Research Methodology

M.L.I.Sc.

DLSM22 – SECOND SEMESTER

Course material prepared by

Dr. P. Balasubramanian,
M.A.,M.B.A.,M.C.A.,M.L.I.Sc.,M.phil., PGDPR,Ph.D.
University Librarian & Head
Department of Library and Information Science,
Manonmaniam Sundaranar University,
Tirunelveli-627 012, Tamil Nadu, India.

MANONMANIAM SUNDARANAR UNIVERSITY
TIRUNELVELI - 627012

Research Methodology

Preamble: This course makes awareness regarding Research Concept - Research Methods - Research Techniques - Statistics and its Applications and Metric Studies and Style Manuals

UNIT – I: Introduction to Research
Research Concept, Need and Purpose ,Research Problem and Research Design Literature Review Hypothesis: Definition, Types, Sources and Functions

UNIT –II: Types of Research Methods
Historical, Survey and Experimental,Case Study, Scientific Research and Statistical Research, etc.

UNIT – III: Research Techniques
Research Techniques and Tools: Questionnaire, Interview, Observation, Schedule and Check-list, etc. Library Records and Reports

UNIT – IV: Statistics and its Applications
Descriptive Statistics – Measures of Central Tendency: & Dispersion, Correlations and linear regression, Chi-Square test, t-test, z-test, f-test. Presentation of Data: Tabular, Graphic, Bar Diagram and Pie Chart, etc. Report Writing Statistical Packages – MS Excel, SPSS, and Web-based Statistical Analysis Tools, etc.

UNIT – V: Metric Studies and Style Manuals
Scientometrics, Infometrics and Webometrics Manual Structure, Style, Contents- ISI, MLA, APA, CHICAGO

References:
Alvesson M and Skoldberg K.(2015) Reflexive methodology: new vistas in qualitative research. Ed. 2 Rev. Sage Publication, London.

Busha C H. (2016). Research methods in librarianship. Academic Press, New York.

Goode W J and Hatt P K (2016). Methods in social research. McGraw-Hill, New York.

Greenfield T. (2017) Research methods: guidance for postgraduates.. Hodder Arnold, London.

Krishan Kumar. (2015). Research methods in library and information science. Rev. Ed. Har-Anand Publications, New Delhi.

Powell R R and Silipigni C L.(2014). Basic research methods for librarians. Ed. 4. Libraries Unlimited, Westport.

UNIT – I: Introduction to Research

Research: Concept, Need and Purpose, Research Problem and Research Design Literature Review Hypothesis: Definition, Types, Sources and Functions

In the modern complex world, every society faces serious social, economic & political problems. These problems need systematic, intelligent and Practical solutions. Problem-solving is a technical process. It requires the accumulation of new knowledge. Research provides the means for accumulating knowledge & wisdom. In other words, research is a systematic effort to gather analysis & interpretation of problems confronted by humanity. It is a thinking process and scientific method of studying a problem and finding a solution. It is an in-depth analysis based on reflective thinking.

Definitions

Research, in common parlance, refers to a search for knowledge. One can also define research as a scientific and systematic search for pertinent information on a specific topic. Research is an academic activity and should be used in a technical sense. a) -William Emory defines research as any organised enquiry designed and carried out to provide information for solving a problem" b) The new Oxford English Dictionary defines research as "the scientific investigation into and study of material, sources etc. to establish facts and the reach new conclusions" c) Redman and Mory defines, research as "a systematised effort to gain new knowledge" d) "A careful investigation or inquiry specially through search for new facts in any branch of knowledge" Advanced Leaner's Dictionary

Characteristics of Research

The above definitions reveal the following characteristics of Research

1. Research is a systematic and critical investigation into a phenomenon.
2. It is not a mere compilation of facts
3. It adopts the scientific method
4. It is objective & Logical
5. It is based on empirical evidence.
6. Research is directed towards finding answers to questions
7. It emphasises the generalisation of theories and principles.

Objectives of Research

The research objectives can be grouped under the following heads

a. To gain familiarity with a phenomenon or to achieve new insights into it.

b. To determine the frequency with which something occurs or is associated with something else.

c. To test a hypothesis or a causal relationship between variables.

Motivations in Research

What makes people undertake research? The answer is as follows.

1. Desire to get a research degree along with its benefits.
2. Desire to face the challenge in solving the unsolved problem.
3. Desire to get intellectual joy from doing creative work.
4. Desire to be of service to society.
5 Desire to get respectability

Importance of Research

All progress is born of enquiry. Doubt is often better than overconfidence, for it leads to enquiry & enquiry leads to the investigation". Research has an important role in guiding social plans. Knowledge of the society & the cultural behaviour of the people requires proper planning for their good development. Because knowledge & cultural behaviour of human beings are interdependent. A piece of reliable knowledge is needed for planning & this is possible only through research. Knowledge is a kind of power with which one can face the implication of a particular Phenomenon. Research provides the basis for all govt. Policies in our economic system. Research helps us in making predictions. E.g. Chornobyl Nuclear, the nuclear plant disaster, and the Bhopal gas disaster Research are equally important in seeking answers to various social problems. In addition, the significance of the research can be understood with the following points

1. To the students who are to write a Ph.D., it is careerism.
2. To Professionals in research methodology, research means a source of live hood.
3. To Philosophers & thinkers, research may mean the outlet for new ideas and insights.
4. To literary man, research means the development of new styles & creative work
5. To the intellectuals, research means the generalisation of new theories.

Research Method & Research Methodology

It is necessary to explain the differences between research methods & research methodology. Research methods may be understood as all those methods & techniques that are used for conducting research. Research methods, thus, refer to the methods the researcher uses in performing the research operations. In other words, the researcher uses all those methods during his research problem, termed research methods. Research methodology is a way to solve the research problem systematically. It may be understood as a science of studying how research is done scientifically. Abraham Kaplan defines research methodology in this way. Research methodology is "the description, explanation & Justification of various research methods". Research Methodology. The scope of the Research Methodology is wider than that of research methods. "Thus, when we talk of research methodology, we not only talk of the research methods but also consider the logic behind the methods we use in the context of our particular method or technique & why we are not using others. So that research results are capable of being evaluated either by the researcher himself or by others" Why a research study has been undertaken, how the research problem has been defined in what way & why the hypothesis has been formulated, what data have been adopted etc. ate usually answered when we talk of Research Methodology

Scientific Method - Research is a scientific endeavour

"The Scientific Method is a systematic step-by-step procedure following the logical reasoning process". (Clover Vernon. T) The scientific method is a means for gaining knowledge of the universe. It is an objective, logical & Systematic Method of analysis of a phenomenon that permits the accumulation of reliable knowledge. It is a systematised form of analysis. An intellectual attitude characterises it. The Scientific Method is based on certain articles of faith. They are

a) Reliance on evidence Truth is established based on an evident conclusion. Itis admitted only when it is based on evidence. The answer to a question is not decided by imagination or guess

b) Commitment to Objectivity: Objectivity is the hallmark of the Scientific method. Objectivity is the willingness & ability to accept the truth with our bias.

c) Ethical Neutrality: Science does not pass normal judgment on facts. It does not say that they are good or bad. Science never imposes anything. Science aims at nothing but making true & adequate statements about the object.

d) Verifiability: The conclusions arrived at by a scientist should be verifiable. He must make known to others how he reached his conclusion. Such verifications help in further research.

e) Logical Reasoning Process: The Scientific Method involves the logical process of reasoning. The reasoning process is used for drawing inferences from the finding of a study or for conclusions.

The Research Process is the Paradigm of the research project. In a research project, there are various scientific activities. The research process is a system of interrelated activities. Usually, research begins with the selection of a problem. The various stages in the research process are explained in the above diagram. Research is a cyclical process. If the Data do not support the hypothesis, research is repeated. In his book, "Research Methodology: Methods & Techniques", C.R. Kothari presents a brief overview of the research process. He has given the following order concerning the Research Process.

1. Formulation of the Research problem.
2. Extensive Literature survey
3. Developing the hypothesis
4. Preparing the research design
5. Determining sample design
6. Collection of Data
7. Execution of the Project
8. Analysis of Data
9. Hypothesis testing
10. . Generalisation & Interpretation
11. . Preparation of the report.

Qualities of Good Research Worker

To a great extent, the success of any Research depends on the qualities of the researcher. The qualities are twofold. 1. General Qualities 2. Specific Qualities 1. **General Qualities a) Scientific attitude** The 1st essential quality of a successful research work is that he must possess a scientific frame of mind. Human beings have certain prejudices, but this should not guide a researcher. He must develop a spirit of science in his mind.

b) Imagination & insight

The researcher must possess a high degree of imagination. He should be able to go deeper & deeper into the area of social phenomena &visualise the intangible aspects of society.

c) Perseverance

The work of scientific research requires unlimited patience. He should not get easily discouraged. He may often face serious difficulties. But he must develop the courage to face the difficulties & work patiently.

d) Quick Grasping Power

1. The researcher should possess the power to grasp the significance of things quickly

e) Clarity of thinking

A good Researcher should have a clear idea about the terminology he will use.

2. Specific Qualities

a) Knowledge of the Subject: The researcher should be an expert in the study of the subject he will research. Hence he should read all texts on the matter & form clear-cut ideas about the subject under study.

b) Knowledge of the Research Technique: The Research worker should also know the techniques he applies to solve the problem.

c) Personal Taste: A Personal taste in the study will inspire him & keep his morale high in times of difficulty.

d) Unbiased Attitude: The Researcher should have no preconceptions about the subject under study. He should maintain an open mind.

Factors Which Hinder Research

1. Tradition in the community is a powerful retarding influence
2. Lack of time, energy & resources.
3. Research is considered to be the business of a few armchair academicians.

1. **Problems faced by Researchers in India**

Researchers in India, particularly those engaged in research in Social Science, face the following problems

1. The lack of scientific training in the methodology of research
2. There is insufficient interaction between the University Research Department and business establishments & govt. Departments
3. In fear of misuse, govt or unwilling to supply basic documents.
4. There does not exist code of conduct for researchers.
5. Another difficulty is insufficient secretarial assistance
6. Library management is not satisfactory in many places
7. Lack of time and money

Social Science Research

Sciences are broadly divided into physical sciences & social sciences. Social sciences include various disciplines dealing with human nature, human life, human behaviour, social groups & Social institutions. Examples include Anthropology, Commerce, Economics, Geography, History, Law, Political science, Psychology, Sociology etc. All these branches are separate but interdependent. Social Science Research is a systematic method of exploring, analysing & conceptualising human life to extend, correct or verify the knowledge of human behaviour & social life. Social Research "Seeks to find explanations to unexplained social phenomena, to clarify the doubtful & correct the misconceived facts social life"- Pauline. V. Young.

Objectives of Social Science Research

a) Social Science Research aims to discover new facts or verify and test old facts
b) It tries to understand human behaviour & its interaction with the environment & Social institutions
c) It tries to find causal connections between human activities and natural laws governing them
d) Social Science Research aims to develop new tools and techniques in social science.

Functions of Social Science Research

- → Discovery of facts & their interpretations - Research answers questions of what, where, when & how of man, social life and institutions. There are half-truths, pseudo-truths, and superstitions. The Discovery of facts enlights us
- → Diagnosis of problems- Developing countries face many problems, such as poverty, unemployment, Social tensions, low productivity etc. Social Science Research helps to discover solutions to these problems.
- → The systematisation of knowledge- The facts discovered through research are part & parcel of the body of knowledge.
- → Prediction- Social Science Research aims to predict social events.
- → Planning - Panning is needed for socio-economic development & Social Science Research provides sufficient data for planning.
- → Social Welfare- Social Science Research unfolds& identifies the causes of social evils & problems.

Good Research

In "criteria of good research", James Harold Fox says that scientific research should satisfy the following conditions.

- ✓ The purpose of research should be clearly defined, and common concepts should be used
- ✓ The Research procedure should be described in sufficient detail to permit another researcher to repeat the research for further advancement, keeping the continuity of what has been attained
- ✓ The research design should be carefully planned to yield as objective as possible results.
- ✓ The Research report should be complete, frank, and without flaws
- ✓ Data collection and analysis should be adequate, and the reliability should be checked carefully.
- ✓ Conclusions should be justified by data 7. Researchers must be experienced people with a goal reputation.
- ✓ Good research is systematic
- ✓ Good research is logical
- ✓ Good research is empirical
- ✓ Good research is replicable (Denny N Ballenger and Burnet. A Green Berg)

Types of Research:

Research is classified into different forms based on intent & methods. The following are the different types of research.

1. Descriptive Vs Analytical Descriptive research includes Surveys or fact-finding enquiries of different kinds. The major purpose of descriptive research is to describe the state of affairs. The main characteristic of this method is that the researcher has no control over the variables; He can only report what has happened or what is happening. In Analytical research, on the other hand, the researcher has to use facts or information already available & analyse this to make a critical evaluation of the material.

2. Applied Vs Fundamental Research can be applied (or action) or fundamental (or pure) research. Applied Research aims to solve an immediate problem facing a society or an organisation. In contrast, Fundamental Research is mainly concerned with generalising and formulating a theory. 'Gathering knowledge for knowledge' is termed pure research. Research studies concerning natural phenomena, human behaviour etc., are examples of Fundamental Research. But research aims at certain conclusions facing concrete social problems is an example of applied research.

3. Qualitative Vs Quantitative: Quantitative Research is based on the measurement of quantity or amount. It applies to a phenomenon that is a phenomenon relating to or involving quality or kind. Qualitative research is especially important in the behavioural sciences, where the aim is to discover the underlying motives of human behaviour.

4. Conceptual Vs Empirical Conceptual Research is related to some abstract ideas for theory. It is generally used by philosophers and thinkers to develop new concepts or to interpret existing ones. On the other hand, Empirical research relies on experiments or observation alone, often without due regard for a system of theory. It is data-based research coming up with conclusions capable of being variable of observation and experiment.

5. One-Time Research or Longitudinal Research In the legal case, the research is confined to a single period. In contrast, in the latter case, the research is carried on over several periods.

6. Laboratory Research and Field setting research This classification is based on the environment in which research is carried out

7. Historical Research is that which utilities historical sources like documents, remains etc., to study events and ideas of the past, including the philosophy of persons and groups at any remote point in time

Survey Research

The survey is a fact-finding study. It is a research method involving collecting data directly from a population or sample thereof at a particular time. It must not be confused with the mere clerical routine of gathering and tabulating figures. It requires expert and imaginative planning, careful analysis, and rational interpretation of the findings.

Definitions of survey research are as follows.

1. Mark Abraham defines a survey as "a social survey processed by which Quantitative facts are collected about the social aspects of community composition and activities

2. Herman N Morse defines It as "a method of analysis on scientific and orderly form for the defined purpose of given social situations and activities.

The Characteristics of Survey

1. It is a field study; It is always conducted in a natural setting.
2. It seeks responses directly from the respondents.
3. It can cover a very large population.
4. A survey involves an extensive and intensive study.
5. A survey covers a definite geographical area, city, district or state. Steps in a Survey

The sequences of the task involved in surveying the 1st stage of planning to the Final stage of preparing the report are presented below.

a) Selection of the problem and formulation,

b) Preparation of the research design,

c) Operationalisation of concepts and construction of measuring indexes and states.

d) Sampling,

e) Construction of tools for data collection and their pre-test.

f) Fieldwork and collection of data

g) Processing of data and tabulation

h) Analysis of data

i) Reporting Purpose of the Survey

Purpose of the survey

1. The purpose of the survey is to provide information to the government, planners, or business enterprises.
2. Many enquiries aim to explain the phenomenon
3. Surveys may be designed to make comparisons of demographic groups.
4. Surveys are useful for making predictions

Types of Survey

1. General or Specific Survey: When a survey is conducted to collect general information about a population institution or phenomenon without any particular object or hypothesis, it is known as a general survey. Specific surveys are conducted for specific problems or to test the validity of some theory or hypothesis.
2. Regular and Adhoc Survey: Some surveys are regular and must be repeated after regular intervals. Such a survey is called Regular Survey.
3. Preliminary and Final Survey: A Preliminary survey is generally known as a 'Pilot study' and is a forerunner of the Final Survey. The final survey is done after the pilot study has been completed. 4. Senses and Sample Survey A survey makes all the units of a given universe then. It is called a sense survey. If the survey covers only a fraction of the universe, it is called a sample survey.

Advantages of Survey:

The major advantages of the survey method are

→ The versatility of the survey method is its greatest strength. It is the only practical way to collect information from individuals, socio-economic data, attitudes, opinions, experiences and expectations.

→ The survey method facilitates drawing generalisations about large populations based on studies of a representative sample.

→ The survey method is flexible to permit various data collection methods.

→ The survey helps the researchers to face unanticipated problems.

→ The survey is useful in verifying theories

Limitations of Survey

1. The survey method is primarily meant to collect data from primary sources. So its success depends upon the willingness and co-operations of the respondents.
2. The survey method depends primarily on verbal behaviour. The respondent can give misleading answers.
3. A sample survey is subject to sampling error.
4. There is a limit to the number of items of information that can be collected in a single survey. There is an optimal length of time for an interview.
5. A survey is very expensive in terms of time and cost.

Research Problem:

In the Research process, the first and foremost step is that of selecting properly and defining a research problem. The researchers must find the problem and formulate it, making it susceptible to research. Like a doctor, a researcher must examine all the symptoms concerning a problem before diagnosing it correctly. "A problem well put is half solved". This saying highlights the importance of proper formulation of the selected problem. The primary task of research is collecting relevant data and analysing data or finding answers to research questions. The proper performance of this task depends upon the identification of correct data and information required for the study. Once the problem is formulated, he can execute the other steps without wasting time and energy. This formulation is a direction and specific focus of the research effort. It helps to delimit the field of enquiry and prevent blind research and indiscriminate data gathering.

A proper formulation helps solve all major research tasks like sampling, collection of data, construction of tools, plan of analysis, etc. What is a Research Problem? A Research problem generally refers to some difficulty the researcher experiences in a theoretical or practical situation and wants to obtain a solution.

"The term problem means a question or issue to be examined" The term problem originates from the Greek word 'Problem' - meaning anything that is thrown forwards, a question proposed for a solution, or a matter stated for examination. What is the formulation? Formulation means "translating and transforming the selected Research problem into a scientifically researchable question".

Problem statements should possess the following attributes:

- → Clarity and precision [a well-written statement does not make sweeping generalisations and irresponsible pronouncements; it also does include unspecific determinates like "very" or "giant"],
- → Demonstrate a researchable topic or issue [i.e., the feasibility of conducting the study is based upon access to information that can be effectively acquired, gathered, interpreted, synthesised, and understood],
- → Identification of what would be studied while avoiding the use of value-laden words and terms,
- → Identification of an overarching question or small set of questions accompanied by key factors or variables,
- → Identification of key concepts and terms,
- → Articulation of the study's boundaries or parameters or limitations,
- → Some generalizability in regards to applicability and bringing results into general use,
- → Conveyance of the study's importance, benefits, and justification [i.e., regardless of the type of research, it is important to demonstrate that the research is not trivial],
- → Does not have unnecessary jargon or overly complex sentence constructions; and,
- → Conveyance of more than the mere gathering of descriptive data providing only a snapshot of the issue or phenomenon under investigation.

Sources of Problems for Investigation

Identifying a problem to study can be challenging, not because there's a lack of issues that could be investigated, but due to the challenge of formulating an academically relevant and researchable problem that is unique and does not simply duplicate the work of others. To facilitate how you might select a problem from which to build a research study, consider these sources of inspiration:

Deductions from Theory

This relates to deductions made from social philosophy or generalisations embodied in life and society that the researcher is familiar with. These deductions from human behaviour are then placed within an empirical frame of reference through research. From a theory, the researcher can formulate a research problem or hypothesis stating the characteristic findings in certain empirical situations. The research asks, "What relationship between variables will be observed if theory aptly summarises the state of affairs?" One can then design and carry out a systematic investigation to assess whether empirical data confirm or reject the hypothesis and hence, the theory.

Interdisciplinary Perspectives

Identifying a problem that forms the basis for a research study can come from academic movements and scholarship originating in disciplines outside your primary study area. This can be an intellectually stimulating exercise. A review of pertinent literature should include examining research from related disciplines that can reveal new avenues of exploration and analysis. An interdisciplinary approach to selecting a research problem offers an opportunity to construct a more comprehensive understanding of a very complex issue that any single discipline may be able to provide.

Interviewing Practitioners

Identifying research problems about particular topics can arise from formal interviews or informal discussions with practitioners who provide insight into new directions for future research and how to make research findings more relevant to practice. Discussions with experts in the field, such as teachers, social workers, health care providers, lawyers, business leaders, etc., offer the chance to identify practical, "real world" problems that may be understudied or ignored within academic circles. This approach also provides practical knowledge that may help design and conduct your study.

Personal Experience

Don't undervalue your everyday experiences or encounters as practical problems for investigation. Think critically about your experiences and frustrations with an issue facing society, your community, your neighbourhood, your family, or your personal life. This can be derived, for example, from deliberate observations of certain relationships for which there is no clear explanation or witnessing an event that appears harmful to a person or group or that is out of the ordinary.

Relevant Literature

A research problem can be selected from a thorough review of pertinent research associated with your general area of interest. This may reveal gaps in understanding a topic or an issue that has been understudied. Research may be conducted to 1) fill such gaps in knowledge; 2) evaluate if the methodologies employed in prior studies can be adapted to solve other problems; or 3) determine if a similar study could be conducted in a different subject area or applied in a different context or to different study sample [i.e., different setting or different group of people]. Also, authors frequently conclude their studies by noting implications for further research; read the conclusion of pertinent studies because statements about further research can be a valuable source for identifying new problems to investigate. The fact that a researcher has identified a topic worthy of further exploration validates the fact it is worth pursuing.

What Makes a Good Research Statement?

A good problem statement begins by introducing the broad area in which your research is centred, gradually leading the reader to the more specific issues you are investigating. The statement need not be lengthy, but a good research problem should incorporate the following features:

1. Compelling Topic

The problem chosen should motivate you to address it. Still, simple curiosity is not a good enough reason to pursue a research study because this does not indicate significance. The problem that you choose to explore must be important to you. Still, it must also be viewed as important by your readers and to a larger academic and social community that the results of your study could impact.

2. Supports Multiple Perspectives

The problem must be phrased in a way that avoids dichotomies and supports the generation and exploration of multiple perspectives. A general rule of thumb in the social sciences is that a good research problem would generate a variety of viewpoints from a composite audience of reasonable people.

3. Research ability

This isn't a real word, but it represents an important aspect of creating a good research statement. It seems a bit obvious, but you don't want to investigate a complex research project and realise that you don't have enough prior research to draw from for your analysis. There's nothing inherently wrong with original research, but you must choose research problems that can be supported, in some way, by the resources available to you

Research Design:

"A Research Design is the logical and systematic planning in directing the research. The design research involves translating a general scientific model into a varied research problem. But in practice, most of the basis is just a study plan. The research design can either be formal or informal. Definition

1. "It constitutes the blueprint for the collection, measurement and analysis of data" - Philips Bernard S

2. It "provides a systematic plan of procedure for the researcher to follow" -Best John N

3. "The design research from controlling general scientific model into varied research procedure"- P.V. Young

4. "A research design is "the programme that guides the investigator in collecting, analysing and interpreting observations". - David and Shava

Does a research design address certain key issues, such as a) What is the problem in the uncle's study? b) What is the major research question? c) What is the area of the study? d) How many people will be studied? e) How will this person be selected? f) What methods and techniques will be used to collect data from them?

The uniqueness of Research Design

a) It is a plan that specifies the study's objectives and the hypothesis to be tested.

b) It is an outline that specifies the sources and types of information relevant to the research question.

c) It specifies the methods adopted for gathering and analysing data.

d) It is a scheme defining the procedure involved in a research process.

Features of a Good Design:

Good design has the following features.

1. Flexibility
2. Efficiency
3. Appropriate
4. Economical
5. Minimum error
6. Maximum reliability
7. Smallest experimental error
8. Maximum information

Why a Research Design

1. Research Design is needed because it helps in the smooth sailing of Research operations.
2. Research without a pre-drawn plan is like an ocean voyage without a mariner's campus.
3. The Research Design helps in providing direction for our study.
4. It prevents welter in a study.
5. The use of Research Design prevents blind search.
6. A Research Design fixes clear-cut boundaries to a research
7. It makes the research systematic 2. Why is the study being made? 3. Where will the study be carried out? 4. What type of data is required? 5. Where can the required data be found?

6. What period will the study include? 7. What will be the sample design? 8. What technique of data collection will be used? 9. How will the data we analyse? 10. In what style will the report be prepared?

In conclusion, it can be said that research design must contain at least: a) a Statement of a problem, b) Procedure and techniques, c) a Sampling frame, d) Processing and analysis of data

Types of Research Design:

1. Exploratory Research Design (Formulate Research) Exploratory Research studies are also termed formative research studies. Exploratory research is a preliminary study of an in-

familiar problem the researcher has little or no knowledge of. It is similar to a doctor's initial investigation of a patient suffering from an in familiar malady to get some clues for identification.

2. Descriptive Research Design Descriptive study is a fact-finding investigation with adequate interpretation. It is the sample type of research. It is more specific than the exploratory study. As it has focused on particular aspects or dimensions of the problem studied. It is designed to gather descriptive information and provides information for formulating more sophisticated studies. Data are collected using appropriate methods.

3. Action Research is a type of evaluation study. It is a concurrent evaluation study of an action programme launched to solve a problem. Action research is otherwise called Apply Research.

The following are the different phases in action research

1. A baseline survey of the pre-action situation.
2. A feasibility study of the proposed action programme
3. Planning and launching the programme.
4. Concurrent evaluation of the programme
5. Making modifications and changes in the programme and its implementation methods in light of research findings.
6. Final Evaluation (The Researcher can design his research depending upon the nature of the research being conducted)

Literature Review:

A literature review surveys books, scholarly articles, and other sources relevant to a particular issue, area of research, or theory, providing a description, summary, and critical evaluation of these works about the research problem being investigated. Literature reviews are designed to provide an overview of sources you have explored while researching a particular topic and to demonstrate to your readers how your research fits within a larger field of study.

A literature review may consist of simply a summary of key sources. Still, in the social sciences, a literature review usually has an organisational pattern and combines both summary and synthesis, often within specific conceptual categories. A summary is a recap of the important information from the source. Still, a synthesis is a re-organisation, or a reshuffling, of that information in a way that informs how you plan to investigate a research problem. The analytical features of a literature review might:

- ✓ Give a new interpretation of old material or combine new with old interpretations,
- ✓ Trace the intellectual progression of the field, including major debates
- ✓ Depending on the situation, evaluate the sources and advise the reader on the most pertinent research, or
- ✓ Usually, in the conclusion of a literature review, identify where gaps exist in how a problem has been researched to date.

The purpose of a literature review:

- → Place each work in the context of its contribution to understanding the research problem being studied.
- → Describe the relationship of each work to the others under consideration.
- → Identify new ways to interpret prior research.
- → Reveal any gaps that exist in the literature.
- → Resolve conflicts amongst seemingly contradictory previous studies.
- → Identify areas of prior scholarship to prevent duplication of effort.
- → Point the way in fulfilling a need for additional research.
- → Locate your research within the context of existing literature

Types of Literature Reviews

It is important to think of knowledge in a given field consisting of three layers. First, there are the primary studies that researchers conduct and publish. Second are the reviews of those studies that summarise and offer new interpretations built from and often extending beyond the primary studies. Third, the perceptions, conclusions, opinions, and interpretations shared informally among scholars become part of the body of epistemological traditions within the field.

In composing a literature review, it is important to note that this third layer of knowledge is often cited as "true", even though it often has only a loose relationship to the primary studies and secondary literature reviews. Given this, while literature reviews are designed to provide an overview and synthesis of relevant sources you have explored, several approaches could be adopted depending on the type of analysis underpinning your study. The important types of literature review are as follows.

Argumentative Review

This form examines literature selectively to support or refute an argument, deeply embedded assumption, or philosophical problem already established in the literature. The purpose is to develop a body of literature that establishes a contrarian viewpoint. Given the value-laden nature of some social science research [e.g., educational reform; immigration control], argumentative approaches to analysing the literature can be a legitimate and important form of discourse. However, note that they can also introduce problems of bias when they are used to make summary claims of the sort found in systematic reviews [see below].

