on literate population. The response rate can also be low many a times as respondents just ignore the questionnaire sent to them.

A questionnaire, as the most prevailing way of gathering primary and quantitative data, makes the procedure of data collection standardized. Thus, it can guarantee a faster and added accuracy in data collection process, and assist in the data processing also. (Krosnick, 2018; Malhotra, 2006).

As the questionnaire is self administered by the researcher, few basic rules must be kept in mind while constructing it, so that the respondents feel comfortable in answering it. Some of the rules are as follows:

a. The questions should be in clear, understandable language depending on the categories of respondents.

b. Those statements should be used which can be interpreted by all the respondents in similar way.

c. Do not use more than one item per question otherwise the respondents will get confused while answering.

d. Questions should not be biased or ambiguous.

e. Questions should not be such which can lead respondents to an answer.

i. PROCESS OF CONSTRUCTING THE QUESTIONNAIRE

a. Define the research objectives: The researcher should refer to the research proposal and make brief notes of all the objectives and what kind of information is required through this study.

b. Identify the target respondents: The background, knowledge level, educational qualification etc should be identified so that the questions are framed according to their level of understanding.

c. Choose the method of reaching the target respondents: The survey can be done either in pen-paper format, mailed questionnaire or sent via email. This has to be decided before conducting the study so that the maximum target population can be reached without wastage of time.

d. Decide on question content and the language of the questionnaire: The question content should be simple, concise and clear so that the respondents can interpret it properly. The language of the question should be as per the understanding of the respondents.

e. Develop the questionnaire into a meaningful order: the questions should be framed in a meaningful way and in a proper sequence so that the respondents can relate with it and feel comfortable in answering it.

f. **Provide well-defined response options for close ended questions:** The responses for close ended questions should be coded and mentioned clearly which makes it easier for the respondents to choose an answer.

g. **Pre-test the questionnaire:** The pretesting of questionnaire is necessary to see that it works out properly. If any changes are needed that can be done after pretesting.

h. **Develop the final survey form:** The final questionnaire is ready after the pretesting and it can now be sent to the respondents.

i. **Administer the Questionnaire**: The questionnaire can now be administered through the selected method for reaching to the respondents.

j. **Covering letter**: The questionnaire can easily be administered on a large size of sample, so it is desirable to attach a Covering letter to the questionnaire. A good covering letter should include the following things:

- A short introduction of the researcher and basic information about the organization involved in the process of research.
- A brief statement about the purpose of the study.
- Give assurance to the respondent that his/her information will be kept confidential.
- If it is sent through postal mail then enclose a self-addressed envelope for the convenience of the respondents in returning the questionnaire.

ii. USES OF QUESTIONNAIRE

Questionnaires can be an excellent way to gather quantitative data if it is administered in a proper way. It allows the researcher to collect a considerable number of data at a comparatively little cost. The respondents have the convenience to fill up the questionnaire from their home, so it saves time both for the researcher as well as the respondents. The uses of questionnaires are as follows:

a. To easily collect information from the respondents who are scattered in wider geographical area.

b. To accumulate realistic data in order to classify people and their circumstances.

c. To collect clear-cut information describing people's behaviour.

d. To find out at the basic attitudes or opinions of a group of people describing a particular issue.

e. To quantify the contentment level of customers related to a product or service.

f. To gather baseline information this can be utilised to keep a track on the changes in attitudes, behaviour, opinion etc.

g. A relatively cost effective way of data collection.

h. The results can be quickly quantified through the use of software package.

i. Respondents can answer from the comfort of their home.

j. Questions on sensitive issues can be answered in a better way as interviewer is not present.

k. No question of interviewer's biasness.

iii. LIMITATIONS OF QUESTIONNAIRE

a. Chances of non response are high.

b. Many people don't even bother to return the filled up questionnaire.

c. Only literate and educated people can answer the questionnaire.

d. Questions can be interpreted differently as no one is present to clarify the doubts.

e. Questionnaire can be filled up by other person who is not actually the respondent.

B. Interview

The interview is one of the significant methods of primary data collection. It is basically a two persons meeting to gather some information on a particular subject matter. In this method the researcher or the interviewer ask relevant questions and the respondents answer the questions. While conducting interviews the researchers can investigate emerging areas by tailoring their questions and probing based on respondent's responses. The researcher can explore the social, cultural and environmental factors which has an impact on the respondent's opinion and viewpoints. The respondents have the advantage of clarifying the questions reducing the chances of misinterpretation of questions.

Interview is a mode of collecting data as well and at the same time obtaining knowledge from respondents. Kvale (1996:14), defined interviews as " … an interchange of views between two or more people on a topic of mutual interest,

sees the centrality of human interaction for knowledge production, and emphasizes the social situations of research data."

Interview is a two-way process which allows an exchange of ideas and information (Prabnat & Meenu, 2015).

This method allows the researcher to directly interact with the respondents which enables them to gather original information. The interviews can be conducted through different ways such as:

a. Face to face interview
b. Telephonic interview
c. Online survey

a. Face to face interview: Face-to-face interviews are structured interviews conducted by skilled interviewers who use a consistent interview procedure. They offer the chance to build rapport with respondents and get a deeper understanding of their experiences. In face-to-face interviews the interactions can be controlled, complex questions can be asked and probing method can be utilized to fetch more information on a particular topic. This can also help in observing the clients behaviour and reactions. Telephone interviews also help in getting better information as there also interviewer can ask questions and clarify all doubts. Face to face interview can be done either in person or via video calls. It is a very popular method of data collection in social science research. This can be used to gather both quantitative and qualitative data. With all its advantages this method is quite expensive and time consuming.

b. Telephonic interview: This kind of interview is conducted via a phone call and is less expensive and time consuming as compared to face to face interview. The quality of data collected here sometimes gets interrupted due to poor connection or bad sound quality. Another major drawback is that one cannot notice or record the non verbal expressions such as their body language or facial expressions. This type of interview may not be suitable for all types of research but sometimes due to financial or time constraints, it is opted as a method of data collection.

c. Online survey: A most popular method of data collection in the age of technology is online survey which is considered to be time saving and cost effective method. This can be administered to a very large number of populations living in any corner of the world. Though it's a convenient method, the response rate may be low.

i. TYPES OF INTERVIEWS

An interview is a technique to get information from an interviewee by asking questions and getting the answers from them. There are three basic types of interview in research.

a. Structured,

b. Semi-structured

c. Unstructured.

a. Structured Interviews

A structured interview is at times known as standardized interview. The identical questions are asked to all the respondents. Corbetta (2003:269) states structured interviews are "… interviews in which all respondents are asked the same questions with the same wording and in the same sequence."

In structured interviews there is a set of standardized questions which are consistent throughout the interview. The interviewer administers the same set of questions on all the respondents. The researcher's duty is to help out respondents give the answer in a casual and friendly way without showing any biasness or imposing their opinions. It consists of both open ended and close ended questions. This type of interview facilitates consistency and allows easy assessment and quantitative analysis of responses. The chance of getting influenced by the pitch of the interviewer diminishes if the questions could be read out in the same tone of voice to all the respondents. (Gray, 2004: 215).

b. Semi-structured Interviews

Semi-structured interviews are considered as non-standardized and are often used in qualitative research. This kind of interviews is not conducted to test a hypothesis rather done to achieve the objectives of the study (David & Sutton, 2004).

In semi structured interview there is a list of topics and questions which needs to be covered while conducting the interview. The topics decided here serves as an interview guide, which helps the researcher to gather information from the respondents in a particular way. It mostly consists of open ended questions so that the respondents can share their viewpoint about any issue. Semi structured interview has a flexible framework and are widely used in qualitative research.

c. Unstructured Interviews

In unstructured interview there is no structured interview guide or sets of questions. The researcher develops rapport with the respondents helping them to open up and express their viewpoint on the issue selected for research purpose. This is the most flexible method of conducting interview having no predetermined set of questions. The discussion happens in a natural way based on respondent's likeness and interest.

ii. PROCESS OF CONDUCTING INTERVIEWS

Conducting interview is a challenging process and the role of the interviewer is very crucial in this context. The interviewer should have a sound knowledge of research methodology, command over the topic of research and good communication skill to deal with different categories of respondents. On the other hand, a standardized format of questions is required to conduct interviews as it is vital to ask questions in the same way to the whole respondents. Otherwise, the way that questions are asked can overshadow the results (Brace, 2018).

The process of conducting a interview involves the following:

a. Planning- The preparation for conducting the interview has to be done before the actual conduct of interview. The whole interview process has to be planned which included the finalization of respondents finalized, orientation for conducting interview, interview kits ready before proceeding for interview.

b. Preparation of interview schedule- It is a set of questions arranged in a proper sequence which will act as a tool for data collection. The questions have to be drafted & redrafted before the final copy is ready. It should be prepared keeping in mind the objectives of the research study.

c. Pilot testing of schedule- Before the actual interview process, it is advisable to pre-test the schedule on a group of people.

d. Conducting the interview- The interviewer has to reach on time, meet and greet the respondent and communicate in a way so that the respondent is comfortable. Initiate the interview process by asking questions having an eye contact with the respondents. Ask the questions in proper sequence and try to get answers of each question mentioned in the schedule.

e. **Probing** – To get adequate response for a question probing is necessary. This method involves asking for additional information, which is required to explain a situation, opinions or feelings.

f. **Recording the response**- The response can be recorded in pen and paper or recorded in an audio device. Sometimes respondents are not comfortable in audio or video recording then pen and paper recording has to be done. The responses have to be recorded immediately. Abbreviations can be used while recording.

g. **Concluding the Interview**: After completion of the entire process of interview, the closing has to be done properly. Don't be in a hurry. Wind up the conversation appropriately. Thank the respondent for their valuable time.

An interview involves the researcher and the respondents. It can only be successful when the researcher is intelligent, efficient, committed and motivated to the work. If these qualities are there in the researcher, it will be easier for him to motivate the client to give relevant answers.

iii. **ADVANTAGES OF INTERVIEW** Some of the advantages of interview method of collecting data are given below:

a. Interview is more apt for complex situations.

b. Through interview detailed and in depth original information can be gathered.

c. Helps in capturing both verbal and non verbal information.

d. Questions can be explained, restructured or reframed as per the convenience of respondents.

e. The language of communication can be adjusted according to the level and understanding of respondents.

f. Probing technique can be utilized to get adequate responses.

g. If required additional questions can be incorporated based on the response of the respondents.

h. While conducting interview, information can be supplemented by observing the respondents and the environment.

i. The chances of non response are minimized in this method.

j. The chances of getting accurate information are high.

k. It can reach a large section of respondents.

iv. DISADVANTAGES OF INTERVIEW METHOD

a. It is a time consuming and costly process.

b. The quality of data collected depends upon the quality of the researcher.

c. The quality of data may vary when more than one researcher are used to collect information.

d. Interviewer's personal biases, attitudes and beliefs can influence the responses.

v. USES OF INTERVIEW SCHEDULES

i) To secure certain information from the subject, this is known only to him and cannot be gathered from any other source.

ii) To study scientifically the verbal behaviour of the subject under given circumstances.

iii) It is self-reporting practice which provides significant flexibility to the interviewer.

iv) Questions can be clarified while conducting the interview.

v) The respondents can modify their answers.

vi) The behaviour of the respondents can be observed and information and information regarding this can be incorporated in the schedule.

vii) It is a strong and effective tool for collecting data.

viii) It can be adaptable, capable of being used with different categories of respondents.

13.6 Types of Questions used in Questionnaire/Schedule

There are two types of questions used in questionnaire or schedule, Open ended questions and close ended questions.

a. Open ended questions – These are those questions which are intended to fetch a meaningful answer based on the respondent's opinion, feelings and understanding. For eg

- What is your opinion about the effects of covid?
- Can you explain your childhood memories?

The advantage of open ended questions is that it can fetch detailed and more specific information. But at the same time giving these types of answers can be quite time consuming.

b. Close ended questions – The close ended questions limits a person's choice of answers. The options are given and the respondent can select answer from it .For eg

i. In which department do you work?

a) Accounts

b) Human Resource

c) Medical

d) Education

The advantage of close ended questions is that the answers are quick and fast. The respondents do not have to do the brainstorming to find an answer as the options are readily available.

13.7 Difference between Questionnaire and Schedule

S.No.	Questionnaire	Schedule
1.	A questionnaire is a structured instrument used for data collection which is sent to or given to the respondents for filling it up.	A schedule is a set of structured questions which is used to conduct interview with the respondents.
2.	Questionnaire method can only be used when the respondents are literate.	The Schedule method of data collection can be used irrespective of the literacy status of the respondents.
3.	Respondents receive questionnaires through emails, posts and the answers will be given as per directions given in the cover letter.	Answers in the Schedule method of data collection are filled by researchers.
4.	In this method, there is less chances of direct personal contact with the respondents.	In this method, there is direct personal contact of the interviewer and interviewee.
5.	It is economical as compared to schedules.	It is expensive than the questionnaire.

6.	The coverage of the Questionnaire method is widespread as it can be sent to any number of respondents easily.	The coverage of this method is relatively small as there are time and money constraints in sending researchers to larger areas.
7.	In the questionnaire, there is a higher chance of getting wrong or incomplete answers due to inability to have a clear understanding of the given question.	In this method the researchers are present to clarify the doubts of the respondents.
8.	No chance of interviewer's biasness.	There are chances of interviewer's biasness.
9.	The response rate in this method is low as compared to the Schedule.	The response rate in the Schedule method of data collection is high.
10.	This is a time consuming method.	This can be conducted well within stipulated time

13.8 Use of Probing Technique

Probing is the exercise of examining, investigating, or exploring something in a deep way. The effectual use of probing in interviews is to extract deep data from the respondents. The probing questions are made to get deeper answers to some specific questions. **These questions have got to be open-ended**, so that the respondents can have the option to discover their thoughts and put their personal feelings into words.

There are some crucial questions in the schedule which has to be handled with care. It happens sometimes that the respondents are not giving satisfactory answers. In such cases probing comes into picture which is done to get proper response.

Probing question may be like:

- Can you give description about what you said just now?
- Give some additional information about that.
- What is your feeling about this situation?
- How should this situation be improved?

Rubin and Rubin (2005) and Evers and De Boer (20013b) draw on a difference between main questions, probes and follow- up questions. The main questions are set prior to the interview. The probes are utilized to get more, more precise answers. Follow-up questions are entirely new questions prevailed upon by the respondent's earlier answer.

It is very important to understand when a probing question is to be asked. Some of the situations can be as follows:

- When the interviewer is getting unclear or incomplete answers.
- When there are chances to get some unexpected responses.
- To get more in depth information

While probing open ended questions are best way to get information. The probes can always be modified based on respondents responses. Probing helps in enriching the quality of information gathered from the respondents. It gives an improved understanding about respondent's viewpoint.

13.9 Conclusion

Questionnaires are necessary tools for gathering valuable insights professionally and in a cost effective way. A well planned questionnaire is necessary for a successful survey. A good questionnaire is considered to be the one which helps directly attain the research objectives and provides complete and correct information. It is easier for respondents to complete and for researchers to make proper interpretation.

Interview is one of the essential and powerful instruments for data collection. The interview technique is an oral method of securing data particularly in research linked with social sciences. It is not only the verbal skill which matters in this kind of research but also the gestures, postures, facial expressions, pattern of speech etc.

13.10 Summary

To summarize, questionnaires and interviews are two basic data collection methods in surveys, both having its advantages and disadvantages. To choose a method for data collection totally depends on the research objectives, research design, and the universe.

Questionnaires are considered to be more effective when the population size is large.

It is good for quantitative analysis of data. Interviews can help in getting more in depth information and can provide enough information for qualitative analysis.

By skillfully utilizing probing techniques, researchers can come across valuable insights, improve data quality, and fortify the results.

It is the researcher's expertise which can help in reducing the biases and make significant contributions in the field of research study.

