RESEARCH METHODOLOGY FOR ECONOMICS

Dr. G. YOGANANDHAM
Professor and Head,
Department of Economics,
Thiruvalluvar University.

Unit – I
SOCIAL SCIENCE RESEARCH

Dr. G. YOGANANDHAM
Professor and Head,
Department of Economics,
Thiruvalluvar University.

Nature of Social Science Research

- Research is systematic and organized effort to investigate a specific problem that needs a solution.
- It contributes to the general body of knowledge.
- It also corrects human knowledge.
- Actually research is simply the process of arriving as dependable solution to a problem through the planned and systematic collection, analysis and interpretation of a data.
- Research is the most important process for advancing knowledge for promoting progress and to enable man to relate more effectively to his environment to accomplish his purpose and to solve his conflicts.
- Although it is not the only way, it is one of the most effective ways of solving problems. The term research consist of two words,' Re'+'Search'.
- "Re" means again and again and "Search" means to find out something.

Definition of Research

- **Black and Champion -** "Research consist of obtaining information through empirical observation that can be used for systematic development of logically related propositions attempting to establish casual relations among variable".

- **Emory** – "Any organized inquiry designed and carried out to provide information for solving a problem".

- **Kerlinger –** " A systematic, controlled, empirical and critical investigation of hypothetical relations among natural phenomena".

- **L.V. Redman and A.V. H. Morry -** " Systematic effort to gain new knowledge we call research".

Manipulation of Things, Concepts and Symbols

- While, dealing with things the scientist remains at the concrete level.
- He is able to purposefully handle things for experimentation.
- But at this level his results are at best limited to the particular thing in a specific situation and none else.
- Therefore the concepts symbolizing the things and their properties are also dealt with, so as to make much sense to conduct controlled inquiries through abstract notions.
- Use of concepts or symbols in the process of manipulation not only reduces the content and load of the things but also provides the scientist with greater facility and effect.

Generalization

- The sole purpose with which manipulation of things, concepts or symbols is undertaken is to arrive at statements of generality. It implies that the findings of controlled investigation should be a conclusion which will enable us to expect that under certain class of conditions influencing a class of things, something will happen in a generalized manner, notwithstanding its degree.
- But in any case the absence is generality cannot characterize science. Therefore the propositions derived on the basis of observations and through manipulation of things, concepts or symbols may vary in their levels of generality, may maintain a high or low degree but should never reach the null point.

Verification of Old Facts

- A major purpose of social research is verification of conclusions which have already been accepted as established facts.
- Since there is no place for complacency in the arena of science, the established system of knowledge always warrant frequentative scrutiny so as to confirm whether or not the observations are in accordance with the predictions made on the basis of the established corpus of knowledge.
- In case it is confirmed, the empirical observation strengthens the established system of knowledge.
- Otherwise in the light of the research outcome, the system of established corpus of knowledge calls for revision or even rejection.

Observe and Identity - I

- Businesses today have so much data that it can be difficult to know which questions to address first.
- Researchers also have business stakeholders who come to them with problems they would like to have explored.
- A researcher’s job is to sift through these inputs and discover the higher-level trends that are worth the investment of resources.
- This often means asking questions and doing some initial investigation to decide which avenues are worth pursuing further.
- That could mean talking to cross-functional teams across your business, or going outside your organization for additional expertise and contextual information from the wider industry.

Observe and Identity - II

- Sometimes, a small-scale preliminary study might be worth doing to help get a better understanding of the business context and needs, and to make sure your research question addresses the most critical problems.
- This could take the form of a few in-depth interviews, an environmental scan, or a literature review.
- **Example:** The sales manager of a sportswear company has a problem: sales of trail running shoes are down year-on-year and she isn't sure why.
- She approaches the company's research team for input and they begin to explore by asking question within the company and reviewing their knowledge of the wider market.

Extension of Knowledge

- As a sequel to generalization the seemingly inconsistencies in the existing corpus of knowledge are brought into light and attempts are made to reconcile these inconsistencies. The new general proposition, established as an outcome of research also identifies gaps in the established system of knowledge. A gap in knowledge implies the inadequacy of the theory as well as the failure of a conceptual scheme to explain and account for certain aspects of a social phenomenon.
- The gap is bridged up in the light of the new empirical observations. Thus knowledge gets expanded. The expansion of systematic knowledge occurs at least in a couple of ways.
- First in cognizing certain aspects of phenomena which were not examined in these terms prior to the advent of the new general proposition.
- Secondly in the light of new observation, the phenomena under investigation may be incorporated in a comparatively large class of phenomena, so as to be governed by a uniform law.
- As a result, the new system of knowledge not only accumulates more units under its conceptual scheme, but also appreciates greater depth of understanding and bettering of predictions.

Knowledge may be used for theory building or practical applications

- By seeking to explain the unexplained social phenomena, clarifying the doubtful one and correcting the misconceived facts relating to it, social science research provides the scope to use the fruits of research in two possible ways:
- **Theory building.**
- **Practical application.**
- In its basic or pure form social research gathers knowledge for the sake of it, for building a theory in order to explain human behaviour in its totality, only for the satisfaction of knowing. For construction of theoretic models, the researcher organizes knowledge into propositions and then meaningfully articulated those propositions to constitute a more abstract conceptual system pertaining to a class of phenomena, influenced by a certain class of conditions.
- In its practical or applied form, social research gathers information regarding the betterment of quality of life in social settings. The findings of social research are used as the means to an end, not construed just as an end in itself From its utilitarian point of view the results of social research provide decision makers with proper guidelines for policy making, social welfare, amelioration of practical problems, mitigation or resolution of social conflict and tensions as well as rectification and removal of social evils.