Integrative Review

Considered a form of research that reviews, critiques, and synthesises representative literature on a topic in an integrated way such that new frameworks and perspectives on the topic are generated. The literature includes all studies that address related or identical hypotheses or research problems. A well-done integrative review meets the same standards as primary research regarding clarity, rigour, and replication. This is the most common form of review in the social sciences.

Historical Review

Few things rest in isolation from historical precedent. Historical literature reviews focus on examining research throughout a period, often starting with the first time an issue, concept, theory, or phenomenon emerged in the literature, then tracing its evolution within the scholarship of a discipline. The purpose is to place research in a historical context to show familiarity with state-of-the-art developments and to identify the likely directions for future research.

Methodological Review

A review does not always focus on what someone said [findings] but on how they came about saying what they say [method of analysis]. Reviewing methods of analysis provides a framework of understanding at different levels [i.e. those of theory, substantive fields, research approaches, and data collection and analysis techniques], how researchers draw upon a wide variety of knowledge ranging from the conceptual level to valuable documents for use in fieldwork in the areas of ontological and epistemological consideration, quantitative and qualitative integration, sampling, interviewing, data collection, and data analysis. This approach helps highlight ethical issues which you should be aware of and consider as you go through your study.

Systematic Review

This form consists of an overview of existing evidence pertinent to a formulated research question, which uses pre-specified and standardised methods to identify and critically appraise relevant research and to collect, report, and analyse data from the studies included in the review. The goal is to deliberately document, critically evaluate, and summarise all of the research scientifically about a clearly defined research problem. Typically it focuses on a very specific empirical question, often posed in a cause-and-effect form, such as "To what extent does A contribute to B?" This literature review is primarily applied to examining prior research studies in clinical medicine and allied health, but it is increasingly used in the social sciences.

Theoretical Review

This form examines the corpus of theory accumulated regarding an issue, concept, theory, or phenomenon. The theoretical literature review helps to establish what theories already exist, the relationships between them, to what degree the existing theories have been investigated, and to develop new hypotheses to be tested. Often this form is used to help establish a lack of appropriate theories or reveal that current theories are inadequate for explaining new or emerging research problems. The unit of analysis can focus on a theoretical concept or a whole theory or framework.

Hypothesis:

When a researcher observes known facts and takes up a problem for analysis, he first has to start somewhere, and this point of starting is the hypothesis. In other words, one has to proceed to formulate a tentative solution. These proposed solutions constitute the hypothesis. Collecting facts (data) will be fruitful if they are either for or against this proposed solution. The tentative explanation or solutions are the very basis for the research process. When to make a Hypothesis? Hypotheses are not given to us readymade. This is so, especially in social sciences. Because there is not a system of highly developed theoretical order in many social sciences. As a result, many social science researchers devote a considerable portion of their research to hypothesis-making. So it should be remembered that research can begin only with a well-formulated Hypothesis. Definition 1. In the words of George A. Lund Burg," a hypothesis is a tentative generalisation the validity of which remains to be tested………In its most elementary stage, the hypothesis may be very bunch, guessing, imaginative data, which becomes the basis for action or investigation 2. Goode and Hatt defined it as " a proposition which can be tested to determine its validity 3. Rummel ", a hypothesis is a statement capable of being tested and thereby verified or rejected".

Need (importance) of Hypothesis

In all analytical and experimental studies, hypotheses should be set up to give a proper direction to them. The hypothesis is useful and guides the Research Process in the proper direction. In addition, to put the theory to the test, a hypothesis has to perform certain other functions. In many ways, it is a guiding point to research. Young says, "Formulation of hypothesis gives the point of enquiry aids in establishing direction in which to proceed and help to delimit the field of enquiry" The use of hypothesis prevents a blind search and indiscriminate gathering of data which may later prove irrelevant to the problem under study. In the data, the collection hypothesis serves as the forerunner. A Researcher based on a hypothesis can save a lot of time and keep the researcher from a considerable amount of confusion. hypothesis helps direct the researcher to find out the order among facts. A hypothesis also has certain practical value to society besides serving as a means for seeking solutions to various problems. They help in understanding social phenomena from the proper perspective.

Sources of Hypothesis

The hypothesis can be derived from various sources.

1. Theory This is one of the main sources of the hypothesis. It gives direction to research by stating what is known. The logical deduction from theory leads to a new hypothesis

2. Observation Hypothesis can be derived from observation. For example, from observation of price behaviour in a market, the relationship b/w the price and demand of an article can be hypothesised.

3. Analogies: Analogies are another source of hypothesis. Julian Huxley has pointed out that causal observation in nature or the framework of another science may be a fertile source of hypothesis.

4. Intuition and personal experiences Intuition and personal experiences may also contribute to the hypothesis formulation. The personal life and experience of a person determine their perception and conception. These may, in terms, direct a person to a certain hypothesis more quickly. The story Newton and falling apple, the flash of wisdom to Sree Buddha under the Banyan tree, Illustrates this accidental process.

5. Findings and Studies Hypothesis may be developed out of the findings of other studies to repeat the test.

6. Culture Another source of the hypothesis is the culture in which the researcher is nurtured. For example, sociology as an academic discipline originated from western culture. Over the past decade, a large part of the hypothesis on American society examined by researchers was connected with violence. Indian society is caste-ridden and riddled with inequalities and privileges.

Characteristics of a Good Hypothesis

What is a good hypothesis? What are the criteria for judging it. An acceptable should fulfil certain conditions.

1. Conceptual Clarity A hypothesis should be conceptually clear. It should consist of clearly defined and understandable concepts

2. Specificity A hypothesis should be specific and explain the expected relation b/w variables and the conditions under which these relations will hold.

3. Testability A hypothesis should be testable and should not be a moral judgement. It should be possible to collect empirical evidence to test techniques

4. Availability of techniques Hypothesis should be related to available techniques. Otherwise, they will not be researchable. Therefore, the research must ensure that methods are available for testing their proposed hypothesis.

5. Consistency Hypothesis should be logically consistent. The propositions derived should not be Contradictory

6. Objectivity Scientific hypothesis should be free from value judgment. The researcher's system of values has no placing Research.
7. Simplicity A hypothesis should be as simple as possible. Simplicity demands insight. The more insight the researcher has into a problem, the simpler their hypothesis will be.

Types of Hypothesis:

1. Descriptive Hypothesis These are repositioning it describe the characteristics of a variable. The variable may be an object, person, organisation, situation or event. For ex. "The unemployment rate among arts graduates is higher than that of commerce graduates."
2. Relational Hypothesis These propositions describe the relationship b/w two variables. The relation suggested may be positive or negative, for example. 'Families with higher income spent more for recreation'. 'Upper-class people have more children than lower-class people
3. Causal Hypothesis states that the existence of, or a change in, one variable causes or leads to an effect on another variable. The 1st variable is called the independent variable later, the dependent variable.
4. Common Sense Hypothesis These represent commonsense ideas. They state the existence of empirical uniformities received through day-to-day observations
5. Analytical Hypothesis These are concerned with the relationship of analytic variables. This hypothesis occurs and a higher level of abstraction.
6. Null Hypothesis Null means 'Zero' When a hypothesis is stated negatively. It is called Null Hypothesis. The object of this hypothesis is to avoid the personal bias of the investigator in the matter of the collection of data. A null hypothesis is used to collect additional support for the general hypothesis.
7. False Hypothesis A hypothesis bound to be unsatisfactory when verified is called a false hypothesis.
8. Barren Hypothesis A hypothesis from which no consequences can be deducted is called a Barren Hypothesis. It is a hypothesis which cannot test. Ex. The child fell ill because a wicked women's eye fell upon it. This is a baseless hypothesis because it cannot be verified. Testing of Hypothesis Science does not admit anything as valid knowledge until a satisfactory test confirms the validity. A hypothesis should be subjected to a regret test. Type I and Type II errors should be eliminated. Concepts are basic elements of the scientific method, but all concepts are

abstractions and represent only certain aspects of reality. In the words of P.V. Young, "Each new class of data, isolated from other classes on the other basis of definite characteristics, their given name, a label in shorthand concept. A concept is, in reality, a definition in shorthand of a class or group of facts" Categories of Concept Concepts are divided into two categories, i.e. Concepts by postulation and concepts by intuition. The concepts following in the first category have meaning except for the specific theory; When these concepts are used in two different theories, these communicate two different meanings, sometimes even different and opposite. On the other hand, concepts by intuition devote something which is immediately apprehended. The meaning of these concepts is

constant, whoever uses them. Both categories of concepts have equal importance and significance in social science research.

Features of Good Concept:

a) The concepts should be clear, definite and precise.

b) The concept should be comprehensive and have clear information and understanding.

c) The concept should avoid multiple meanings and, as far as possible, should convey exactly what was intended when the concept was coined.

a) Concrete concepts: - Symbolise material objects which can be seen, touched and, e.g. book, table Characteristics of concepts

1. Concepts are symbols we attach to the meanings we hold.

Concepts represent only one part of reality.

3. Different people hold different concepts of the same thing

4. Concepts also represent various degrees of abstraction. (A concept is thus an abstract symbol representing an object, a property of an object, or a certain phenomenon.) Variable The concrete, observable events representing abstract concepts or constructs are called variables.

Review Questions:

1. What is research, and explain the characteristics of research?
2. Briefly explain the objectives and the importance of the research.
3. Narrate a few words about the process of research.
4. State the various types of research.
5. What do you understand about research design, and discuss its features.

Reference:

1. Andrews, L.; Higgins, A.; Andrews, M. W. and Lalor, J.G. (2012), Classic Grounded Theory to Analyze Secondary Data: Reality and Reflections, The Grounded Theory Review, 11(1): 12–26
2. Bond, T. (2003), Validity and Assessment: A Rasch Measurement Perspective, Metodoliga de las Ciencias del comportamento, 5(2): 179–194.
3. Bryman A. & Bell, E. “Business Research Methods” 4th edition, Oxford University Press, 2015.
4. C.R. Kothari, Research Methodology Methods and Techniques, ISBN (13): 978-81-224-2488-1, 2004.
5. The irony, S.I. and Rose, A.A. (2005) “Designing a Strategic Information Systems Planning Methodology for Malaysian Institutes of Higher Learning (isp-pita), Issues in Information System, Volume VI, No. 1, 2005
6. Patel M, Patel N. Exploring research methodology: review article. International Journal of Research and Review. 2019; 6(3):48-55.

UNIT –II: Types of Research Methods

Historical, Survey and Experimental, Case Study, Scientific Research and Statistical Research, etc.

Research is a serious academic activity with objectives to explain or analyse or understand a problem or find the solution(s) for the problem(s) by adopting a systematic approach to collecting, organising and analyzing the information relating to the problem. Research is a multidimensional activity. It comes in various forms and is used in all social, behavioural, educational, economical and management sciences. According to the approach and method involved in research, one can classify the following types of research.

I. Fundamental (Or) Basic Research:

1. Pure or Basic research is a search for broad principles and synthesis without immediate utilization objectives. It is not concerned with solving practical policy problems but with designing fascinating analysis tools and discovering underlying and, if possible, universal laws and theories. E.g. John Robinson's imperfect competition and chamberlains' monopolistic competition.

2. Applied (or)Action Research: Applied research, also known as action research, is associated with a particular project and problem. Being of practical value, such research may release to current activity (or) immediate practical situation it aims at finding a solution for an immediate problem facing society. Practically all social science research undertaken in India is of the applied variety, particularly of the type which helps formulation of policy.

3. Descriptive Research: It is designed to describe the demographic characteristics of consumers who use the product. It is designed to describe something, such as the demographic characteristics of consumers who use the product. It deals with determining the frequency with which something occurs or how two variables vary. An initial hypothesis also guides this study. For example, an investigation of the trends in the consumption of soft drinks about ratio economic characteristics such as age, sex, ethnic group, family income, education level, geographic location, and so on would be a descriptive study

Merits:

- → This approach helps test the conclusion and findings based on laboratory studies. Using this approach, it is possible to substantiate existing theories and conclusions by modifying them.
- → Direct contact between the researcher and the respondent is brought about in this approach. This is very significant because the researcher could understand the problem to be studied.
- → With the possibility of a direct contract with the respondent, the researcher can elicit all the relevant information and eliminate irrelevant facts.

Limitations:

- ↘ Unless the researcher is experienced, there is every possibility of the approach being misused.
- ↘ Hurried conclusions and generalizations may be formed based on inaccurate field data.
- ↘ As this approach involves the collection of field data, enormous time and effort is required to plan and execute the field survey
- ↘ This approach also involves incurring high costs for data collection.
- ↘ Collecting data through this approach is impossible unless the respondents are cooperative.

II. Historical Research:

As the name suggests in this approach, historical data is important to analyse and interpret the results. Following this approach, a researcher would collect past data for his research. A scholar using this approach must rely on libraries to refer to magazines or periodicals for collecting data.

Merits:

- ✓ This approach alone is relevant in certain types of research work. For example, to understand the trend in India's exports. One has to collect the export data for, say, 20 years and analyze it similarly to study the impact of the liberalisation policy. One must collect information from 1991 till date.

- ✓ The historical approach makes research possible as it is firmly believed that once we understand the past, our understanding of the present and expectations of the future can be predicted to some extent. Hence historical research provides insight into the past and facilitates looking into the future.

Limitations:

- ✓ The personal bias of those who have written about historical events or incidents cannot be misled.
- ✓ Researchers tend to over-generalize their results using a historical approach.
- ✓ Persons using this approach should be conscious that historical data can be taken to indicate the past. Still, the formulation of solutions on that basis and applying them in the current period is not correct.

III. Exploratory Research:

Most marketing research projects begin with exploratory. It is conducted to explore the possibilities of doing a particular project. The major emphasis is on the discovery of ideas and insights. For example, a soft drinks firm might conduct an exploratory study to generate possible explanations. The exploratory study is used to split the broad and vague problem into smaller, more precise sub-problem statements in the form of a specific hypothesis. An exploratory study is conducted in the following situations.

- → To design a problem for investigation and to formulate the hypothesis.
- → To determine the priorities for further research.
- → To gather data about the practical problems for researching particular conjectural statements.
- → To increase the interest of the analyst towards the problems and
- → To explain the basic concepts.

An exploratory study is more flexible and highly informal. There is no formal approach in exploratory studies. Exploratory studies do not employ detailed questionnaires. These studies will not involve probability sampling plans. The following are the usual methods of conducting exploratory research

- Literature Survey
- Experience Survey and
- Analysis of insight-stimulating cases.

IV. Literature Survey:

The literature searches quickly and economically for researchers to understand better a problem area in which they have limited experience. In this regard, a large volume of published and unpublished data is collected and scanned relatively quickly. Generally, sources include books, newspapers, Government documents, trade journals, professional journals and soon. These are available in libraries, and company records such as these are kept for accounting sales analysis purposes; reports of previous research projects conducted problems incompletely but will be of great help to provide a director to further research.

V. Experience Surveys;

In this method, people with expert knowledge and ideas about research subjects may be questioned. Generally, the company executives, sales managers, other relevant people of company salesman, wholesalers, retailers who handle the product or related products and consumers are concentrated. It does not involve a scientifically conducted statistical survey, and rather, it reflects an attempt to get available information from people who have some particular knowledge of the subject under investigation.

VI. Case study approach (analysis of insight-stimulating cases).

The case study approach to research is a recent development. This approach focuses on a single organization or unit, an institution, a district or a community. As the focus is on a single unit, it is possible to undertake an in-depth analysis of the single unit. It is a problem-solving approach and the following characteristics of the case study method. The study of the whole unit: In this study, a large variety of units are selected for study, and the unit size may be quite large to cover an entire community in a word. This method treats an individual, an institution, or a group of persons.

Intensive study:

It aims at the deep and thorough study of a unit. It deals with every aspect of a unit and is studied intensively. The following methods are undertaken in the case study;

Determination of Factors:

First of all, the collection of materials about each of the units or aspects is essential.

- → Determining factors may be of two types, certain factors and General factors.
- → Statement of the problem:
- → In this process, the defined problem is studied intensively, and the data are classified into various classes.
- → Analysis and conclusion: After classifying and studying the factors, an analysis is made

Advantages:

- ❖ As this approach involves a focused study, there is much scope for generating new ideas and suggestions.
- ❖ It may provide the basis for developing sound hypotheses.
- ❖ As the researcher studies the problem from his point of view, very useful and reliable findings may be obtained.

Limitations:

- → A significant limitation of this approach is that unless the researcher is experienced, he might ignore very important aspects.
- → This approach also depends on the infirm furnished by the respondents. Unless the firm is accurate, the conclusions are bound to be irrelevant.
- → It is often said that case studies are based on the observations of the researcher

VII. Experimental Research:

This is a very scientific approach. In this approach, the researcher first determines the problem to be studied. Then he identifies the factors that cause the problem. The problem tobe probed is quantified and taken as the dependent variable. The factors causing the problem will be taken as the independent variable. Then the researcher studies the causal relationship between the dependent and independent variables. He is also able to specify to what extent the dependent variable is. He can also specify to what extent each independent variable influences the dependent variable. For example, suppose food production is considered a research study problem. Then the scholar would determine the factors that will affect food production. Viz size of the land cultivated (x) rainfall (y) quantity of fertilizer applied (z) etc. These factors,x,y and z, are independent variables, and Food production [A] is called the dependent variable. Then, collecting data regarding all four [A,x,y and z]. The researcher can state what percentage change in the final food (A) is explained by x,y and z. The effect of x on A, y on A and z on A is also studied. In this manner, the researcher can successfully indicate to what extent various factors included in the study are important.

Merits of Experimental Approach (Research)

- → This approach provides social scientists with a reliable method it observes under given conditions to evaluate various social programmes.
- → This is one of the best methods of measuring the relationship between variables.
- → This approach is more logical and consistent than the conclusions drawn, but research based on this approach is well received.
- → It helps to determine the cause-effect relationship very precisely and clearly
- → Following this approach researcher could indicate the areas of future research

Limitations of Experimental Approach (Research)

- → This approach cannot be easily followed unless a researcher is well-experienced and trained in model building.
- → By relying more on models, this approach may not add anything significant to the knowledge
- → A serious limitation of this approach is that it relies on sampling and collection of data.
- → Unless these are properly planned and executed. The outcome of the analysis will not be accurate

VIII. Diagnostic Study;

This is similar to a descriptive study but with a different focus. It is directed towards discovering what is happening, why, and what can be done about it. It aims to identify a problem's causes and possible solutions for it. A diagnostic study may also discover and test whether certain variables are associated.E.g., are persons from rural areas more suitable for manning rural branches of banks? (or) Do more villagers than city voters vote for a particular party?

IX. Evaluation studies;

An evaluation study is one type of applied research. It is made for assessing the effectiveness of social or economic programmes implemented (e.g. family planning scheme) or for assessing the impacts of developmental projects (e.g., irrigation projects) on the area's development. An evaluation study may be defined as the "determination of the results attained by some activity(whether a program, a drug or a therapy or an approach) designed to accomplish some valued goal or objective".

X- Analytical Study:

The analytical study is the system of procedures and techniques of analysis applied to quantitative data. It may consist of mathematical models (or) statistical techniques applicable to numerical data. Hence it is also known as the statistical method. This method is extensively used in business and other fields in which quantitative numerical data are generated. It is used for measuring variables, comparing groups and examining the association between factors. Data may be collected from either primary sources or secondary sources.

XI- Surveys Research:

The survey is a fact-finding study. It is a research method involving data collection directly from a population or a sample at a particular time. It must not have confused me with the more clerical routine of gathering and tabulating figures. It requires expertise and careful Analytical knowledge. The data analysis may be done using simple or complex statistical techniques depending on the study's objectives. This type of research has the advantage of greater scope in the sense that a larger volume of information can be collected from a very large population

XII. Other types of research

Ex-post Facto Research;

Ex-post Fact research is based on the observation made by inquiry in which the researcher does not have direct control of independent variables because their outcome has already occurred. This research is based on a scientific and analytical examination of dependent and independent variables. The ex-post facto research findings may become riskier by improper interpretations.

Panel Research:

Generally, the survey research is valid for a one-time period, known as the „study period", and they do not reflect changes occurring in time. Consumer attitudes toward purchasing a particular product are not static and changing. For example, it is not possible to study the changes occurring in these attitudes over a period in response to changes in the particular products marketing min. Measuring change over time is known as longitudinal analysis, which uses panels. This method is generally used in sales forecasting by consumer preferences for various products measuring audience size and characteristics for media programmes testing new products.

Advantages;

→ It considers the changes in time.
→ It provides more control
→ It has greater co-operation
→ It offers more analytical Data from respondents.

Errors in Research:

The errors in research will happen in so many stages. Some of them are discussed below:

1. Questionnaire Studies

➢ Using a questionnaire to work with problems is better than other research techniques.
➢ Not giving enough care to developing the questionnaire and not pretesting it.

- Asking too many questions, thus making unreasonable demands on the respondents. Time.
- Overlooking details of format, grammar, printing, and so on can influence respondents.
- First impression.
- Not checking a sample of non-responding subjects for possible bias in the questionnaire.

2. Interview Studies

→ Not adequately planning the interview or developing the interview guide.
→ Not conducting sufficient practice interviews to acquire needed skills.
→ Failing to establish safeguards against interviewer bias.
→ Not making provisions for calculating the reliability of the interview data.
→ Using language in the interview that the respondents won't understand.
→ Asking for information that the respondents cannot be expected to have.

3. Experimental Studies

- Inadvertently or otherwise, treating the experimental and control groups differently, thus leading to biased findings.
- Using too few cases leads to large sampling errors and insignificant results.
- Failing to divide the main groups into subgroups in situations where subgroup analysis may produce valuable knowledge.
- Matching the subjects in the experimental and control groups on criteria that have little to do with the variables being studied.
- Attempting to match control and experimental groups on so many criteria that, in the process, you lose many subjects who cannot be matched.

4. Content Analysis Studies:

→ Selecting content that is easily available but is not an unbiased sample.
→ Selecting some content that is not related to the research objectives.
→ Failing to determine the reliability of the content-analysis procedures.
→ Using classification categories that are not specific yet comprehensive.

5. Observational Studies:

- → Not sufficiently training observers and thus obtaining unreliable data.
- → Using an observation procedure that demands too much of the observer.
- → Failing to safeguard against the observer's disturbing or changing the observed situation.
- → Attempting to evaluate behaviour that occurs so infrequently that reliable data cannot be obtained through observations.
- → Relationship (Correlation) Studies
- → Assuming that a correlation between pieces of data is proof of a cause-and-effect relationship.
- → Using a sample in correlation research that differs on so many variables that comparisons of groups are not interpretable.
- → Putting the cart before the horse: trying to build a correlational study around conveniently available data instead of collecting the data needed to do a worthwhile study.
- → Selecting variables for correlation that have been found unproductive in previous studies.
- → Failing to use appropriate disciplinary theory in selecting variables to study.
- → Using simple correlation techniques in studies where partial correlation or multiple correlations are needed to obtain a clear picture of how the variables operate.

Review Questions:

1. Write a short note on historical research and ex-post facto research.

2. State the merits and demerits of experimental research?

3. What do you understand about the case study approach?

4. Narrate the importance of the literature review.

5. List out the possible errors in research and how to overcome them.

Reference:

1. Babbie, E. R. (2013). The Basics of Social Research. Cengage Learning

2. Bernard, H. R. (2011). Research methods in anthropology: Qualitative and quantitative approaches. Rowman Altamira.

3. Holden, M. T., & Lynch, P. (2004). Choosing the appropriate methodology: Understanding research philosophy. The marketing review, 4(4), 397-409.

4. Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded source book (2nd ed.). Newbury Park, CA: Sage.

5. Newman, I., & McNeil, K. A. (1998). Conducting survey research in the social sciences. University Press of America.

UNIT – III: Research Techniques

Research Techniques and Tools: Questionnaire, Interview, Observation, Schedule and Check-list, etc. Library Records and Reports

The search for answers to research questions is called a collection of data. Data are facts, and other relevant materials, past and present, serving as bases for study and analyses. The data needed for social science research may be broadly classified into

(a) Data about human beings,

(b) Data relating to organizations and

(c) Data about territorial areas.

Personal data or data related to human beings consist of

(1) Demographic and socio-economic characteristics of individuals: Age, sex, Race, social class, religion, marital status, education, occupation, income, family size, location of the house hold, life style, etc.,

(2) Behavioral Variables: Attitudes, opinions, awareness, Knowledge, practice, intentions, etc. Organizational data consist of data relating to an organization's origin, ownership, objectives, resources, functions, performance and growth. Territorial data are related to geophysical characteristics, resources endowment, population, occupational pattern, infrastructure, structure, degree of development, etc., of spatial divisions like villages, cities, taluks, districts, states and the nation.

Importance of data

The data serve as the bases or raw material for analysis. Without an analysis of factual data, no specific inferences can be drawn on the questions under study. Inference based on imagination or guesswork cannot answer research questions correctly. The relevance, adequacy, and reliability of data determine the quality of the findings of a study. Data from the basis for testing the hypotheses formulated in a study. Data also provide the facts and figures required for constructing measurement scales and tables, which are analysed with statistical techniques. Inferences on the results of statistical analysis and tests of significance provide the answers to research questions. Thus, the scientific process of measurements, analysis, testing and inferences depended on relevant data availability and accuracy. Hence, the importance of data for any research study.

Sources of data

The sources of data may be classified into (a) primary sources and (b) secondary sources. Primary sources are sources from which the researcher directly collects data that have not been previously collected. Primary data are first-hand information collected through various methods such as interviewing, mailing, observation etc. Secondary sources contain data which have been collected and compiled for another purpose. The secondary sources consist of readily available compendia and already compiled statistical statements and reports whose data may be used by researches for their studies. E.g., census reports, annual reports and financial reports. Secondary sources consist of not only published records and reports but also unpublished records.

Objective data and subjective data

Objective data is independent of any person's opinion, whereas subjective data can be an individual's opinion or dependent upon the researcher.

Qualitative data and quantitative data

Qualitative data is the description of things made without assigning numeric value. For example, facts generated from unstructured interviews. It needs the researcher's interpretation. Quantitative data entails measurements in which numbers are used directly to represent properties of things. It is ready for statistical analysis. A larger sample is required in quantitative data, and with proper sampling design, the ability to generalize is also high.

Main methods of Data collection

Most research studies collect new data from the respondents even though already existing data are utilized to develop the research design or supplement the collected data. There are various methods of data collection. 'Method is different from a 'Tool' while a method refers to the way or mode of gathering data. A tool is an instrument used for the method. For example, a schedule is used for interviewing.

The important methods are

(a) observation, (b) interviewing, (c) mail survey (D) scheduling.

Observations involve gathering data about the selected research by viewing and listening. Interviewing involves a face-to-face conversation between the investigator and the respondent. Mailing is used for collecting data by getting questionnaires completed by respondents. Experimentation involves the study of independent variables under controlled conditions. Experiments may be conducted in a laboratory or the field in a natural setting. Simulation involves the creation of an artificial situation similar to the actual life situation. Projective methods aim to draw inferences on respondents' characteristics by presenting their stimuli. Even the method has its advantages and disadvantages. A researcher can select one or more methods considering the above factors. No method is universal. Each method's unique features should be compared with the needs and conditions of the study; thus, the methods' choice should be decided.

Observation has no specific purpose. But observation in a method of data collection differs from such causal viewing. Observation may be defined as a systematic viewing of a specific phenomenon in its proper setting for gathering data for a particular study. Observation as a method includes both 'seeing' and 'hearing'. It is accompanied by perceiving as well. Observation is a classical method of scientific inquiry.

Observation also plays a major role in formulating and testing social sciences hypotheses. Behavioural scientists observe interactions in small groups; political scientists observe the behaviour of political leaders and institutions. Observation may serve a variety of research purposes. It can be used in exploratory research to develop a preliminary understanding of social phenomena. It can be applied to study real-life situations and conduct experimental research. Again, it can be used to collect supplementary data in support of other data collection tools. Observation includes the most causal and uncontrolled experiences and exact recording as done in experimentation. Observation is useful for studying simpler as well as complex research problems.

Observation becomes scientific when it

(a) serves a formulated research purpose,

(b) is planned deliberately,

(c) is recorded systematically, and

(d) is subjected to checks and controls on validity and reliability.