13.11 Questions

1) What do you understand by the term data? Discuss about the classification of data.
2) What are the types of survey?
3) What is questionnaire? Discuss the process of construction of a questionnaire.
4) What is the use and limitations of a questionnaire?
5) What is interview? Discuss the types of interview.
6) What is the process of conducting interviews?
7) What are the advantages and disadvantages of the interview method?
8) Discuss about the types of questions used in a questionnaire/ schedule.
9) Differentiate between the questionnaire and schedule.
10) How will you use probing techniques to extract information from the respondents?

13.12 References

1. Brace, I. (2018). Questionnaire design: How to plan, structure and write survey material for effective market research: Kogan Page Publishers.
2. Corbetta, P. (2003). Social Research Theory, Methods and Techniques, London: SAGE Publications.
3. David, M., & Sutton C.D. (2004). Social Research the Basics. London: SAGE Publications.
4. Gray, D. E. (2004). Doing Research in the Real World. London: SAGE Publications
5. Krosnick, J. A. (2018). Questionnaire design The Palgrave handbook of survey research, Springer, 439-455.

6. Kvale, S. (1996). Interviews: An introduction to qualitative research interviewing. Thousand Oaks, CA: Sage.
7. Prabnat, P & Meenu, M.P. (2015). Research Methodology: Tools and Techniques, Bridge Centern.
8. Rubin, H. J., & Rubin, I. (2005). Qualitative Interviewing : The Art of Hearing Data (2nd ed.). Thousand Oaks: Sage Pubications.
9. S. Sreejesh & Sanjay Mohapatra & M. R. Anusree, 2014. "Questionnaire Design," Springer Books, in: Business Research Methods, edition 127, chapter 10, pages 143-159, Springer.

13.13 Suggested Readings

- Goode. W.J. and Hatt. P.K. (1952), Methods of Social Research, McGraw Hill, New York
- Kothari, C.R., 1985, *Research Methodology- Methods and Techniques,* New Delhi, Wiley Eastern Limited.
- Lal Das, D. K. (2000). Practice of Social Research: Social Work Perspectives: Jaipur : Rawat Publications.
- Lal Das, D. K. (2005). Designs of Social Research : Jaipur : Rawat Publications.
- Lundberg, G. A. (1946), Social Research, Longman, New York.
- Pavline V. Young: Scientific Social Surveys & Research, Prentic, Hall of India Private Ltd., 1984.
- Sharma B.A.V , Ravindra Prasad , D, Satyanarayana . PResearch Methodology, in social sciences, sterling Publishers, 1984.
- Wilkinson, T.S. and Bhandarkar, P.L. (1977), Methodology and Techniques of Social Research, Himalaya Publishing, Bombay.

MODULE - V

Descriptive Social Statistics: Measures of Central Tendency and Dispersion

Unit 14 □ Quantitative data analysis

Structure

14.1 Learning Objectives

- To understand the basics of statistics before delving into the discussion of Descriptive Social Statistics.
- To learn about the ways of a careful statistician while dealing with numbers and data, so that one can build up a critical outlook towards statistic and its usages.
- To learn about the relevance of statistical methods in our time, in academic and other fields.
- Getting to know about the history and development of statistics as a distinct discipline.

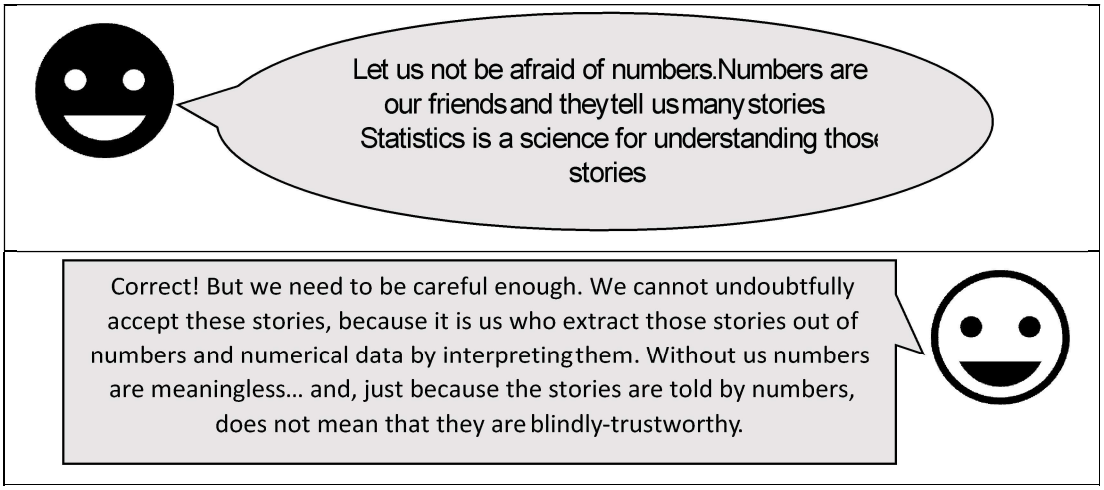
- To have a preliminary understanding of different scales used for measurement, types of variables and types of different statistical methods.
- Knowing as Descriptive social statistics and its types.
- Knowing about two different operations of Descriptive Statistics – Organization and Summarization – and the tools/measures employed for fulfilling these operations.
- To grasp the role of human interpretation in descriptive statistical method.
- Understanding the practicality of statistical methods in the age of 'big data' and internet.
- To be able to envision the moments where sociology meets statistics.

14.2 Introduction

Whenever a study employs statistical methods to present its findings, a certain degree of "scientific rigor" is attached to it, as if the statistician possesses some mystical authority to validate any conclusion. In other words, the sense of precision and objectivity associated with mathematical principles often inspire a feeling of awe, leading to a perception of the study's validity when statistical methods are employed. For this same reason, students are often intimidated to engage with numerical data. Clarifying what statistics entails means to dispel such fears and to foster a rather "humanely" and approachable connection with numbers. It also means moving away from the boring and strict use of statistics and instead critically evaluating the possibilities and openness that statistics can offer. To rephrase, we should not view statistics merely as a science of facts that contrasts with the realm of interpretation. With these previously mentioned precautions in mind, we now delve into a more detailed discussion of statistics.

"It is now clearly recognized that statistical methods are in no sense 'opposed' to the qualitative analysis of case studies but that the two approaches are complimentary" writes Blalock (Blalock, 1960, P.4). In the discussion of statistics, he deliberately took this stance to distinguish his own literary endeavors from older textbooks that so very stringently conformed to an almost institutionalized divide between quantitative and qualitative methods. We have learned that quantitative research strictly deals with numbers and quantifications, turning every possible thing into countable entities, while qualitative methods attempt to avoid such techniques of manipulation, trying to understand things as they are in reality. In

other words, both of these sides adhere to an objective approach in distinct ways. This very adherence to objectivity not only posits an ultimate limitation that they both end up failing to understand the object in its totality (which defeats their intention), but also mitigate the existing gap by bringing them closer on the ground of this purposive failure. Statistics is not devoid of such limitations. In collection, organization and summarization of data, something is already lost that no statistical or scientific venture can retrieve. Furthermore, statistics is again employed to deliver generalizations based on those limited numerical expressions. We now have a minimal exposure to the limited setup on which a researcher or a statistician has to base her/his work. This, however, cannot rob statistics off of its academic popularity. For quite some time, statistics has managed to retain its popularity in several disciplines, providing essential tools for analyzing data, reaching conclusions and making informed decisions. We can very well imagine the weightage of a statement that is supported by statistical measurements in comparison to the one that is not backed by "data". For a similar reason journalists provide "stats" in news articles and reports; "stats" and graphs arouse our bias toward mathematical principles, making the quantitative propositions a little more trustworthy. This calls for an example. Let us compare two statements in terms of their reliability – Statement 1: "The conviction rate in crimes against the SCs went up from 28.5% in 2018 to 32.1% in 2019 and 42.4% in 2020, and then the rate dropped to 36% in 2021 and 34% in 2022", Statement 2: "Over the last five years, even as cases of SC, ST atrocities went up, the conviction rate remained abysmally low". Both of these statements are excerpts from a similar source (The Quint, 12 Dec, 2023) and express similar concerns in a different manner, the former contains numerical expressions and obtains trustworthiness by the virtue of enumerations, while the latter does not, leaving scopes for doubt. It is odd that numerical symbols can facilitate a statement with such mystical attributes that the common man remains un-skeptically ignorant towards it. A statistician, however, must not be blindly apathetic and gullible toward numerical expressions. Instead, s/he should be cautious and mindful that a data set not only has limitations, that it can no way be an accurate reflection of phenomenal reality, but that it can also be interpreted in multiple, often incompatible ways. A careful statistician must also be aware of the sampling strategies, survey-questionnaires and experimental methods (and their limitations), which have been employed to extract a set of data. Furthermore, s/he should apply statistical analysis to imbue the numerical expressions with meaning with the help of available theoretical and conceptual tools. In other words, statistics and sociology intersect under the condition that the social statistician can speak to numbers, without unthinkingly accepting whatever they convey, and always bearing in mind that the numbers have more to say.



14.3 Why Statistics?

Though, now, statistics cannot be held as a standard for authenticating any findings, it surely provides a framework for making informed decisions based on data, reducing reliance on intuition or guesswork. Its principles are applicable across a wide range of fields, including biology, psychology, sociology, engineering, and more, making it a versatile tool. In fields like economics, medicine, and social sciences, statistics helps analyze data to identify trends, correlations, and causal relationships. After all we cannot deny that we live in the time of big-data. If data is simultaneously a numerical expression and characteristics of a certain entity, then big-data signifies the quantified set of attributes of a large number of entities. When we open accounts in banks, post offices, job websites, social media platforms, when we apply for our desired courses in our desired colleges, when we purchase something on online stores (and sometimes during offline purchases) we provide data about ourselves, i.e. our names, ages, nationalities, our linguistic preferences, DOBs, gender, religion, often information about our political affiliations, opinions etc. In the vast virtual hub known as the internet, which is composed of binary numbers (1s and 0s), we become quantifiable entities possessed by attributes. Based on these attributes, we are grouped, separated, entered, and stored in databases, often without our knowledge. Every online activity, from social media interactions to e-commerce transactions, contributes to this vast pool of information. As a result, there is a growing need for sophisticated methods to analyze and interpret this information. The surge in popularity of data science and statistics, on one hand, can be attributed to the rapid expansion and increasing complexity

of this digital landscape. On the other hand, for quite a long-time, the discipline of statistics has been providing researchers with the techniques and methods to understand and make sense of numerical information. Moreover, statistics is integral to predictive analytics, which involves forecasting future trends based on historical data. This capability is valuable across various sectors and disciplines.

14.4 Defining Statistics

In the book *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Box, Hunter and Hunter define statistics as "…the science of learning from data" which "provides methods for collecting, analyzing, and interpreting data, and it is concerned with understanding variability and making inferences about underlying processes" (Box, Hunter & Hunter, 1978, P.1). To learn from data, a statistician needs to build an amicable relation with data, which indeed is not an easy task to do. Blalock also mentions about the "embarrassing" moments when a researcher fails to "absorb" the vastness of the information that s/he has collected. However, a learned statistician ultimately manages to overcome such perplexity.

The term "statistic" was first introduced in 1589 by the Italian scholar Girolamo Ghilini. Ghilini, a Catholic priest and writer, used the term to describe a collection of facts and information about the state. It is highly likely that the term "statistics" in English evolved from this association with "state." In Europe, the field of statistics gained prominence in the 18th century during the Enlightenment, as it responded to the emerging needs of industrial nation-states. Early applications of statistical thinking were primarily linked to demography, cartography, and policy-making. Statistical tools became essential not only in measurements and surveys, but also in the whole process of governing the population through policy and law-making. For instance, with the help of statistics, it became easier for governing elites to identify, quantify, and separate 'adversaries' from 'supporters,' and to manipulate existing policies and laws for their own benefit. Statistical comparison facilitated the measurement and assessment of the strengths and weaknesses of rival military forces. This highlighted the discipline's predictive capacity to estimate potential victories or losses, aiding in the development of war strategies. We can also observe the use of statistics in maintaining numerical records of deaths, illnesses, and injuries during wars, famines, and plagues, which marks its importance in the health-sector. From this we can understand the significant role played by statistics in the work of maintenance and regulation. It is not

surprising that statistics emerged as a discipline during the same period when military, health, legal, prison and other modern institutions became prominent in the West. Sociology emerged as a discipline shortly thereafter, when the conflict between the two epistemological camps—rationalism and empiricism—diminished and became less intense. Durkheim, to whom sociologists owe much of the foundational methodology, believed that the conflict between rationalism and empiricism could be resolved by incorporating reason while still acknowledging the importance of observable empirical data. The science of statistics also mitigates this conflict as it applies rational mathematical formulations on inductive empirical data to grant descriptions and generalizations. Based on these functions, statisticians differentiate between two types of statistics – Descriptive and Inferential/Inductive. In this chapter, we discuss about the former one extensively.

> ➢ **Take Note 1.0 Statistics in India**
>
> Statistics, as a modern discipline, came to India with European advent and evolved with response to imperial needs. It was a "perfect" science to some, to understand the demographic distributions of the orient. The first Synchronous Census survey was conducted under the British rule, in 1881 by W.C. Plowden (Census Commissioner). During the colonial rule, statistical methods became the strategic means of "artful" governance. The collected data was used to categorize the population by caste, region, and religion, with the aim of understanding, manipulating, and even appeasing the desired percentages. Interest in swift socio-economic growth and technological advancement introduced another dimension to the latter development of social statistics in independent India. P. C. Mahalanobis is regarded as the principal architect of modern statistical methods in the Indian subcontinent. However, it would be incorrect to view Europeans as the sole pioneers of statistics in India. Long before the British, the Mughals maintained statistical records of standardized measurements, land classifications, and seasonal crop yields. Abul Fazl was recognized as a statistician during Akbar's reign. Prior to that, the *Arthashastra* by Kautilya, written during the Mauryan period, detailed the system of data collection related to agriculture, population, and economic censuses in villages and towns.

14.5 Descriptive Statistics

Descriptive statistics refers to the branch of statistics that focuses on summarizing and presenting the main features of a dataset in a clear and concise manner. It involves methods for organizing,

displaying, and describing the characteristics of data. After conducting experiments and surveys and gathering data through appropriate sampling methods for a particular study, the researcher is often faced with a large volume of numerical data that may appear confusing and disorganized. This type of data is known as raw data. Here, fortunately we have descriptive statistics providing us with the tools for two major operations – organizing the numbers in meaningful order, and interpreting or summarizing the data by employing conceptual and theoretical means.

> ➢ **Take Note 2.0 Univariate and Bivariate Descriptive Analysis**
>
> Understanding the relation between variable and data is crucial to the discussion of different forms of descriptive analysis. Variables represent attributes or properties of the phenomena being studied, that can take different values, or to put it simply, vary. Data are the actual values that variables take on. We can define data as quantifiable expressions of abstractions of a certain phenomenon. Let us consider, 'Social control' - as a phenomenon it is never fully visible or completely understood. To study and comprehend it, we must abstract and focus on specific features, which may vary (by taking different values) in their expression. One such feature is the legal or punitive system, which can range from repressive to correctional approaches. In this context, the system of law itself becomes a variable, showing different characteristics depending on its nature and application. In research, studies frequently focus on a single variable or explore the relationships between two or more variables. Descriptive statistics, which help summarize and describe data, are commonly used in studies that look at how one variable changes. Univariate descriptive analysis specifically examines a single variable to understand its distribution and characteristics. Its key measures include central tendency, dispersion and distribution shape. On the other hand, examines the relationship between two variables to understand how they interact with each other. However, the measures of bivariate descriptive analysis further help in making predictions and generalizations, which associate them with inductive/inferential statistics. In the latter section of this chapter, we shall discuss about these measures in detail.

14.6 Organization

Organizing a dataset involves classifying, tabulating, frequency distribution and presenting the data using graphs or diagrams.