Validity refers to the extent to which the recorder observations accurately reflect the construct they are intended to measure. Validity is assessed by examining how well the observations agree with alternative measures of the same construct. Reliability entails consistency and freedom from measurement error.

Characteristics of the observation method

Observation as a method of data collection has certain characteristics.

1. It is both a physical and mental activity. The observing eye 'catches' many slight things, but attraction is focused on data pertinent to the given study.

2. Observation is selective. The researcher does not observe anything and everything but selects the range of things to be observed based on the nature, scope and objectives of his study

3. Observation is purposive and not casual. It is made for the specific purpose of nothing things relevant to the study

4. It captures the natural social context in which a person's behaviour occurs.

5. It grasps the significant events and occurrences that affect social relations of the participants.

6. Observation should be exact and be based on standardized research tools such as observation schedules, social-metric scales, and precision instruments if any.

Types of observation

Observation may be classified in different ways. The investigator's role may be classified into (a) participant observation and (b) non-participant observation. In terms of mode of observation, it may be classified into (c) direct observation and (d) indirect observation. Regarding the rigour of the system adopted, observation is classified into (e) controlled and (f) uncontrolled observations.

Participant observation

In this observation, the observer is a part of the observed phenomena or group and acts as both an observer and a participant. Observing persons should not be aware of the researcher's purpose. Then only their behaviour will be 'natural'. The concealment of the research objective and the researcher's identity is justified on the ground that it makes it possible to study certain aspects of the group culture which are not revealed to outsiders makes it possible to study certain aspects of the group's culture which are not revealed to outsiders.

The advantages of participant observation are:

1. The observer can understand the emotional reactions of the experimental group and get a deeper insight into their experiences.
2. The observer will be able to record context which gave meaning to the observed behaviour and heard statements.

Disadvantages

1. The participant observer narrows his range of observation.
2. To the extent that the participant observer participates emotionally, objectivity is lost. Because of the above limitations, participant observation is generally restricted to those cases where non-participant observation is not practical.

Non-participant observation

In this method, the observer stands apart and does not participate in the phenomenon observed. Naturally, there is no emotional involvement on the part of the observer. This method calls for skill in recording observations in an unnoticed manner.

Direct observation

This means observation of an event personally by the observer when it takes place. This method is flexible and allows the observer to see and record subtle aspects of events and behaviour as they occur. He can be free to shift places and change the focus of the observation. A limitation of this method is that the observer's perception circuit may not be able to cover all relevant events when the latter moves quickly, resulting in the incompleteness of the observation.

Indirect observation

This does not involve the physical presence of the observer, and mechanical, photographic or electronic devices do the recording. This method is less flexible than direct observation but less biased and erratic in recording accuracy. It also provides a permanent record for analysing different aspects of the event.

Controlled observation

This involves standardization of observational technique and exercise of maximum control over extrinsic and intrinsic variables by adopting experimental design and systematically recording observations. Controlled observation is earned either in the laboratory or in the field. It is typified by clear and explicit decisions on what, how and when to observe. It is primarily used for inferring causality and testing causal hypotheses.

Uncontrolled observation

This does not involve control over extrinsic and intrinsic variables. It is primarily used for descriptive research. Participant observation is a typical uncontrolled one.

Planning of observation

The use of the observation method requires proper planning. First, the researcher should carefully examine the relevance of the observation method to the data needs of the selected study. Second, he must identify the investigative questions for the observation method. These determine the data collected. Third, he must decide the observation content, viz., specific conditions, events and activities that must be observed for the required data. The observation content should include the relevant variables. Fourth, for each variable chosen, the operational definition should be specified. Fifth, the observation setting, the subjects to be observed, the timing and mode of observation, recording, procedure, recording instruments to be used, and other details of the task should be determined. Last, observers should be selected and trained. The persons to be selected must have sufficient concentration powers, strong memory power and intrusive nature. Selected persons should be imparted both theoretical and practical training.

Observation Tools and Recording Devices

Systematic observation requires using an observation schedule (or observation are), a diary and various mechanical recording devices.

Schedule:

The data requirements are identified by analyzing the core of the problem and the objectives

information easily and correctly. Enough space should be provided for recording observations each time. The item should appear in logical groupings and in the order in which the observer would observe them.

Field observation:

This may take the form of a diary or cards. Each item of observation is recorded in the appropriate sub-heading. At the time of observation, rough noting may be made, and the full log may be made at the end of the day. The card system is flexible and facilities the arrangement and rearrangement of items in any desired order.

Mechanical devices:

These may include cameras, tape recorders, videotapes and electronic devices. Still, motion, sound, colour and time-lapse cameras give a permanent record of events. Microscopic and telescopic lenses may be used in cameras.

Advantages of observation

Observation has certain advantages

1. The main virtue of observation is its directness; it makes it possible to study behaviour as it occurs. The researcher need not ask people about their behaviour and interactions; he can watch what they do and say.

2. Data collected by observation may describe the observed phenomena as they occur in their natural settings. Other methods introduce elements or artificiality into the researched situation.

3. Observation is more suitable for studying subjects unable to articulate meaningfully.

4. Observation is less demanding of the subjects and has a less biased effect on their conduct than questioning.

Limitations of study

1. Observation is of no use in studying past events or activities. One must depend upon people's documents or narration to study such things.

2. It is not suitable for studying opinions and attitudes.

3. Observation poses difficulties in obtaining a representative sample. 4. Observation is a slow and expensive process. Requiring human observation and costly surveillance types of equipment.

Interview:

Interviewing is one of the well-known methods of data collection. An interview is a face-to-face interaction between two individuals in which a person asks questions from another person to gather information. Interview emerged as a data collection tool by the turn of the last century and has become an integral part of social research. During earlier times, interviews were conducted more like probing conversations. Guided by a careful observer, this method was used as a powerful instrument for obtaining information. It involves not only conversation but also learning from the respondent's gestures, facial expressions and pauses, and environment. Interviewing requires face-to-face contact or contact over the telephone and calls for interviewing skills. It is done by using a structured schedule or an unstructured guide.

Importance of interview;

Interviewing may be used either as a main method or as a supplementary one in studies of persons. Interviewing is the only suitable method for gathering information from illiterate or less educated respondents. It is useful for collecting a wide range of data, from factual demographic data to highly personal and intimate information relating to a person's opinions, attitudes, values, beliefs, past experiences and future intentions. Interviewing is required when qualitative information is required or probing is necessary to draw out fully. Where the area covered for the survey is compact, or when a sufficient number of qualified interviews are available, a personal interview is feasible. The interview is often superior to other data-gathering methods. People are usually more willing to talk than to write. Once rapport is established, even confidential information may be obtained. It permits probing into the context and reasons for answers to questions. Interviews can add flesh to statistical information. It enables the investigator to grasp the behavioural context of the data furnished by the respondents. It permits the investigator to seek clarifications and brings to the forefront those questions that respondents do not want to answer for one reason or another.

Characteristics of interview

Interview as a method of data collection has certain characteristics.

1. The participants- the interviewer and the respondent- are strangers. Hence, the investigator has to get him appropriately introduced to the respondent.

2. The relationship between the participants is a transitory one. It has fixed beginning and termination points. The interview proper is a fleeting, momentary experience for them.

3. Interview is not a causal conversational exchange but a conversation with a specific purpose, viz., obtaining information relevant to the study.

4. Interview is a mode of obtaining verbal answers to questions put verbally

5. The interaction between the interviewer and the respondent need not be face-to-face because the interview can be conducted over the telephone.

6. Although the interview is usually a conversation between two persons, it need not be limited to a single respondent. It can also be conducted with a group of persons, such as family members, children or customers, depending on the requirements of the study.

7. Interview is an interaction process. The interaction between the interviewer and the respondent depends upon how they perceive each other.

Types of interview

The interviews may be classified into

(a) structured or directive interview,

(b) unstructured or non-directive interview,

(c) focused interview, and

(d) clinical interview and

(e) depth interview.

Structured, directive interview:

This is an interview made with a detailed standardized schedule. The same questions are put to all the respondents in the same order. Each question is asked in the same way in each interview, promoting measurement reliability. This type of interview is used for large-scale formalized surveys.

Unstructured or non-directive interview:

This is the least structured one. The interviewer encourages the respondent to talk freely about a topic with minimal prompting or guidance. In this type of interview, a detailed pre-schedule is not used. Only a comprehensive interview guide is used.

Focused interview

This is a semi-structured interview where the investigator attempts to focus the discussions on the actual effects of a given experience to which the respondents have been exposed. It occurs with the respondents who have been involved in a particular experience.

Clinical interview

This is similar to the focused interview but with a subtle difference. While the focused interview is concerned with the effects of a specific experience, the clinical interview is concerned with general underlying feelings or motivations or the course of the individual's life experiences.

Depth interview

This is an intensive and searching interview aiming at studying the respondent's opinion, emotions or convictions based on an interview guide. This requires much more training in interpersonal skills than structured interviewing. This deliberately aims to elicit unconscious and extremely personal feelings and emotions.

Advantages of interview

There are several real advantages to a personal interview

First, the greatest value of this method is the depth and detail of information that can be secured.

Second, the interviewer can do more to improve the percentage of responses and the quality of the information received than another method. He can note the conditions of the interview situations and adopt appropriate approaches to overcome such problems as the respondent's unwillingness, incorrect understanding of questions, suspicion, etc.

Third, the interviewer can gather other supplemental information like economic level, living conditions etc. Through observation of the respondent's environment.

Fourth, the interviewer can use special scoring devices and visual materials to improve the interview quality.

Fifth, the accuracy and dependability of the answers given by the respondent can be checked by observation and probing.

Last, the interview is flexible and adaptable to individual situations. Even more, control can be exercised over the interview situation.

Limitations of Interview

First, the Interview is not free from limitations. Its greatest drawback is that it is costly both in money and time.

Second, the interview results are often adversely affected by the interviewer's mode of asking questions and interactions, incorrect recording, the respondents' faulty perception, memory, inability to articulate etc.

Third, certain types of personal and financial information may be refused in face-to-face interviews. Such information might be supplied more willingly on mail questionnaires, especially if they are to be unsigned.

Fourth, the interview poses the problem of recording information obtained from the respondents; no foolproof system is available. Note-taking invariably distracts both respondent and the interviewer and affects the conversation thread.

Last, the interview calls for highly skilled interviewers. The availability of such persons is limited, and the training of interviewers is often a long and costly process.

Schedule

A schedule refers to a set of questions related to a subject, printed or typed in a definite order. It is a device for securing information that asks a person to answer the given questions. The schedule is an important research tool which facilities the collection of data from large, diverse and widely scattered groups of people. It can be used to collect quantitative data and secure g information of qualitative nature. In most empirical studies, the primary research tool is the schedule.

The research investigator administers the schedule. It is not self-administered. In the case of the schedule, the answers are obtained from the respondent in a face-to-face situation. The interviewer notes down the responses or answers. Moreover, if required, the interviewer can act as a stimulus or provide on-the-spot clarifications to the respondent. Thus, a schedule presupposes a face-to-face interaction between the interviewer and the respondent. When the questions are self-administered and require the respondent to answer all items. When the questions are self-administered and require the respondent to answer all questions by himself, it is called a mailed questionnaire. Such questions are often sent by mail/post to the respondent. Sometimes they are distributed to a group of people who may have come to attend a conference, with the request that they may fill it up and return the same.

Usually, a schedule contains structured items. Structured is meant that questions have fixed wording. They are also typed or printed in a definite order. Thus a particular set of questions having the same wording and sequence is administered to all the respondents.

Important considerations in the construction of a schedule

A schedule should contain a limited number of questions. Only such questions as are extremely important to fulfil the requirements of a study should be included. Questions which may elicit some already known or obvious information should be deleted. Generally, respondents do not like to devote sufficient time and energy to a long schedule. A schedule is usually divided into several sections. Each section must contain questions related to a particular item or theme.

In light of the requirement of the research problem, some rationale must be developed to include each item or theme. Questions relating to the same general theme must be placed together. Questions placed at the beginning of a schedule should be such that they can draw the respondents' complete attention. They should be able to evoke his interest. However, the opening question should be neutral. In other words, the beginning should not contain controversial issues as the respondents may develop a negative orientation towards the whole questionnaire. This may lead to tardy responses or even outright refusals.

The researcher should carefully determine the sequence of the different themes covered in a schedule. A gradation of themes, starting from simpler ones and gradually leading to complex questions, should be made. There should be a logical sequence of themes as well as questions. The transition from one theme to the other should not be abrupt. Complex questions requiring serious thinking are preferably placed somewhere in the middle of the schedule. The respondent may become fatigued and not give such questions the attention they deserve. A crucial aspect of a schedule relates to the formulation of questions. These should be framed so that the researcher may logically expect the answers to be significant for his research problem.

A good schedule grows from a sound study of the problem and a literature review. Every item in the schedule must be related to the study's objectives. A tentative list of areas on which questions are to be asked is drawn up. It is useful to consult knowledgeable people and conduct interviews to achieve greater clarity. Initially, efforts should be made to cover as many items as possible. Gradually the researcher may detect omissions, gaps or ambiguities. Also, we may determine which items are most important and which are not. Any research venture can adequately cover only a limited number of themes directly related to the research objectives. Most of the questions should focus sharply on such important themes. The actual process of formulating questions requires great skill and expertise. A detailed discussion on this aspect runs beyond the scope of this unit. The amount of space needed for answering open-ended questions has to be determined. For example, the researcher may suddenly find that an open-ended question needs more space for the answer than has been provided in the printed schedule. Some preliminary exercise helps in assessing this requirement as well.

Telephone interviewing

Telephone interviewing is a non-personal method of data collection. It may be used as a major method or supplementary method. It will be useful in the following situations: 1. When the universe is composed of those persons whose names are listed in telephone directories, e.g., business houses, business executives, doctors, and other professionals. 2. When making the respondents are widely scattered and when there are many callbacks to make. 3. When the subject is interesting or important to respondents. 4. When the survey must be conducted in a very short period, provided the units of study are listed in the telephone directory. 1. The survey can be completed at a very low cost because a telephone survey does not involve

The advantages of telephone interviews are:

- → Travel time, cost, and all calls can be made from a single location.
- → Information can be collected in a short period. 5 to 10 interviews can be conducted per hour.
- → The quality of the response is good because interviewer bias is reduced as there is no face-to-face contact between the interviewer and the respondent.
- → It has a higher response rate
- → 5. It has greater sample control.

The telephone interview has several limitations.

1. There is a limit to the length of the interview. Usually, a call cannot last over five minutes. Only five or six simple questions can be asked. Hence, the telephone cannot be used for a longer questionnaire.

2. The type of information to be collected is limited to what can be given in simple, 0short war of words. Hence, the telephone is not suitable for complex surveys, and there is no possibility of obtaining detailed information.

3. If the questions cover personal matters, most respondents will not cooperate with the interviewer.

4. The respondent's characteristics and environment cannot be observed.

5. Establishing rapport between the respondent and the interviewer is difficult.

Mail survey

The mail survey is another method of collecting primary data. This method involves sending questionnaires to the respondents with a request to complete them and return them by post. This can be used in the case of educated respondents only. The mail questionnaires should be simple so the respondents can easily understand and answer the questions. It should preferably contain mostly closed-end and multiple-choice questions so that it can be completed within a few minutes. The distinctive feature of the mail survey is that the questionnaire is self-administered by the respondents themselves, and the responses are recorded by them and not by the investigator, as in the case of the personal interview method. It does not involve a face-to-face conversation between the investigator and the respondent. Communication is carried out only in writing, requiring more respondent cooperation than verbal communication.

Procedure

The researcher should prepare a mailing list of the selected respondents by collecting the addresses from the telephone directory of the association or organization to which they belong. A cover letter should accompany a copy of the questionnaire. It must explain to the respondent the purpose of the study and the importance of his cooperation to the project's success. Anonymity may be assured.

Alternative modes of sending questionnaires There are some alternative methods of distributing questionnaires to the respondents. They are:

(1) Personal delivery,

(2) Attaching questionnaire to the product.

(3) Attaching questionnaire in a newspaper or magazine and

(4) News stand inserts.

Personal delivery: The researcher or his assistant may deliver the questionnaires to the potential respondents with a request to complete them at their convenience. After a day or two, he can collect their completed questionnaires. Often referred to as the self-administered questionnaire method combines the advantages of the personal interview and the mail survey. Alternatively, the questionnaires may be delivered in person, and the completed questionnaires may be returned by mail by the respondents.

Attaching questionnaire to a product:

A firm test- a firm test-marketing a product may attach a questionnaire to a product and request the buyer to complete it and mail it back to the firm. The respondent is usually rewarded with a gift or a discount coupon.

Advertising the questionnaire:

The questionnaire with the instructions for completion may be advertised on a magazine page or in a section of newspapers. The potential respondent completes it, tears it out and mails it to the advertiser.

News-stand inserts: This method involves inserting the cover letter, questionnaire and self-addressed reply-paid envelope into a random sample of newspaper or magazine newsstand copies.

The advantages of mail surveys are:

1. They are less costly than personal interviews, as the cost of mailing is the same throughout the country, irrespective of distance.

2. They can cover extensive geographical areas.

3. Mailing is useful in contacting persons such as senior business executives who are Difficult to reach in any other way.

4. The respondents can complete the questionnaires at their convenience

5. Mail surveys, being more impersonal, provide more anonymity than personal interviews.

6. Mail survey is free from interviewer bias, as there is no personal contact between the respondents and the investigator.

7. Certain personal and economic data may be given more accurately in an unsigned mail questionnaire.

The disadvantages of mail surveys are:

1. The scope for mail surveys is very limited in a country like India, where the percentage of literacy is very low.

2. The response rate of mail surveys is low. Hence, the resulting sample will not be a representative one.

3. It is difficult to determine the degree of representativeness of a sample obtained by mail. 4. The causes for inadequate and non-responses cannot be known, and no probing is possible.

5. Information on the respondent's characteristics and environment cannot be secured.

6. Respondents may not cooperate if the mail questionnaire is long or complex. 7. Several returned questionnaires may contain unanswered questions and incomplete responses.

Tools for Data collection

The various methods of data gathering involve the use of appropriate recording forms. These are called tools or instruments of data collection. They consist of- Questionnaire, observation schedule, interview guide, interview schedule and mailed questionnaire. Each of the above tools is used for specific methods of data gathering: Observation schedule for the observation method, interview schedule and interview guide for interviewing, and questionnaire for the mail survey.

Functions

The data collection tools translate the research objectives into specific questions/items, the response to which will provide the data required to achieve the research objectives. To achieve this purpose, each question/item must convey to the respondent the idea or group of ideas required by the research objective research objects. Each item must obtain a response that can be analyzed for fulfilling the research objectives. Information gathered through the tools provides descriptions of individuals, institutions or other phenomena under study. The characteristics may help to explain differences in behavioural patterns and performance of objects under study. Information gathered through the tools serves another purpose also. It is useful for measuring the various variables in the study. The variables and their interrelationships are analyzed to test the hypothesis or to explore the content areas set by the research objectives.

Questionnaire

The questionnaire depends upon research objectives. For each objective or research question, list all the questions a researcher wants to answer through the study. Then the information required to answer them is listed, and the questions are listed. A questionnaire consists of questions presented to a respondent for answers. The questionnaire is used during structured surveys or interviews. The respondents read the questions, interpreted what was expected, and then wrote down the answers. It is also called an interview schedule, when the researcher asks the questions and records the respondents' replies on the interview schedule. Here, the researcher may have to explain questions to the respondents. There are many options before the researchers adopt this method, but questionnaires should be developed and tested carefully before being administered on a large scale. There are three basic types of questionnaires, closed-ended open-ended and a combination of both.

1. **Closed-ended questionnaire:** closed-ended questionnaires generally include multiple-choice or scale questions. This type of questionnaire can be administered to many respondents or sample sizes. As there is a set format, the data generated from the questionnaire can be easily fed into a computer system for analysis.

2. **Open-ended questionnaire:** open-ended questionnaires offer the flexibility to respondents to answer in their own words. It may leave a blank section to write an answer. Closed-end questionnaires might be used to find out how many people use the metro rail service in New Delhi, but open-ended questionnaires might be used to find out what people think about the quality of service.

3. **Combined questionnaire:** in this method, it is possible to find out how many people use a service and what they think of the service in the same form. The combined questionnaire may begin with a series of closed-end questions, with boxes to tick or scales to rank, and then finish with a section of open-ended questions or more detailed responses.

Observation schedule or Observationnaire

This is a form on which each unit observation for observations of an object or a phenomenon is recorded. This item to be observed is determined concerning the nature and objectives of the study. They are grouped into appropriate categories and listed in the schedule in the order in which the observer would observe them. The items are structured with possible alternatives. Space is each unit observation for encircling, checking, or recording, as the case may be. Provision is made for the correct identifications of each case observed and the observer. The schedule should be constructed to make it possible to record the observations easily and correctly and to tabulate and analyse them effectively.

The schedule must be devised to provide the required verifiable and quantifiable data and avoid selection bias and misinterpretations of observed items. The observation units must be simple and meticulously worded to facilitate precise and uniform recording.

Interview Guide

This is used for non-directive and depth interviews. It does not contain a complete list of which information has to be elicited from a respondent; it just contains only the broad topics or areas to be covered in the interview. The interview guide serves as a suggestive reference or promoter during the interview. It aids in focusing attention on salient points relating to the study and in securing comparable data in different interviews by the same or different interviewers.

There is considerable flexibility regarding the manner and order of language In which the interviewer asks the questions. If the interviewer has to refer to the guide very often, it will defeat its purpose. The interviewer cannot listen closely and analytically if his attention rests on the guide. He may fail to respond to the cues and implications of the interviewee's remarks.

Interview schedule and mailed Questionnaire

Both these tools are widely used in surveys. Both are complete lists of questions on which information is elicited from the respondents. The basic difference between them lies In recording responses. While the interviewer fills out a schedule, a questionnaire is completed by the respondent.

Library Records and Reports:

Libraries maintain different types of records to exercise control over their routine functions as per established norms. Some records are created before the library becomes operational; the others are created and captured in the course of the conduct of library operations and services. As a library science student, you must understand the nature and type of such records and their functions in library work. In this Unit, we will study library records, their meaning, need kinds and the roles that records play in various library sections.

Records are created and captured by individuals or organisations in the course of conduct of their affairs. Most records are evidence of administrative, executive, service or work-related transactions. Traditional records formats include letters, minutes, memoranda and reports, etc. The electronic record formats include spreadsheets, databases, e-mails and facsimiles. In some ways, what technology has done is redefine traditional records formats. For example, an e-mail is very similar to a letter. It is the fact that it is transmitted electronically that differentiates e-mail from letters

Records provide information in the form of descriptions, statistics, or routines carried out by an individual or a group working in an institution or organisation. While all records convey information, not all information sources are necessary. For example, a published book or an externally provided database (online or offline) will not be a record, although information selected from it and reused in a new context may become a record.

The records keep on accumulating as the transaction proceeds. By the time the activity is over, records take the shape of documents providing complete details of the activity's history, process and outcome. The activity may be about making a policy, carrying out an operation, or introducing a new service. Records are perceived as a ‘snapshot’ of an action or event in this context. They offer a picture of something that happened.

Records arise from actual happenings. For instance, an exhibition of rare books would generate a series of records from questions that visitors may put to the staff managing the exhibition and the opinions they may record about the exhibition in the visitor book before they leave. Records that may arise from such an event include some files, the exhibition brochure, the bibliography of the books exhibited, the visitor’s book and so on. Some records have both current and future uses, while others are required to be used only in future as source material. For example, the files created for the exhibition will be used mainly in future when another exhibition is organised. On the other hand, the bibliography of the books may be used at present and in the future.

The term record has been in use in the context of archives for a long time. But now it is a widely used term in industrial houses, institutions, organisations and libraries. In library science, it refers to the documents containing details of processes, activities and results of some function of the whole organisation or a part of it. ISO 15489: International Standard on Records Management defines a record as: “recorded information in any form, including data in a computer system, created, received and maintained as evidence and information by an organisation or person, in the transaction of business and kept as evidence of such activity”. A library is an organisation where sources of information in various forms, such as books, serials, microfiches, CDs, etc., are acquired, processed and made available to the clientele for their use. In making provision for resources and, later on, for offering various services, libraries generate many records in descriptive and statistical tables.

Need for Records:

The records contain information that serves as evidence of functions executed and activities performed. As such, they are a valuable source of knowledge about how and why decisions are made. Given that records are of value to any organisation, their proper creation and management are necessary to ensure current and future accountability and to support future actions. The need to properly manage records is also evidenced by the crucial role that records play when they are needed to protect the library's interests during crisis periods and when they are required to meet official audits, investigations and reviews from outside agencies. We can understand that the library needs records for a variety of reasons, such as follows:

→ Records are needed to document actions and decisions taken in a library and to conduct library business in an orderly, efficient and accountable manner.

→ Records are needed to provide consistency, continuity and productivity in library management and administration.

→ Records are needed to deliver services consistently and equitably as records help to bring and maintain uniformity in the routines.

→ Records are needed to support and document policy formulation and managerial decision-making and to promote informed decision-making.

→ Records provide benchmark measures to monitor and evaluate the progress in performance, efficiency and achievements.

→ Records support the formation and revision of policies, programs, systems and long-term strategic plans.

→ Records lend support to various library proposals such as expansion in the building, demand for additional staff, proposals for more funds for acquisitions

→ and new services/services, etc.

→ Records serve as source materials to construct library history and maintain library memory.

→ Records, in the form of a catalogue, serve as a tool for efficient access to library resources.

→ Records as proof of budget utilisation to fulfil audit requirements.

→ Records of workload help in the rational allocation of the workforce across various departments.

→ Records as data of utilisation of services to justify library expenditure to the parent management and the society.

Kinds of Library Records:

Unlike records of government departments or industrial houses, which are mostly files, library records are created and captured in several different forms depending upon the nature of activities in various library departments. In a fully computerised library, such records are computer generated by the relevant software modules. In non-automated libraries, the dominant records forms are ledgers, registers, files, cards and statistical sheets. Records may be categorised by

i) their life span,
ii) content and
iii) form of presentation of information

Records as per their Life Span These records are categorised as follows:

i) Permanent,
ii) Semi-permanent, and
iii) Temporary.

Permanent Records –

Records which are retained permanently include library policy documents, library procedures and guidelines, documents relating to library origin and growth, staff sanction approvals, accession registers, user registration forms, audit reports, inventory of furniture, fittings and types of equipment, building plans, electrical and sewage systems plans, etc.

Semi-permanent Records – These include funds utilisation register, agreements with firms about terms and conditions of supply of books and access to databases and library circulation records, in-house usage records and online usage records, etc. Financial records are retained till the audit is over. Similarly, records about the terms and conditions of supply of a firm need not be retained once dealing with the firm is over.

Temporary Records – Records which are generated for a specific work are to be destroyed upon completion of the work. For example, book selection cards/slips are usually destroyed once the books have been catalogued. These are not destroyed in case they are used as accession records. The serial registration card is also a temporary record. Once it is full, it is replaced by a new card.

Records as per their Content

These records include the following:

i) records of resources,

ii) financial records,

iii) service records and

iv) administrative records.

Records of Resources – These records pertain to books and non-book materials such as audiobooks, audio-visual works, brochures, computer software, dissertations, globes, maps, microfiches, microfilms, photos, postcards, posters, reports, serials, slides and stamps.

Financial Records – Accounts books such as budget registers, registers showing section-wise allocation of funds and bill registers are financial records.

Library Service Records include inter-library loans, circulation, reference service records, etc.

Administrative Records – The number of staff varies from library to library. There are libraries which are run by a single staff. Others have staff varying from a few to more than a hundred. Every staff has a service book that maintains records of his full service starting from his application for the service. Confidential reports of every staff written by their seniors are also carefully maintained. These are all important administrative records

Records by Form of Presentation

These records are of two types – descriptive and statistical. Annual reports of a library, the stated library policies, minutes of various meetings, etc., are descriptive records. A library purchases several books in a year, processes them, issues them to the readers, replaces them after their return, and so on. Every day several readers use the library. Some of them ask questions, want some services and so on. At the end of the year, cumulated records are analysed, and data is generated indicating the number of books purchased, processed, issued, and some queries answered. Photocopies supplied, etc. and progress made over the previous years. These are statistical records.