14.6.1 Classification is the process of organizing data into distinct groups based

on well-defined and unambiguous attributes of entities. These groups are referred to as mutually exclusive because the members of one group cannot belong to another. Nominal scale of measurement is used to categorize data in such groups. A limitation of this process is that each entity can be categorized into only one specific group (e.g., either delinquent or non-delinquent, either educated or uneducated), whereas social reality often includes entities that may fit into multiple categories.

14.6.2. Tabulation is the process of organizing and summarizing data into a table or grid format to facilitate analysis and interpretation. It involves arranging data into rows and columns, often with headings and categories, to make it easier to identify patterns, trends, and relationships. Tabulation helps in presenting data in a clear and structured way, allowing for straightforward comparison and analysis. In recent days statisticians use software such as Microsoft Excel, SPSS (Statistical Package for the Social Sciences), SAS (Statistical Analysis System), etc. to efficiently tabulate data and store the tables in computational devices. Below is an example of a hypothetical table created after organizing a dataset. It showcases the number of married working women and housewives in Community A and Community B.

	Community A	Community B
Total number of respondents	50	50
Number of married female respondents	25	19
Married working women (working outside house)	10	15
Housewives (working in indoor setting)	15	4

Table 1.0

In the table displayed above, the categories of "Married working women (working outside the home)" and "Housewives (working in an indoor setting)" can be considered mutually exclusive groups. For classification and quantification purposes, members of the former group cannot belong to the latter. However, in reality, married working women often

simultaneously fulfil the duties of a housewife.

14.6.3 ***Frequency Distribution*** is another method to tabulate a data set in understandable order by displaying the number of observations or occurrences (frequency) of each unique value or category. The term frequency distribution was first used by Karl Pearson in 1895. In statistical language, frequency is denoted either with an 'f_i' or an 'n_i' in short. It helps in understanding the distribution and pattern of data by showing how frequently each value or category appears. In construction of frequency distribution, those variables which can take all possible values are called continuous variables (e.g. age, weight, height etc.). In contrast, discrete variables are those which cannot take all possible values in a given range of data (e.g. number of students, marks, etc.). Discrete variables are not incorporated into grouped frequency distributions.

If the data set is relatively smaller or contains discrete variables, statisticians prefer to list all the values either in ascending or descending order, counting how many times they appear in the data-list and placing a tally (/) mark on each count. This particular type is regarded as an *ungrouped frequency distribution* which generally begets a *discrete series table*. In discrete series, the values are presented by a single number. With an example, this shall be clarified. Data set A (the raw data) expresses that in an under-graduate classroom, there are thirteen students with varying marks in exam. By sorting their total marks in ascending order, we get data set B: 185, 185, 185, 190, 190, 190, 190, 190, 190, 200, 201, 211, 221 (given in simple series). Based on data set B we can distribute the frequency of scores received and create the following table.

Marks	Tally	Frequency/ number of students (f_i)
185	///	3
190	~~////~~ /	6
200	/	1
201	/	1
211	/	1
221	/	1
	Table 2.0	

The preceding table is a discrete series table, product of ungrouped frequency distribution.

The variables here are all unique values and they do not belong to any groups of variables. In data set B, total observations of 185 is three, which has been symbolized with three tally (/) marking. It is important to know if the counts exceed four observations, for instance in the case of age 190, the tally marks are to be crossed with a strikethrough mark, symbolizing five (𝍸).

In case of a data containing wide range of values, researchers break data sets into equal interval to create groups. Each group consists of an upper limit and a lower limit. The interval between the upper and the lower limit is regarded as range (denoted by i). The range of each group are to be kept same in case of a *grouped frequency distribution*. In general, grouped frequency distribution produces *continuous series*. In continuous series, values are presented in different classes with equal intervals. There is no strict rule regarding the numbers of groups or the range. However, the range needs to be equal for all of the groups. An example of grouped frequency distribution is given below.

Ages (years) of 27 actors who act in mainstream Bollywood movies have been collected: 45, 55, 54, 27, 26, 34, 38, 38, 39, 22, 48, 50, 55, 43, 42, 26, 28, 29, 60, 40, 50, 56, 59, 24, 30, 45, 51.

Given that the lowest age is 22, and the highest age is 60, age 20 has been taken as the lower limit of the first group, and age 70 is taken as the upper limit of the last group. An interval of 10 is taken as range (20-30). Groups are formed using the *exclusive method,* meaning the upper limit of the preceding class is same as the lower limit of the next class (for instance, 20-30, 30-40). In exclusive method, the value equal to the upper limit of a particular class is excluded from that class (for example, the age of 30 shall be counted in the 30-40 class, and not in the 20-30 class, since in the latter one 30 is the upper limit). Adhering to these conditions the following table has been formed:

Groups/ class interval (i) (age in years)		Tally	Frequency (f_i)
Group 1	20-30	𝍸 //	7
Group 2	30-40	𝍸	5
Group 3	40-50	𝍸 /	6
Group 4	50-60	𝍸 ///	8
Group 5	60-70	/	1
Total number of actors / total frequency (N)			27

Table 2.1

In social statistics, researchers commonly work with 4 types of frequency measures:

(i) **Absolute frequency** is actual count of the number of occurrences of each distinct value in a given data set. The frequency measures (f_i) shown in Table 2.0 and Table 2.1 are examples of absolute frequency.

(ii) **Relative frequency** is the proportion of the total number of observations that a specific value represents. Expressed in a mathematical formula, the relative frequency (R) is calculated as (R) = f_i/N, where f_i represents the absolute frequency of the specific value and N denotes the total number of observations. Relative frequency is often measured in percentage [(R) = (f_i ×100)/N]. The information from Table 2.1 has been slightly restructured in the table below to illustrate the calculation of relative frequency more clearly.

	Groups / class interval (i) (age in years)	Tally	Absolute Frequency (f_i)	Relative Frequency (R) = f_i/N
Group 1	20-30	𝍸//	7	7/27 = 0.26 or 26%
Group 2	30-40	𝍸	5	0.19 or 19%
Group 3	40-50	𝍸/	6	0.22 or 22%
Group 4	50-60	𝍸///	8	0.30 or 30%
Group 5	60-70	/	1	0.037 or 3.7%
Total number of actors / total frequency (N)			27	

Table 2.2

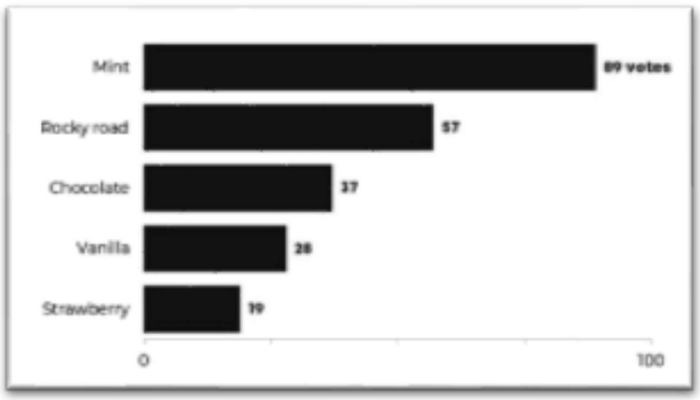
(iii) **Cumulative frequency** is the sum of the frequencies for all values up to a certain point in an ordered list. It shows the accumulation of the frequencies as we move through the data. The graphical representation of a specific cumulative frequency distribution is called an Ogive. In short, it is denoted by c.f. Cumulative frequency is calculated by adding the f_1 (frequency of the first-class interval) to the f_2 (frequency of the second-class interval) and then, by adding their sum to the f_3 (frequency of the third-class interval) and so on. Once again, table 2.1 has been slightly modified into the following table for a better understanding.

	Groups/ class interval (i) (age in years)	Absolute Frequency (f_i)	Mathematical formulation	Cumulative frequency (c.f.)
Group 1	20-30	7	f_1	7
Group 2	30-40	5	$f_1 + f_2$	7+5 = 12
Group 3	40-50	6	$f_1 + f_2 + f_3$	18
Group 4	50-60	8	$f_1 + f_2 + f_3 + f_4$	26
Group 5	60-70	1	$f_1 + f_2 + f_3 + f_4 + f_5$	27
Total frequency (N)		27		

Table 2.3

(iv) **Relative cumulative frequency** is the cumulative frequency expressed as a proportion of the total number of observations. It is calculated by dividing the cumulative frequency by the total number of observations. For example, if the cumulative frequency (c.f.) of a certain class is 18 and the total number of observations (N) is 27, the relative cumulative frequency would be [c.f./N] 18/27 = 0.67.

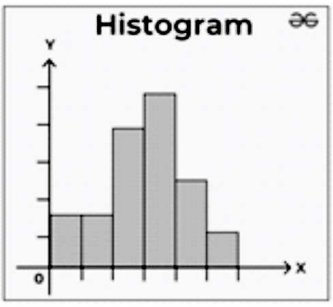
14.6.4 ***Graphical representation*** of data refers to the visual display of information using charts, graphs, maps, or other visual tools. Graphs can be excellent tools for not only organizing data in a comprehensible way but also for summarizing the data, as they condense large volumes of data into a visual format, making complex information more accessible and easier to understand. Graphical representations are pictorial tools which can depict different types of frequency distribution. Common types of graphical representation include Bar Chart, Histogram, Ogive, Frequency Polygon, Line graphs, Pie or Donut chart, Scatter Plots, etc.



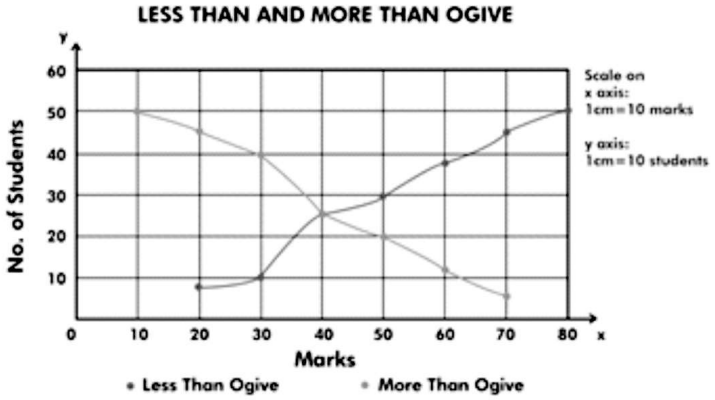
Bar Chart

Coloum Chart

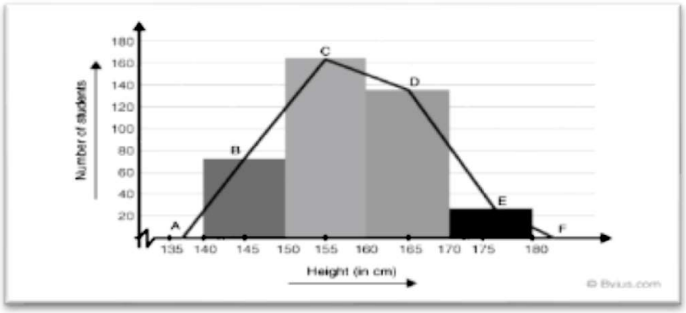
ii. **Histograms** are graphical representations of the distribution of numerical data. It uses adjacent bars to show the frequency of data points within specified intervals. Histograms are ideal for visualizing the distribution of a continuous variable and identifying patterns such as skewness, modality, and the presence of outliers.



iii. Ogives are graphical representations (line graphs) of cumulative frequency measures of a given data set. They are helpful for identifying the number of observations that fall below a specific value and for comparing various distributions. Ogives are frequently utilized in pedagogical researches to analyse assessment scores and other graded evaluations. There are two types of ogive curves – i. Less than Ogive and ii. More than Ogive, depending upon the order (either 'top to bottom' or 'bottom to top') in which we calculate the cumulative frequency.

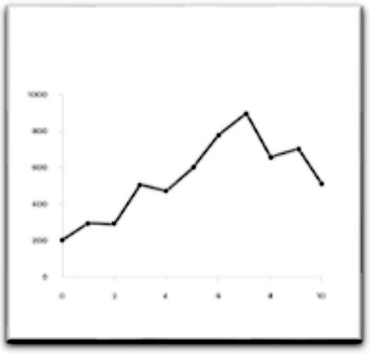


iv **Frequency polygons** are line graphs that show the frequencies of different values or ranges by connecting the midpoints of the tops of the bars in a histogram with straight lines. Frequency polygons are used to compare the distribution shapes of different datasets and to identify trends over intervals. They are particularly useful when comparing two or more distributions.



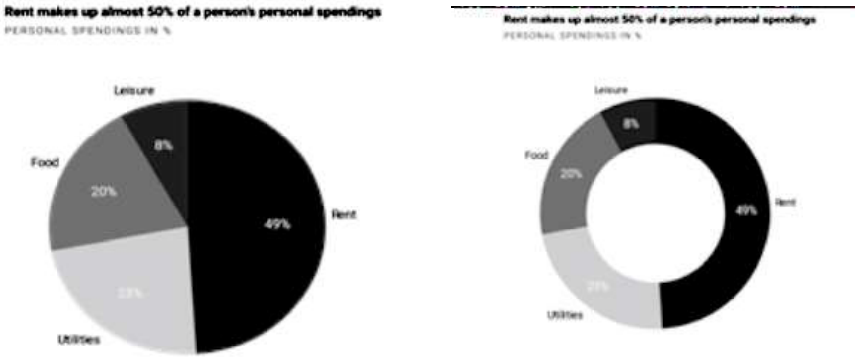
Frequency Polygon

v. **Line graphs** are plot-points connected by lines which show how a variable changes over time or in response to some other continuous factor. Line graphs are ideal for showing temporal issues and the degree of dynamism pertinent to certain social trends, such as market prices, temperature changes, population growth. Line graphs can be misleading if not properly scaled. They are less effective when displaying categorical data or non-continuous variables.



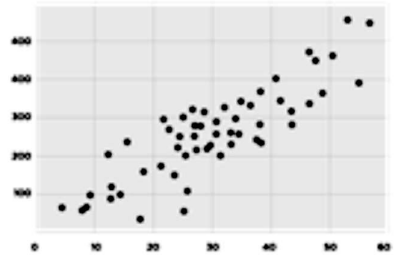
Line Graph

vi. **Pie charts and Donut charts** are circular graphs divided into slices, where each slice represents a proportion of the whole. The size of each slice corresponds to the relative magnitude of the category it represents. These charts are used to reflect the composition of a whole in categories. They are useful for showing proportions and percentages.



Pie chart Donut chart

vii. **A Scatter plot** is a type of data visualization that displays the relationship between two numerical variables. Each point on the scatter plot represents an observation in the dataset, with its position determined by the values of the two variables. Scatter plots can also help identify outliers or data points that deviate significantly from the general pattern of the data. Sometimes, scatter plots include a trend line (also known as a line of best fit) to highlight the overall direction or pattern of the data.



Scatter plot

➢ **Take Note 3.0 Scales of Measurement**

Before we jump into the discussion of data summarization, it will help to learn a little about the scales used in social research for both organization and summarization purposes. These scales constitute a system for categorizing and quantifying variables. They are closely connected with both aspects of descriptive social statistics, namely data organization and data summarization. The scales of statistical measurement include nominal scale, ordinal scale, interval scale and ratio scale. Nominal scale is the simplest form of measurement which is used for categorizing data and then labelling those categories. The numbers used in a nominal scale serve no mathematical purpose; they are purely symbolic. For instance, in a dataset tracking classroom attendance for a specific semester, '1' might represent the category 'present,' while '0' indicates 'absent.' We can only tally the occurrences of '1's and '0's, but they can't be used for operations like addition or subtraction. The ordinal scale is used to categorize data and arrange the categories in a ranked order, though the intervals between subsequent rankings are not necessarily equal. For instance, educational qualifications can be organized in this sequence: Primary, Secondary, Higher Secondary, Graduate, Master's Degree, Ph.D. Interval scale can categorize, rank and group data with equal intervals. However, there is no true zero point in interval scale. Having a true zero point in the context of measurement scales refers to the possibility of complete absence of the quantity being measured. It allows for meaningful comparisons and calculations, such as determining ratios. Temperature, Distance, etc. are measured in interval scale. The ratio scale is the only scale that includes all the features of the interval scale and also has a true zero point. a true zero means that zero represents "none" or "nothing" of the measured attribute, such as zero height meaning no height or zero income meaning no income. This allows for statements like "twice as much" or "half as much" to be accurate. Ratio scale allows for the calculation of ratios and it is statistically utilized to mathematically deal with the categories like age, height, income, etc.