Records Maintained by the Individual Sections of Libraries:

Libraries organise their activities under a departmental structure. The actual organisational structure in a library depends upon the size of its collection, the span of library functions or the administrative needs of the parent body. The following structure, with marginal variations, is more common among libraries.

i) Acquisition Section

ii) Technical Processing Section

iii) Circulation-cum-Reading Room Section

iv) Reference and Documentation Service Section

v) Periodicals Section

vi) Maintenance Section

vii) Administration Section

Each section creates and maintains records to help conduct its routines in a systematic order,

1. Acquisition Section

This section is responsible for building up the library collection. The process of collection building begins with the formulation of a library collection development policy and, subsequently, its implementation on the following lines:

- → Receiving suggestions from library members, staff and authorities,
- → Recording suggestions for each title on book selection slips,
- → Checking slips with various records for resources available in the library holdings to ensure that no title is duplicated unintentionally,
- → Placing an order with an appropriate supplier,
- → When the ordered material is received - it is checked for correct supply, entered in the accession record and transferred to the Technical Section for further action,
- → The invoice for the supplied material is recorded in a register, adjusted against financial provisions and passed on to the finance department responsible for payment.

The above description of the acquisition of materials through the purchase order method necessitates the creation of the following types of records:

Ledgers – There are different ledgers for record keeping as described below:

→ Ledgers for the recording of invoices, i.e. bill registers for every category of library materials such as books, serials subscriptions and non-book materials, including electronic sources.
→ Ledger for expenditure adjustment against funds allocation already made for each subject/department, i.e. budget allocation register.
→ Ledger for accessioning of materials, i.e. accession register/accession cards tray.

Files – Libraries create and maintain several files as records for different purposes. These can be categorised hereunder,

Correspondence File – This type of file is used to keep all correspondence with the suppliers of library materials. Correspondences may be orders, reminders, complaints regarding the supply of damaged or wrong books, etc. Correspondence files may be more than one. A file for the purpose is required if a library obtains periodicals through consortia and books, etc., through institutional membership. A library may receive books, periodicals, etc., through exchange or gift. In that case, also, another file is required for the purpose.

Suggestions File – This file contains suggestions received for books, periodicals, etc., from library members, staff of various departments and authorities.

Legal File – This file contains agreement documents such as terms and conditions of supply of available or rare books, serials, electronic sources, exchange of publications and receipt of gifts, etc.

Cards and Slips in Trays – These trays hold suggestion cards/ slips for books on order and books awaiting technical processing.

Statistical Sheets – These contain records as to the number of book selection slips prepared, checked and filed in the order tray, removed from the order tray, etc.

2. Technical Processing Section:

The department has three distinct functions: classification, cataloguing and physical processing. The first two functions are done with the help of classification schemes and cataloguing codes adopted for the purpose, often with certain deviations in constructing class numbers or providing details in the catalogue entry for a document. This is done in response to meeting the local needs of the library. The library may decide to drop providing added entries for editors or class books on biography under 920 (DDC number) only. Such policy decisions are recorded in the authority file for future reference.

Authority File - The authority file is the only file maintained in the department for recording deviations adopted by the library in constructing class numbers, providing details in a catalogue card entry or in preparing a maximum of several cards for a book. Such deviations are called local variations.\

Cards – Catalogue cards and shelf lists represent the library collection. These are the two records that the technical department generates. The shelf list is usually retained in the department, and the card catalogue is placed close to the entrance lobby or reading area. If any book is lost or withdrawn from the library, the relevant cards are removed from the catalogue cabinet and shelf list.

Statistical Sheet – The sheet records the number of catalogue cards prepared, filed, and updated by adding information for additional copies of a book, cards replaced or withdrawn, etc

3. Circulation Section:

The main function of the circulation section is the charging and discharging of books. The department has to perform many activities to carry out this function successfully. These are reservation routines, bespoken card writing, requisitioning materials on interlibrary loans, etc. Although binding is a distinct administrative function, it is dealt with by the circulation department in some libraries as it involves keeping a record of books going out of the library. These activities call for the creation and retention of several records in all three formats - ledger, files and card:

Ledgers – The ledgers maintain records of interlibrary loan materials, overdue charges,
receipt books for overdue charges.

Files – There may be one or more files keeping various documents such as letters, lists of books, etc. The circulation section will have files relating to a library circulation policy, registration of library members, inter-library loan service and correspondence on overdue books or reserved books, binders for getting the books and periodicals bound, user complaints, etc. If the correspondence volume is small, one file may serve the purpose for all these; otherwise, different files may be opened depending on the need.

Cards – The issue record is based on a membership card in many libraries. A member's membership card is filed in the tray when a book is issued to a member. When the book is returned, the card is removed from the tray and returned to the member. Membership cards filed in the tray are valuable records indicating the books issued.

Book cards of books sent for binding, repair, digitisation, etc., are also maintained systematically. When the books come back to the library after binding, repair, digitisation, etc., the book cards are inserted into the relevant books, and then they are shelved

Membership application forms or cards are also important records. Basing the data in the cards or forms, it is possible to know the total number of members, types of members by age group, level of education, etc.

Statistical Sheet:

The statistical sheet records the number of books issued, shelved and repaired, reminders sent, interlibrary loan requests fulfilled, etc. These records provide valuable indicators of the workload, types of materials in demand, number of books shelved, number of books repaired, etc.

4. Reference and Documentation Service Section:

The nature and level of reference service vary with the nature and size of a library. In public libraries, it is mostly a question-answer type of service. In academic and special libraries, in addition to such queries, the staff is also expected to prepare a documentation list and offer current awareness services. In special libraries, the staff is also expected to provide selective dissemination of information (SDI) services to individual scholars. Since much of the dealings in the reference section with users are about providing reference service with the help of reference tools, not many records are generated or maintained in this department. The following types of records are maintained in the department:

Cards – Scholars' profiles are maintained on specially designed cards for providing SDI service to them.

Files – Reference enquiries received through various telephone, e-mail, SMS, letters and so on. In most cases, replies are also given using these channels. A
copy of the reply is kept in the file, which may help in replying to queries in future if it is on the same topic.

Statistical Sheets – Statistics of i) queries attended, ii) abstracts of articles prepared,
iii) bibliographies compiled, iv) SDI services provided, etc., are maintained on a daily
basis and cumulated every year.

5. Periodical Section;

Different libraries have different systems for organising the acquisition and maintenance of periodicals. Libraries with large numbers of serials subscriptions organise the entire routines in an independent section. Smaller libraries often include subscription and related routines, along with books, in the acquisition section and leave the receipt and display, etc., only fora separate section. Various records are maintained in this section relating to ordering, receipt and non-receipt of periodicals, issuing of reminders and binding of periodicals, etc.

Ledgers – Some libraries use Kardex for recording the receipt of periodicals. Many libraries record the receipt of periodicals in a register. If the serials department places orders for periodicals independently, it will also have to maintain a bill register.

Files – Depending on the need, one or more files may be required. The following papers or documents are kept in the files:

Correspondence relating to the subscription of periodicals. It may include quotation papers, price lists from the publishers, lists of periodicals, suggestions received from users, etc.

→ An organisation publishing periodicals get many periodicals in exchange for their periodicals. All correspondences and related papers are to be maintained properly for the smooth functioning of the activity.

→ Many libraries are members of one or more consortia. This also results in the incorrespondence and accumulation of other papers, such as the list of periodicals according to publishers and the list of consortia members.

→ Non-receipt of issues, receipt of defective copies, duplicate copies, or unsubscribed copies also leads to correspondence.

6. Maintenance Section

The section is primarily concerned with the upkeep and maintenance of library materials in good condition and in an order that helps users easily access the material. The section puts the collection in different sequences and specific orders to achieve this objective. It provides the stack area with helpful guides and identifies titles needing repair or binding. It makes periodic checks to identify books which need weeding out from the collection. As per library policy, the section also conducts stock verification at fixed intervals. The records in the section are mainly in the form of a list of books kept in files. The list may include the books withdrawn from the shelves for weeding out or repair. The notes in the files may relate to acquiring new furniture, removing old or unused furniture, requesting equipment and materials for cleaning, etc. Preserving rare and important documents is also an important library activity. This also generates a lot of records.

7. Administration Section

The administration section looks after a library's administrative activities, which involve the library workforce, finance and accounts, library building, and library assets such as furniture and equipment. It also functions as the nodal point for contact with the administration of the parent body of the library. The records in administration are mostly in the form of files, registers and ledgers.

Manpower Records – The human resources record comprises, among others, the attendance register, service books of employees, leave records and staff deployment records. Salary and wage records are also maintained where the library is the wage disbursing authority. You should note that some of these records are in the form of registers.

Statistical Records – Statistical sheets collected from each department are consolidated, whereby the overall scenario of the library comes up.

Financial Records – These records comprise correspondence relating to the annual budget, budget register, budget allocation and utilisation register, etc. The last mentioned record is very useful as it always provides information about money already spent and the money that remains to be spent. When anything needs to be purchased, this record is always very useful.

Records of Library Assets – Every library has various assets comprising bookshelves, almirahs, tables, chairs, fans, etc. Modern libraries have computers, photocopying machines, microfilm reader printers, videos and so on. All these furniture items, fittings, machines, tools, etc., are recorded in a register. When something is issued to someone, a record is maintained. Record is also maintained for those items that are written off.

Records concerning Library Development, Policy, etc. – These records include the annual report of the library, proceedings of library committee meetings, development plans and proposals of the library, library rules and regulations, contracts and agreements
and so on.

Library Building Records – Many libraries have their buildings. There are records relating to the library building plan, maintenance and repair. A library usually grows to require the construction of new rooms, storeys and, at times, new buildings. Security is of great concern for a library as it contains many rare books, manuscripts and other valuable materials. Hence flawless security arrangement is made. All these also generate records. These records are usually maintained in files.

Disaster Management Manuals – Disasters like fire, flood, earthquakes and tsunamis sometimes afflict libraries. That is why many libraries prepare disaster management manuals that act as guides to help librarians to save valuable books and other materials in the library when a disaster strikes. Records are usually maintained in the record room by the record keeper, who functions under the head of the administration department.

Report for a library:

Any organization is expected to present its annual report. The annual report of a library is the survey of work carried out during the preceding year. It summarizes the activities and achievements of the library. It is a report by the librarian to the higher authorities. An annual report is a survey of the previous year's work. An annual reveals the books accessioned and processed, services rendered, new departments or services introduced, financial and other resources and their utilization, assessment of performance and the like.

Purpose of the Annual Report:

- → To acquaint the Board of Trustees and other administrative authorities about the real financial position and other aspects of the services rendered by a library.
- → To serve as a public media towards attaining the users.
- → Helps to compare annual reports of different years, enabling the librarian to determine the strong and weak points of the library. This will help him to improve the library.

→ The staff position can be reviewed, and staff can be reallocated in the library.

→ To access the actual needs and to find ways and means of the provisions more funds.

Contents of Annual Reports:

The most important feature of a report is its contents. A report may include the following main heads.

1. **Area/ Population served**: The area/population served shows the extent of the library's services in terms of the supply of books and non-book materials per head of the civil population.

2. **Stocks of books and non-book materials**: Stocks of books and non-book materials like illustrations, prints, manuscripts, slides, phono-records, etc. The nature of the book and non-book resources, including the fields of specialization as reported in the Annual Report, will help access the library's resources and capacity to serve the community's various needs. The report must show the total number of volumes in stock at the beginning of the year under the report, the number of volumes withdrawn, the number of volumes added during the years, and the total number of volumes at the end of the year. Besides, the number of current periodicals subscribed to must be brought to the readers' notice.

3. Services (Technical and Readers service):

Technical services include classification, cataloguing, and documentation; readers' services include reference and circulation, photo-reproduction services, etc. The annual report should show what classification scheme is followed, the physical and inner form of catalogue services rendered, and the staff position revealed in the annual report will show if the staff is insufficient or otherwise.

In special libraries, technical activities like indexing, abstracting and documentation should be mentioned and evaluated as far as possible. A statistical table may be appended to show the number of books subject-wise and the per capita circulation of such books, etc.

Reader's advisory service is the most important factor of modern library service. Proper statistics with the description of such services are necessary for insertion in the annual report so that the authorities may be prompted to finance more improved services with the latest devices and enhanced reference resources.

4. Clientele: The number of registered readers, adults, children, blind, etc., is to be clearly shown along with statistics showing the percentage of readers to the population. The statistics of the clientele and their reading interests, as shown in the circulation statistics, will enable the librarian to judge the usefulness of the library.

5. Finance: The annual report must indicate the actual financial condition of a library so that it may be found out if the budget allocation is enough and needs augmentation. Sources of income from the Govt. fund, library authority fund, membership fees, fines, and other specified sources, and the total income should be shown. Expenditure on salaries of staff, books and non-book materials, periodicals, furniture, heating, cooling, lighting, cleaning, etc., and total expenditure is to be shown in the annual report.

6. Staff: The number of staff,

Librarians and Assistants

Typists and Clark

Caretakers, Sorters and Janitors

Cleaner

The annual report should show binding staff and the total number of staff.

7. Miscellaneous: The activities like extension services (lectures, film shows, training programmes. Seminars, conferences, study circles, etc.) and other social and adult education activities must be given in the annual report due to publicity.

Examples of an Annual Report for University Library:

A. Background Information:

1. Name of the library: …………….
2. Nature of the Library: …………….
3. Year of establishment: …………………..
4. Number of days close in the year: ………………
5. Working hours per day: …………………

B. Clientele:

1. Total number of clientele: ………..

a) Male: ………..

b) Female: …………..

2. Total number of active members: ……………..

3. New members added during the year: …………..

C. Collection of library materials:

1. Number of total books: ……………..

2. Number of total journals: ……………….

3. Number of total AV materials: ……………

4. Number of books added during the year: …………….

5. Number of books withdrawn during the year: …………………

D. Technical section:

1. Whether the materials are classified or not: ……………………

2. What classification scheme is followed: …………………………

3. Which catalogue code is used: ……………………….

4. Physical form of catalo0gue: …………………………

E.

(1) Circulation section:

a. Issue system: ………………………

b. Total number of books issued and returned daily: ……….

(2) Reference section:

a. Total

no. Of reference volumes in stock: ……………………..

b. No of

the total books consulted daily: …………………………

c. No of

the total users daily: …………………..

F. Staff:

The total number of staff:

a. Professional: ……………..

b. Non- professional: ………..

c. Unskilled: ……………….

G. Financial statement:

1. Income:

a. Govt. fund: …………….

b. Library authority: ………………

c. Other sources: ………………….

2. Expenditure:

a. Salaries of staff: ……………….

b. Books, periodicals: ……………

c. Miscellaneous charges: …………

Review Questions:

1. Explain the different data collection methods for research.

2. What do you understand about the questionnaire, and describe their importance in empirical research.

3. State the need for various records to be maintained in the library.

4. Enumerate the various library records to be maintained in the library.

5. what is the library's annual report, and state its contents?

Reference:

1. Akintunde, S.A. (2004). Libraries as tools for ICT Development in a paper presented at the National Library Association 42nd National Conference and Annual General Meeting. Akure, Nigeria: June 20 - 25.

2. Higgins, Gavin (ed). (1983). Printed Reference Material. 2nd ed. London: Library Association.

3. Robert d. 2007, stueartLibrary and information centre management (library and information science text series)

4. Olanlokun, S. O., &Salisu, T. M. (1993).Understanding the library: A handbook on library use. Lagos: the University of Lagos Press.

5. Subramanyam K. (1981). Scientific and Technical information Sources. N.Y.: Marcel Dekker.

UNIT – IV: Statistics and its Applications

Descriptive Statistics – Measures of Central Tendency: & Dispersion, Correlations and linear regression, Chi-Square test, t-test, z-test, f-test. Presentation of Data: Tabular, Graphic, Bar Diagram, Pie Chart, etc. Report Writing Statistical Packages – MS Excel, SPSS, Web-based Statistical Analysis Tools, etc.

The word "Statistics" has been derived from the Latin word "Status", or Italian word "Statista", or the German word "Statistika". Each of these words means Political State. Initially, Statistics was used to collect information about the state's people about their income, health, illiteracy, wealth, etc. But now a day, Statistics has become an important subject having useful applications in various fields of daily life. In the plural sense, Statistics refers to information in terms of numbers or numerical data such as population Statistics, Employment Statistics etc. However, any numerical information is not statistics.

According to Bowley - "Statistics are numerical statements of facts in any department of inquiry placed about each other."

Characteristics of Statistics:

The main characteristics of Statistics in terms of numerical data are as follows:

(1) **An aggregate of Facts –** A single number does not constitute Statistics. We cannot draw any conclusion from a single number. We can draw any conclusion by the aggregate number of facts. For example, if it is stated that there are 1,000 students in our college, then it has no significance. But if it is stated that there are 300 students in arts, 400 students in commerce and 300 in science in our college. It makes statistical sense as this data conveys statistical information. Similarly, if it is stated that the population of India is 130 crore or the value of total exports from India is `11 66,439 crores, then these aggregate facts will be termed as Statistics.

(2) **Numerically Expressed**: Statistics are expressed in terms of numbers. Qualitative aspects, like small or big, rich or poor, are not statistics. For instance, if we say Irfan Pathan is tall and Sachin is short, this statement has no statistical sense. However, if it is stated that the height of Irfan Pathan is 6 ft and 2 inches and the height of Sachin is five ft. and4 inches, these numerals will be called Statistics.

(3) **Affected by Multiplicity of Causes** – Statistics are not affected by any single factor but by many factors. For instance, a 30% rise in prices may have been due to several causes, like a reduction in supply, an increase in demand, a shortage of power, a rise in wages, a rise in taxes, etc.

(4) **Reasonable Accuracy** - A reasonable degree of accuracy must be kept in view while collecting statistical data. This accuracy depends on the purpose of the investigation, its nature, size and available resources.

(5) **Pre-determined Purpose -** Statistics are collected with some pre-determined objective. Any information collected without a definite purpose will only be a numerical value, not Statistics. Suppose data about the farmers of a village is collected. In that case, there must be some pre-determined objective, whether the statistics are collected to know their economic position or distribution of land among them or their total population. All these objectives must be–determined.

(6) **Collected Systematically** – Statistics should be collected systematically. Before collecting the data, a plan must be prepared. No conclusion can be drawn from data collected haphazardly. For instance, data regarding the marks secured by college students without any reference to the class, subject, examination, maximum marks, etc., will lead to no conclusion.

The subject matter of statistics:

The subject matter of statistics includes two components:

- ✓ Descriptive Statistics
- ✓ Inferential Statistics

Descriptive Statistics: Descriptive Statistics refers to the methods used to collect, present, and analyse data. These methods relate measurement of central tendencies, dispersion, measurement of correlation etc. For Example, Descriptive statistics are used when you estimate the average height of the secondary students in your school. Descriptive statistics are also used when you find the students' marks in science and mathematics in all classes are intimately related.

Inferential Statistics: Inferential Statistics refers to all methods by which a conclusion is drawn related to the universe or population based on a given sample. For example: If your class teacher estimates the average weight of the entire class based on the average weight of only a sample of students, then we use inferential statistics.

Important terminology in statistics:

Population: By population, we mean a well-defined set or group of all the objects for a particular study. The objects may be persons, plants, books, fishes in ponds, shops etc. the population will consist of certain elements like the plants of a certain kind in a specified field, the fishes in a pond, the unemployed person in India, books in a library and so on. For instance, if we want to study the properties of students in a school, then the population consists of all the students of the school. For instance, if we want to study books in a library, then the population includes all the books from the library. If the number of elements is limited, then the population is finite. On the other hand, if the number of elements is not limited, then the population is infinite. Mostly we deal with
finite population.

Sample: It is a part of the population selected by some sampling procedure. The process of selection of a sample is known as sampling. The number of objects in the sample is called the sample size. It is believed that a sample is the best representative of the population.
For instance, suppose research is required to study the weight of fish in a pond after a particular period of growth. For this purpose, suppose that there are 3,000 fish in the pond, he may either measure the weight of all the fish in the pond, or he may decide to select a small group of fish and measure their weights. The first approach to measuring the weight of all fish is called complete enumeration or census. Another approach in which only a small group of fish is considered is called a sample survey. In brief, we can say that in complete enumeration, information is collected on all the units of the universe and in sample surveys, only a part of the universe is considered.
Variable: A property of objects is known as the variable which differs from object to object and is expressible numerically in terms of numbers. For instance: the marks in Mathematics of

students in a class can be expressed in the marks obtained by the students. So it is a variable property which is expressible quantitatively.

Attribute: A property and characteristic of objects are known as attributes which are not expressible quantitatively in number. We can express the data qualitatively. For example, smoking, colour, honesty etc.

Stages in a Statistical Investigation:

There are five stages in statistics investigation. They are as follows.

→ Data Collection

→ Organisation

→ Presentation

→ Analysis

→ Interpretation

Collection: This is the primary step in a statistical study, and the investigator should collect data carefully. If data are faulty, the conclusions drawn can never be reliable. The data may be available from existing published or unpublished sources or collected by the investigator himself. First-hand data collection is a statistician's most difficult and important task.

2. Organization: Data collected from published sources are generally organised. However, a large mass of figures collected from a survey frequently needs organization. In organizing, there are three steps (A) Editing, (B) Classify (C) Tabulation.

(A)Editing: The collected data must be edited carefully so that the omissions, inconsistencies, irrelevant answers and wrong computation in the returns from a survey may be corrected or adjusted.

(B) Classify: Classification is arranging the data according to some common characteristics of the items constituting the data.

(C)Tabulation: To arrange the data in columns and rows. Hence collected data is organized properly so that the desired information may be highlighted and undesirable information avoided.

3. Presentation: Arranged data is not capable of influencing a layperson. Thus, it is necessary that data may be presented with the help of tables, diagrams and graphs. With these devices, facts can be understood easily.

4. Analysis: A major part of it is developed to the methods used in analyzing the presented data, mostly in a tabular form. Several statistical tools are available for this analysis, such as averages, correlation, regression, etc.

5. Interpretation: the interpretation of data is a difficult task and necessitates a high degree of skills and experience in the statistical investigation because certain decisions are made based on conclusions drawn.

Scope of Statistics:

In the early stages, the scope of statistics was very limited. It was mainly confined to government administration and was called the 'Science of Kings'. But in modern times, the scope of statistics has widened; usually, all those facts come in the purview of statistics, which are expressed in quantitative terms directly or indirectly. That is why Croxton & Cowden observed, "Today there is hardly a phase of endeavour which does not find statistical devices at least occasionally useful." It is not unfair to say, science without statistics bears no fruit and statistics without science has no root." The applications of statistics are so numerous that it is often remarked, "Statistics is what statisticians do." Now let us examine a few fields or areas in which statistics is applied.

1. Statistics and the State: in recent years, the functions of the state have increased tremendously. The state's concept has changed from simply maintaining law and order to that of a welfare state. Statistical data and statistical methods are of great help in promoting human welfare. Most countries' governments are the biggest collectors and users of statistical data. These statistics help in framing suitable policies.

2. Statistics in Business and Management: with growing size and increasing competition, the problems of business enterprises have become complex. Statistics is now considered an indispensable tool in analysing business, commerce and industry activities. The object can be achieved by properly conducting market surveys and research, which greatly depends on statistical methods. The trends in sales and production can be determined by statistical methods like time-series analysis which are essential for future planning of the phenomena. Statistical concepts and methods are also used in controlling the quality of products to the satisfaction of the consumer and the producer. The bankers use the objective analysis furnished by statistics and then temper their decisions based on qualitative information.

3. Statistics and Economics: R.A. Fisher complained of "the painful misapprehension that statistics is a branch of economics." Statistical Data and methods are of immense help in properly understanding economic problems and the information on economic policies. In the field of exchange, we study markets, the law of prices based on supply and demand, cost of production, banking and credit instruments etc. The development of various economic theories owes greatly to statistical methods, e.g., 'Engel's law of family expenditure', 'The Malthusian theory of population'. The impact of mathematics and statistics has led to the development of new disciplines like 'Econometrics'' and 'Economic statistics'. The concept of planning so vital for the growth of nations would not have been possible without data and proper statistical analysis.

4. Statistics and Psychology and Education: Statistics has been widely applied in psychology and education. Statistical methods are used to measure human ability, such as; intelligence, aptitude, personality, interest etc., by tests. The theory of learning is also based on Statistical Principles. Applications of statistics in psychology and education have led to the development of a new discipline called 'Psychometrics'.

5. Statistics and Natural science; Statistical techniques have proved to be extremely useful in the study of all-natural sciences like biology, medicine, meteorology, botany etc. for example- in diagnosing the correct disease, the doctor has to rely heavily on factual data like the temperature of the body, pulse rate, B.P. etc. In botany- the study of plant life- one must rely heavily on statistics in conducting experiments about the plants, the effect of temperature, soil type, etc. In agriculture- statistical techniques like 'analysis of variance' and 'design of experiments' are useful for isolating the role of manure, rainfall, watering process, seed quality etc. It isn't easy to find any scientific activity where statistical data and methods are not used.

6. Statistics and Physical Science: The physical sciences in which statistical methods were first developed and applied. It seems to be increasingly using statistics, especially in astronomy, chemistry, engineering, geology, meteorology and certain branches of physics.

7. Statistics and Research; statistics are indispensable in research work. Most of the advancement in knowledge has occurred because of experiments conducted with the help of statistical methods. Statistical methods also affect research in medicine and public health. There is hardly any research work today that one can find complete without statistical methods.

8. Statistics and Computer: The development of statistics has been closely related to the evolution of electronic computing machinery. Statistics is a form of data processing that converts data into information useful for decision-making. Computers can process large amounts of data quickly and accurately. This greatly benefits businesses and other organizations that must maintain records of their operations. Processing of row data is extensively required in applying many statistical techniques.

CLASSIFICATION OF STATISTICS:

Statistics can be divided into three parts;

- → **Descriptive statistics**
- → **Inferential statistics**
- → **Applied statistics**

1. Descriptive Statistics: Descriptive statistics is related to numerical data or facts. Such data are collected either by counting or by some other measurement process. It is also related to those methods, including editing of data, classification, tabulation, diagrammatic or graphical presentation, measures of central tendency, measures of dispersion, correlation etc., that help to make the description of numerical facts simple, systematic, synoptic understandable and meaningful.

2. Inferential Statistics: Inferential statistics help make generalizations about the population or universe based on the study of samples. It includes the process of drawing proper and rational conclusions about the universe. Among these methods, probability theory and different techniques of sampling tests are important.

3. Applied Statistics involves applying statistical methods and techniques to problems and facts. For example-statistics related to national income, industrial and agricultural production, population, price etc., are called applied statistics. It can be divided into two parts-(1) Descriptive Applied Statistics- which deals with the study of the known and related data. Its main objective is to provide descriptive information, either the past or present, for any area. For example- price index numbers and vital statistics come under descriptive applied statistics. (2)Scientific Applied Statistics- under this branch of statistical science, statistical methods are used to formulate and verify scientific laws. For example, an economist tries to establish the law of demand, quantitative theory of money, trade circle, etc. These are established and verified with the help of scientific applied statistics.

Functions of Statistics:

Statistics performs the function of making the numerical aspects of facts simple, precise, Comparable and reliable. The various functions performed by statistics are the basis of its utility. **R.W. Burgess** says, "The whole gospel of statistics is to push back the domain of ignorance, prejudice, rule of thumb, arbitrary and premature decisions, tradition & dogmatism and to increase the domain in which decisions are made. Principles are formulated based on analyzed quantitative facts."

1. Numerical and definite expression of facts: The first function of statistics is collecting and presenting facts in numerical form. We know that numerical presentation helps us better understand the nature of a problem. One of the most important functions of statistics is to present general statements in a precise and definite form. Statements and facts conveyed in exact quantitative terms are always more convincing than vague utterances.

2. Simplifies the data (condensation): It presents statistics facts in a definite form and helps condense the mass of data into a few significant figures. According to A.E. Waugh, **"**the purpose of a statistical method is to simplify great bodies of numerical data."The human mind cannot follow huge, complex and scattered numerical facts. So these facts are made simple and precise with the help of various statistical methods like averages, dispersion, graphic or diagrammatic, presentation, classification, tabulation etc., so that a common person understands them easily.