For summarizing an organized set of data, statisticians generally use measures of central tendency, measures of dispersion, and graphical representations. In the following section we will be discussing about the first two -

14.7 Central Tendency

The term "central tendency" suggests that numerical values in a data set tend to cluster around a central value. Now, the central value differs according to the measure taken in the process of summarization. Measures of central tendency, such as the mean, median, and mode, summarize information by providing a single value that represents the typical value of a data set. These measures help to condense and simplify the data, making it easier to understand and interpret the overall distribution and central characteristics of the data. They provide a quick overview of where most data points are located and can be useful for comparing different data sets or identifying trends. We've touched on the different types of series (simple, discrete, and continuous) and discussed about frequency distribution in detail. In relation to central tendency, these organizational measures are crucial.

In statistics, a simple series, also known as a raw or ungrouped data series, consists of individual data points listed in no particular order. This type of data series presents values without any grouping or classification, making it a straightforward collection of observations. Each data point represents an individual observation, and the simple series does not involve any frequency distribution or categories. For example, 2, 10, 5, 8, 4 years are the ages of five siblings. A set of data presented like this can be understood as simple series.

Whereas a discrete series consists of data points that are distinct and separate, often taking specific, countable values. In this type of series, each value corresponds to a particular category or classification, and the data points are typically integers or whole numbers. A discrete series often includes a frequency distribution, where each unique value is associated with a frequency count indicating how often it occurs in the data set. Table 2.0 can be taken as an example of data arranged in discrete series.

A continuous series in statistics refers to data that can take any value within a given range or interval. Unlike discrete data, continuous data can include fractions and decimals, and the values are not restricted to whole numbers. Continuous data is often grouped into class intervals, and the frequency distribution shows how many observations fall within each interval. We can take table 2.1 as an example of this.

Now, we may start to discuss about mean, median and mode.

- ***Mean*** – Many of us may already be familiar with calculating an average. In statistics, the mean is essentially the same concept. The mean is determined by summing all the values in a dataset and then dividing this sum by the total number of values (N). Let's say that the mean income of thirty workers working in a mill is 15,000 Rs in a month. Not only does the amount of 15,000 represent the income group, but this also implies that if the total income were evenly distributed among all thirty workers, each would receive 15,000 Rs. This property of the mean makes it an important tool in sociological analysis, as it allows for the comparison between an individual's actual income and the average, thereby indicating the level of economic disparity within a group. The mean is typically symbolized by $\bar{X}$ (X-bar). To find the mean of a simple data series - $X_1, X_2, X_3 \ldots X_n$ – we sum all the values and then divide by the total number of values, N. To put it in mathematical formula, Mean $\left(\bar{X}\right) = \frac{(X_1 + X_2 + X_3 + \ldots X_n)}{N}$, or $\frac{(\sum X)}{N}$. For discrete and continuous series, however, we are in need of slight modification in the formula. In case of a discrete series data, we consider the values as mid points (x) and further multiply them with corresponding class frequencies (x_1 with f_1, x_2 with f_2, and so on). Further, summation of all the products ($\sum fx$) is divided by total frequency (N). On the other hand, the mid-points (x) in a continuous series have to be calculated by adding the limits (l_1 and l_2) of each class and dividing them by 2 - $\left[\frac{l1 + l2}{2}\right]$.

$$\text{Mean } \left(\bar{X}\right) = \frac{\sum fx}{N}$$

Advantage and Disadvantage of Mean:

Advantages

- Simplicity and Ease of Calculation.
- Unlike the median or mode, the mean takes into account all the values in the data set, providing a comprehensive measure that reflects the overall dataset.

- The mean is often used in various statistical analyses and calculations, such as variance and standard deviation, which measure the spread of data.
- The mean reflects every value in the data set, so it changes if any data point is altered, making it a sensitive measure.
- The mean is applicable in a wide range of fields, including economics, sociology, psychology, and natural sciences, as it provides a central value that can be used for comparison and analysis.

Disadvantages

- The mean can be significantly affected by outliers, which are extreme values that differ greatly from the rest of the data.
- In skewed distributions, where the data is not symmetrically distributed, the mean may not accurately reflect the central tendency of the data.
- When data points are highly variable or have different levels of dispersion, the mean may not provide a meaningful central value.

- ***Median*** – Median can be defined as the middle position or middle value in a given data set. The steps to calculate the median for a simple series involve some preliminary steps i.e. A. *arranging the data either in ascending or descending order (3, 3, 6, 9, 12, 12, 16),* B. *removing the repeated values (3, 6, 9, 12, 16),* C. *counting if the observations are in even or odd number.*

  If the number of observations(N) is odd (like 3, 6, 9, 12, 16), the value seated in $\left(\frac{N+1}{2}\right)^{th}$ position is taken as the median value. However, when the number of observations(N) is even (like 3, 6, 9, 12, 16, 18), then the values in $\left(\frac{N}{2}\right)^{th}$ and $\left(\frac{N}{2}+1\right)^{th}$ positions are located, added and divided by 2. Median (Mdn) = $\frac{\left(\frac{N}{2}\right)\text{th position} \left(\frac{N}{2}+1\right)\text{th position}}{2}$

  If the median income for a group of factory workers is 13,000 Rs, it means that 50% of the workers earn less than 13,000 Rs, while the other 50% earn more than 13,000 Rs. In case of continuous series median is calculated with the help of cumulative frequency.

Advantage and Disadvantage of Median

Advantages

- The median is not affected by extreme values or outliers, making it a more reliable measure in skewed distributions.
- The median can be used with ordinal data. Unlike the mean, the median does not require numerical data values to be meaningful, making it useful for ordered categories or ranks.

Disadvantages

- The median is determined solely by the middle position and does not consider the other data points.
- Small changes in the data that do not affect the middle value(s) will not change the median, which can be a disadvantage in detecting subtle trends or variations in the whole data.

- ***Mode*** - The mode is a statistical measure that represents the most frequently occurring value(s) in a dataset. It is one of the measures of central tendency, alongside the mean and median. The mode is particularly useful in datasets where certain values appear more frequently than others, providing insight into the most common data points. A dataset is unimodal if it has only one value that occurs most frequently. A dataset can be multimodal if there are two or more values that are equally frequent and occur more often than others. If the modal value among the incomes of a group of factory workers is 12,000 Rs, it means that most of the workers are earning 12,000 Rs per month.

Advantages and Disadvantages of Mode

Advantages

- The mode can be used with both numerical and categorical data. It is the only measure of central tendency that can be used with nominal data, such as categories or labels.
- The mode provides a direct understanding of the most frequent or popular item in a dataset, which can be particularly useful in social surveys and preference studies.

Disadvantages

- In datasets with multiple values having the same highest frequency, the mode may not be unique, which can make interpretation less clear.
- The mode only considers the most frequent values, ignoring the distribution and values of the rest of the data.

14.8 Measures of Dispersion

Dispersion is a statistical concept that describes the extent to which values in a dataset differ from one another and deviate from the central values. Measures of dispersion summarize the data by quantifying the spread or variability of the data points in a dataset. Understanding dispersion is crucial for assessing the risk and reliability of data. High variability might indicate greater uncertainty or risk. Some important measures of dispersion are –

- ***Range*** - It is the simplest measure of dispersion. The range is the difference between the maximum and minimum values in a dataset. The range gives a quick sense of the overall spread of the data, showing the extent between the smallest and largest values.
- ***Variance*** - Variance measures the average squared deviation of each data point from the mean. It gives a sense of how much the data points vary around the mean. A higher variance indicates greater variability and a more spread-out dataset. Variance = (σ^2) or $(\text{standard deviation})^2$.
- ***Standard Deviation (σ)*** – Standard deviation is defined as the root squared mean-deviation from mean. It is also the square root of variance that provides a measure of dispersion in the same units as the original data. It indicates the average distance of each data point from the mean, offering a more intuitive understanding of variability.

  Standard Deviation $(\sigma) = \sqrt{\left[\frac{\sum fx2}{n} - \left(\frac{\sum fx}{n}\right)^2\right]}$.

- ***Interquartile Range and Quartile Deviation*** - Quartiles are statistical measures that divide a dataset into four equal parts, providing insights into the distribution of data. The difference between the third quartile (Q_3) and the first quartile (Q_1) is known as Interquartile range (Q_r). Quartile deviation, also known as the semi-interquartile range, is a measure of statistical dispersion that describes the spread of the middle

50% of data in a dataset. It provides a summary of the variability around the median. Quartile Deviation $(Q_d) = \frac{Q3 - Q1}{2}$.

- ***Mean Deviation*** – The measure of mean deviation, also known as the mean absolute deviation, represents the average distance between each data point in a set and the mean of the set. Unlike standard deviation, which squares the differences before averaging them, mean deviation uses the absolute values of the differences, making it a straightforward measure of spread.

14.9 Conclusion

The discipline of statistics, through its historical development and methodological advancements into different disciplines, have always played a critical role in both theoretical and applied fields. Due to its wide range of measurement tools, statistics have been essential in quantitative research. We have discussed about descriptive statistic which is a fundamental branch of the field that provides essential techniques for organizing, summarizing and presenting data. By organizing data through classification and tabulation, descriptive statistics facilitate a clearer understanding of complex datasets. This branch of statistics plays a pivotal role in transforming raw data into interpretable summaries, thereby enhancing the ability to identify patterns, trends, and insights.

14.10 Summary

Without proper descriptive and interpretative tools, sociologists cannot effectively align their research objectives with the raw data they have collected. In this context, descriptive social statistics play an indispensable role in sociological research by providing the foundational tools necessary to distill complex datasets into meaningful insights. By offering a clear snapshot of social phenomena, descriptive statistics ensure that sociological inquiries are grounded in robust, interpretable data, thereby enhancing the rigor and relevance of research outcomes.

14.11 Questions

Have you completed reading the whole chapter? Try writing answers for these questions.

i. What is Statistics? Write a brief history of the development of statistics in Europe.

ii. Can statistics validate all our findings? Show reason behind your answer.

iii. What is Descriptive statistics? Mention its types.

iv. What are the two main operations of Descriptive social statistics? Discuss about the operational tools or measures.

v. Define central-tendency? What are the measures of central tendency? Discuss.

vi. What advantages and disadvantages do a statistician face while dealing with the measures of central tendency?

vii. Write a brief note on the measures of dispersion.

viii. What is frequency distribution? Write a brief note on the types of frequency distribution.

ix. What is Raw data? Why do we need to organize and tabulate a raw data?

x. Why is interpretation important in statistics?

xi. Discuss briefly about the role of statistics in sociology.

14.12 Reference

- Blalock, H. M., Jr. (1960). *Social statistics*. McGraw-Hill.
- Box, G. E. P., Hunter, W. G., & Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building* (1st ed.). Wiley.
- Dong, Y. (2023). *Descriptive Statistics and Its Application.* Highlights in Science, Engineering and Technology. Volume 47. AMMMP 2023.
- Ghosh, J. K., Maiti, P., Rao, T. J., & Sinha, B. K. (1999). *Evolution of Statistics in India.* International Statistical Review. Volume 67.
- **Image Sources:** Google, Datylon, byjus.com, Spotfire

Unit 15 □ Inferential Statistics in Social Research: Hypothesis Testing and Statistical Inference

Structure

15.1 Learning Objectives

a) To understand the importance of statistical tools in respect of social science research.

b) To learn about the qualitative data as well as quantitative data.

c) To have a quick recapitulation of how to analyze quantitative data with the help of descriptive statistical methods.

d) To understand mathematical problems and their solutions associated with social science research.

e) To learn Inferential Statistics measures i.e. estimation and test of hypothesis.

f) To know how to test of hypothesis with the help of test statistics tools i.e. t-test, Z-test, ANOVA Chi-Square**.**

g) To be able to know overall picture of descriptive statistics used in social science research.

15.2. Introduction

Statistics may be defined as numerical facts or science or methods or techniques from groups of people. It is a methodology of organizing of numerical facts and figures with help of some logical tools and techniques. It is a process of collecting; summarizing; analyzing and interpreting of numerical data from a given population. It is now applied in every corner of our human life. It is hugely used in our modern society basically in research work like agricultural research; medical science research; humanities research; social science research

etc. In the field of health science, it is used to study long-term effect of diseases. Statistical data are used to study customers' demand and choice; level of pricing; customer's behavior pattern and so on in respect of business sector. In social science research, statistical process and techniques are used to analysis people's knowledge and attitudes; level of living in the society and so on. Thus, statistics is most important part and useful tools in all discipline of subjects. However, statistical method is a method which deals with analysis of data in a systematic way. It is used to organize, present, and interpret the large or small amount of numerical data.

Statistical methods can be classified into two categories i.e. i) Descriptive Statistics and ii) Inferential Statistics. When statistical methods are used to describe and to summaries the data, it is known as descriptive statistics. Though we have already learnt about descriptive statistics, we shall have a recapitulation and learn it more elaborately here. On the other side, when statistical methods are used to draw inferences about a group of people (called population) based on samples; it is called as inferential statistics. When two methods are used on a particular study, it is called as applied statistics. Descriptive statistical tools are two types – a) univariate analysis which are i) frequency distribution; ii) measurement of central tendency and iii) measurement of dispersion or variability and b) bivariate analysis. Frequency distribution shows an idea about how the observations are distributed across different categories of a variable in a population. The measurement of central tendency and dispersion are dealt with only numerical variables (i.e. discrete and continuous) and not to code (i.e. ordinal and nominal) and the method is a) A. M.; b) Median and c) Mode. Measurement of dispersion is deal with S.D., Mean deviation; Quartile deviation etc. So these are part of the univariate data analysis. On the other hand, bivariate data analysis are i) Correlation and Regression; ii) Normal distribution curve etc. Correlation refers to show the relationship between two variables. Regression is the measure of the average relationship between two or more variables and it is used to predict the most likely values of one variable for specific values of the other variables. The normal distribution curve is based on the law of probability and it is described with two parameters i.e. Mean and S.D. This curve is bell-shaped.

Data are values which are observations or evidences and variables are any type of observation which is different values from person to person at different times. Data can be divided into qualitative and quantitative nature. The qualitative data cannot be measurable as numerical values but quantitative data have numerical values. From the view of quantitative data analysis with regard to social science research, bivariate data analysis as well as univariate data analysis methods have to be followed to get fruitful solution. Inferential Statistics is a branch of statistics to express as critical approach with regard to summarization predictions on a given sample of data. Apart from other side, inferential

statistics is to show inferences about the population of a given problem based on the characteristics of the sample. It is also to make predictions, draw conclusions, and generalize insights from a sample of population. There are two types of procedures under inferential statistics, namely a) estimation and b) hypothesis testing. The researcher will make an estimation that needs to be close to the actual or true population value. Hypothesis testing is a statistical tool used to test assumptions or supposition about a population parameter. The parametric test and non-parametric test are used for test of hypothesis and here two tests are revealed i.e. null hypothesis and alternative hypothesis at 5% or 1% level of significant associated with degree of freedom. If the calculated value is greater than table value, then null hypothesis is rejected and alternative hypothesis is accepted. That is there is a significant difference is revealed both the elements. So, inferential statistics is most important to test of hypothesis of a research process. In modern social science research, each and every researcher should have some knowledge about application and implication of statistical methods for the benefits of data analysis and interpretation. Social science research deals with always raw data. So, application of statistical method has to be followed to get research validation scientifically.

15.3. Importance of statistical tools and techniques in social science research

There are some following importance is revealed when the statistical tools and techniques are used in social science research.

a) **Systematic Analysis:** Statistical tools are more important to show the structural method in respect of systematic data analysis.

b) **Deeply Understanding:** It helps to identify patterns and trends of the dataset. So the researcher can understood deeply about social phenomena and behavior.

c) **Meaningful Conclusions:** With the help of application of statistical tools, researchers can draw meaningful conclusions from their study.

d) **Logical Reasoning:** It helps to enhance the logical phenomena about the social science data.

e) **Reliable Findings:** Analysis of statistical data is more helpful for standardization and the research findings are more reliable that enhancing our knowledge for future course of study.

f) Validity: Statistical data analysis provides evidence to make a decision in the organization. These data is more valid for fruitful solution.

g) Continuous Up-gradation: With the help of new statistical tools and techniques, researcher can enhance the quality and the standardization of their research work to great extent. It is an ongoing up-gradation process in the field of social science research.

15.4. Data

The data is from Latin word. It means anything that is given. It includes facts, figures, letters, symbols, words, charts and graphs that represent an idea, objects or condition. It is an identification of self-designed instrument. Data are values which are observations or evidences and variables are any type of observation which is different values from person to person at different times. In science, data can be represents as graphic nature. The data has some characteristics-

a) Data must be amenable to use.

b) Data should have clarification with its nature.

c) Data should have proper accuracy.

d) Data should have essence of values.

15.4.1 Types of statistical data:

i) Qualitative Data and Quantitative Data:

Data can be divided into qualitative and quantitative nature. The qualitative data cannot be measurable as numerical values, it expressed as descriptive. It is called as attributes. Example- honesty, motivation, confidence, preference, socio-economic structure etc.

Quantitative data can be measurable as numerical values. It is called as variables. Example- age, height, weight etc.

It is possible to transform qualitative data into quantitative data which represents numerical values. To convert qualitative data into quantitative data, you need to create categories and assign numerical values to each category. For example, a question is "use of mobile in respect of online education system under COVID-19 pandemic", if you conducted an interview and asked participants to rate their satisfaction on a scale of 1-5, you could assign numerical values to each rating, 1 could represent very dissatisfied, 2 could represent somewhat dissatisfied, 3 could represent neutral, 4 could represent

somewhat satisfied, and 5 could represent very satisfied. By assigning numerical **values** to each category, you can analyze the data statistically and draw conclusions based on numerical data.

ii) Ungrouped Data and Grouped Data:

After conducting an investigation, the researcher should have collected raw data for his/her research work and these data are not in a systematic manner. These are ungrouped data. Example- Marks obtained in English for class-VII for last year by 20 students are- 86, 75, 96, 55, 68, 47, 39, 57, 62 34, 80, 72, 62, 45, 92, 13, 71, 24, 94, and 63

When the data are analyzed and represented by systematic way i.e. use to statistical methods, these are grouped data. The frequency distribution table represents at least one variable from a sample and this representation is called as grouped data. Example-

Marks in English	No of Students
00-10	0
11-20	1
21-30	1
31-40	2
41-50	2
51-60	2
61-70	4
71-80	4
81-90	1
91-100	3
	Total 20

iii) Univariate Data, Bivariate Data and Multivariate Data:

Univariate data refers to a type of data in each observation deals with one variable and it measures with a single characteristic for each individual. Example- average height of 50 students of class-VII.

Bivariate data refers to a type of data where two different variables are there and it shows the relationship between two variables. Example- Production of paddy and wheat of West Bengal in 2023.

Multivariate data refers to a type of data where more than two variables are involved in each observation. These variables can represent different aspects, characteristics, or measurements related to the observed aspects. Example- A study of agricultural production for last five years in West Bengal.

15.5. Descriptive Statistical Tools

A) Univariate Analysis:

i) Frequency Distribution:

After conducting an investigation, we get raw data. These data will not have any meaning and these are arbitrary manner. So we have to classify and to arrange of these in a tabular form. When one character (variable) of the population at a time is involved in a distribution table, it is called as frequency distribution. It is clear that frequency distribution deals with one variable at a time. So it is called as one-way tabulation sheets.

The following data given from 54 families from Burdwan District

Table-1: Number of children per family

0	1	4	4	3	2	2	3	1	2	4	3	0	2	1	1	2	2
1	1	3	2	2	4	0	0	4	2	2	3	1	1	2	3	2	2
2	0	3	4	2	1	3	2	2	3	4	4	1	0	3	2	1	1

From Table-1, it can be seen that, the minimum and the maximum numbers of children per family are 0 and 4, respectively. Apart from these numbers, it is impossible, without further careful study, to extract any exact information from the data. By breaking down the data into the form of Table, however, certain features of the data become apparent. For instance, from Table-2, it can easily be seen that, most of the 54 families selected have two children. This information cannot easily be obtained from the raw data in Table-2

Table-2: Frequency Distribution of the data in Table-1

Number of children	Tally	Frequency
0	//// /	**6**
1	//// //// //	**12**
2	//// //// //// ///	**18**
3	//// ////	**10**
4	//// ///	**8**
		Total = 54

Grouped Frequency Distribution:

Table-3: Body masses (in kilograms) of 22 patients

60	45	72	55	42	65	54	68	74	50	78
70	58	48	67	64	68	52	60	58	75	83

It is shown from above table-3 that the minimum and the maximum body masses are 42 kg and 83 kg, respectively. A frequency distribution giving every body mass between 42 kg and 83 kg would be very long and would not be very informative. The problem is overcome by grouping the data into classes. If we choose the classes 41–49, 50–58, 59–67, 68–76 and 77–85 which are 9 class intervals, we obtain the frequency distribution given in Table-4.

Table-4: Grouped frequency distribution of the data in Table-3

Mass (kg)	Tally	Frequency
41 – 49	///	3
50 – 58	~~////~~ /	6
59 – 67	////	5
68 – 76	~~////~~ /	6
77 – 85	//	2
		Total = 22

Table-4, represents the frequency of each group or class; it is therefore called a grouped frequency table or a grouped frequency distribution. Using this grouped frequency distribution, it is easier to obtain information about the data than using the raw data in Table 2.3. For instance, it can be seen from Table 2.4, that 17 of the 22 patients have body masses between 50 kg and 76 kg (both inclusive). This information cannot easily be obtained from the raw data in Table-3.

It should be noted that, even though Table-4 is concise, some information is lost. For example, the grouped frequency distribution does not give us the exact body masses of the patients. Thus the individual body masses of the patients are lost in our effort to obtain an overall picture. However, Table-4 is far more comprehensible and its contents are easier to grasp than Table-3.

We now define the terms that are used in grouped frequency tables.

a) **Class Limit:** The two end-values of a class interval are called class limit

There are two class limit viz. i) Lower Class Limit and ii) Upper Class Limit

Lower Class Limit- 41, 50, 59, 68 and 77

Upper Class Limit- 49, 58, 67, 76 and 85

a) Class Boundaries:

The raw data in Table-3 were recorded to the nearest kilogram. Thus, a body mass of 49.5 kg would have been recorded as 50 kg, a body mass of 58.4 kg would have been recorded as 58 kg, while a body mass of 58.5 kg would have been recorded as 59 kg. It can therefore be seen that, the class interval 50–58, consists of measurements greater than or equal to 49.5 kg and less than 58.5 kg. The numbers 49.5 and 58.5 are called the lower and upper boundaries of the class interval 50–58. The class boundaries of the other class intervals are given in Table-5.

a) Class Mark or Mid Value:

Class Mark or mid value is equal to $\frac{(\text{Lower class limit} + \text{Upper class limit})}{2}$

Table-5: Body masses of 22 patients (to the nearest kg) Class interval

Class Interval	Frequency	Tally	Class Boundaries		Class Mark	Class Limit	
			Lower	Upper		Lower	Upper
41 – 49	3	///	40.5	49.5	45	41	49
50 – 58	6	~~////~~ /	49.5	58.5	54	50	58
59 – 67	5	~~////~~	58.5	67.5	63	59	67
68 – 76	6	~~////~~ /	67.5	76.5	72	68	76
77 – 85	2	//	76.5	85.5	81	77	85
	Total = 22						

ii) Graphical representation of quantitative data:

In the above discussion, we find out how to analysis and to interpret of numeric data from a frequency distribution table. Now we discuss the Graphical representation of quantitative data. Graphical representation of quantitative data is easier to understand and to clearly visualize than tabular form. Graphical representations are pictorial tools that can depict different types of frequency distribution. It shows visual indications of groups, trends and pattern in respect of data analysis. It also helps to show the comparison between the variables. A diagram is a visual representation of statistical data analysis. It includes different types of devises like circle, maps, bars, pictorial and cartograms etc.

Advantages:

a) It helps to easy understand, clear visualize and simple and attractive.

b) It shows clear comparison between the variables.

c) It gives to find out quick information from comparison data.

d) It extends the clear idea and stable effect of the researcher's mind.

Disadvantages:

a) It deals with approximate value of the data.

b) It cannot display detail and full value. So decision making cannot be successful in future.

c) Some diagrams have limited and complex nature. Thus it cannot be helpful for further analysis.

Types:

a) One-dimensional diagram i.e. line and bar

b) Two-dimensional diagram i.e. rectangle, square and circle

c) Three-dimensional diagram i.e. cube, sphere, cylinder

d) Pictorial and Cartograms

Bar Diagram:

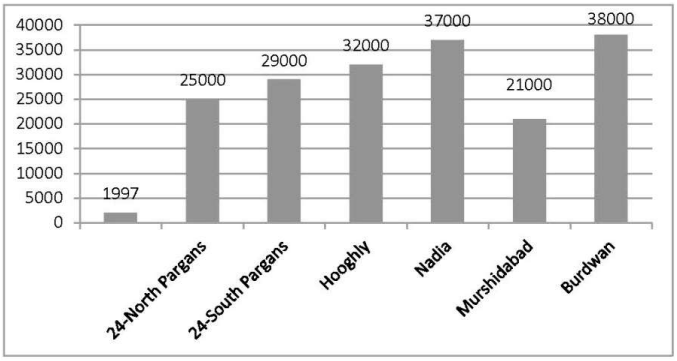
A bar chart is a diagram consisting of a series of horizontal or vertical bars of equal width. The bars represent various categories of the data. There are three types of bar charts, and these are simple bar charts, component bar charts and grouped bar charts.

Example:

i) Draw a bar diagram from given data sheet-

District	**Production (Paddy)**
24-North Pargans	25000
24-South Pargans	29000
Hooghly	32000
Nadia	37000
Murshidabad	21000
Burdwan	38000
Medinipur	17000

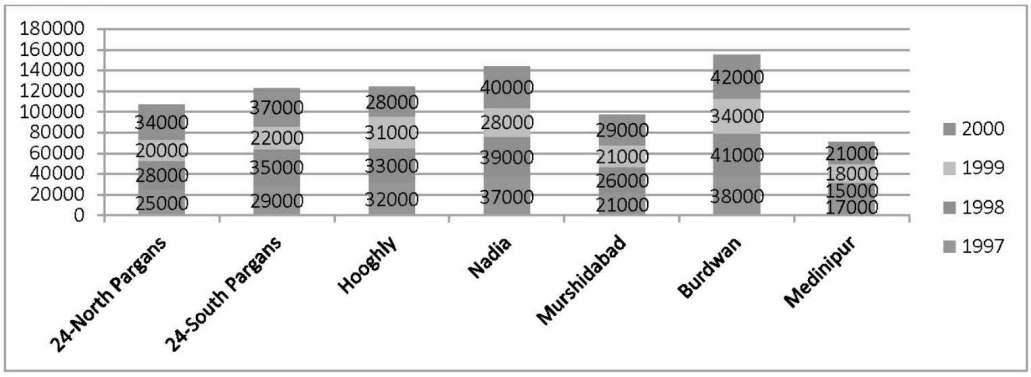
Solution:



ii) Draw a bar diagram from given data sheet-

District	**Production (Paddy)**			
	1997	**1998**	**1999**	**2000**
24-North Pargans	25000	28000	20000	34000
24-South Pargans	29000	35000	22000	37000
Hooghly	32000	33000	31000	28000
Nadia	37000	39000	28000	40000
Murshidabad	21000	26000	21000	29000
Burdwan	38000	41000	34000	42000
Medinipur	17000	15000	18000	21000

Solution:



Line Diagram:

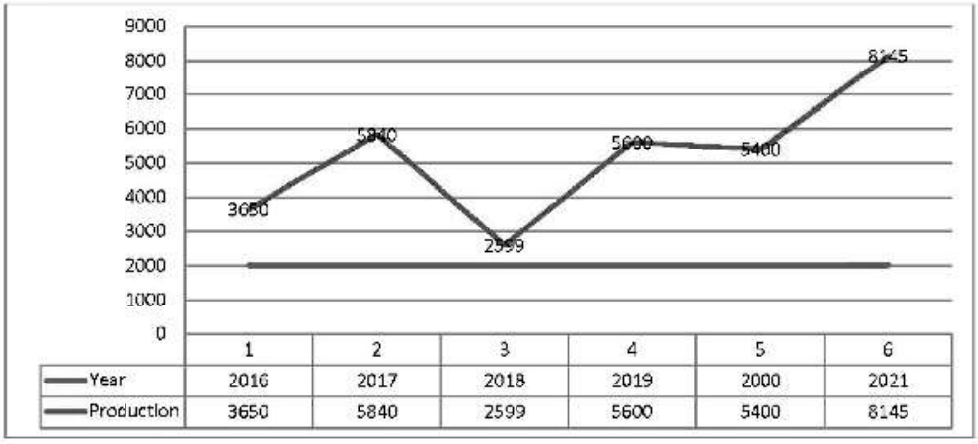
Line diagram is mostly used to represent the statistical data in terms of their trends. It constructed with the help of graph like X axis and corresponding to Y axis. It is mostly used in business sector. It is simple to represent time of occurrence of the collected data and to show the slap wise distribution points (decrease or increase) of two variables.

Example:

i) Draw a line diagram from given data sheet-

Year	Production
2016	3650
2017	5840
2018	2599
2019	5600
2000	5400
2021	8145

Solution:



Pie Chart:

A pie chart is a circular graph divided into sectors, each sector representing a different value or category. The angle of each sector of a pie chart is proportional to the value of the part of the data it represents.

Different the steps for constructing of a pie chart –

(1) Find the sum of the category values.

(2) Calculate the angle of the sector for each category, using the following result:

Angle of the sector for category $X = \frac{\text{value of category X}}{\text{sum of category value X}} \times 360^0$

(3) Construct a circle and mark the centre.

(4) Use a protractor to divide the circle into sectors, using the angles obtained in step 2.

(5) Label each sector clearly.