3. Comparison of facts: Paddington states, "The essence of the statistics is not only counting but also comparison." The function of comparison does help in showing the relative importance of data. For example- the pass % of examination results of a college may be appreciated better when compared with the result of other colleges or the results of previous years of the same college.

4. Establishment of relationship b/w two or more phenomena; investigating the relationship b/w two or more facts is the main function of statistics. For example-demand, the demand and supply of a certain commodity, prices and wages, temperature, and germination time of seeds are interrelated.

5. Enlarges individual experiences: In the words of Bowley**,** "the proper function of statistics indeed is to enlarge individual experience." Statistics is a master key to solving humanity's problems in every field. It would not be an exaggeration to say that many fields of knowledge would have remained closed to humanity forever but for the efficient and useful techniques and methodology of the science of statistics.

6. Helps in formulating policies: statistics help formulate policies in different fields, especially in economic, social and political fields. The government policies like industrial policy, export-import policy, taxation policy and monetary policy are determined based on statistical data and their movements, and plan targets are also fixed with the help of data.

7. Helps in forecasting: statistical methods provide helpful means of estimating the available facts and forecasting for the future. Here *Bowley'sstatement*is relevant: "a statistical estimate may be good or bad, accurate or the reverse; but in almost all cases, it is likely to be more accurate than a casual observer's impression."

8. Testing of hypothesis: statistical methods are also employed to test the hypothesis in theory and discover newer theories. For example, the statement that the average height of students in the college is 66 inches is a hypothesis. Here students of college constitute the population. It is possible to test this statement's validity by using statistical techniques.

Limitations of Statistics:

Newsholme states, "Statistics must be regarded as an instrument of research of great value but has several limitations which are not possible to overcome, and as such, they need careful attention."

1. Statistics does not study qualitative facts: Statistics means the aggregate of numerical facts. It means that in statistics, only those phenomena studied can be expressed in numerical terms directly or indirectly. Such as (1) directly in numerical terms like age, weight and income of the individual (2) not directly but indirectly like the intelligence of students and achievements of students (3) neither directly nor directly like morality, affection etc., such types of facts don't come under the scope of statistics.

2. Statistics doesn't study individuals: According to **W.I. King**, "Statistics from their very nature of the subject cannot and will never be able to take into account individual causes. When these are important, other means must be used for their study." These studies compared the general behaviour of the group at different points in time or the behaviour of different groups at a particular point in time.

3. Statistical results are true only on average: The statistical laws are not completely true and accurate like the law of physics. For example – the law of gravitational forces is perfectly true & universal, but statistical conclusions are not perfectly true. Such as, the average age of a person in India is 62 years. It does not mean that every person will attain this age. Based on statistical methods, we can say only in terms of probability and not certainty.

4. Statistics as lack of complete accuracy: According to Conner, "Statistical data must always be treated as approximations or estimates and not as precise measurements."Statistical result is based on sample or census data and is bound to be true only approximately. For example, according to the population census 2001, the country's population is 1,02,70,15,247, but the entire population may not be more or less a hundred, two hundred and so on.

5. Statistics is liable to be misused: Statistical deals with figures and can be easily manipulated and distorted by inexpert and unskilled persons. It is very much likely to be misused in most cases. In other words, the data should be handled by experts. Technically sound persons must use Thus it.

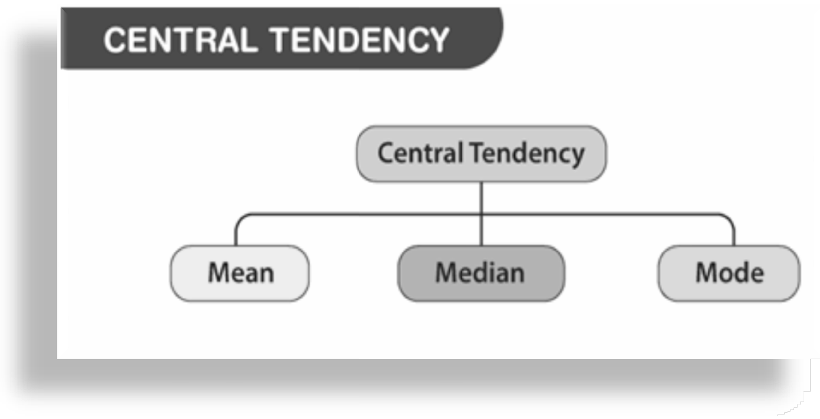
6. Statistics is only one of the methods of studying a phenomenon; According to Croxton& Cowden, "It must not be assumed that the statistical method is the only method to be used in research; neither this method is considered the best attack for every problem."The conclusions arrived at with the help of statistics must be supplemented with other evidence.

7. Statistical results may be misleading; Without any reference, statistical results may provide tentative conclusions. For example – based on the increasing no. of prisoners in prison, it may be concluded that crime is increasing. But it may be possible that due to the rude behaviour of the police administration, the number of prisoners is increasing, but crime is decreasing.

Therefore, it is worth mentioning that every science is based on certain assumptions and limitations. This does not reduce the subject's importance but emphasises that precautions should be taken while dealing with statistical analysis and interpretations.

Measures of Central Tendency:

The objective of central tendency is to find one representative value which can be used to locate and summarise the entire set of varying values. This value can be used to make many decisions concerning the entire set. We can define measures of central tendency (or location) to find some central value around which the data tend to cluster.



Significance of measures of central tendency:

Measures of central tendency, i.e. condensing the mass of data in one single value, enable us to get an idea of the real data. For example, it is impossible to remember the individual incomes of millions of earning people in India. But if the average income is obtained, we get one value representing the entire population. Measures of central tendency also enable us to compare two or more data sets to facilitate comparison. For example, the average sales figures for April may be compared with those of previous months.

Properties of a good measure of central tendency

[1] It should be easy to understand and calculate.

[2] It should be rigidly defined.

[3] It should be based on all observations.

[4] It should be least affected by sampling fluctuation.

[5] It should be capable of further algebraic treatment.

[6] It should be least affected by extreme values.

[7] It should be calculated in case of open-end interval.

Following are some important central tendency measures commonly used in business and industry.

- Arithmetic Mean
- Weighted Arithmetic Mean
- Median
- Quantiles (quartiles, deciles and percentiles)
- Mode
- Geometric Mean
- Harmonic Mean

Arithmetic Mean:

The arithmetic mean (or mean or average) is the most commonly used and readily understood measure of central tendency. In statistics, the term average refers to any measures of central tendency. Ungrouped data/Raw data arithmetic mean is equal to the sum of the numerical values of every observation divided by the total number of observations.

The mean is the sum of the value of each observation in a dataset divided by the number of observations. This is also known as the arithmetic average.

Looking at the retirement age distribution again:

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

The mean is calculated by adding all the values (54+54+54+55+56+57+57+58+58+60+60 = 623) and dividing by the number of observations (11), which equals 56.6 years.

Merits Of Mean

[1] It is easy to understand and calculate.

[2] It is rigidly defined.

[3] It is based on all observations.

[4] It is least affected by sampling fluctuation.

[5] It is capable of further algebraic treatment.

Demerits Of Mean

1) It is highly affected by extreme values.

2) It is not calculated in the case of an open-end interval.

Median:

A second measure of central tendency is the median. The median is that value which divides the distribution into two equal parts. Fifty per cent of the observations in the distribution are above the value of the median, and the other fifty per cent of the observations are below this value of the median. The median is the value of the middle observation when the series is arranged in order of size or magnitude

Looking at the retirement age distribution (which has 11 observations), the median is the middle value, which is 57 years:

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

When the distribution has an even number of observations, the median value is the mean of the two middle values. In the subsequent distribution, the two middle values are 56 and 57. Therefore the median equals 56.5 years:

Merits Of Median:

[1] It is easy to understand and calculate.

[2] It is rigidly defined.

[3] It is not affected by extreme values.

[4] It is calculated in the case of open-end intervals.

[5] It is located graphically, also

Demerits Of Median:

[1] It is not based on all observations.

[2] It is affected by sampling fluctuation.

[3] It is not capable of further algebraic treatment.

Mode:

The mode is the typical or commonly observed value in a data set. It is the value that occurs most often or with the greatest frequency. The dictionary meaning of the term mode is most usual. For example, in the series of numbers 3, 4, 5, 5, 6, 7, 8, 8, 8, 9, the mode is eight because it occurs the maximum number of times.

The mode is the most commonly occurring value in a distribution.

Consider this dataset showing the retirement age of 11 people in whole years:

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

This table shows a simple frequency distribution of the retirement age data.

Frequency distribution table

Table showing the frequency distribution of the above data

The most commonly occurring value is 54. Therefore, the mode of this distribution is 54 years.

Merits Of Mode

[1] It is easy to understand and calculate.

[2] It is not affected by extreme values.

[3] It is calculated in the case of open-end intervals.

[4] It is located graphically also.

Demerits Of Mode:

[1] It is not based on all observations.

[2] It is highly affected by sampling fluctuation.

[3] It is not capable of further algebraic treatment

[4] It is not rigidly defined.

Limitations of the mode

The are some limitations to using the mode. In some distributions, the mode may not reflect the centre of the distribution very well. When the distribution of retirement age is ordered from the lowest to the highest value, it is easy to see that the centre of the distribution is 57 years, but the mode is lower at 54 years.

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

It is also possible for there to be more than one mode for the same distribution of data (bi-modal or multi-modal). The presence of more than one mode can limit the ability of the mode to describe the centre or typical value of the distribution because a single value to describe the centre cannot be identified.

In some cases, particularly where the data are continuous, the distribution may have no mode (i.e. if all values are different).

In such cases, it may be better to consider using the median or mean or group the data into appropriate intervals and find the modal class.

Median

The median is the middle value in distribution when the values are arranged in ascending or descending order.

The median divides the distribution in half (50% of observations are on either side of the median value). In a distribution with an odd number of observations, the median value is the middle value.

Looking at the retirement age distribution (which has 11 observations), the median is the middle value, which is 57 years:

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

When the distribution has an even number of observations, the median value is the mean of the two middle values. In the subsequent distribution, the two middle values are 56 and 57. Therefore the median equals 56.5 years:

52, 54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

Advantage of the median

The median is less affected by outliers and skewed data than the mean and is usually the preferred measure of central tendency when the distribution is not symmetrical.

Limitation of the median

The median cannot be identified for categorical nominal data, as it cannot be logically ordered.

Mean

The mean is the sum of the value of each observation in a dataset divided by the number of observations. This is also known as the arithmetic average.

Looking at the retirement age distribution again:

54, 54, 54, 55, 56, 57, 57, 58, 58, 60, 60

The mean is calculated by adding all the values (54+54+54+55+56+57+57+58+58+60+60 = 623) and dividing by the number of observations (11), which equals 56.6 years.

Advantage of the mean

The mean can be used for both continuous and discrete numeric data.

Limitations of the mean

The mean cannot be calculated for categorical data, as the values cannot be summed.

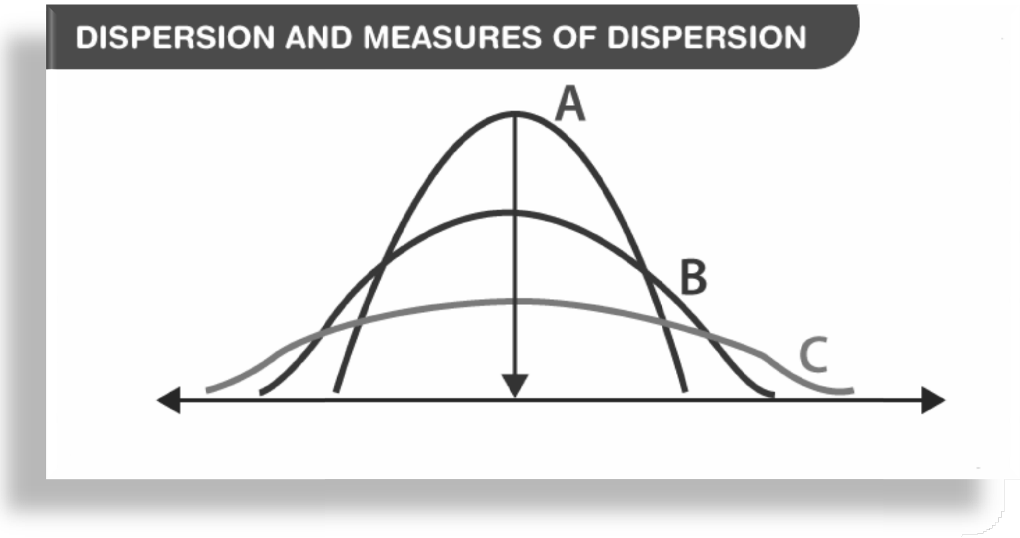
As the mean includes every value in the distribution, the mean is influenced by outliers and skewed distributions.

Another thing about the mean

The population mean is indicated by the Greek symbol μ (pronounced 'mu'). When the mean is calculated on a distribution from a sample, it is indicated by the symbol $\bar{x}$ (pronounced X-bar).

Dispersion

Dispersion is the state of getting dispersed or spread. Statistical dispersion means the extent to which numerical data is likely to vary about an average value. In other words, dispersion helps to understand the distribution of the data.



The field of statistics is used across every sector and industry to help people better understand and predict potential outcomes. In finance, investors often turn to statistics to understand how returns on certain assets, or groups of assets, could be distributed.

This range of possible investment returns is called dispersion. In other words, dispersion refers to the range of potential outcomes of investments based on historical volatility or returns.

There are two important ways to measure dispersion—alpha and beta—which calculate risk-adjusted returns and returns relative to a benchmark. Investors can understand the risk inherent in a particular security or investment portfolio by considering the dispersion of possible investment returns and values such as alpha and beta.

Key Takeaways

- ❖ Dispersion refers to the range of potential outcomes of investments based on historical volatility or returns.
- ❖ Dispersion can be measured using alpha and beta, which calculate risk-adjusted returns and returns relative to a benchmark index.
- ❖ Generally speaking, the higher the dispersion, the riskier an investment is, and vice versa.

Types of Measures of Dispersion

There are two main types of dispersion methods in statistics which are:

- ❖ Absolute Measure of Dispersion

- ❖ Relative Measure of Dispersion
- ❖ Absolute Measure of Dispersion

An absolute measure of dispersion contains the same unit as the original data set. The absolute dispersion method expresses the variations in terms of the average of observations, like standard or means deviations. It includes range, standard deviation, quartile deviation, etc.

The types of absolute measures of dispersion are:

Range: The difference between the maximum and minimum values given in a data set. Example: 1, 3,5, 6, 7 => Range = 7 -1= 6

Variance: Deduct the mean from each data set, square each of them, add each square, and divide them by the total number of values in the data set to get the variance. Variance ($\sigma 2$) = $\sum(X-\mu)2/N$

Standard Deviation: The square root of the variance is known as the standard deviation, i.e. S.D. = $\sqrt{\sigma}$.

Quartiles and Quartile Deviation: The quartiles are values that divide a list of numbers into quarters. The quartile deviation is half the distance between the third and the first quartile.

Mean and Mean Deviation: The average of numbers is known as the mean, and the arithmetic mean of the absolute deviations of the observations from a measure of central tendency is known as the mean deviation (also called mean absolute deviation)

Examples

Question 1: Find the mean of the following data set.

10, 20, 36, 12, 35, 40, 36, 30, 36, 40

Solution:

Given,

xi = 10, 20, 36, 12, 35, 40, 36, 30, 36, 40

n = 10

Mean = $\sum xi/n$ Or

= (10 + 20 + 36 + 12 + 35 + 40 + 36 + 30 + 36 + 40)/10

= 295/10

= 29.5

To find the median, place all the numbers in ascending order and find the middle.

Example 1:

Find the Median of 14, 63 and 55

Solution:

Could you put them in ascending order: 14, 55, 63?

The middle number is 55, so the median is 55.

Example 2:

Find the median of the following:

4, 17, 77, 25, 22, 23, 92, 82, 40, 24, 14, 12, 67, 23, 29

Solution:

When we put those numbers in the order, we have:

4, 12, 14, 17, 22, 23, 23, 24, 25, 29, 40, 67, 77, 82, 92,

There are fifteen numbers. Our middle is the eighth number:

The median value of this set of numbers is 24.

Example 3:

Rahul's family drove through 7 states on summer vacation. The prices of Gasoline differ from state to state. Calculate the median of gasoline cost.

1.79, 1.61, 2.09, 1.84, 1.96, 2.11, 1.75

Solution:

By organizing the data from smallest to greatest, we get the following:

1.61, 1.75, 1.79, 1.84, 1.96, 2.09, 2.11

Hence, the median gasoline cost is 1.84. There are three states with greater gasoline costs and 3 with smaller prices.

Therefore, the mean of the given data set is 29.5.

Ungrouped data

Example: The following table represents the number of wickets taken by a bowler in 10 matches. Find the mode of the given set of data.

Match No.	1	2	3	4	5	6	7	8	9	10
No. of Wickets	2	1	1	3	2	3	2	2	4	1

It can be seen that two wickets were taken by the bowler frequently in different matches. Hence, the mode of the given data is 2.

Grouped data

Example 1: A survey conducted on 20 households in a locality by a group of students resulted in the following frequency table for the number of family members in a household:

Family Size	1 – 3	3 – 5	5 – 7	7 – 9	9 – 11
Number of Families	7	8	2	2	1

Find the mode for the above data.

Solution 1: Here, the maximum class frequency is 8, and the class corresponding to this frequency is 3 – 5. So, the modal class is 3 – 5.

Now

– Modal class = 3 – 5, lower limit (l) of modal class = 3, class size (h) = 2
– Frequency (f_1) of the modal class = 8, frequency
– (f_0) of class preceding the modal class = 7,
– Frequency (f_2) of class succeeding the modal class = 2.

Putting the values in the formula:

Mode = l +f1−f02f1−f0−f2×h�1−�02�1−�0−�2×*h*

Mode = 3 +8−72×8−7−2×28−72×8−7−2×2 = 3.286.

Correlations and linear regression

Correlation and regression are the two most commonly used techniques for investigating the relationship between quantitative variables. Here regression refers to linear regression. Correlation gives the relationship between the variables, whereas linear regression uses an equation to express this relationship. That is, Correlation and regression are statistical measurements that are used to give a relationship between two variables. For example, suppose a person is driving an expensive car. Then it is assumed that she must be financially well. To numerically quantify this relationship, correlation and regression are used.

Correlations

Correlation can be defined as a measurement that is used to quantify the relationship between variables. If an increase (or decrease) in one variable causes a corresponding increase (or decrease) in another, then the two variables are directly correlated. Similarly, if an increase in one causes a decrease in another or vice versa, then the variables are said to be indirectly correlated.

Regressions

Regression can be defined as a measurement used to quantify how the change in one variable will affect another variable. Regression is used to find the cause and effect between two variables. Linear regression is the most commonly used type because it is easier to analyze than the rest. Linear regression finds the best-fit line to establish a relationship between variables.

Correlations and Regression

Question 1: Calculate the linear correlation coefficient for the following data. X = 4, 8,12, 16 and Y = 5, 10, 15, 20.

Solution:

Given variables are,

X = 4, 8 ,12, 16 and Y = 5, 10, 15, 20

To find the linear coefficient of these data, we first need to construct a table for the required values.

x	y	x^2	y^2	XY
4	5	16	25	20
8	10	64	100	80
12	15	144	225	180
16	20	256	400	320

Σ x = 40	Σ y =50	480	750	600

According to the formula of linear correlation, we have,

$$r(xy) = \frac{(4\times600)-(40\times50)}{\sqrt{4(480)-40^2}\sqrt{4(750)-50^2}}$$

$$r(xy) = \frac{2400-2000}{\sqrt{1920-1600}\sqrt{3000-2500}}$$

$$r(xy) = \frac{400}{\sqrt{320}\sqrt{500}}$$

$$r(xy) = \frac{400}{17.89\times22.36}$$

$$r(xy) = \frac{400}{400} = 1$$

Therefore, r(xy) = 1

Problem

Calculate the regression coefficient and obtain the lines of regression for the following data

X	1	2	3	4	5	6	7
Y	9	8	10	12	11	13	14

Solution:

X	Y	X^2	Y^2	X^Y
1	9	1	81	9
2	8	4	64	16
3	10	9	100	30
4	12	16	144	48
5	11	25	121	55
6	13	36	169	78
7	14	49	196	98
$\sum X = 28$	$\sum Y = 77$	$\sum X^2 = 140$	$\sum Y^2 = 875$	$\sum XY = 334$

$\overline{X} = \frac{\Sigma X}{N} = \frac{28}{7} = 4$,

$\overline{Y} = \frac{\Sigma Y}{N} = \frac{77}{7} = 11$

The regression coefficient of *X* on *Y*

$$b_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\,\Sigma Y^2 - (\Sigma Y)^2}$$

$$= \frac{7(334) - (28)(77)}{7(875) - (77)^2}$$

$$= \frac{2338 - 2156}{6125 - 5929}$$

$$= \frac{182}{196}$$

$$b_{xy} = 0.929$$

(i) Regression equation of *X* on *Y*

$$X - \overline{X} = b_{xy}(Y - \overline{Y})$$

$$X - 4 = 0.929(Y - 11)$$

$$X - 4 = 0.929Y - 10.219$$

$\therefore$ The regression equation X on Y is $X = 0.929Y - 6.219$

(ii) Regression coefficient *of Y* on *X*

$$b_{yx} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{N\,\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{7(334) - (28)(77)}{7(140) - (28)^2}$$

$$= \frac{2338 - 2156}{980 - 784}$$

$$= \frac{182}{196}$$

$$\therefore \quad b_{yx} = 0.929$$

(iii) Regression equation of *Y* on *X*

$$Y-\overline{Y} = b_{yx}(X-\overline{X})$$

$$Y-11 = 0.929\,(X-4)$$

$Y = 0.929X - 3.716 + 11$

$= 0.929X + 7.284$

The regression equation of Y on X is $Y = 0.929X + 7.284$

Problem 2:

Calculate the two regression equations of X on Y and Y on X from the data given below, taking deviations from an actual means of X and Y.

Price(Rs.)	10	12	13	12	16	15
Amount demanded	40	38	43	45	37	43

Estimate the likely demand when the price is Rs.20.

Solution: Calculation of Regression equation

X	$x=(X-13)$	x^2	Y	$y=(Y-41)$	y^2	xy
10	−3	9	40	−1	1	3
12	−1	1	38	−3	9	3
13	0	0	43	2	4	0
12	−1	1	45	4	16	−4
16	3	9	37	−4	16	−12
15	2	4	43	2	4	4
$\sum X = 78$	$\sum x = 0$	$\sum x^2 = 24$	$\sum Y = 246$	$\sum y = 0$	$\sum y^2 = 50$	$\sum xy = -6$

(i) Regression equation of *X* on *Y*

$$X-\overline{X} = r\,\frac{\sigma_x}{\sigma_y}(Y-\overline{Y})$$

$$\overline{X} = \frac{78}{6} = 13\,,\ \overline{Y} = \frac{246}{6} = 41$$

$$b_{xy} = r\,\frac{\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma y^2} = \frac{-6}{50} = -0.12$$

$$X-13 = -0.12\,(Y-41)$$

$$X-13 = -0.12Y+4.92$$

$$X = -0.12Y+17.92$$

(ii) Regression Equation of Y on X

$$Y-\overline{Y} = r\,\frac{\sigma_y}{\sigma_x}(X-\overline{X})$$

$$b_{yx} = r\,\frac{\sigma_y}{\sigma_x} = \frac{\Sigma xy}{\Sigma x^2} = -\frac{6}{24} = -0.25$$

$$Y-41 = -0.25\,(X-13)$$

$$Y-41 = -0.25\,X+3.25$$

$$Y = -0.25\,X+44.25$$

When *X* is 20, *Y* will be

$= -0.25\ (20)+44.25$

$= -5+44.25$

$= 39.25$ (when the price is Rs. 20, the likely demand is 39.25)

Chi-Square test

The Chi-Square test is a statistical procedure for determining the difference between observed and expected data. This test can also be used to determine whether it correlates to the categorical variables in our data. It helps to determine whether a difference between two categorical variables is due to chance or a relationship between them.

Main Purpose :

It is to identify whether a disparity between actual and predicted data is due to chance or to a link between the variables under consideration.

Usage:

- The Chi-squared test can be used to see if your data follows a well-known theoretical probability distribution like the Normal or Poisson distribution.
- The Chi-squared test allows you to assess your trained regression model's goodness of fit on the training, validation, and test data sets.

Two types

1. Independence
2. Goodness-Of-Fit

Independence

The Chi-Square Test of Independence is a derivable (also known as inferential) statistical test which examines whether the two sets of variables are likely to be related to each other or not. This test is used when we have counts of values for two nominal or categorical variables and is considered a non-parametric test. The required criteria for conducting this test are a relatively large sample size and observation independence.

E.g.:

In a movie theatre, suppose we made a list of movie genres. Let us consider this as the first variable. The second variable is whether or not the people who came to watch those genres of movies bought snacks at the theatre. Here the null hypothesis is that the genre of the film and whether people bought snacks are unrelatable. If this is true, the movie genres don’t impact snack sales.

Goodness-Of-Fit

In statistical hypothesis testing, the Chi-Square Goodness-of-Fit test determines whether a variable is likely to come from a given distribution. We must have a set of data values and an idea of the distribution of this data. We can use this test when we have value counts for categorical variables. This test demonstrates a way of deciding if the data values are a " good enough" fit for our idea or if it is representative sample data from the entire population.

E.g.:

Suppose we have bags of balls with five different colours in each bag. The given condition is that the bag should contain an equal number of balls of each colour. The idea we would like to test here is that the proportions of the five colours of balls in each bag must be exact.

Example

Let's say you want to know if gender has anything to do with political party preference. You poll 440 voters in a simple random sample to determine which political party they prefer. The results of the survey are shown in the table below:

	Republican	Democrat	Independent	Total
Male	100	70	30	200
Female	140	60	20	220
Total	240	130	50	440

To see if gender is linked to political party preference, perform a Chi-Square test of independence using the steps below.

Step 1: Define the Hypothesis

H0: There is no link between gender and political party preference.

H1: There is a link between gender and political party preference.

Step 2: Calculate the Expected Values

Now you will calculate the expected frequency.

$$\text{Expected Value} = \frac{(Row\ Total) * (Column\ Total)}{Total\ Number\ Of\ Observations}$$

For example, the expected value for Male Republicans is:

$$= \frac{(240) * (200)}{440} = 109$$

Similarly, you can calculate the expected value for each of the cells.

Expected Values				
	Republican	**Democrat**	**Independent**	**Total**
Male	109	59	22.72	200
Female	120	65	25	220
Total	240	130	50	440

Step 3: Calculate (O-E)2 / E for Each Cell in the Table

Now you will calculate the (O - E)2 / E for each cell in the table.

Where

O = Observed Value

E = Expected Value

$(O - E)^2/E$				
	Republican	Democrat	Independent	Total
Male	0.74311927	2.050847	2.332676056	200
Female	3.33333333	0.384615	1	220
Total	240	130	50	440

Step 4: Calculate the Test Statistic X2

X2 is the sum of all the values in the last table

= 0.743 + 2.05 + 2.33 + 3.33 + 0.384 + 1

= 9.837

Before you can conclude, you must first determine the critical statistic, which requires determining our degrees of freedom. In this case, the degrees of freedom is equal to the table's number of columns minus one multiplied by the table's number of rows minus one, or (r-1) (c-1). We have (3-1)(2-1) = 2.

Finally, you compare our obtained statistic to the critical statistic in the chi-square table. As you can see, for an alpha level of 0.05 and two degrees of freedom, the critical statistic is 5.991, less than our obtained statistic of 9.83. You can reject our null hypothesis because the critical statistic is higher than your obtained statistic.

This means you have sufficient evidence to say there is an association between gender and political party preference.

Critical values of the Chi-square distribution with *d* degrees of freedom

	Probability of exceeding the critical value						
d	0.05	0.01	0.001	*d*	0.05	0.01	0.001
1	3.841	6.635	10.828	11	19.675	24.725	31.264
2	5.991	9.210	13.816	12	21.026	26.217	32.910
3	7.815	11.345	16.266	13	22.362	27.688	34.528
4	9.488	13.277	18.467	14	23.685	29.141	36.123
5	11.070	15.086	20.515	15	24.996	30.578	37.697
6	12.592	16.812	22.458	16	26.296	32.000	39.252
7	14.067	18.475	24.322	17	27.587	33.409	40.790
8	15.507	20.090	26.125	18	28.869	34.805	42.312
9	16.919	21.666	27.877	19	30.144	36.191	43.820
10	18.307	23.209	29.588	20	31.410	37.566	45.315

INTRODUCTION TO POPULATION GENETICS, Table D.1

t-test

The t-test is a test mainly used to compare the mean of two groups of samples. It is meant to evaluate whether the means of the two data sets are statistically significantly different from each other.

There are many types of t-tests. Some of these are:

The one-sample t-test is used to compare the mean of a population with a theoretical value.

The unpaired two-sample t-test compares the mean of two independent given samples.

The paired t-test compares the means between two groups of related samples.

T-test Formula

The T-test formula is given below:

t = x1¯−x2¯(√s21n1+s22n2)

Where,

t	t-test value
x1¯	Mean of the first set of values
x2¯	The mean of the second set of values
s1	The standard deviation of the first set of values
s2	The standard deviation of the second set of values
n1	Total number of values in the first set
n2	Total number of values in the second set.

Also,

The formula for standard deviation is given below:

s=(√∑(x−x¯)2n−1)

Where,

s	The standard deviation for a data set
x	Values are given in the data set
$\bar{x}$	The mean value of the data set
n	Total number of values in the data set

Solved Examples

Q.1: Find the t-test value for the following given two sets of values:

7, 2, 9, 8 and

1, 2, 3, 4?

Solution: For the first data set:

Number of terms in the first set, i.e. n_1 = 4

Calculate the mean value for the first data set using the formula:

$\bar{x}1=\sum x1n1$
i.e. $\bar{x}1=7+2+9+84$
i.e. $\bar{x}1=6.5$

Construct the following table for standard deviation:

x1	$x1-\bar{x}1$	$(x1-\bar{x}1)2$
7	0.5	0.25

2	-4.5	20.25
9	2.5	6.25
8	1.5	2.25

Thus , ∑((x1−x¯1)2)=29

Now, compute the standard deviation using the formula,

s1=(√∑(x1−x¯1)2n1−1)
i.e. s1=(√294−1)
i.e. s1=(√9.66)
s1=3.11

Therefore, the standard deviation for the first set of data: s_1 = 3.11

For the second data set:

Number of terms in the second set, i.e. n2=4

Calculate the mean value for the second data set using the formula:

x¯2=∑x2n2
i.e. x¯2=1+2+3+44
i.e. x¯2=2.5

Construct the following table for standard deviation:

x2	x2−x¯2	(x2−x¯2)2
1	-1.5	2.25
2	-0.5	0.25
3	0.5	0.25

4	1.5	2.25

Thus, ∑((x2−x¯2)2)=5

Now, compute the standard deviation using the formula,

s2=(√∑(x2−x¯2)2n2−1)
i.e. s2=(√54−1)
i.e. s1=(√1.66)
s1=1.29
Therefore, the standard deviation for the second set of data: is s2=1.29

Now, apply the formula for the t-test value:

t=x1¯−x2¯(√s21n1+s22n2)
t=6.5–2.5(√3.1124+1.2924)
=4(√9.36674+1.6674)

t = 2.38

Hence t-test value for the two data sets is = 2.38

z-test

A z-test is conducted on a population with a normal distribution with independent data points and a sample size greater than or equal to 30. It checks whether the means of two populations are equal when the population variance is known. The null hypothesis of a z-test can be rejected if the z-test statistic is statistically significant compared to the critical value.

A Z test is a form of inferential statistics. It uses samples to conclude populations.

z-test is a statistically significant test for Hypothesis Testing

There are three steps in Hypothesis Testing:

1. State Null and Alternate Hypothesis
2. Perform Statistical Test
3. Accept and reject the Null Hypothesis

Usage of z-test

- Population variance is unknown
- The sample size is greater than 30

Importance

- Z test is a statistical test conducted on normally distributed data to check if there is a difference in means between two data sets.
- The sample size should be greater than 30, and the population variance must be known to perform a z-test.
- The one-sample z-test checks if there is a difference in the sample and population mean,
- The two-sample z-test checks whether the means of the two groups are equal.

Formula

A one-sample z-test is used to check if there is a difference between the sample mean and the population mean when the population standard deviation is known. The formula for the z-test statistic is given as follows:

$$z = \frac{\overline{x} - \mu}{\frac{\sigma}{\sqrt{n}}} . \ \overline{x}$$

Is the sample mean, μ� is the population mean, σ� is the population standard deviation, and n is the sample size.

A two-sample z-test is used to check if there is a difference between the means of the two samples. The z-test statistic formula is given as follows:

$$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} . \ \overline{x_1}\ , \mu_1, \sigma_1^2$$

are the sample mean, the population mean and the population variance, respectively, for the first sample. ‾‾‾x2�2‾, μ2�2, σ22�22 are the sample mean, the population mean and the population variance, respectively, for the second sample.

f-test

An “F Test” is a catch-all term for the tests using the F-distribution. When people talk about the F-Test, they are talking about the F-Test Compare Two Variances. But the f-statistic is used in various tests such as regression analysis, the Chow test, and the Scheffe test. If running an F-Test, we may use many kinds of technology to run the test. Because doing an F-test by hand, including variances, is a complex and time-consuming task.

If we are using an F Test using technology, the following steps are there:

- State the null hypothesis with the alternate hypothesis.
- Calculate the F-value using the formula.
- Find the F Statistic, which is the critical value for this test. This F-statistic formula is the ratio of the variance of the group means divided by the mean of the within-group variances.
- Finally, support or reject the Null Hypothesis.

A Statistical F Test uses an F Statistic to compare two variances, σ1andσ2, by dividing them. The result will always be a positive number because variances are always positive. Thus, the equation for comparing two variances with the F-test is:

$$F - value = \frac{variance\ 1}{variance\ 2}$$

$$\text{i.e. } F - value = \frac{\sigma_1^2}{\sigma_2^2}$$

The F-test formula is used to compare the variances of two sets of values. Applied to F distribution under the null hypothesis, first, we have to compute the mean of two given observations and then calculate their variance.

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

σ2	Variance
X	Values are given in a set of data
X¯	Mean of the data
n	The total number of values.

We always test that the population variances are equal while running an F-Test. In other words, we always assume that the variances are equal to 1. Thus, our null hypothesis will always be that the variances are equal.

Solved Examples for F Test Formula

Q.1: Conduct an F-Test on the following samples:

Sample-1 having variance = 109.63, sample size = 41.

Sample-2 having Variance = 65.99, sample size = 21.

Solution:

Step-1:- First, write the hypothesis statements as:

H_0: No difference in variances.

H_a: Difference in variances.

Step-2:- Calculate the F-critical value. Here take the highest variance as the numerator and the lowest variance as the denominator:

FValue=σ21σ22

FValue=109.6365.99

FValue=1.66

Step-3:- Calculate the degrees of freedom as:

The degrees of freedom in the table will be the sample size -1, so for sample-1, it is 40, and for sample-2, it is 20.

Step-4:- Choose the alpha level. As no alpha level was given in the question, we may use the standard level of 0.05. This needs to be halved for the test, so use 0.025.

Step-5:- We will find the critical F-Value using the F-Table. We will use the table with 0.025. Critical-F for (40,20) at alpha (0.025) is 2.287.

Step-6:- Compare the calculated value to the common table value. We may reject the null hypothesis if our calculated value is higher than the table value. Here, 1.66 < 2 .287. So, we cannot reject the null hypothesis.

Presentation of Data:

1. Tabular

   A tabular presentation of data helps the viewer understand and interpret the information better. Take, for example, your annual report card that is presented in a tabular format. You have your subjects written in one column of the table and your grades in the other. The third column mentions any teachers' remarks. A glance at your report card lets you easily read through the grades, subjects, and remarks.

Now think about what would have happened if the same information had been presented as a paragraph. You would have to go through each line to know your grade and the teachers' remarks on a particular subject. This would make it tedious and also confusing to understand the report card.

Presentation of Data

Data must be presented properly. If the information pleases the eyes, it immediately gets attention. Data presentation is about using the same information to exhibit it in an attractive and useful way that can be read and interpreted easily. Data presentation is of three broad kinds. These are:

- Textual presentation.
- Data tables.
- Diagrammatic presentation.

On this presentation of data Class 11 page, you will understand the textual and tabular data presentation or the data tables.

Textual Presentation

Data is first obtained in a textual format. It is a vague and raw format of the data. The data is mentioned in text form, usually written in a paragraph. The textual presentation of data is used when the data is not large and can be easily comprehended by the reader just when he reads the paragraph.

This data format is useful when some qualitative statement is to be supplemented with data. The reader does not want to read volumes of data to be represented in a tabular format. Does he want to understand the data in a diagrammatic form? All that the reader wants to know is the data that provides evidence for the statement written. This is enough to let the reader gauge the intensity of the statement.

The textual data is evidence of the qualitative statement, and one needs to go through the complete text before he concludes anything.

For example, the coronavirus death toll in India today is 447. The reader does not need a lot of data here. The entire text of the state-wise breakup is accumulated to arrive at the national death figure. This is enough information for the reader.

Data Tables or Tabular Presentation

Data Tables or Tabular presentation of data is the arrangement of certain values recorded in tables that are easy to manage and read. It is mostly done for a reader to understand the data without making it too complicated. The data presentation can be used for a proper matter, which is informative and creative at the same time.

What is Data Presentation?

If the reader has to interpret a lot of data, then this has to be organized in an easy-to-read format. The data should be laid out in rows and columns so the reader can get what he wants at a glance. Data tables are easy to construct and also easy to read, which makes them popular.

Components of Data Tables

Below are the key components of the data table.

- Table Number - Each table has a table number that makes it easy to locate. This number serves as a reference and leads one to a particular table.
- Title - The table should also have a title that lets the reader understand what information the table provides. The place of study, the period, and the nature of data classification are also mentioned in the title.
- Headnotes - The headnotes give further information. It provides the data unit in brackets mentioned at the end of the title. The headnote aids the title to offer the reader more information to interpret the data.
- Stubs - These are the titles that tell you what the row represents. In other words, the stubs give information about the data contained in each row.
- Caption - The caption is the column title in the data table. It gives information about what is contained in each column.
- Body or Field - The body or the field is the entire content of the table. Each item that is present in the body is the cell.
- Footnotes - Footnotes are not commonly used, but these are used to supplement the table title if needed.

- Source - If the data used in the table is taken from a secondary source, that must be mentioned in the footnote.

Construction of Data Tables

- Tabular can be constructed in many ways. Here are some ways that are commonly followed.
- The table's title should be able to reflect the table content.
- If two rows or columns have to be compared, these should be placed adjacent.
- If the rows in the table are lengthy, then the stub can be placed on the right-hand part of the table.
- Headings should always be in the singular.
- Footnotes are not compulsory and should be provided only if required.
- The column size should be symmetrical and uniform.
- There should be no abbreviations in the headings and subheadings.
- The units should be specified above the column.

The Advantages of Tabular Presentation

- Makes the representation of data easy.
- Makes it easy to analyze the data.
- Makes it easy to compare data.
- The data is represented in a readable manner which saves space and the reader's time.

Classification of Data and Tabular Presentation

Classification of data and Tabular presentation is needed to arrange complex, heterogeneous data in a more simple and sophisticated manner. This is done for the audience's convenience in studying the data, making the values easy to distinguish. One can classify the data and Tabular presentation in four ways. These are as follows.

Qualitative Classification

In qualitative classification, the data is classified based on its qualitative attributes. This is when the data has attributes that cannot be quantified. These could be boys-girls, rural-urban, etc.

Quantitative Classification

In quantitative classification, the data is classified based on quantitative attributes. These could be marks where the data is categorized into 0-50, 51-100, etc.

Temporal Classification

In this tabular presentation, the data is classified according to the time. The data is represented in various time frames, such as 2016, 2018, etc.

Spatial Classification

In this method of classification, the data is classified according to location, like India, Pakistan, Russia, etc.

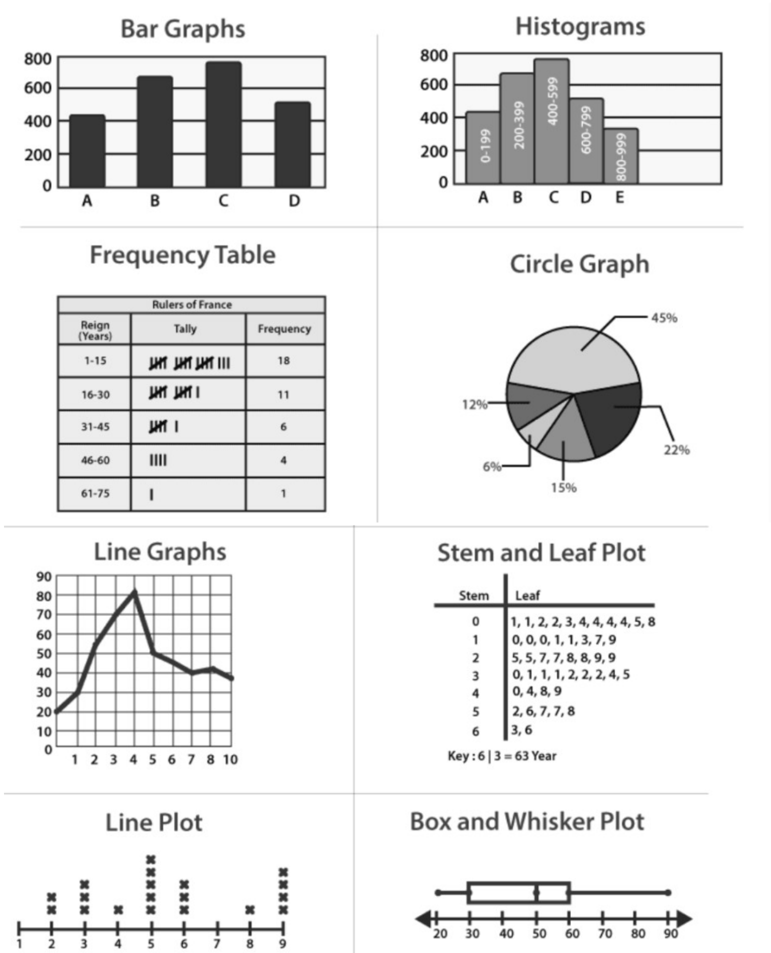
2. Graphic

Graphical Representation is a way of analysing numerical data. It exhibits the relation between data, ideas, information and concepts in a diagram. It is easy to understand and is one of the most important learning strategies. It always depends on the type of information in a particular domain. There are different types of graphical representation. Some of them are as follows:

- **Line Graphs** – A linear graph displays continuous data and is useful for predicting future events over time.
- **Bar Graphs** – Bar Graph displays the data category and compares the data using solid bars to represent the quantities.
- **Histograms** – The graph that uses bars to represent the frequency of numerical data organised into intervals. Since all the intervals are equal and continuous, all the bars have the same width.
- **Line Plot –** It shows the frequency of data on a given number line. ‘ x ‘ is placed above a number line each time when that data occurs again.
- **Frequency Table** – The table shows the number of data pieces that fall within the given interval.
- **Circle Graph** – Also known as the pie chart that shows the relationships of the parts of the whole. The circle is considered 100% and the categories occupied are represented with that specific percentage like 15%, 56%, etc.

- **Stem and Leaf Plot –** The data are organised from the least value to the greatest value in the stem and leaf plot. The digits of the least place values from the leaves and the next place value digit forms the stems.
- **Box and Whisker Plot** – The plot diagram summarises the data by dividing it into four parts. The box and whisker show the range (spread) and the middle (median) of the data.

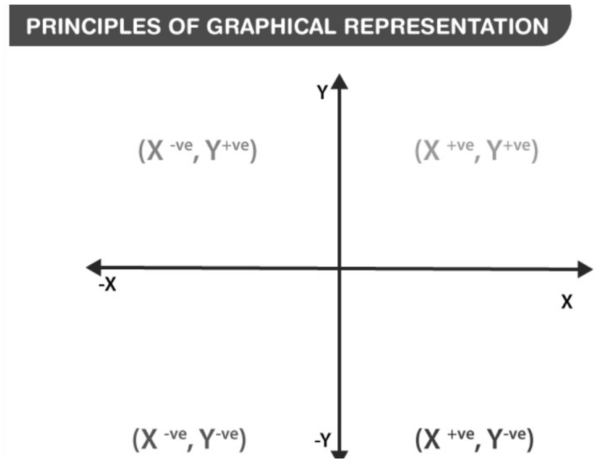
Types of graphical presentation



General Rules for Graphical Representation of Data

There are certain rules to present the information in the graphical representation effectively. They are:

- **Suitable Title:** Make sure that the appropriate title is given to the graph, indicating the presentation's subject.
- **Measurement Unit:** Mention the measurement unit in the graph.
- **Proper Scale:** To represent the data accurately, choose a proper scale.
- **Index:** Index the appropriate colours, shades, lines, and designs in the graphs for better understanding.
- **Data Sources:** Include the source of information wherever it is necessary at the bottom of the graph.
- **Keep it Simple:** Construct a graph easily so everyone can understand.
- **Neat:** Choose the correct size, fonts, colours etc., in such a way that the graph should be a visual aid for the presentation of information.



Problem :

Draw the frequency polygon for the following data

Class Interval	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
Frequency	4	6	8	10	12	14	7	5

Solution :

Mark the class interval along the x-axis and frequencies along the y-axis.

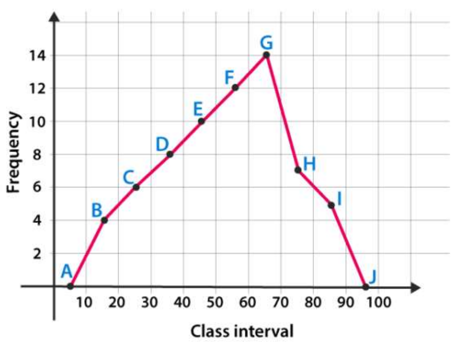
Let's assume that class intervals 0-10 with frequency zero and 90-100 with frequency zero.

Now calculate the midpoint of the class interval.

Class Intervals	Midpoints	Frequency
0-10	5	0
10-20	15	4
20-30	25	6
30-40	35	8
40-50	45	10
50-60	55	12
60-70	65	14
70-80	75	7
80-90	85	5
90-100	95	0

Using the midpoint and the frequency value from the above table, plot the points A (5, 0), B (15, 4), C (25, 6), D (35, 8), E (45, 10), F (55, 12), G (65, 14), H (75, 7), I (85, 5) and J (95, 0).

To obtain the frequency polygon ABCDEFGHIJ, draw the line segments AB, BC, CD, DE, EF, FG, GH, HI, and IJ, and connect all the points.



3. **Bar Diagram**

   The pictorial representation of grouped data in the form of vertical or horizontal rectangular bars, where the lengths of the bars are equivalent to the measure of data, are known as bar graphs or bar charts.

   The bars drawn are of uniform width, and the variable quantity is represented on one of the axes. Also, the measure of the variable is depicted on the other axes. The bars' heights or lengths denote the variable's value, and these graphs are also used to compare certain quantities. The frequency distribution tables can be easily represented using bar charts which simplify the calculations and understanding of data.

   The three major attributes of bar graphs are:

   The bar graph helps easily compare the different data sets among different groups.
   It shows the relationship using two axes, in which the categories are on one axis and the discrete values are on the other.
   The graph shows the major changes in data over time.

Types of Bar Graphs

The bar graphs can be vertical or horizontal. The primary feature of any bar graph is its length or height. If the length of the bar graph is more, then the values are greater than any given data.

Bar graphs normally show categorical and numeric variables arranged in class intervals. They consist of an axis and a series of labelled horizontal or vertical bars. The bars represent frequencies of distinctive values of a variable or, commonly, the distinct values themselves. The scale is called the number of values on the x-axis of a bar graph or the y-axis of a column graph.

The types of bar charts are as follows:

1. Vertical bar chart
2. Horizontal bar chart

Even though the graph can be plotted horizontally or vertically, the most common type of bar graph used is the vertical bar graph. The orientation of the x-axis and y-axis are changed depending on the vertical and horizontal bar chart. Apart from the vertical and horizontal bar graphs, the two different types of bar charts are:

- Grouped Bar Graph
- Stacked Bar Graph

Now, let us discuss the four different types of bar graphs.

Vertical Bar Graphs

When the grouped data are represented vertically in a graph or chart with the help of bars, where the bars denote the measure of data, such graphs are called vertical bar graphs. The data is represented along the graph's y-axis, and the height of the bars shows the values.

Horizontal Bar Graphs

When the grouped data are represented horizontally in a chart with the help of bars, then such graphs are called horizontal bar graphs, where the bars show the measure of data. The data is depicted here along the graph's x-axis, and the bars' length denotes the values.

Properties of Bar Graph

Some of the important properties of a bar graph are as follows:

- All the bars should have a common base.
- Each column in the bar graph should have equal width.
- The height of the bar should correspond to the data value.
- The distance between each bar should be the same.

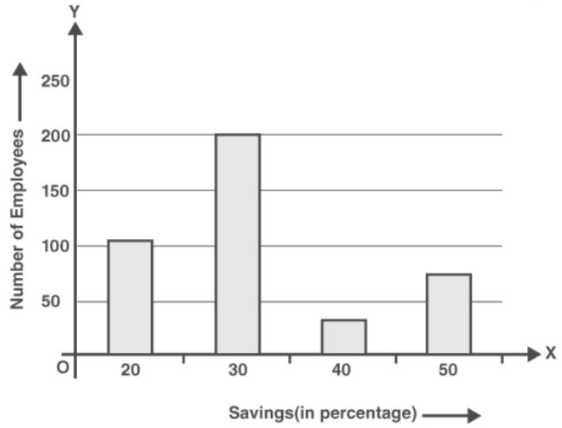
Example 1:

- In a firm of 400 employees, the percentage of monthly salary saved by each employee is given in the following table. Represent it through a bar graph.

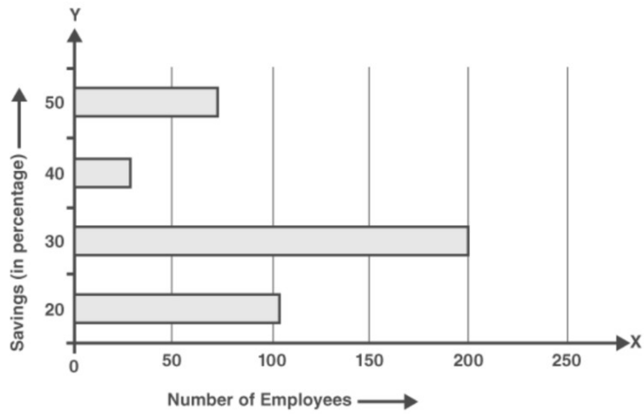
Savings (in percentage)	Number of Employees(Frequency)
20	105
30	199
40	29
50	73
Total	400

Solution:

- The given data can be represented as



This can also be represented using a horizontal bar graph as follows:



4. **Pie Chart**

A **pie chart** is a type of graph in which a circle is divided into sectors representing a proportion of the whole. Pie charts are a useful way to organize data to see the size of components relative to the whole, and they are particularly good at showing percentage or proportional data. While pie charts are popular data representations, they can be hard to read, and comparing data from one pie chart to another cannot be easy. Pie charts are a useful way to visualize information that might be presented in a small table.

Pie Chart Example

Let us look at the following example of the following pie chart that represents the ingredients used to prepare a butter cake.

Example: The whole pie represents a value of 100. It is divided into ten slices or sectors. The various colours represent the ingredients used to prepare the cake. What would be the exact quantity of each ingredient in specific colours in the following pie chart?

Solution: As we can see, the pie is divided into ten slices or sectors. To calculate the number of ingredients added to the cake, we divide the whole sector's value, i.e., 100, by the number of sectors. So, 100 ÷ 10 = 10. Hence, looking at the colour divisions made in the pie chart, we can conclude that:

Quantity of Flour	30
Quantity of Sugar	20
Quantity of Egg	40
Quantity of Butter	10

Pie Chart Formula

We know that the total value of the pie is always 100%. It is also known that a circle subtends an angle of 360°. Hence, the total of all the data is equal to 360°. Based on these, there are two main formulas used in pie charts:

- To calculate the percentage of the given data, we use the formula: (Frequency ÷ Total Frequency) × 100
- To convert the data into degrees, we use the formula: (Given Data ÷ Total value of Data) × 360°

We can work out the percentage for a given pie chart using the steps given below,

- Categorize the given data and calculate the total
- Divide the different categories
- Convert the data into percentages
- Calculate the degrees

RESEARCH REPORT

MEANING

A report is a detailed description of what has been done and how that has been done concerning a particular area or topic. A research report is a presentation of findings in the form of a report. It is a necessary part of the research process. It is the oral or written presentation of evidence. Research report writing is the culmination of the research investigation. Reporting is the end product of research activity. (Indeed, its practical application will follow."

Need/purpose of the research report

1. It helps to communicate the methodology and study results to the interested person. 2. It serves as a means for presenting the problem studied, methods and techniques used for collecting and analysing data findings, conclusions and recommendations in an organised manner. Thus, it helps to evaluate the researcher's ability and competence to research 3. It serves as a basic reference material for future use m developing research proposals in the same or relevant area 4. It serves to judge the quality of the completed research project 5. It provides a solid base for

formulating policies and strategies relating to the subject matter studied 6. It provides systematic knowledge of problems and issues analysed

Types of Reports

Research reports may be classified into two types-a& oral and written reports

A.Oral Report

In this, the researcher uses spoken words for communicating his study, e.g., in seminars, conferences etc. It helps to have two-way communication between the researcher and the audience. However, no permanent record concerning the research details is available.

B.Written Report

In this, the researcher uses written words to present his study. Written reports are of six types

1. Technical Report/Thesis

This is a full comprehensive report of the research process. It is primarily meant for the academic community, i.e., the scientists and other researchers. It is a long formal report covering all the aspects of the research process. The problem studied, the objectives of the study methods and techniques used, a detailed account of the sampling field and other research procedures, analysis, detailed findings, conclusions, and suggestions. There is also a technical appendix for methodological details, copies of measuring instruments and the like. It is comprehensive and complete. It is written by using technical language, following a specified pattern

2. Popular Report

This type o report is designed for an audience of executives/administrators and other non-technical users. The reader is less concerned with methodological details but more interested in quickly studying the major findings and conclusion. It should present broad facts, findings and recommendations. It must be interesting, simple and lucid. It must avoid all technical jargon and details in the method of investigation. The style may be more journalistic but precise, encouraging rapid reading and quick comprehension. More headlines, underlining, pictures and graphs may be used.

3. Interim Report

When there is a long time gap between data collection and the presentation of the results, the study may lose significance and usefulness. To avoid such eventualities, a short report containing (a) the first results of the analysis or (2) the outcome of the analysis of some aspects completely analysed is presented. Such a report is called an interim report. This type of report is required, particularly when the study was undertaken for a sponsor; whose interest may lose it and there is an inordinate delay in giving a report. It helps the sponsor to take action without waiting for the full report.

4. Summary report

A summary report is prepared for the consumption of the lay audience, viz., the general public. It is written in non-technical and simple language. It is a short report of two or three pages. It briefly references the study's objective, its major findings and implications.

5. Research Abstract

Research Abstract is a summary of the technical report. Doctoral students usually prepare it in the event of submitting their thesis. It contains a brief presentation of the problem statement, the objectives of the study methods and techniques used and an overview of the report. A summary of the results of the study may also be added.

6. Research Article

The research article is designed for publication in a professional journal. It must be written in concise and unambiguous language. It must be logically organised, progressing from a statement of the problem and the purpose of study, through the analysis of evidence, to the conclusions and implications

Steps/Stages in report writing

1. Plan the project, and fix the target and final date for completing the report.
2. Prepare a layout of the structure of the report. Arrange the data, document, bibliography etc., in conformity with the report's structure.
3. Prepare the outline for the report. The outline should be based on all main points and sub-points.
4. Prepare a rough report of what one has done in his studies. He has to write down the procedure he adopted in collecting the material, the techniques or analysis he adopted, the broad findings, generalisations, and suggestions. This forms a rough report.
5. Keep the rough report for a few days for careful reading and then revise it based on thinking and discussing it with others. Expert guidance and experienced person's help can be sought for the purpose and revised accordingly
6. Rewrite the report based on the revision made and corrections effected on the first report. Eliminate irrelevant aspects

7. Prepare the final bibliography. The bibliography is a list of books referred to or consulted pertinent to the research 8. Write the final draft of report 11 should be written in a concise and objective style and in simple language

Planning report writing

After the data analysis is over, the planning stage begins. At this stage, the researcher deters various basic questions, viz., who says what, to whom, in which way and with what effect of the report.