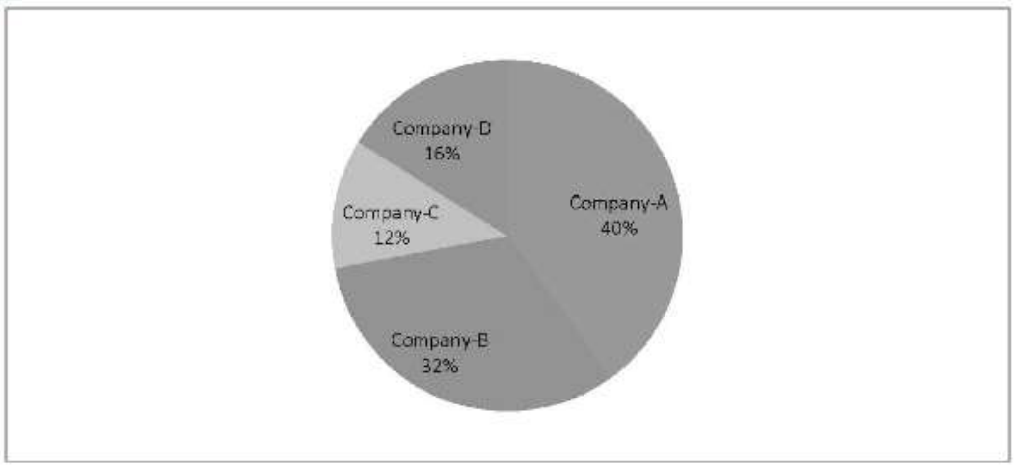
Example:

i) Draw a Pie diagram from given data sheet-

Name of Company	**Production**
Company-A	38400
Company-B	30720
Company-C	11520
Company-D	15360
Total	96000

Solution:

Name of Company	**Percentage**
Company-A	38400 / 96000 x100 % = 40 %
Company-B	30720 / 96000 x100 % = 32 %
Company-C	11520 / 96000 x100 % = 12 %
Company-D	15360 /96000 x100 % = 16 %



iii) Measurement of Central Tendency:

From frequency distribution, we get general idea about data. It is not indicates the actual view of data. Now we have to describe characteristics of population and simple we use averages method. In statistics, the term average is so many types and it is not indicated that only one is appropriate and specific for describing the character (typical vale) of the sample. Most of the frequency distribution shows typical vales of the variables are lie near central part of the sample and other values are cluster around these central value. This character of the data is called as central tendency. Mean, Median and Mode are the mostly used to analysis the central tendency of data.

a) Mean:

The Mean of a distribution is the arithmetic average. The Mean (Arithmetic Mean) is the ratio of the sum of all observations to the total number of observations. It is computed by dividing the sum of all the observations by the total number of observation. If *X1 + X2 + X3 + ...Xn* are the N observations, the formula for computing the Mean (X) is given

$$\textbf{Mean } (\mathrm{X}) = \frac{X1+X2+X3+...Xn}{N} \frac{\sum \mathrm{X}}{N}$$

$$\text{Mean, } \overline{\mathrm{X}} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

Example: Find the AM of 4, 20, 23, 18 and 20.

Sum of values = 4+ 20+23+18+20 = 85

Number of values = 5

Mean = $\frac{85}{5}$ = 17

Arithmetic Mean of Ungrouped Data

If $x_1, x_2, x_3, \ldots\ldots, x_n$ be the observations with the frequencies $f_1, f_2, f_3, \ldots\ldots, f_n$, then the arithmetic mean is given $\bar{x} = (x_1f_1+x_2f_2+\ldots\ldots+x_nf_n) / \Sigma f_i$ So, Mean (X) = $\frac{\Sigma fx}{\Sigma fi}$

Where, Σf_i is the summation of all the frequencies.

Example: Find the mean of given distribution :

x	10	20	30	40	50
f	3	2	4	5	1

Solution:

x	**f**	**fx**
10	3	10 x 3 = 30
20	2	20 x 2 = 40
30	4	30 x 4 = 120
40	5	40 x 5 = 200
50	1	50 x 1 = 50
	$\Sigma f = 15$	$\Sigma fx = 440$

A.M. = $\bar{x} = \Sigma fx / \Sigma f = 440/15 = 29.33$

Example: Marks obtained in English subject, Calculate A.M.

Marks (%)	60-64	65-69	70-74	75-79	80-84
Number of student	2	15	25	14	4

Solution:

Marks (%)	**Class mark (x)**	**Frequency (f)**	**fx**
60 – 64	62	2	124
65 – 69	67	15	1 005
70 – 74	72	25	1 800
75 – 79	77	14	1 078
80 – 84	82	4	328
		$\Sigma f = 60$	$\Sigma fx = 4335$

A.M. = $\bar{x}$ = Σfx / Σ f = 4335 / 60 = 72.25 %

ii) Median:

Median is the value of the central item which divides the series into two equal parts. If the number is arranged with ascending or descending order of magnitudes, the median is the middle term according to even or odd event.

Example: Find median.

110, 90, 40,125, 65,100, 50

Solution:

Arrange with ascending order-

40, 50, 65, 90, 100, 110, 125

Median is 4th value i.e. 90

Example: Find median.

54, 61, 91, 57, 72, 88, 33, 29, 70, 86

Solution:

Arrange with ascending order-

29, 33, 54, 57, 61, 70, 72, 86, 88, 91

Here we see that 10 numbers is there. So middle term may be two i.e. 5th and 6th = 61 and 70

Hence Median = (61+70) / 2 = 65.5

In case of data are grouped in frequency distribution, we have

$$\text{Median} = M_d = 1 + \frac{N/2 - F}{f} \times i$$

l= Exact lower limit of the class interval upon which the median lies.

i= Width of class interval in which the median falls.

f = Frequency within the class interval upon which the median lies.

F = Sum of all the frequencies below 1.

N/2 = One-half of the total number of observations.

Example: Find median value.

X	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	Total
f	1	3	4	7	9	11	8	4	2	1	50

Solution:

Median = M_d = $l + \frac{(N/2-F)}{F} \times i = 39.5 + \frac{(50/2)-9}{11} \times 5 = 44.05$

iii) Mode:

The mode is the value which occurs with the greatest frequency. So it is most common value. It can be seen that the mode of a distribution may not exist, and even if it exists, it may not be unique.

Example: Marks obtained by different students in English subject. Find mode.

Name	A	B	C	D	E	F
Marks Obtained	73	80	73	70	73	65

The maximum frequency observation is 73, The mode is 73.

Example:

(a) The mode of 1, 2, 2, 2, 3 is 2.

(b) The modes of 2, 3, 4, 4, 5, 5 are 4 and 5.

(c) The mode does not exist when every observation has the same frequency. For example,

The following sets of data have no modes:

(i) 3, 6, 8, 9; (ii) 4, 4, 4, 7, 7, 7, 9, 9, 9.

In case of data are grouped in frequency distribution, we have

$$\textbf{Mode} = l + \frac{f - f1}{2f - f1 - f2} \times i$$

Where,

l = Frequency of the modal class or maximum frequency

i= Width of class interval

f_1= Frequency of class just preceding the modal class

f_2= Frequency of class just succeeding the modal class

Example: Find mode.

Marks Obtained	00-09	10-19	20-29	30-39	40-49	50-59	60-69	70-79
No of Students	9	20	24	38	48	27	17	6

Solution:

$$\text{Mode} = 1 + \frac{f-f1}{2f-f1-f2} \times i = 39.5 + \frac{48-38}{2\times48-38-27} \times 10 = 42.73$$

Relationship of Mean, Median and Mode is

Mean-Mode = 3 (Mean- Median)

iv) Measurement of Dispersion or Variability:

Measurement of central tendency is to highlight the central value of a set of data. It also represents the typical value or quality of whole group of dataset by a single number. This typical value is so different with the comparison of achievements of two groups. Let, there three family distributions are -

a) 50, 50, 50; b) 40, 50, 60 and c) 20, 50, 80

Here, we see the all mean is 5 but the size of three families is so different i.e. the average is different as the sample. Thus we need some measures to calculate variation of this data. The procedures that measure the variation in the data are called measurement of dispersion or variance.

Different measure of dispersion or variance:

The common methods are a) Range; b) Quartile Deviation; c) Mean Deviation and d) Standard Deviation

a) The Range: The range is the difference between largest and smallest values of a variable.

b) Quartile Deviation: Quartile Deviation is denoted by Q. So the formula is

$$\text{Quartile Deviation } (Q_d) = \frac{Q3-Q1}{2}$$

Where Q_3 is upper quartile; Q_1 is lower quartile and Q_3-Q_1 inter-quartile range. So, Quartile Deviation (Q) is semi- quartile range of the dataset.

c) Mean Deviation: Mean deviation is to compute the average deviation from the average value of a dataset. Formula is

$$\textbf{Mean Deviation} = \frac{\sum|x-\bar{x}|}{N}$$

Where,

x = denotes each value in the data set

$\bar{x}$ = denotes the mean value of the data set

N = total number of data values

| | = represents absolute value, i.e. it ignores the sign

d) Standard Deviation: Standard deviation is the square root of the average of the squares of the deviation of each from the mean.

$$\text{Standard Deviation (ó)} = \sqrt{[\frac{\sum fx2}{n} - (\frac{\sum fx}{n})^2]}.$$

Example: Calculate standard deviation from the data-

49, 63, 46, 59, 65, 52, 60 and 54

Solution:

X	**X= (x- $\bar{x}$) = x-56**	Square of deviation= **(x- $\bar{x}$)²** i.e. x^2
49	-7	49
63	7	49
46	-10	100
59	3	9
65	9	81
52	-4	16
60	4	16
54	-2	4
∑x= 448		x^2 = 324

Mean = $\frac{\Sigma x}{N}$ = **448 / 8 = 56**

Standard Deviation (ó) = $\sqrt{\frac{\sum x^2}{n}}$ $\sqrt{\frac{324}{8}}$ = 6.36 (approx.)

Example: Calculate the mean, variance and standard deviation for the following data:

Class Interval	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	27	10	7	5	4	2

Solution:

Class Interval	**Frequency (f)**	**Mid Value (x_i)**	**fx_i**	**fx_i^2**
0 – 10	27	5	135	675
10 – 20	10	15	150	2250
20 – 30	7	25	175	4375
30 – 40	5	35	175	6125
40 – 50	4	45	180	8100
50 – 60	2	55	110	6050
			$\sum fx_i = 925$	$\sum fx_i^2 = 27575$

$N = \Sigma f = 55$

Mean = $(\Sigma fx_i)/N = 925/55 = 16.818$

Variance = $1/(N - 1)\ [\Sigma fx_i^2 - 1/N(\Sigma fx_i)^2]$

$= 1/(55 - 1)\ [27575 - (1/55)\ (925)^2]$

$= (1/54)\ [27575 - 15556.8182]$

$= 222.559$

Standard deviation = √variance = √222.559 = 14.918

Short-cut method for calculation of SD-

Example: Calculate Standard Deviation from the following data:

X	100-102	103-105	106-108	109-111	112-114	115-117	118-120	121-123	124-126	127-129
f	1	1	2	3	4	6	1	3	2	1

Solution:

X	**f**	**X=Mid Value**	$x' = \frac{X-A}{i}$	fx'^2	fx'^2
100-102	1	101	-5	-5	25
103-105	1	104	-4	-4	16
106108	2	107	-3	-6	18
109-111	3	110	-2	-6	12
112-114	4	113	-1	-4	4
115-117	6	116 (A)	0	0	0
118-120	1	119	1	1	1
121-123	3	122	2	6	12
124-126	2	125	3	6	18
127-129	1	128	4	4	16
	N=24			$\Sigma fx'^2 = -8$	$\sum fx'^2 = 122$

Where A = Assume Mean; i = Class Interval i.e. 3 class interval

So SD (O) $= \sqrt{\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2} \times i \sqrt{\frac{122}{4} - \left(\frac{-8}{24}\right)^2} \times 3$

$= 6.69$

B) Bivariate analysis:

Bivariate analysis is a statistical analysis where two variables are observed. One

variable is dependent and another one is independent. It shows the changes occurred between the two variables and to what extent. It shows the relationship of two variables (dependent and independent).

a) Correlation:

Correlation is to establishing the process of the relationship between two variables. It shows a general idea about whether two variables are related or not. In this case correlation is the most appropriate method for measurement of two variables.

b) Correlation Co-efficient:

Methods of correlation summarize the relationship between two variables in a single number called the correlation coefficient. The correlation coefficient is usually represented using the symbol r, and it ranges from -1 to +1. A correlation coefficient quite close to 0, but either positive or negative, implies little or no relationship between the two variables. A correlation coefficient is close to plus 1 means a positive relationship between the two variables, with increases in one of the variables being associated with increases in the other variable.

There are two methods to calculate the coefficient of correlation

a) Rank correlation method

Formula $p = 1 - \frac{6\sum d^2}{N(N^2 - 1)}$

b) Product moment method

Formula $r = \frac{\text{cov}(x,y)}{\sigma x \sigma y} = \frac{\sum\left[(x - \bar{x})\,(y - \bar{y})\right]}{\sigma x \sigma y}$

Example:

Calculate correlation of coefficient under rank correlation method

Student	A	B	C	D	E	F	G	H	I	J	K
Marks in Bengali	80	45	55	56	58	60	65	68	20	75	85
Marks in English	82	86	50	48	60	62	64	65	70	74	90

Solution:

Student	Marks in Bengali	Marks in English	Rank in Bengali	Rank in English	Difference in Rank (d)	d^2
A	80	82	2	3	1	1
B	45	86	11	2	9	81
C	55	80	10	10	0	0
D	56	48	9	11	2	4
E	58	60	8	9	1	1
F	60	62	7	8	1	1
G	65	64	6	7	1	1
H	68	65	5	6	1	1
I	20	70	4	5	1	1
J	75	74	3	4	1	1
K	85	90	1	1	0	0
N= 11						$\sum d^2 = 92$

So, $p = 1 - [6\Sigma d^2 / N(N^2-1)] = 1- [6x92 / 11(11^2-1)] = 0.58$ (approx.)

Example: Calculate correlation of coefficient under product moment method

Marks in Bengali	39	65	62	90	82	75	25	98	36	78
Marks in English	47	53	58	86	62	68	60	91	51	84

Solution:

Marks in Bengali (X)	Marks in English (Y)	$x = x - \bar{x}$	$y = y - \bar{y}$	x^2	y^2	xy
39	47	-26	-19	676	361	494
65	53	0	-13	0	169	0
62	58	-3	-8	9	64	24
90	86	25	20	625	400	500
82	62	17	-4	289	16	-68
75	68	10	2	100	4	20
25	60	-40	-6	1600	36	240
98	91	33	25	1089	625	825
36	51	-29	-15	841	225	435
78	84	13	18	169	324	234
$\sum X = 650$	$\sum Y = 660$			$\sum x = 5398$	$\sum y = 2224$	$\sum xy = 2704$

Where $\bar{x} = \Sigma x / n = 650 / 10 = 65$

$\bar{y} = \Sigma y / n = 660 / 10 = 66$

So, $r = \Sigma xy / (\sqrt{\Sigma x} \times \sqrt{\Sigma y}) = 2704 / (\sqrt{5398} \times \sqrt{2224}) = 0.78$ (approx.)

15.6. Inferential Statistics

The fundamental statistic is to represent the relationship of two or more data. Co-variance and correlation and probability distributions are the techniques for these measurements. Now it is need to move into Inferential Statistics. It is a branch of statistics to express as critical approach with regard to summarisation predictions on a given sample of data. The first step is data collection, and then it is organised and summarised in systematic way. After that, inferential statistics can be uses to analyse the data and draw conclusions and make inferences. Inferential statistics is to show inferences about the population of a given problem based on the characteristics of the sample. So that inferential

statistics is to make predictions, draw conclusions, and generalize insights from a sample of population. The following steps are used for inferential statistics.

1st Step: Selection of sample

2nd Step: Collection of data

3rd Step: Analysis of data

4th Step: Draw conclusions and make inferences about the population

15.6.1 Process for Inferential Statistics:

There are two types of procedures under inferential statistics, namely a) estimation and b) hypothesis testing.

i) Estimation: the researcher will make an estimation that needs to be close to the actual or true population value. It is two types- point estimation and interval estimation.

Point estimation: This is a type of estimation in which the value is a single point. For example the estimation for sample mean is made as 46.8, that is expected to e equal to the population mean. Point estimate comprises of sample mean and sample proportion. The population mean is 'μ' the sample mean will be $x^{\llcorner}$. In similar manner, if the population proportion is "p' then sample proportion will be 'p'.