1. The target audience.

The form and style of reporting and other aspects depend upon the type of reader for whom the report *s intended. The target audience may be (a) the academic /scientific community, b) the research sponsors or c) the general public. In each situation, the form and content of the report would be different. For instance, where the target audience is the academic community, a technical-type report will best serve the purpose. However, a popular' report must be resorted to where the general public constitutes the intended audience.

2. The communication characteristics of the audience.

The level of knowledge and understanding of the selected audience should be considered. The kind of language (scientific or journalistic), their interests etc., determine the scope, form and reporting style.

3. The intended purpose

It may be for evaluation by experts for the award of a degree or diploma, for references by researchers and fellow scientists or for implementation by a user/ organisation. It also determines the type of the report and its contents and form of presentation.

4. **The type of report**

The type may be technical, popular or summary, based on the intended use.

5. The scope of the report

The scope of the report's contents is based on the type of report land and its intended purpose. For example, a research thesis or dissertation to be submitted for the award of a degree or diploma should narrate the entire research process and experience, the state of the problem, a review of review studies, objectives of the study, methodology, findings, conclusion and recommendations.

6. The style of reporting

It may be simple and clear or elegant and pompous; it is decided concerning the target audience.

7. The format of the report should be designed as explained below. An outline/ table of contents should be prepared for each of the proposed chapters of the report. An outline lends cohesiveness and direction to report writing work. Research Report Format (layout/ Structure /contents of a report)A Research report contains three sections viz., I. Preliminaries II. The Text III. Reference Materials I. The preliminaries include the following.

1. The title page

The title page of a research report carries the title of a thesis, the name of the candidate, the name and designation of the supervisor, the degree for which the thesis is presented, the name of the facility and university month and year the thesis is presented

2. Preface

The preface includes the writer's purpose of the study, a brief resume of the background, scope and general nature of the research and acknowledgements., Acknowledgement recognizes persons to whom the researcher is indebted for providing guidance and assistance during the study

3. Table of contents

The table of contents includes major divisions of the thesis viz., introduction, chapters with subsections, bibliography, and appendix. It provides an analytical overview of the material included in the study. Respective page numbers are also given.

4. List of tables

A list of tables gives numbers to different tables.

5. List of Figures

A list of figures gives different numbers figures.

6. The text

It is an important part of a thesis. The researcher presents his argument here. It may consist of five components

1. Introduction:

It provides the reader with background information to grasp the study. It helps to identify the central issues the study addresses, summarise previous research and provide specific reasons for the study conducted. It introduces the reader to the study. It also contains definitions of major concepts employed, references made to their books etc

2. Research procedure

It explains the study's methodology, basic design, experimental manipulations, data collection methods, questions, interview experience, etc. It also explains the samples used, who were subjects, the number of subjects, how they were selected, generalisation from particular aspects

3. Conclusion

It contains *z* ascription of the data like means, standard deviation and statistical analysis. It guides the reader through findings and gives clear and complete information.

4. Summary

It should be concluded with a summary, recalling the problem, procedure, major finding and major conclusions

5. References

References in the text part give references to someone else's published work. It attempts to relate our study to the existing literature. It should give the name of the author, year of publication, edition, page number etc

III Reference Materials

Reference materials include two components

1. Bibliography

The bibliography lists in alphabetical order all published and unpublished references Used by the writer in preparing the report. All books, articles, reports, and other documents may be presented in one common list in the alphabetical order of their authors. Alternatively, the bibliography may be classified into books, articles, reports and other documents, and in each section, relevant references may be arranged in alphabetical order.

2. Appendices

The following documents are included in appendices a) Copies of data collection, instruments like interview schedules or questionnaires, b) Technical details of the sampling plan, c) Complex and long primary tables, d) Supporting documents and any other evidence that may be important as backup details for the report

Principles of writing

The writing of a research report is governed by certain principles / standard practices. These are described below

1. Organisation of the report

The research report requires clear organization. Each chapter may be divided into two or more sections with appropriate headings, and in each section, margin headings and paragraph headings may be used to indicate subject shifts. A page should not be fully filled-in from top to bottom. Wider margins should be provided on both sides and top and bottom.

2. Style

A research report is a formal presentation of an objective, unbiased investigation and should be written at an elevated level of Standard English. It does not require elegant word usage and allusion. It needs a plain discourse with accuracy, clarity, coherence, conciseness and readability. a) Accuracy: The report should be factual with objective presentation Exaggeration and superlatives should be avoided. b) Clarity: The presentation should be made using familiar terms, common words, and

unambiguous statements c) Coherence. Each sentence must be so linked with other sentences that the writer's thoughts move smoothly and naturally from one statement to the next. d) Conciseness: The statements must be concise and precise e) Readability: They should be easily understandable. Technicalities should be translated into a language understandable by the reader interested in the study's results. Readability can be achieved by using active verbs, correct and exact names, references, facts and figures, and simple words and sentences.

3. Unclear writings

To avoid unclear writing, the following aspects should be considered: 1. Avoid the jargon and pretentious style 2. Avoid offensive words, e.g. adding the suffix "wise to a noun or adding 'size" to nouns/adjectives to make verbs. 3. Omit needless words that cause verbosity. E.g., instead of a large number, say 'many' and the like. 4. Avoid superfluous phrases. E.g.' authorities agree that.

5. Avoid abstract words and use concrete words. 6. Avoid words that exaggerate, such as spectacular, immeasurable, gigantic, awfully, and dreadfully. 7. Avoid tautology, or repetition, e.g., ln phrases, like ‘joint partnership’, return’ the common word is redundant and should be omitted.

4. Grammar, Spelling

The presentation should be free from spelling and grammatical errors. Each word must be spelt correctly. The rules of punctuation should be carefully observed. Standard practices for capitalising words in English should be followed when working *m* English or quoting English titles. Principal words of titles and parts of specific works are capitalised. Do not *use* masculine nouns and pronouns when content refers to both genders.

5. Words and Numerals

Certain conventions determine how and when to use numbers, such as using the per cent symbol (%) only in tables and figures. Use numerals to express page, street, telephone number, dates and quantities combined with abbreviations and symbols. Express reference to the table and exhibit figure numbers numerically

6. Documentation

Every quotation used, either direct or indirect, should be acknowledged through a footnote.

7. Ellipses

Omissions in the quoted matter are permitted if the original meaning is not altered. Such omissions are to be indicated by the sign of ellipses.

8. Abbreviations

Do not abbreviate words in the text. Spell them out in full. This rule does not apply to materials included in the footnotes, appendices, bibliographies, tables and figures where abbreviations are desirable.

Documentation: Footnotes and Bibliography

Documentation

The researcher should give credit for borrowed words, ideas, symbols or other forms of expression. Their sources should be stated in the text or footnotes. There are two alternative modes for documenting sources of ideas and information. 1. Footnotes 2. References - cited format 1. Footnotes:

Footnotes are of two kinds. Content and reference. Content notes contain explanatory materials. Reference notes serve as documentation of sources or as means for cross-references.

Purposes of Footnotes

1. To acknowledge indebtedness - To another writer whose passage is paraphrased or whose quotation is used.

2. To amplify/clarify the ideas or information presented in the text.
3. To establish the validity of evidence
4. To refer the reader to further sources of information on the subject under discussion.
5. To give the original version of the translated material into the text.
6. To provide a cross-reference to various parts of the thesis

Bibliography

The bibliography to a list of references relating to a topic or subject. It is located in the main body of the report. It contains all the information in the first footnote relating to a work. It lists in alphabetical order the references used by the writer. The references in the bibliography are arranged alphabetically, sometimes by topics, geographical location or other plans.

Writing the Report

The researcher's report needs several revisions and rewriting before reaching the final form. The researcher should arrange the following detailed outlines for chapters, note cards arranged in the order of chapters, source cards arranged in alphabetical order, statistical tables, charts and results of analysis, each in separate sheets of paper and a good pen.

First Draft

In the first draft, the researcher should concentrate on substance, i.e., the fullness of facts, as per the planned outline. The entire first draft work should be completed without any stopping for editing.

Divisions

The first draft should be read carefully again and again and edited thoroughly, and revised. Any writing improves upon revision. In revising the first draft, attention should be given to form, language, readability, clarity and transparency. With an open and critical mind, the researcher must correct, carve, cut, add and polish.

Final Stage

The final stage of the work consists of A. Adding the following demerits to the report (1) Title page (2) acknowledgement/preface (3) Table of contents (4) List of tables and charts (5) Bibliography (6) Appendices B. The final editing of the revised and completed report. The final editing should focus on the relationship between the original research questions and the report.

Typing the Report

The final manuscript of the report should be given for typing to a professional typist with experience in typing research reports. The writer must submit an accurate and acceptable draft to the typist. The final should be correct in all grammatical conventions, capitalisation, punctuation, spelling, compound words, hyphenation and paragraphing. The writer should give clear instructions on the requirements of margins, word division, indents, documentation placement, spacing headings, tables, charts and quotations

Briefing

Briefing involves an oral presentation of the lengthy complex report in a condensed summarised form. The presentation may take 15 to 30 minutes, followed by questions and discussion. The scope of the briefing varies according to its situation and purpose. In the case of a briefing before the sponsor organisation's executives, the presentation's focus will be on the conclusions based on the findings and the recommendations. In an academic presentation, the focus will be on the entire research process, the methodology, and the study's contribution to the wealth of knowledge. Academic people are keenly interested in knowing the problem formulation and conceptualization, sampling, methods of and tools for data collection, their validity and reliability, the plan of analysis, and the reliability of the findings.

Evaluation of a Research Report

A research report may be evaluated or reviewed When a) A doctoral research thesis or dissertation is submitted to a University for the award of a Ph.D. Degree evaluated by a Board of Examiners consisting of academic experts. b) A critical analysis of the research report by research students, selected from the published/unpublished report in the university library or research abstracts published in journals c) Research promotion bodies like ICSSR may evaluate the reports Evaluation may be done about the following

a) The appropriateness of the title

b) Importance of the problem

c) Problem formulation

d) Review of related interactions and earlier studies

e) Soundness of methodology

f) Data analysis

g) Contribution or study conclusions and recommendations h) Presentation

Statistical Packages:

One of the features of the development of the modern world is the development of the capacity to convert observations into numbers. The science that deals with numbers are statistics. It crunches the numbers and organises them meaningfully so that information is generated. This information builds up knowledge, and thus, the development goes on. Advances in computing have come in handy in this as they help in doing this part of the job accurately, timely, effectively and convincingly. The computer can help immensely in statistical analysis. Numerous statistical tools exist, and the need is to identify their actual usage. Many statistical procedures require prior knowledge and insight, even with a statistical package. In this Unit, we have tried to buildup a case for using statistics by first defining statistics and what it can do to your data. You will further find the definition of the data types. An explanation of the common tasks performed in a preliminary analysis is also given. This Unit also describes two popular packages (MS Excel and SPSS) and gives a glimpse of some other statistical packages.

Data Measurement:

Statistical data is generally obtained in many formats, such as spreadsheets (e.g. MS Excel) or databases (e.g. MS Access). Data may also be received in various open formats such as typically tab-delimited text (*.dat, *.tab, *.txt), comma-separated text or fixed-width text data (*.dat, *.txt). The data could be of two types, qualitative and quantitative. Most of the statistical methods are based on quantitative data. The quantitative variable is a variable whose values are numbers with real numeric meanings. It consists of mainly two types of data, viz. discrete and continuous. A set of data is said to be discrete if the values/observations belonging to it are distinct and separate, i.e. they can be counted, e.g. several books in a library. A set of data is said to be continuous if the values/observations belonging to it may take on any value within a finite or infinite interval. There are four well-known levels of measurement scales, i.e. nominal, ordinal, interval, and ratio. There is a relationship between the level of measurement and the appropriateness of various statistical procedures. For example, it would be impractical to compute the mean of nominal measurements. Data must be measured on an interval or a ratio scale for the computation of means and other statistics to be valid. Therefore, if data are measured on an ordinal scale, the median, not the mean, can serve as a measure of central tendency. Let us have a brief discussion of what these scales are:

1) **Nominal Scale:**

The nominal measurement consists of assigning items to groups or categories. No quantitative information is conveyed, and no ordering of the items is implied. Nominal scales are, therefore, qualitative rather than quantitative. Religious preference, race, and sex are all examples of nominal scales. Frequency distributions are usually used to analyse data measured on a nominal scale. The main statistic computed is the mode. Variables measured on a nominal scale are often called categorical or qualitative variables. Nominal variables allow for only qualitative classification. That is, they can be measured only in terms of whether the individual items belong to some distinctively different categories, but we cannot quantify or even rank order those categories

2) **Ordinal Scale:**

Measurements with ordinal scales are ordered because higher numbers represent higher values. However, the intervals between the numbers are not necessarily equal. Ordinal variables allow us to rank order the items we measure in terms of which has less and which has more of the quality represented by the variable, but still, they do not allow us to say "how much more."

3) **Interval Scale (Cardinal Scale)**
4) On interval measurement scales,

one unit on the scale represents the same magnitude of the trait or characteristic being measured across the whole scale range. However, interval scales do not have a "true" zero point; therefore, it is impossible to make statements about how many times higher one score is than another. True interval measurement is somewhere between rare and nonexistent in the behavioural sciences. A good example of an interval scale is the Fahrenheit scale for temperature. Interval variables allow us to rank order the measured items and quantify and compare their differences.

5) **Ratio Scale**

Ratio scales are like interval scales, except they have true zero points. Ratio variables are very similar to interval variables. In addition to all the properties of interval variables, they feature an identifiable absolute zero point. Thus, they allow for statements such as x is two times more than y. A typical example of a ratio scale is the measure of time or space. Interval scales do not have the ratio property. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales.

FEATURES OF SOME STATISTICAL PACKAGES:

Technological advances, especially the personal computer (PC), have made computing a game for the commoners. Today one has computing power as one can easily load software of his choice or need into his PC. There is a plethora of read-made computer packages available today. Now one can find different statistical packages for applications to different disciplines. We will describe two ready and available packages and popular and user-friendly. We will also glimpse some other packages in the subsequent section.

Microsoft Excel:

Microsoft Excel is a big worksheet (it can take data rows in thousands across 256 columns). This worksheet can be used for data entry and calculations by clicking buttons. It has a "paste function where you can paste any formula from a big list of inbuilt functions. MS Excel can create tables and graphs and perform statistical calculations. The work done in MS Excel can be easily copied and pasted to many window-based programs for further analysis. According to Potter, "Spreadsheets are a useful and popular tool for processing and presenting data. Microsoft Excel spreadsheets have become somewhat of a standard for data storage, at least for smaller data sets. The fact that the program is often being packaged with new computers, which increases its easy availability, naturally encourages its use for statistical analysis. However, many statisticians find this unfortunate since Excel is not a statistical package. There is no doubt that Excel has never claimed to be one. But one should face that due to its easy availability, many people, including professional statisticians, use Excel daily for quick and easy statistical calculations. Therefore, it is important to know the flaws in Excel, which, unfortunately, still exist today! …. ''Excel is not an adequate statistics package because many statistical methods are simply unavailable. This lack of functionality makes it difficult to use it for more than computing summary statistics and simple linear regression and hypothesis testing". However, in MS Excel 2003, some statistical functions, including rounding results and precision, have been enhanced. The MS Excel worksheet is a collection of cells. As we said, 65,000 (rows) X 256 (columns) cells are in an MS Excel worksheet. Each row or column can enter data belonging to one category. Data entry in MS Excel is as simple as writing on paper. MS Excel assigns each column a field depending upon the type of data. It supports various data formats; one can choose a format by formatting the cells.

Once the type of cells is defined, it is easy to enter the data without taking care of the format. MS Excel can perform usual calculations on the data so entered. It has an insert function (f x) icon that contains many inbuilt functions like sum, count, max/min, standard deviation etc. It has many built-in functions that perform special calculations without even typing the formula. To perform a calculation, one has to select a function and specify the range of values to which it has to be applied. These functions are known as paste functions.

MS Excel has a built-in statistical package for taking you into further details of data analysis. It provides a set of data analysis tools called the Analysis Tool Pak, which you can use to save steps when you develop complex statistical analyses. You provide the data and parameters for each analysis; the tool uses the appropriate statistical macro functions and then displays the results in an output table. Some tools generate charts in addition to output tables.

Pivot Tables:

A Pivot Table is an Excel tool for summarizing a list into a simple format. It helps you analyze all the data in your worksheet to make better business decisions. Excel can help you by recommending and automatically creating PivotTables, a great way to summarize, analyze, explore, and present your data.

A pivot table may be used as an interactive data summarization tool to automatically condense large datasets into a separate, concise table. You can use it to create an informative summary of a large dataset or make regional comparisons between brand sales.

Descriptive Statistics

The Descriptive Statistics tool in the Data Analysis add-in can be used on an existing data set to get up to 16 different descriptive statistics without entering a single function on the worksheet. Descriptive Statistics gives you a general idea of trends in your data, including:

→ The mean, mode, median and range

→ Variance and standard deviation

→ Skewness

→ Sample Variance

→ Kurtosis and Skewness

→ Count, maximum and minimum

Descriptive Statistics is useful because it allows you to take a large amount of data and summarize it. For example, you may want to represent the incomes of a community. Instead of showing it in excel, you may summarize it, and it becomes useful: an average wage, or a median income, is much easier to understand and then analyze the data.

Review Questions:

1. State the functions of statistics.
2. Discuss the properties of good measures of central tendency.
3. State the key components of the data table.
4. Given the various types of graphical presentations of data.
5. Write short notes on the research report.

Reference:

1. Box, George E.P., William G. Hunter, and J. Stuart Hunter (1978). Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building. New York: John Wiley and Sons

2. Levin R.I., Rubin, D.S. (2001). Statistics for Management. Prentice Hall of India Private Limited: New Delhi

3. Pottel Hans. Statistical flaws in Excel. Innogenetics NV, Technologiepark 6, 9052 Zwijnaarde, Belgium, http://www.mis.coventry.ac.uk/~nhunt/pottel.pdf. Downloaded 08/11/05

4. SPSS for Windows Release 21.0. Help Documentation

5. Wegman Edward J., Solka Jeffrey L. (2005). Statistical Software for Today and Tomorrow. Center for Computational Statistics, George Mason University Fairfax, VA 22030.

UNIT – V: Metric Studies and Style Manuals Scientometrics, Infometrics and Webometrics Manual Structure, Style, Contents- ISI, MLA, APA, CHICAGO

Introduction:

The library and information science (LIS) domain has a long history of measuring an organisation's or individual's research output using quantitative techniques derived from mathematics and other subfields. Sociology of science, Iibrametrics, bibliometrics, scientometrics and Informetrics are some of the terms used often for measuring the research output. The term bibliometrics was first used in 1969 by Pritchard. He defined bibliometrics as "the application of mathematical and statistical methods to books and other media of communication". By 1970 the term bibliometrics had gained considerable attention and was mentioned in LIS literature, including Library and Information Science Abstract and other LIS literature (Hood & Wilson, 2001).

Scientometrics was first coined by Vassily V. Nalimov& Z. M. Mulchenko in 1969. The term 'scientometrics' is the Russian equivalent of 'naukometriya'. Scientometrics though overlapping with the term 'bibliometrics', has a wider ambit when compared to bibliometrics. Scientometrics has been defined as "the study of the quantitative aspects of science as a discipline or economic activity. It is part of the sociology of science and has application to science policymaking. It involves quantitative studies of scientific activities, including, among others, publication, and so overlaps bibliometrics to some extent" (TagueSutcliffe, 1992, p. 1)

As bibliometrics and scientometrics expanded, the new term 'informetrics' was introduced to include information science research in the early 1990s. The definition provided by Tague-Sutcliffe (1992, p. 1) defines Informetrics as "the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists. Thus it looks at the quantitative aspects of informal or spoken communication, as well as recorded information needs and uses of the disadvantaged, not just the intellectual elite. It can incorporate, utilise, and extend the many studies of the measurement of information that lie outside the boundaries of both bibliometrics and scientometrics. …Two phenomena that have not, in the past, been seen as a part of bibliometrics or scientometrics but fit comfortably within the scope of Informetrics are the definition and measurement of information and types and characteristics of retrieval performance measures."

The above definitions help identify the similarities between the terms, such as bibliometrics and scientometrics and also distinguish between the terms. Since the 1990s, "informetrics" has often been used in the quantitative analysis of scholarly communications. As the web gained momentum and became a major source of information and scholarly communication shifting its boundaries from print to online, researchers in scientometrics and Informetrics have tried to quantify web pages or sites; because of this, the new term "webometrics" was coined in 1997 by Almind&Ingwersen (1997). Today 'Webometrics' has become one of the major areas of research in the LIS domain. The subsequent section of this paper discusses the origin of Webometrics and its applications in information science and other areas of social sciences.

Bibliometrics has seen major changes over the last five decades as a tool to measure scientific research output. It has emerged as an established field of research in library and information schools. It has arisen as a recognised scientific specialism, taught in universities as part of information science courses in Europe and America. It has a substantial body of techniques, some theories and an international group of specialist science evaluators (Thelwall, 2008). In 1906, Cattell launched the biographical directory of American men of science, published every five years and the directory collected information on thousands of American scientists active in research (Godin, 2007). Catell introduced two dimensions to the measurement of science, quality and quantity, and these two dimensions still largely define the field of bibliometric studies today. Quantity or productivity, as he called it, was counting the number of scientists a nation produces. At the same time, quality or performance was defined as contributions to the advancement of science and was measured by averaging the per rankings of colleagues (Godin, 2007). Although bibliometrics was used as a standard for measuring the publication output of scientists almost a century ago, it was largely the work of Eugene Garfield in the 1960s and his Science Citation Index through his newly developed Institute for Scientific Information (ISI) which made possible the quantitative analysis of scientific research output (Garfield, 1979). The two major changes in publishing are the computerization of the printing process, and the conversion of the entire publishing cycle, that is, the submission of an article, refereeing and publication to the internet, allowing for faster and possibly cheaper communication thought

Bibliometrics is a quantitative analysis of publications to ascertain specific phenomena (Herubel, 1999). It encompasses the measurement of properties of documents and document-related processes.(Borgman and Ferner (2002)). It uses mathematical and statistical methods to analyse and measure the output of scientific publications. Most bibliometric studies have been devoted to scientific and technological disciplines. It is important to note that since E.W. Hulme wrote his famous study in 1923, the measurement of published scholarship and scientific research has developed momentum and evolved its terminology. From statistical bibliography to bibliometrics to scientometrics and Informetrics to webometrics, this type of publication has become instrumental for library and information science and scholarly communication (Sengupta, 1992). Researchers can examine kinds of literature and establish characteristics of disciplines, obsolescence of scholarship, institutional affiliations and relationships, and types of materials constituting scholarly pursuits. Bibliometrics is used as a methodology in many fields of science, first and foremost, to map the publication pattern in different disciplines. For instance, bibliometrics is an indispensable tool for historians studying a discipline's intellectual heritage and evolution.

As pointed out above, the origin of bibliometrics Cattell is generally associated with the first systematic collection of statistics on science (Goddin, 2007). He laid the foundation for others who introduced the systematic use of bibliometrics. (Sengupta, 1992; Hood and Wilson, 2001). Catell used his bibliographical directory to study scientists and their activity in research in the United States. From the data, Catell produced statistics on the number of scientists and their geographical distribution and ranked scientists according to their performance. Catell can thus be credited for launching scientometrics or the systematic measurement of science.

Catell was followed by other psychologists like Buchner, who started his series of reviews on psychology. In these reviews, he included a discussion of recent papers, the number of psychologists, a list of new journals, statistics on publications, the percentage distribution of papers appearing in the Psychological Index and the interests of the psychologists (Godin,2006). But it was S W Weinberger of the University of Pennsylvania who developed the statistics on publication. He looked at the evolution of membership, placed increasing emphasis on publishing as a criterion for eligibility, and discussed finances, journals of the Association, the organisation and its meetings. He charted the number of papers presented at each meeting since

1892. He looked at the productivity of universities at these meetings and what he called the consistency of publication and fields of interest. He found that 19 universities produced 53% of all papers. Thus, Fernberger put forward the concepts of productivity and the index for measuring scientific productivity.

Bibliometrics analysis predates the development of the Science Citation Index (SCI). Still, the advent of SCI and specifically the availability of electronic access (online, CD-ROM and web-based) to the Institute for Scientific Information's (ISI) massive datasets has had a catalytic effect on the popularity, scope and ambition of bibliometric research, both within and beyond the information community. SCI was created as a database of references made by authors to earlier publications which will lead the readers to other similar articles and encourage them to work on similar topics. The ISI also developed other databases, such as the Social Sciences Citation Index (SSCI) and Arts and Humanities Citation Index (AHCI), and, along with them, new and varied statistics. These statistics reckons Thelwall (2008) include the number of citations to all journal articles by an author, research group, or country. Some are further developed into named indicators with supporting theories and reasonably well-accepted standard interpretations. The most well-known is the journal impact factor (JIF). Since the advent of SCI, three types of bibliometric applications have arisen: descriptive, relational and evaluative (Borgman and Ferner, 2002). Descriptive bibliometrics places emphasis on the characteristic features of the document.

In contrast, relational bibliometrics seeks to illuminate relationships within research, such as the cognitive structure of research fields, the emergence of new research fronts, or national or international co-authorship patterns. Evaluative bibliometrics seeks to assess the impact of scholarly work and compares the relative contributions of two or more individuals or groups (Thelwall, 2008). Descriptive Bibliometrics Descriptive bibliometrics describes the characteristics or features of literature and is used to measure the productivity of scientists and information scientists. The research is divided into geographic areas, periods and departments and disciplines. The descriptive bibliometric includes the study of the number of publications in a given field or the productivity of literature in the field to compare the research in different institutions/countries and periods.

Evaluative bibliometrics:

Evaluative bibliometrics use citations as the source of its / their raw data (Thelwall, 2008). This theory stems from Robert Merton's (1973) sociology of science, which indicates that citations are how scholars acknowledge influential prior work. Based on this, citation counting is therefore used as an indicator of research of scientific value. Subsequent research has shown that Merton's perspective somewhat oversimplifies reality. There are many and varied reasons to cite articles

Relational Bibliometrics:

Relational Bibliometrics are used to examine relations within scientific research using ISI data. This was not possible in the early days due to a lack of computing power and experience in technology. Even so, these early relational analyses produced interesting insights into the structure of science research through simple means, such as network diagrams of the flow of citations between key sets of articles (Cawkell, 2000). According to Telwall (2008), this idea came from geneticist Allen in 1960, who sent his citation diagram to Garfield (Cawkell, 2000). Journal citation diagrams could illustrate the connections between journals within a field, both central and peripheral. Garfield was credited with–citation as a measure of similarity, i.e., if two documents often appear together in reference lists (co-cited), they are likely to be similar in some way. This means that if collections of documents are arranged according to their co-citation counts, it should produce a pattern reflecting cognitive scientific relationships.

The importance of Bibliometrics as a research tool:

Modern bibliometrics as a research tool has been largely inspired by Derek de Solla Price and the seminal work carried out by him in the middle of the 20th century. In his book "Little ScienceBig Science", published in 1963, he analysed research communication and presented several quantitative evaluation techniques. He was the first to examine the growing trend of collaboration among chemistry researchers by using bibliometrics. Since bibliometrics has developed into a research field in its own right, it has given rise to a community of specialised experts called biometricians. According to Mattison (2008), Bibliometrics is used as a methodology in many other fields of science, mainly to map the publication pattern in different disciplines. In economics and sociology, the main interest has been for cognitive purposes, that is, studying researchers' publication behaviour. Bibliometrics has gained increasing importance in science policy and management in the last decade, specifically in research evaluation, where it plays a prominent role.