Interval estimation: An interval estimate is an interval or two numbers within which the population parameter could lie. Thus, for population mean 'μ' the interval estimate will be a<x< b. The interval estimate is greater than 'a' but lesser than 'b'. For example, an interval estimate could be 45- 47 within which it is expected that the population mean will lie. As the researcher has an interval, he/ she is thus able to trust that the estimate is close to the population value with 95% or 99% level of confidence. Interval estimate comprises of confidence interval for mean and confidence interval for proportions.

ii) Hypothesis Testing:

Hypo+Thesis= Hypothesis; Hypo means tentative or subject to the verification. Thesis means statement about solution of a given problem. It is a tentative supposition or provisional guess to given a research problem. It is an assumption or supposition whose validity is to be tested. Thus hypothesis means a mere assumption or some supposition to be proved or disproved. **Hypothesis testing** is a statistical tools used to test assumptions or supposition about a population parameter. It is widely used for **evaluating the significance of results.**

For example:

- Testing whether a new feature significantly improves model performance.
- Comparing the accuracy of two models to determine if one is statistically better than the other.
- Validating assumptions about data distributions (e.g., normality) before applying certain algorithms.

It involves **formulating Hypotheses, using Statistical Tests and calculating p-Values:**

a) **Null Hypothesis:** A default assumption (e.g., "There is no difference between two models").

b) **Alternative Hypothesis**: The claim you want to test (e.g., "Model X performs better than Model Y").

- Statistical tests like z-tests, t-tests, or chi-square tests are used to determine whether there is enough evidence to reject hypothesis (These are the tests used to measure assumptions.). It is also called as research hypothesis.
- The **p-value is a number that shows how strong the evidence is against an assumption. A smaller** p-value (i.e., < 0.05) indicates strong evidence to reject the assumption (null hypothesis), while a large p-value means there isn't enough proof to reject it.

Difference between Null hypothesis and Alternate hypothesis:

Null Hypothesis	Alternative Hypothesis
There is no relationship between the two variables in null hypothesis.	In the alternative hypothesis, there is some relationship between the two variables i.e. They are dependent upon each other.
It is rejected or disprove by the researcher.	Researchers may try to accept or approve the null hypothesis
If it is accepted by researchers, then they have to make changes in their opinions and statements.	If it is accepted by researchers, then they not have to make changes in their opinions and statements.
Here no effect can be observed i.e. it does not affect output.	Here effect can be observed i.e. it affects the output.
Here the testing process is implicit and indirect.	Here the testing process is explicit and direct.
This hypothesis is denoted by H_0.	This hypothesis is denoted by H_a or H_1.
Here the p-value is smaller than the significance level.	Here the p-value is greater than the significance level.

The equation are-

The null hypothesis is H_0: E(X) = μ

The alternative hypothesis is H_1: E(X) ‘“ μ

Not equal to is either E(X) < μ or : E(X) > μ

Type-I Error: Rejecting null hypothesis (H_0) when it should have been accepted.

Type-II Error: Accepting (H_1) when it should have been accepted.

In hypothesis testing, null hypothesis is tested (E(X) = μ) against one of the following alternative hypothesis testing:

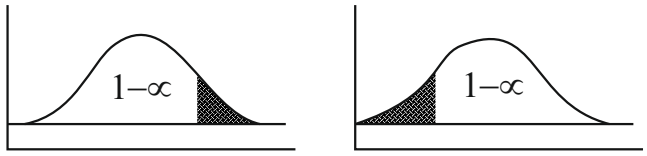
Condition-1: E(X) < μ

Condition-2: E(X) > μ

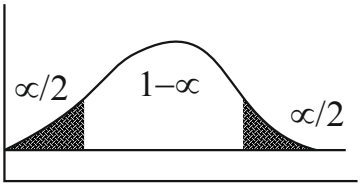
Under such conditions, critical values are given by-

$$\text{Condition - 1 : } \rho\left\{\left|\frac{x-\mu}{\sigma/\sqrt{n}}\right| \geq Z\alpha\right\} = 1-\propto$$

$$\text{Condition - 2 : } \rho\left\{\left|\frac{x-\mu}{\sigma/\sqrt{n}}\right| \geq Z\lambda\right\} = 1-\propto$$



a) E (x) > μ b) E (x) < μ



c) E (x) ≠ μ

In critical region is selected such that it lies either in left side or right side of the curve, then it is known as one-side test. Otherwise it is as two-side test.

15.6.2 Parametric Test and Non-Parametric Tests:

Statistics tests may be classified into two categories i.e. a) parametric test and b) non-parametric test, which are commonly used to test of hypothesis. Statistic test represents numerical value calculated from sample data in accordance with a hypothesis test. It measures how far the observed data deviates from the null hypothesis. Thus, it shows whether hypothesis is rejected or accepted with the help of a critical value or used to calculate a p-value.

a) **Parametric test:** It is a statistical test to make assumptions about the parameters of the population. It is assume that the data are to be known and specific nature, i.e. the normal distribution. The variance within the groups is homogeneous nature. Examples: T-test, Z-test, ANOVA (Analysis of Variance)

b) **Non-parametric test:** It is free test and not to required any assumption under data sheets. When data is not normally disturbed and the sample size is very small, then this test is suitable. Basically it is used to test for ordinal data sheet. Examples: Chi-Square Test, Mann-Whitney U test, Kruskal-Wallis test.

Difference between parametric test and non-parametric test:

Parametric test	Non-parametric test
It is based on normal distribution	It is based on skewed distribution
The data are known as quantitative data	The data are known as qualitative data
Here scale of measurement is METRIC scale (i.e. interval / ratio scale)	Here nominal or ordinal scale are used
It compare mean and standard deviation	It compared as percentage(%), proportion, or ratio et
It is more powerful tools	It is comparatively less powerful tools

p-Value:

The p-value is known as the probability. It is calculated under the assumption that the null hypothesis is true, of observing a value from the test statistic at least as extreme as the one that was actually observed.

Thus, P-value is the chance that the presence of difference is concluded when actually there is none.

a)When the p value is between 0.05 and 0.01 the result is usually taken as significant.

b)When p value is less than 0.01, result is taken as highly significant.

c)When p value is less than 0.001 and 0.005, result is taken as very highly significant.

The common tests and their test statistics are given below-

Hypothesis Test	**Test Statistics**
T-test	T statistics
Z-test	Z statistics
ANOVA (Analysis of Variance)	F- statistics
Chi-square tests	Chi-square statistics

Degree of freedom (d.f.):

The number of degree of freedom in a distribution is the number of observations or values that are indecent of each other and cannot be deducted from each other.

15.6.2.1 Parametric test- t-test, Z-test, ANOVA:

A) t-test: t-test is used to make compare the mean of two given samples. It is developed by Prof. W.S Gossett in 1908, who publishes statistical papers under the pen name of "student." Thus the test is known as Student's "t" test. It is assumes a normal distribution of the sample. When we don't know the population parameters (mean and standard deviation), we use t-test.

Use of t-test:

1. Data utilized is Quantitative.
2. The population is normal or approximately normal
3. The variance of population is not known
4. Sample variances are homogeneous groups under the study.
5. The size n of the sample is small (i.e. $n < 30$)

Types of t-tests:

a) One sample t-test: Tests the mean of a single group against a known value.

b) Independent two-sample t-test: Compares mean for two groups.

c) Paired sample t-test**:** Compares means from the same group at different times.

i) t-test: Test for a specified mean:

t-test: Test for a specified mean

Two tailed test hypothesis: H_0 : ì=ì$_0$ and H_1 : $\mu \neq \mu_0$

Test Statistic is $$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N-1}}}$$

Where, $\bar{x}$ = sample mean

μ = population mean, = standard error

Inference:

- Table Value: (n-1) is the degrees of freedom for the distribution. This value is used to find the table value for the given level of significance.
- If the calculated value is less than the table value at 5% or 1% significance value, Null Hypothesis is accepted.
- If the calculated value is more than the table value at 5% or 1% significance value, Null Hypothesis is rejected. So, the alternative hypothesis will be accepted in that case.
- Note: One tailed test is performed the same way the difference is with the representation in hypothesis and the table values for significance level.

Example:

A machine is designed to produce insulting washers for electrical devices of average thickness of 0.025cm. A random sample of 10 washers was found to have an average thickness of 0.24cm. with S.D. of 0.002 cm. Test the significance of deviation. Vale of t for 9 degree of freedom at 5% level is 2.262.

Solution:

We have μ = 0.025. n = 10, $\bar{x}$ = 0.024, S = 0.002 where S is the sample S.D. Here sample size is small (i.e n< 30) and population S.D. is not known.

Thus $$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{N-1}}} = \frac{0.0024 - 0.25}{\frac{0.002}{\sqrt{10-1}}} = \frac{-0.001}{\frac{0.002}{\sqrt{2}}} = \frac{-0.003}{0.002} = -1.5$$

The tabulated value of t for 9 d.f. (i.e. n-1 = 10-1 = 9) at 5% level = 2.262.

So, $= |t| < 2.262$

Since calculated value of $|t|$ < the tabulated value, so we accept the Null Hypothesis at 5% level of significance. The conclusion is that the deviation is not significant.

ii) t-test: Test of significance for the difference between two population means when the population SD's are unknown:

Two tailed test hypothesis: $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu$

If the two population S.D. σ_1 and σ_2 are equal, then $\sigma_1 = \sigma_2 = \sigma$ (say) and unbiased estimate σ^2 of is given by

$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$, Where S_1^2, S_2^2 are the two sample variance.

Test Statistic is $$t = \frac{\bar{x}_1 - \bar{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n2}}}$$

Inference :

- Table Value: (n-1) is the degrees of freedom for the distribution. This value is used to find the table value for the given level of significance.
- If the calculated value is less than the table value at 5% or 1% significance value, Null Hypothesis is accepted.
- If the calculated value is more than the table value at 5% or 1% significance value, Null Hypothesis is rejected. So, the alternative hypothesis will be accepted in that case.
- Note: One tailed test is performed the same way the difference is with the representation in hypothesis and the table values for significance level.

Example:

A group of 5 patients treated with medicine A weight 42, 39, 48, 60 and 41 kgs.; a second group of 7 patients from the same Hospital treated with medicine B weight 38, 42, 56, 64, 68, 69, and 62 kgs. Do you agree with the claim that medicine 'B' increases the weight significantly? (the value of t at 5% level of significance for 10 degree of freedom is 2.2281.)

Solution:

Null Hypothesis $H_0: \mu_1 = \mu_2$

i.e. there is no significant difference between A and B as regards their effect no increase in weight.

So the Alternative Hypothesis is $H_1: \mu_1 < \mu_2$ i.e. B increases the weight significantly

Since the sample size is small and the population S.D. are unknown, we use t test.

Let H_0 be true, we have

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}, \text{ and } t = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n2}}}$$

Medicine A			**Medicine B**		
x_1	$x_1 - \bar{x}_1$	$(x_1 - \bar{x}_1)^2$	X_2	$X_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$
42	-4	16	38	-19	361
39	-7	49	42	-15	225
48	2	4	56	-1	1
60	14	196	64	7	49
41	-5	25	68	11	121
---	--	--	69	12	144
---	--	--	62	5	25
$\sum x_1 = 230$	--	$\sum(x_1 - \bar{x}_1)^2 = 290$	$\sum x_2 = 399$	--	$\sum(x_2 - \bar{x}_2)^2 = 926$

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = \frac{230}{5} = 46; \bar{x}_2 = \frac{\sum x_2}{n_2} = \frac{399}{57} = 57$$

$$n_1 S_1^2 = \sum (x_1 - \bar{x}_1)^2 = 290; n_2 S_2^2 = \sum (x_2 - \bar{x}_2)^2 = 926$$

$$s = \frac{290 + 926}{5 + 7 - 2} = \frac{1216}{10} = 121.6$$

$$s = \sqrt{1221.6} = 11.03$$

$$t = \frac{46 - 57}{11.03\sqrt{\frac{1}{5} + \frac{1}{7}}} = \frac{-11}{511.03 \times 0.585} = \frac{-11}{6.45} = -1.71$$

So, $|t| = 1.71 < 2.2281$

Since the calculated value of $|t|$ < tabulated value of t at 5% level with 10 d.f., we accept the Null Hypothesis H_0. The conclusion is that the medicine A and medicine B do not differ significantly in increasing the weights of the patients.

iii) t-test: Paired Observations:

Condition of independence may not hold good for all samples. When the samples are related to each other t-test can be performed for small samples on converting the given sample into single data type by taking the difference. So, the formula will be:

Test Statistic is $t = \frac{\bar{d}}{S^2/\sqrt{n}}$ where $\bar{d} = \frac{\sum d}{n}$ and $S^2 = \frac{\sum d^2 - (\bar{d})^2 \times n}{n-1}$

Note: If S^2 value is taken from the sample then the denominator will be $S^2 / \sqrt{n-1}$

• Inference is similar to the previous t-tests discussed.

Example:

A certain stimulus to each of the 12 patients resulted in the following increases of blood pressure:

5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4 and 6.

Can it be concluded that the stimulus will, in general, be accompanied by an increase in blood pressure? (The value of 't' at 1% level of significance for 11 d.f. is 2.2)

Solution:

Null Hypothesis $H_0 : \mu_1 = \mu_2$

i.e. there is no significant difference in blood pressure before and after administrating the stimulus. Alternative Hypothesis: $H_1 : \mu_1 \neq \mu_2$

Assuming that H_0 is true, the test statistics $t = \frac{\bar{d}}{S^2/\sqrt{n}}$ where $\bar{d} = \frac{\sum d}{n}$ and $S^2 = \frac{\sum d^2 - (\bar{d})^2 \times n}{n-1}$

So, d = 5, 2, 8, -1, 3, 0, -2, 1, 5, 0, 4 and 6

and d^2 = 25, 4, 64, 1, 9, 0, 4, 1,25, 0, 16, 36

$$\bar{d} = \frac{\sum d}{n} = \frac{31}{21} = 2.58 \text{ and } \Sigma d^2 = 185$$

$$S^2 = \frac{\sum d^2 - (d^2) \times n}{n-1} = \frac{185 - (2.58^2) \times 12}{12-1} = \frac{185 - 79.87}{11} = 9.55$$

$$S = \sqrt{9.55} = 3.09$$

$$t = \frac{\bar{d}}{S^2/\sqrt{n}} = \frac{2.58}{3.09/\sqrt{12}} = \frac{2.58 \times \sqrt{12}}{3.09} = 2.09$$

d.f. = n-1 = 12-1 = 11

Since the calculated value of t > the tabulated value i.e. 2.09 > 2.2 with d.f. at 1% level, we have reject the Null Hypothesis and conclude that the stimulus will, in general, accompanied by an increases in blood pressure.

Example:

IQ test was administrated to 5 persons before and after they were trained. The results are given below:

Test whether there is any change in IQ after the training programme. (The value of 't' at

Name of Candidate	A	B	C	D	E
IQ before training	110	120	123	132	125
IQ after training	120	118	125	136	121

1% level of significance for 4 d.f. is 4.6)

Solution:

Null Hypothesis $H_0 : \mu_1 = \mu_2$

i.e. there is no significant effect of the training.

Alternative Hypothesis: $H_1 : \mu_1 < \mu_2$

Assuming that H_0 is true, the test statistics is $t = \frac{\bar{d}}{s/\sqrt{n}}$, where $\bar{d} = \frac{\sum d}{n}$, d = y-x

and $S^2 = \frac{\sum(d-\bar{d})^2}{n-1} = \frac{\sum d^2 - (\bar{d})^2 \times n}{n-1}$

Name of Candidate	IQ before training (x)	IQ after training (y)	d = y-x	d^2
A	110	120	10	
B	120	118	-2	
C	123	125	2	
D	132	136	4	
E	125	121	-4	
Total	----	----	$\sum d = 10$	$\sum d^2 =$

$\bar{d} = \frac{\sum d}{n} = \frac{10}{5} = 2$

and $S2 = \frac{\sum d^2 - (\bar{d})^2 \times n}{n-1} = \frac{140 - (2)^2 \times 5}{5-1} = \frac{140-20}{4} = 30$ $S = \sqrt{30}$

$t = \frac{\bar{d}}{s/\sqrt{n}} = \frac{2}{\sqrt{30}/\sqrt{5}} = \frac{2}{\sqrt{6}} = \frac{2}{2.45} = 0.82$

d.f = n-1 = 5-1 = 4 Thus t = 0.82 < 4.6 at 1% level with 4 d.f.