Developing performance indicators to respond to science policy questions has been the most common application. Indicators used for this purpose include productivity analyses measuring the output and volume share of a specific actor, e.g. a country's world share of publications or citations; research impact analysis using citations; and relational indicators studying heterogeneity of collaboration patterns between different actors (Mattson, 2008). One of the major focus areas in bibliometrics is research collaboration which receives increasing attention from policy-makers and e general users. Modern research is regarded as increasingly complex and specialised, making it impossible for an individual researcher to master all the knowledge and technical skills needed. In collaboration, different skills complement each other and, so doing, contribute to the stimulation of knowledge sharing and the generation of innovation and new ideas. As a result, collaborative research activities, besides enabling the pooling and sharing of resources for enhanced efficiency, also contribute to the quality of the research outcome (Mattson et al., 2008). Funding agencies and institutions, therefore, increasingly encourage collaborative research. Grants awarded by many different funding institutions and for many different disciplines often seek to encourage and, at times, require, as a condition, collaborations between different countries, research fields or institutions. The National Science Foundation does research (where South Africa??) found that research done by multiple institutions has increased from 40% to 61% between 1988 and 2008. (NSF S&E indicators)

Scientometrics:

It is a common misbelief that scientometrics is the publication and citation-based gauging of scientific performance or compiling cleaned-up bibliographies on research domains extended by citation data. Scientometrics studies the quantitative aspects of science as a discipline or economic activity. It is part of the sociology of science and has application to science policymaking. It involves quantitative studies of scientific activities, including, among others, publication, and so overlaps bibliometrics to some extent. The term scientometrics became the name of a journal founded by T. Braun in 1977, originally published in Hungary and now in Amsterdam (Jean Tague-Sutcliffe, 1992). The main subjects of scientometrics are individual scientific documents, authors, scientific institutions, academic journals, and regional aspects of science. Van Raan (1994) reckons that there is a rapid addition of scientometrics but not bibliometric data, such as data on human resources, infrastructural facilities, and funding. In information science-oriented scientometrics, in contrast to the economy, sociology or psychology of science, aspects of information and communication are examined. These aspects may include productivity (documents per year), subjects of the documents (words or co-words), reception (readers of the documents) and formal communication, references and citations, and co-citations (Juchem, Schlogl and Stock, 2006).

Scientometrics is a multifaceted research strategy encompassing structural, dynamic, evaluative and predictive subareas. Structural scientometrics came up with results like re-mapping the epistemological structure of science-based, for instance, on co-citation," bibliographic coupling" techniques or co-word techniques. On the other hand, Dynamic scientometrics constructed sophisticated models of scientific growth, obsolescence, citation processes, etc. These models are not only of theoretical interest. Still, they can also be usefully applied in the evaluation and prediction of what??? . Beyond policy-relevant applications of scientometrics results, there are recently important applications in the context of studying the linkage between science and technology or applications to related fields such as the library and information science and more recently also Webometrics. Examples of the latter are ongoing projects such as EICSTES (European Indicators, Cyberspace and the Science-Technology-Economy System) and WISER (Web indicators for science, technology and innovation research(Ganzel, 2003).

Informetrics:

Informetrics is the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any social group, not just scientists. It looks at informal or spoken communication and recorded information needs (Tague –Sutcliffe, 1992). The quantitative study of recorded discourse may relate to any medium, although until recent decades, print media have dominated informetric research. The wider availability of documentary resources and electronic discourse formats, particularly through machine-readable databases and, more recently, the internet, informetric research based on electronic data sets has become commonplace. It incorporates and utilizes many studies of measurements of information that lie outside the boundaries of both bibliometrics and scientometrics. As such, 'informetrics' is the broad term comprising all metrics and studies related to information science. However, the field of Informetrics can be traced back to the first half of the twentieth century through the works of Lotka (1926), Bradford (1934), and Zipf (1949). The term was popularized by Blackert and Siegel (1979) and Nacke (1979) in the late nineteen nineties. From here on, the concept steadily gained popularity through the organisation of international informetrics conferences.

Growth and expansion of Informetrics. According to Egghe, there is a fast multidisciplinary expansion (growth) of the field of Informetrics, mainly due to the new topics that have been included in Informetrics, such as the quantitative study of networks, including the internet" (Stock and weber, 2006). (Egghe, 2006). This is confirmed by the request on "Web of Science" and the use of its ANALYSE function. A search for "TS= informetrics OR bibliometrics OR scientometrics OR webometrics OR retrieval evaluation" returned the results as shown in Table 1 below. From this, it is clear that Informetrics is a rapidly growing research field. / Or field of contemporary research.

Webometrics:

Webometrics is the quantitative study of web-related phenomena (Thelwall, Vaughan &Björneborn, 2005). The growth of the web as a major source of information attracted many researchers to apply traditional bibliometrics and scientometrics techniques to a new form of communication media, the World Wide Web. According to Björneborn&Ingwersen (2004), webometrics is "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the Web drawing on bibliometric and informetric approaches." Mike Thelwall is one of the pioneers in the field of webometrics. According to him, webometrics is "the study of web-based content with primarily quantitative methods for social science research goals using techniques that are not specific to one field of study" (Thelwall, 2009), which emphasizes a small subset of relatively applied methods for use in the wider social sciences. The purpose of giving this alternative definition is to help publicise appropriate methods outside of the information science discipline rather than to replace the original definition within information science.

Björneborn and Ingwersen have proposed a differentiated terminology. Distinguishing between studies of the web and studies of all internet applications. They used an information science-related definition of webometrics as "the study of the quantitative aspects of the construction and use of information sources, structures and terminologies on the world wide Web drawing on bibliometric and informetric approaches" (Björneborn and Ingwersen, 2001). This definition thus covers quantitative aspects of both the construction and usage sides of the web, embracing the four main areas of webometric research –(1)Web page content analysis, (2)Web link structure analysis, (3)Web usage analysis,(e.eg., exploiting log files of users' searching and browsing behaviour), (4) Web terminology analysis (including search engine performance)" (Thelwall, Vaughan and Björneborn, 2004). Pirolli, Pitkov and Rao (1996) explored web analysis techniques for automatic categorization using link graph topology, text content, metadata similarity, and usage data. All four main research areas include longitudinal studies of changes on the dynamic web, for example, page contents, link structures, and usage patterns.

Moreover, the web has its citation indexes in the form of commercial search engines, so it is ready for researchers to exploit (Egghe, 2005). One of the most visible outputs of webometrics is the ranking of world universities based on their websites and online impact (Aguillo et al., 2006). Webometrics includes explains Telwall (2008), 'link analysis, web citation analysis, search engine evaluation and purely descriptive studies of the web together with the recent addition of the web analysis of web 2.0 phenomena'.

Key Areas of Webometrics

Thelwall, who has contributed immensely to the growth of webometric research, has identified key areas of webometrics (Thelwall, 2009; 2012). They are:

1. Link Analysis,
2. Web Citation Analysis,
3. Search Engine evaluation, and
4. Web Data Analysis or Measuring Web 2.0.

Link Analysis:

Link analysis is the quantitative study of hyperlinks between web pages. Early webometric studies have conducted this type of research extensively. Link analysis has been used in webometrics studies for identifying the patterns in scholarly web spaces. Link analysis has direct analogies to traditional bibliometric studies. In traditional bibliometric-related studies, the researcher often used to do studies by measuring or quantifying an organisation's or researchers' research contributions; similarly, link analysis can be done by identifying how popular a website is. Which link of the website or page is more popular and why? This kind of research can be done using link analysis in webometrics. The web crawling tools such as SocSciBot (www.socscibot.wlv.ac.uk) introduced by the Statistical Cybermetrics Research Group at the University of Wolverhampton, UK and other freely available link analysis web crawlers have made the process of crawling web data simpler. Some of the major studies in this field of research have been reviewed by Kousha (2005) in his paper entitled "webometrics and scholarly communications: an overview".

The 'Webometrics Ranking' is one of the best examples for link analysis of academic university websites. Webometrics Ranking of World Universities (www.webometrics.info/en) has been conducted since 2006 to identify the web presence of world universities through link analysis.

Web Citation Analysis:

Another key area of webometrics research is web citation analysis. Web citation analysis is nothing but counting online citations to published items such as journal articles. Earlier, bibliographic citations found in traditional citation indexing databases such as SCI, SSCI, and A&HCI were considered major sources for identifying the impact of research publications. The web provided new methods to overcome the limitations of the bibliographic citation databases by helping in finding web citations for articles published online, largely for open-access journals. Research on web citations has found a correlation between the citations found in bibliographic citations and citations found on the web for particular research papers (Vaughan and Shaw, 2003). The introduction of Google Scholar has helped biometricians use this citation database to identify the impact of research publications easily. Because of the introduction of tools such as Google Scholar and Altemetrics, research in this area has seen considerable growth in recent times. Two of the research publications mentioned below highlight how this study can be conducted using webometrics. These studies were done by Kousha&Thelwall (2005, 2007) on Google Scholar and Google Web/URL citations.

Search Engine Evaluation:

Another key area of research that has been conducted in webometrics is search engine evaluations. The major webometrics studies have concentrated on the coverage of the web and the accuracy of the reported results. It was a known fact that none of the search engines had indexed the full information on the web. Still, 70% of the information on the web has not been indexed by any search engines available today (Metayer, 2010). The reliability and validity of the data obtained through commercial engines were also very contentious issues in webometrics; many studies have raised the validity and reliability of data obtained through search engines (Bar-Ilan, 2004; Kousha, 2005). Many of the search engines such as Alta Vista (which was initially used by many webometrics for link analysis), Yahoo and Bing now have stopped providing advanced link search commands feature, which was used extensively at the beginning of webometric research (Thelwall, 2012). The development of web crawling programs and the availability of API (Application Programming Interface) from different information processing industries have now helped to expand webometrics research to other social science domains

Web Data Analysis or Measuring Web 2.0

As the web evolved over the years, webometrics has expanded its research base by measuring academic websites to other research areas of social sciences. The availability of web crawlers and API (Application Programming Interface) has helped extract a larger cache of data easily on the web. Recent studies done by Thelwall and his team (see Sugimoto &Thelwall, 2013; Kousha, Thelwall&Abdoli, 2012; Wilkinson &Thelwall, 2002; Shema, Bar-Ilan&Thelwall, 2012) have shown how webometrics can be used in other areas of the subject apart from library and information science domain, which is the traditional forte of webometric research. Webometrics has been used in many other areas of study apart from the library and information science domain. Today Webometrics is not only used for research evaluation but also in other social sciences, mainly examining web phenomena. As the information on the web increases, webometrics has become an important research method for studying web data.

Cybermetrics or Webometrics:

Cybermetrics or Webometrics, is there any difference? The web, a large repository of information, has been considered one of the important mediums for research. The availability of structured and non-structured data provided a fertile ground for using bibliometrics, scientometrics and Informetrics. Many researchers have proposed various terms to define new areas of research. The terms that were proposed: were neoteric, geometry, interferometric, webometrics and cyber metrics, web bibliometry and web metrics (Thelwall, Vaughan, &Björneborn, 2005). Bjorneborn&Ingewersen (2004) have made an effort to differentiate these two terms. Webometrics is "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the WWW drawing on bibliometrics and informatic approaches." On the other hand, Björneborn&Ingewersen (2004) have defined cyber metrics as "the study of the quantitative aspects of the construction and use of information resources, structures and technologies on the whole Internet, drawing on bibliometrics and informetrics approaches." Cybermetrics encompasses statistical studies of discussion groups, mailing lists and other computer-mediated communication on the internet."

Going through the above definitions, one can easily distinguish between webometrics and cyber metrics. Cybermetrics includes all aspects of Internet-related studies, including computer-mediated communications, and covers quantitative aspects of the web, including studying web network properties. Figure 1 briefly depicts the difference in Informetrics, bibliometrics, scientometrics, cyber metrics and webometrics and their overlapping nature of each other.

Network Analysis or Nettometrics:

Network analysis has become one of the important areas of research because of the availability of huge data sets through the internet. Though network analysis has its origin long back in mathematics, sociology and another field of human interactions, it is in recent times that network analysis has gained momentum because of people's interactions through technological means (on the web). A social network is a social structure made up of individuals (or organizations) called "nodes", which are tied(connected) by one or more specific types of interdependency, such as friendship, kinship, common interest, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige (Wikipedia, 2013).

Social network analysis views social relationships in terms of network theory consisting of nodes and ties (also called edges, links, or connections). Nodes are the individual actors within the networks, and ties are the relationships between the actors.

Social network analysis has become a key technique in sociological research. It has been in other fields of study, such as anthropology, anthropology, biology, communication studies, economics, geography, information science, organizational studies, social psychology, and sociolinguistics. In information science, the social network analysis technique is used to study the authors' citation pattern, co-citation analysis and co-authorship pattern (Otte & Rousseau, 2002).

The study on "Co-authorship networks in the digital library research community" by Xiaoming Liu and others (Liu, Bollen, Nelsson, & Van de Sompel, 2005) is one of the examples of how social network analysis can be used to identify co-authorship network pattern among the digital library research community. There is a whole lot of opportunity to apply social network analysis to different domains of subject areas. The development of network analysis tools and visualization software has allowed researchers to investigate large networks like the internet and understandably interpret them.

Citation:

The term citation is used synonymously with the term bibliographic reference. In contrast, the references are intended to identify earlier contributions (concepts, methods, theory, empirical findings, etc.) upon which the present contribution was built and against which it positions itself. Thus, a basic feature of the scientific article is that it contains several such references and that these references are attached to specific points in the text. De Solla Price states a distinction between citations and references: "…if Paper R contains a bibliographic footnote using and describing Paper C, then R contains a reference to C, and C has a citation from R

Oxford dictionary defines the term citation as "a quotation from or reference to a book, paper, or author, especially in a scholarly work" and "Law a reference to a previous case, used as guidance in the trying of comparable cases or support of an argument" (Oxford Advanced Learner's Dictionary).

In Merriam Webster dictionary, the following definitions are provided for citations: an official summons to appear (as before a court); an act of quoting; especially: the citing of a previously settled case at law; a formal statement of the achievements of a person receiving an academic honour; specific reference in a military dispatch to meritorious performance of duty (Merriam-Webster). Citation in Library and Information Science has different views. Sadison

(59-64) believes that "Every citation represents a decision by one author that he wishes to draw attention to the work of another as being relevant to his theme at a particular point in the document he is writing". That definition leads to many factors to be considered, revealing that each citation comes from a pool of citable items, and the sizes of the pools represent the denominator's incorrect statements of the units counted. Such counts must be recognised as ratios which can only be validly compared if the denominators are the same and if the procedures have provided representative and unbiased samples. Horri (65-95) explained the term referring to earlier conversations or documents. Egghe and Rousseau also defined the term as referring to earlier literature relating to the theme of the study. Citations have increasingly been applied as performance indicators in science policy and research evaluation. The basic assumption underlying such applications is that number of citations can be regarded as a measure of scientific quality or impact. Wouters (39-55) believed that because citations may be regarded as the mirror images of the references, the use of citations as indicators of research performance needs to be justified or grounded in the referencing behaviour of the scientists. Citation represents the pool of archival knowledge from which authors retrieve established ideas and, in turn, generate new research ideas. This knowledge may be disseminated within an area and across disciplinary boundaries (Sharif 172-76). Citation count and Impact factors can be easily manipulated (Gorman 581-84). Citation impact can be used to measure an article's impact within its particular field. An article being widely read and cited is an indication that it has had an influence on other researchers within the field (Turk 65-74)

The development of citation analysis has been marked by the invention of new techniques and measures, the exploitation of new tools, and the study of different units of analysis. These trends have led to rapid growth in the number and types of studies using citation analysis. The easiest technique is a citation count, determining how many citations have been received by a given document or set of documents over time from a particular set of citing documents. Citation counts provide researchers and administrators with a reliable and efficient indicator for assessing the research performance of authors, projects, programmes, institutions, and countries and the relative impact and quality of their works (Cronin 16-24; Van Raan 59-63). Using citation counts for evaluating research is based on the assumption that citations give credit to and recognise the value, quality, and significance of an author's work (Borgman and Furner 3-72; Van Raan 397-420). When this count is applied to articles appearing in a particular journal, it can be refined by calculating the impact factor, the average number of citations received by articles published in a journal during a specified period. This measure allows one to compare the "impact" of journals which publish different numbers of articles (Smith 83-106). Citation analysis essentially involves counting the number of times a scientific paper or scientist is cited, and it works on the assumption that influential scientists and important works will be cited more frequently than others. Many governments, funding agencies and tenure and promotion committees use citation data to evaluate the quality of a researcher's work, partly because they prefer not to rely on peer review and publication output alone

Zunde (1-18) noted citation analysis has three main applications: i. qualitative and quantitative evaluation of scientists, publications, and scientific institutions; ii. modelling of the historical development of Science and Technology; and iii. information search and retrieval

Citation analysis has been used since the mid-20th century to measure scientific articles' impact and visibility. Citation analysis is a powerful means of mapping the flow of ideas between speciality groups, disciplines and nation-states. It can be considered a tool to monitor a subject trend and to evaluate the scientific impact of a given researcher or institution (Bauer and Backkalbasi). Citation analysis was made popular by the works of Garfield (Presley and Caraway 67-80), who created three indexes to record citations for articles: the Science Citation Index, the Social Science Citation Index and the Humanities Index. These three print resources are combined into a database, Web of Science, which constitutes a powerful interdisciplinary research tool. Web of Science builds on citation analysis to determine journal impact factors calculated as the average number of citations to articles in a journal for the last two years per article published (Garfield 471-479).

Web Citations:

Nowadays, the web plays an important role in information dissemination and affects the print and online resource usage (Kousha 394-406). The outburst of e-contents, such as e-journals, e-books, etc., facilitates access to scholarly information. Thus, the nature of citations is susceptible to change. Print-to-web and Web-to-print citations are now fairly common; thus, it seems inevitable that Web resources are becoming favourable in scholars' communication. The cited content is considered available if it can be found either at the URL included in the sample citation (persistent) or elsewhere on the web (accessible) (Casserly and Byrd 300-17). This progressive transition of scientific literature publishing from print to the Web environment has been a key factor in motivating information professionals to explore scholarly communication patterns on the web. In particular, many have considered whether methods of bibliometrics, such as citation analysis, can be applied to the Web environment (Thelwall et al. 81-135).

Citations Indexes:

The modern history of citation indexing and citation indexes began with Eugene Garfield's construction of the Science Citation Index in 1963, published by the Institute for Scientific Information (ISI), followed by the Social Sciences Citation Index in 1972 and the Arts and Humanities Citation Index in 1980. Garfield (1962-1973) believed that systematically recording and indexing the authors' references to other literature would be more useful in identifying publishing trends and emerging new research areas. Summarising the rationale behind establishing a citation index for Science periodicals, he later wrote (Garfield 1962-1973): "The basic concept of citation indexing is based on the observation that when one article cites another article, there must be a subject relationship between the two articles. [emphasis added] Each bibliographic citation … is an unambiguous descriptor or symbol which designates the subject matter discussed in some aspect of the citing work. For this reason, citation indexing provides the scientist with a unique and efficient method for conducting, in a minimum of time, highly specific searches, since each cited reference becomes an entry point in the index." Citation indexes are powerful tools that allow searching documents which cite a given set of other documents. The application of citation indexes for information retrieval is based on the implicit definition of subjects that the relationship between citing and cited documents is subject-related. In this way, the traditional view of what a subject is somewhat challenged (Bauer and Backkalbasi).

The web has had a huge impact on citation-analysis research. Since the turn of the century, dozens of databases, such as Google Scholar, have appeared, allowing the citation pattern of academic papers to be studied with unprecedented speed and ease. This could mark the beginning of the end of the 40-year monopoly of citation analysis held by the US-based firm

Thomson Scientific, formerly known as the Institute for Scientific Information (ISI) (Meho and Yang 2105-125). ISI Web of Science The ISI citation databases – which include the Arts and Humanities Citation Index (A&HCI), Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) – have for decades been used as a starting point and often as the only tools for conducting citation analyses. Since their origin in the 1960s and 1970s, these databases have grown drastically in size and influence. Today, they contain about 40 million records from more than 8,700 world's most prestigious research journals.

Garfield launched his Science Citation Index in the early 1960s. After creating the Science Citation Index, ISI, within a few years, began publishing separate indexes for the Social Sciences and Humanities. Garfield had exploited the vast amounts of citation data accumulated by ISI over several decades and has revealed patterns across many types of scholarly publishing. In 1964, Garfield began to show how citation analysis could track ideas' emergence, propagation and development (Garfield, Sher, and Torpie). Until 1988 these indexes exist only in print form, although searching them online has been possible since the mid-1970s using a third-party information-retrieval system known as Dialog. In 1988, the ISI supplemented its indexes with CD-ROM editions, and in 1997 the databases finally migrated online with the launch of the Web of Science. The move to an online interface, which can analyse thousands of records in a few seconds, has given the ISI's databases an even greater stranglehold in the field of citation analysis (Meho 32-36). But at the same time, the web has produced new publication avenues and competitors that challenge the wisdom of continuing to use the Web of Science exclusively. Web of Science builds on citation analysis to determine journal impact factors being calculated as the average number of citations to articles in a journal for the last two years per article published (Garfield 471-479). This is a controversial but highly influential measure (Seglen 498-502; Walter et al. 280-281). Web of Science is designed to cover only the most important journals, thus covering the greatest percentage of citing activity efficiently. (Garfield 471-479; Seglen 498-502).

Features of the ISI Web of Science database include:

→ The opening screen in ISI Web of Science allows the user to do a general search, a cited reference, or an advanced search. The general search has boxes for the user to fill in (topic, source title, etc.)and allows limits by language and document type. The advanced search allows the user to fill in a box using Boolean, again allowing the user to limit by language or document type. The Cited ReferenceSearch allows the user to search by the cited author, title or year.

→ In the General Search area, a box is available for author searching. Cited Reference Search is one of the options in Web of Science for searching from the opening screen. The user can choose from cited Author, Cited Work or Cited Year.

→ Results in Web of Science are presented in citation format and can be sorted by date, relevance, times cited, author or source title.

→ In Web of Science, when viewing one of the retrieved records by clicking on "Find Related Records", the results are presented in relevancy ranking based on the number of shared references. The number of cited and shared references (with the original article) are listed. The user can view the shared references for each citation from this screen. These results are in a table format, but the complete citation can be viewed for those citations in the Web of Science database.

→ Once citations have been marked, they can be printed, saved, exported, e-mailed or ordered.

→ In Web of Science, after a list of citations is retrieved, the user can click on the Analyze Records option to view rankings and histograms of the authors, journals, etc., for the set of records (Scopus vs Web ofScience).

Scopus Elsevier (http://www.elsevier.com/), a publisher of scientific, technical and medical information products and services, launched SciVerse (http://www.sciverse.com/). This innovative platform integrates the company's key products and encourages the scientific community to collaborate on the development of customised search and discovery applications. SciVerse included SciVerse Hub beta, a module that integrates ScienceDirect (http://www.sciencedirect.com/), Scopus (http://www.scopus.com/) and targeted web content from Scirus Elsevier's science-specific Internet search engine. SciVerse Hub beta allows for a single search across its integrated content with results ranked by relevance and without duplication, saving valuable researcher time (About Scopus). Scopus, launched in 2004, offers coverage of articles from journals in Science and Social Sciences, along with citation tracking. Scopus indexes more journals than the Web of Science and has greater coverage of open-access and international journals (About Scopus). Reviewers generally credit Scopus with having a superior interface, but it lacks depth coverage in years of scientific journals and has no humanities area coverage (Deis and Goodman; Jacso b; LaGuardia 40-42). It is updated daily, offers nearly 18,000 titles from 5,000 international publishers, including coverage of 19,500 peer-reviewed journals (including 1,200 Open Access journals), 600 trade publications, 350 book series, extensive conference proceedings coverage (3,6 million conference papers), 38 million records, of which 19 million records back to 1996 with references, and 19 million pre-1996 records (About Scopus).

Scopus has captured references post-1996 and provides a matching of references to citations. Scopus also provides links to the "cited by" count next to every article in a search result set; this count is updated daily. Furthermore, on the abstract record Scopus not only provides the full citation of the first three most recent citations plus a drop-down list of all citations to that article. The references in the article record also each show a cited by count. Scopus also provides the Scopus Citation Tracker, which allows the user to get an instant overview of citations to a particular article set (by author, group of authors, subject area etc.). It shows the number of citations each article received each year (1996 onward), and each of these counts is linked to a result set showing the full records of each citing article (Database: Scopus).

Features of Scopus include:

→ Links to citing and cited documents allow the user to go forward and backwards in time.

→ Open Access titles are included in the index.

→ Indexes web pages and patents, claiming over 167 million relevant web pages.

→ Runs an entitlement check before returning a full-text image if the article is available to the user.

→ Link to the publisher's website to view the document.

→ Developers claim that "citation accuracy is achieved using state-of-the-art technology, with 99% of citing references and citing articles matched exactly."

→ For statistics on usage, Scopus delivers customer-specific usage reports, which will be COUNTER (=Counting Online Usage of NetworkedElectronic Resources- an international standard for the recording and exchange of online usage statistics) compliant.

→ Offers both on- and off-site training, web-based training, and online tutorials. The database website offers quick reference guides, tips, etc. (in English and other languages) and online technical support.

→ Works equally well with Internet Explorer

→ , Netscape and Firefox(About Scopus).

Knowing the Styles and When to Use Them

In academic writing, how you present your information (technically) is often seen as important as the ideas you are putting forth. Proper citing, quoting and referencing source material allows you to convey your breadth of research in a language commonly shared by others in your discipline. Giving others a chance to review and compare your work under these established guidelines enables your instructors to better see the work on its own merits instead of getting sidetracked by technical inefficiencies.

You MUST follow the rules like every other student: this is not an area where you want to stand out for doing things your way. Writing for any academic purpose carries with it certain expectations and formatting consistencies, and a failure to properly understand how or why you cite your sources in a specific way can negatively affect your written projects and communications.

The Big Three: APA, MLA, and CMS

Three main "Schools of Style" are used to format an academic paper properly, referred to as APA, MLA, or CMS.

- **APA style**: These are the official guidelines of the American Psychological Association, now in its sixth edition. This is the preference of the social sciences, so if you are studying sociology, psychology, medicine, or social work, you will know the APA style.
- **MLA style**: The Modern Language Association provides guidelines you will be familiar with if you are focused on the Humanities: so artists, English majors, and theatre students will know MLA as they have used this style now for more than half a century.
- **CMS style**: These are the guidelines in the *Chicago Manual of Style*, now in its 16th edition. CMS style is predominantly seen in the humanities, particularly with literature students and those who study advanced segments of history and the arts.

While these formatting methods share many characteristics, such as margins and spacing, how they attribute references to source materials is the main differentiator. For example, APA lists "references" while MLA calls the same thing "works cited" - a small but important distinction that might affect your grade.

Typically, you are going to use one style for most of your classes and communications, but there is certainly the possibility that you'll need to know how to use any one of these three common styles. The good news is it is not hard to get up-to-speed on any one of them and use them properly.

Review Questions:

1. State the importance of Bibliometrics.

2. Narrate the key area of Web metrics

3. What do you understand about Web Data Analysis

4. short Writ note on Citations Indexes

5. Briefly discuss the features of the ISI Web of Science database

Reference:

1. Agarwal, S., Choubey, L., & Yu, H. (2010). Automatically classifying the role of citations in biomedical articles. In AMIA Annual Symposium Proceedings (Vol. 2010, p. 11). American Medical Informatics Association

2. Cohen, A. M., Hersh, W. R., Peterson, K., & Yen, P. Y. (2006). Reducing workload in systematic review preparation using automated citation classification. Journal of the American Medical Informatics Association, 13(2), 206-219

3. Evert, S., Proisl, T., Greiner, P., & Kabashi, B. (2014). SentiKLUE: Updating a Polarity Classifier in 48 Hours. SemEval, 551

4. Tanguy, L., Lalleman, F., François, C., Muller, P., & Séguéla, P. (2009). RHECITAS: citation analysis of French humanities articles. In Corpus Linguistics 2009 (pp. http-ucrel)

5. Teufel, S., Siddharthan, A., & Tidhar, D. (2006, July). Automatic classification of citation function. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (pp. 103- 110). Association for Computational Linguistics.