Since the calculated value of t < the tabulated value with 4 d.f. at 10% level, we accept H_0 at 10% level and conclude that there is no significant change in IQ after training programme.

B) Z-test:

The sample statistics is the sample mean of the data, $\bar{x}$. If the population SD is known, the distribution of the sample means is $\bar{x} \sim N\ \mu 0, \left(\frac{\sigma}{\sqrt{n}}\right)$, where ì$_0$ is the population mean assumed in the null hypothesis. The test statistics is a Z-score: $Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$. The p-value is the tail area under the $\bar{x}$ normal curve beyond $\bar{x}$ in the direction of the alternative hypothesis, which is the same as the tail area under Z ~ N (0,1) beyond Z.

Z test for large samples (n>30):

Z-test is a statistical test where normal distribution is applied and is basically used for dealing with problems relating to large samples when the frequency is (greater than or equal to) > 30. It is used when population standard deviation is known**.**

Assumptions:

1. Population is normally distributed
2. The sample is drawn at random

Let $x_1, x_2, \ldots\ldots x_n$ be a random sample size of n from a normal population with mean μ and variance σ^2 . Let $\bar{x}$ be the sample mean of sample of size "n"

Null Hypothesis: Population mean (μ) is equal to a specified value μ_0

$$H_0: \mu = \mu_0$$

Under H_0, the test statistic is $Z = \frac{|\bar{x} - \mu|}{\frac{s}{\sqrt{n}}}$

If the calculated value of *Z* < table value of Z at 5% level of significance, H_0 is accepted and hence we conclude that there is no significant difference between the population mean and the one specified in H_0 as ì$_i$.

C) ANOVA:

ANOVA (Analysis of Variance), is a statistical method **used to test the differences between means of** two or more groups. It is developed by Ronald Fisher in the early 20th century. It is used to determine whether or not the means of three or more groups are

equal. The ANOVA test is used to look for heterogeneity within groups as well as variability across groupings. The f -test returns the ANOVA test statistic.

Formula of ANOVA:

Source of Variation	Sum of Squares	Degree of Freedom	Mean Squares	F Value
Between Groups	$SSB = \Sigma nj(\bar{X}_j - \bar{X})^2$	$df_1 = k - 1$	$MSB = SSB / (k - 1)$	$f = MSB / MSE$ or, $F = MST/MSE$
Error	$SSE = \Sigma nj(\bar{X} - \bar{X}_j)^2$	$df_2 = N - k$	$MSE = SSE / (N - k)$	
Total	$SST = SSB + SSE$	$df_3 = N - 1$		

- F = ANOVA Coefficient
- MSB = Mean of the total of squares between groupings
- MSW = Mean total of squares within groupings
- MSE = Mean sum of squares due to error
- SST = total Sum of squares
- p = Total number of populations
- n = The total number of samples in a population
- SSW = Sum of squares within the groups
- SSB = Sum of squares between the groups
- SSE = Sum of squares due to error
- k = number of samples
- s = Standard deviation of the samples
- N = Total number of observations
- df = N-1 i.e. (Number of items in all the sample)-1

Types of ANOVA:

a) One-Way ANOVA:

This test is used to see if there is a variation in the mean values of three or more groups. Such a test is used where the data set has only one independent variable. If the test statistic exceeds the critical value, the null hypothesis is rejected, and the averages of

at least two different groups are significant statistically.

b) Two-Way ANOVA:

Two independent variables are used in the two-way ANOVA. As a result, it can be viewed as an extension of a one-way ANOVA in which only one variable influences the dependent variable. A two-way ANOVA test is used to determine the main effect of each independent variable and whether there is an interaction effect. Each factor is examined independently to determine the main effect, as in a one-way ANOVA. Furthermore, all components are analyzed at the same time to test the interaction impact

15. 6. 2 .2 Non-parametric test-Chi-Square Test:

A chi-square test is used when we want to see if there is a relationship between two categorical variables from a single population. It is a test of independence and is used to estimate the likelihood that some factor other than chance accounts for the observed relationship. Since the null hypothesis states that there is no relationship between the variables under study, the Chi-square test merely evaluates the probability that the observed relationship results from chance.

The formula for Chi-square (χ^2) is $\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$

f_0 = frequency of the occurrence of observed or experimentally determined facts

f_e = expected frequency of occurrence

The number of degrees of freedom df = (r - 1) (c - I), in which r is the number or rows and c is the number of columns in which the data are tabulated.

Example:

The following data of 500 subject who have been categorized into three groups, A, B, C on the basis of age and their preference or four non-veg. items, mutton, chicken, fish and egg.

Group	Mutton	Chicken	Fish	Egg	Total
A	40	50	35	45	170
B	35	42	44	39	160
C	38	41	36	55	170
Total	113	133	115	139	500

1) Calculation of expected frequencies (f_e)

Age	Red	Blue	Yellow	Green
A (21-30)	(113x170)/500 = 38.42	(133x170)/500 = 45.22	(115x170)/500 = 39.10	(139x170)/500 = 47.26
B (31-40)	(113x160)/500 = 36.16	(133x160)/500 = 42.56	(115x160)/500 = 36.80	(139x160)/500 = 44.48
C (41-50)	(113x170)/500 = 38.42	(133x170)/500 = 45.22	(115x170)/500 = 39.10	(139x170)/500 = 47.26

2) Chi-square value $\chi^2 =$" $\frac{(f_o - f_e)^2}{f_e}$

$\chi^2 = (40-38.42)^2/38.42 + (50-45.22)^2 / 45.22 + (35-39.10)^2 / 39.10 + (45-47.26)^2 / 47.26 + (35-36.16)^2 / 36.16 + (42-42.56)^2 / 42.56 + (44-36.80)^2 / 36.80 + (39-44.48)^2 / 44.48 + (38-38.42)^2 / 38.42 + (41-45.22)^2 / 45.22 + (36-39.10)^2 / 39.10 + (55-47.26)^2 / 47.26$

$\chi^2 = 5.182$

df = (r-1) x (c-1) = (3-1) x (4-1) = 8

The χ^2 critical values for 8 d.f. as given in are 15.507 and 20.090 respectively for .05 and .01 levels of significance and, the obtained value, 5.182, is less than the table value even at .05 level.

So that, there is no relationship between the age and the colour preference and thus the hypothesis that group and non-veg. items preference are essentially independent may be accepted at 0.05 level of significance.

15.7. Limitations of statistical measures

i) Statistical tools may be misused or misinterpreted. It may be highlighted to erroneous conclusions or wrong results.

ii) Assumptions on the basis of statistical methods may not be true in our real life, it is shown incorrect predictions or fully biased.

iii) Statistical analyses may be incorrect for missing of data. Thus, it required more careful for data processing.

iv) It should have more specialized knowledge and techniques are required to analysis the data. Statistical training programme of the researcher is required

v) Data privacy and confidentiality are carefully addressed when data are processing with the help of statistical tools.

15.8. Conclusion

From the above discussion we may point out the different types of statistical tools which are used for quantitative data analysis in respect of social science research. Statistical tools and techniques are playing pivotal role in respect of social science research that enable researcher to analyze dataset, test hypotheses and highlight valid conclusions about social phenomena and behavior. With the help of statistical tools, researchers can have familiar to solve the complex research questions and highlight valid research findings. Social science research deals with social phenomena and behavior in respect of qualitative and quantitative nature. Qualitative data have no numerical value, but it may be converted into quantitative nature, it is needed to create categories and to assign numerical values to each category. On the other side, quantitative data have numerical vale. For quantitative data analysis, we may use descriptive statistical tools for the benefit of more information can be given in future research. Thus, social science researcher should have basic knowledge about statistical tools and techniques for data analysis, interpretation and for fruitful solution.

15.9. Summary

Statistics is most important part and useful tools in all discipline of subjects. It is broadly used in our modern society basically in research work like agricultural research; medical science research; humanities research; social science research etc. Each and every researcher should have some knowledge about statistical methods and tools for analysis and interprete of research data as well as their research work. Statistics is the science of two aspects namely (a) the collection, organization, analysis and interpreting of numerical facts and (b) the drawing of inferences about a group of people. Statistical methods can be classified into two categories i.e. i) Descriptive Statistics and ii) Inferential Statistics. When two methods are used on a particular study, is called as applied statistics. Descriptive statistical tools are two types – a) univariate analysis which are i) frequency distribution; ii) measurement of central tendency and iii) measurement of dispersion or

variability and b) bivariate analysis which are i) Correlation and Regression; ii) Normal distribution curve etc. Data are values which are observations or evidences and variables are any type of observation which is different values from person to person at different times. Data can be divided into qualitative and quantitative nature. Qualitative data have no numerical value, but it may be converted into quantitative nature, it is needed to create categories and assign numerical values to each category. On the other side, quantitative data have numerical vale. In the view of quantitative data analysis with regard to social science research, the most important common operation is applied statistics tools and techniques which are used for collecting, summarizing, analyzing and organizing of data on a particular research problem. On the other hand, inferential statistics is most important to test of hypothesis of a research process. If we choose the hypothesis for a particular research work then it should be tested with the help of test statistics measures. There are so many test statistics tools are available i.e. t-test, f-test, z-test ANOVA, Chi-Square test etc. If the calculated value is greater than table value, then null hypothesis is rejected and alternative hypothesis or research hypothesis is accepted. That is there is a significant difference is revealed both the elements. This chapter discussed the analysis of quantitative data with the help of descriptive statistical measures like frequency distribution; measurement of central tendency; correlation etc. It also discussed the inferential statistics measures for test of hypothesis. Some mathematical problems and their solutions have been discussed here for better understanding of the learners.

10. Questions

1. Define statistics? Why statistical tools are used in the field of social science research?
2. What is data? Discusses the different types of data.
3. What do you mean by descriptive statistical tools? Mention the difference between descriptive statistics and inferential statistics.
4. How to analysis univariate data analysis?
5. What is bivariate data analysis?
6. Why do you use the techniques of measurement of central tendency? Write short notes on Mean, Median and Mode.
7. How to prepare a) line diagram and b) pie diagram from a data sheet.
8. What is standard deviation? How to calculate? What is correlation? How to calculate **correlation coefficient under '**product moment method'?

9. What is Inferential Statistics? What are the **steps used of inferential statistics?**

10. Define Hypothesis. Discuss the difference between Null Hypothesis and Alternative Hypothesis.

11. What is parametric test and non-parametric test? Discuss the difference between parametric test and non-parametric test.

12. Write short note on p-value; t-test; z-test; ANOVA and Chi-square test.

13. The age of 30 workers of a certain factory are given below-

42	32	21	44	30	33	51	30	44	26
54	25	36	32	45	39	49	26	41	56
35	28	34	28	38	43	37	41	31	43

i)Arrange the data in a frequency distribution in 10 class-intervals;

ii) Find the class-boundaries; mid-point and cumulative frequencies 'from below' and 'from above'.

14. Draw line and bar diagram of the following data**:**

Year	2011	2012	2013	2014	2015	2016	2017	2018
Production (in million tons)	120	170	380	450	680	240	170	980

15. Draw a pie-chart from the following data:

Revenue of the Central Govt.

a) Customs——————————160 Cores
b) Excise——-——————————500 Cores
c) Income Tax——————————330 Cores
d) Corporation Tax————110 Cores
e) Other sources——————100 Cores

Total———— 1,200 Cores

16. Find Mean; Median and Mode of the following numbers:

7, 4, 3, 5, 6, 3, 3, 2, 4, 3, 4, 3, 3, 4, 4, 3, 2, 2, 4, 3, 5, 4, 3, 4, 3, 4, 3, 1, 2, 3.

Ans: Mean= 3.47; Median= 3and Mode= 3

17. The A.M. calculated from the following frequency distribution is known to be 28.8. find the missing frequency:

Marks	0-10	10-20	20-30	30-40	40-50	50-60
Frequency	4	6	20	?	7	3

Ans: 10

18. Calculate the Mean, Median and Mode of the following frequency distribution:

Marks	10-14	15-19	20-24	25-29	30-34	35-39
Frequency	5	8	10	15	8	4

Ans: Mean= 24.5; Median= 28.83 and Mode= 37.49

19. Find the A.M. and S.D. from the following distribution table:

Daily Wages (in Rs.)	141-150	151-160	161-170	171-180	181-190	191-200	201-210
No. of workers	5	8	15	25	20	17	10

Ans: A.M. = Rs. 179.30 and S. D. = Rs. 15.99

20. Calculate the coefficient of rank correlation from the following data:

Student	A	B	C	D	E	F	G	H	I	J
Test-X	48	33	40	9	16	16	65	24	16	57
Test-Y	13	13	24	6	15	4	20	9	9	19

Ans: r = 0.73

21. A certain diet newly introduced to each of the 12 pigs resulted in the following increase in body weight:

6, 3, 8, -2, 3, 0, -1, 1, 6, 0, 5 and 4.

Can you conclude that the diet is effective in increasing the weight of pigs? (The value of 't' at 5% level of significance for 11 d.f. is 2.20)

Ans: Yes, effective.

22. Two sets of the students selected at random from a college were taken; one was given memory tests as they were and the other set was given a memory tests after two weeks' training and score were given below:

Set A	10	8	7	9	8	10	9	6	7	8
Set B	12	8	8	10	8	11	9	8	9	9

Do you think there is any significant effect due to training? (The value of 't' at 5% level of significance for 18 d.f. is 2.10)

Ans: No significant effect.

23. Apply χ^2 (Chi-Square) test to find out if the following figures provide evidence of the effectiveness on inoculation.

	Attacked	Non Attacked	Total
Inoculated	20	300	320
Non- inoculated	80	600	680
Total	100	900	1000

(Given, the value of χ^2 at 5% level with 1 d.f. is 3.84)

Ans: The inoculation is quite effective; since χ^2 is 7.353

15.11. References

1. Begum, Kousar Jaha and Azeez Ahmed. (2015).The Importance of Statistical Tools in Research Work. International Journal of Scientific and Innovative Mathematical Research (IJSIMR), v. 3; No.12, p. 50-58
2. Gupta, Rupa Z. (2024). Statistical Tools for Social Sciences: A Comprehensive Review and Analysis. Journal of Emerging Technologies and Innovative Research

(JETIR), v. 11; No. 4, p. 34-38

3. [online] https://egyankosh.ac.in/bitstream/123456789/53663/1/Block-4.pdf

4. [online]https://www.jsscacs.edu.in/sites/default/files/Files/STATISTICAL_METHODS_FOR_THE_SOCIAL_SCIENCES_Academia.pdf

5. [online] https://byjus.com/maths/

6. Panda, J.R. and others. (2019). Self-Learning Material: M.A. Sociology-Methodology of Social Research. Directorate of Distance Education, University of Jammu: Jammu

7. Rajaretnam, T. (2016). Statiscs for Social Sciences, Sage: New Delhi

8. Saha, Suranjan. (1988). Business Statistics. New Central Book Agency: Kolkata

9. Saha, Suranjan. (1991). Business Statistics: for I.C.W.A.I intermediate course. 4th ed. New Central Book Agency: Kolkata

10. Sain, C. R. and others. (2017). MLIS Study Material: Research Methodology, Netaji Subjas Open University: Kolkata

11. Singh, Yogesh Kumar. (2006). Fundamental of Research Methodology and Statistics. New age international Publisher: New Delhi, p. 286-310

12. [online] file:///C:/Users/user/Downloads/Block-1.pdf

13. [online] https://gurunanakcollege.edu.in/files/commerce-management/STATISTICS-UNIT-3.pdf

14. [online] https://mgcub.ac.in/pdf/material/20200519073328de6dca404c.pdf

15. [online] https://egyankosh.ac.in/bitstream/123456789/33587/1/Unit-13.pdf

16. [online] https://www.geeksforgeeks.org/what-is-inferential-statistics/

[online] https://www.geeksforgeeks.org/anova-formula