Define the Population

- Research projects can focus on a specific group of people, facilities, park development, employee evaluations, programs, financial status, marketing efforts, or the integration of technology into the operations.
- For **example**, if a researcher wants to examine a specific group of people in the community, the study could examine a specific age group, males or females, people living in a specific geographic area, or a specific ethnic group.
- Literally thousands of options are available to the researcher to specifically identify the group to study. The research problem and the purpose of the study assist the researcher in identifying the group to involve in the study.
- In research terms, the group to involve in the study is always called the population. Defining the population assists the researcher in several ways. First, it narrows the scope of the study from a very large population to one that is manageable.
- Second, the population identifies the group that the researcher's efforts will be focused on within the study. This helps ensure that the researcher stays on the right path during the study.
- Finally, by defining the population, the researcher identifies the group that the results will apply to at the conclusion of the study.

Develop the Instrumentation Plan

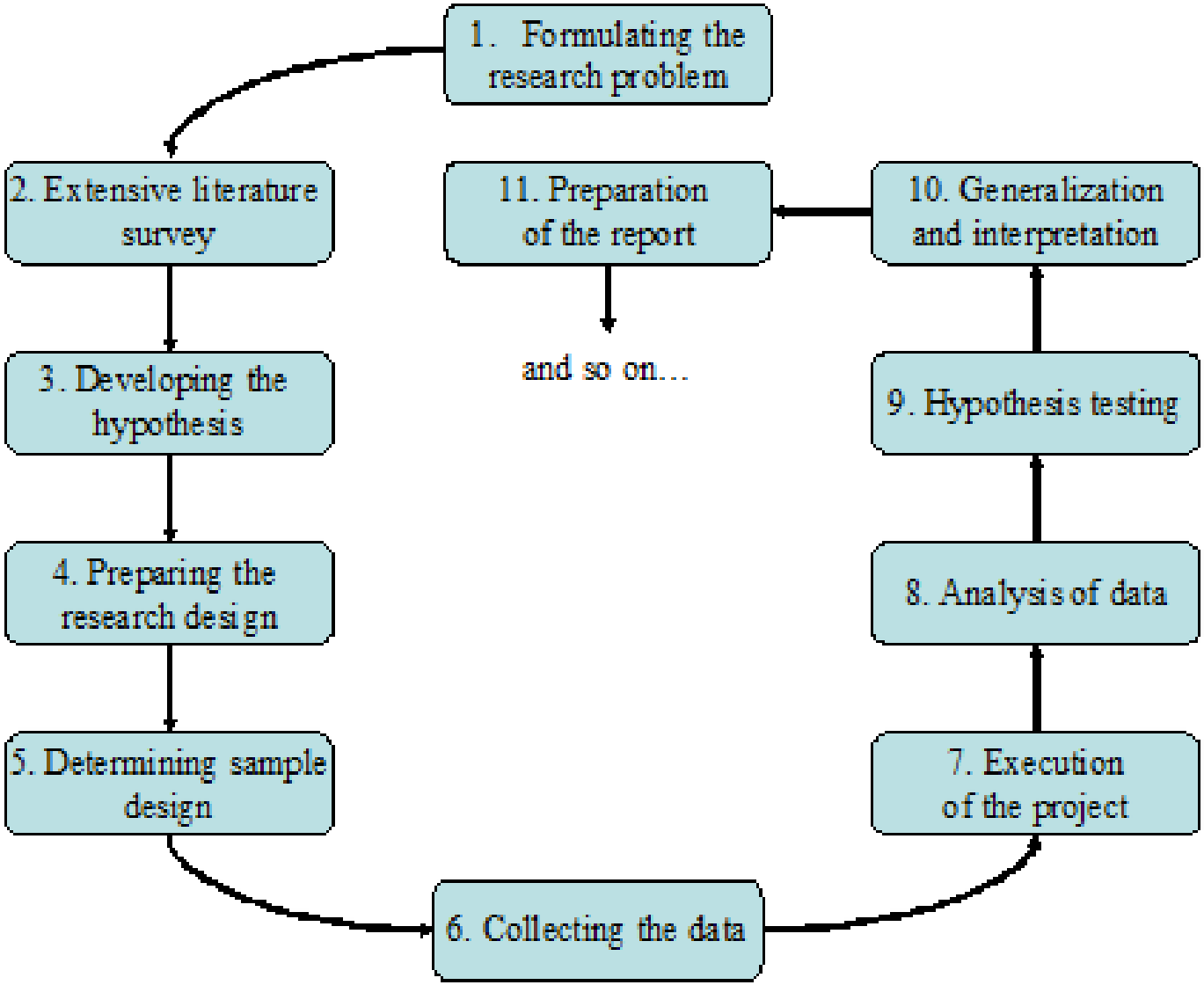
- The plan for the study is referred to as the instrumentation plan. The instrumentation plan serves as the road map for the entire study, specifying who will participate in the study; how, when, and where data will be collected; and the content of the program.
- This plan is composed of numerous decisions and considerations that are addressed in chapter 8 of this text. In the obesity study, the researcher has decided to have the children participate in a walking program for six months.
- The group of participants is called the sample, which is a smaller group selected from the population specified for the study. The study cannot possibly include every 10- to 12-year-old child in the community, so a smaller group is used to represent the population.
- The researcher develops the plan for the walking program, indicating what data will be collected, when and how the data will be collected, who will collect the data, and how the data will be analyzed.
- The instrumentation plan specifies all the steps that must be completed for the study. This ensures that the programmer has carefully thought through all these decisions and that she provides a step-by-step plan to be followed in the study.

7 Steps of Research Process

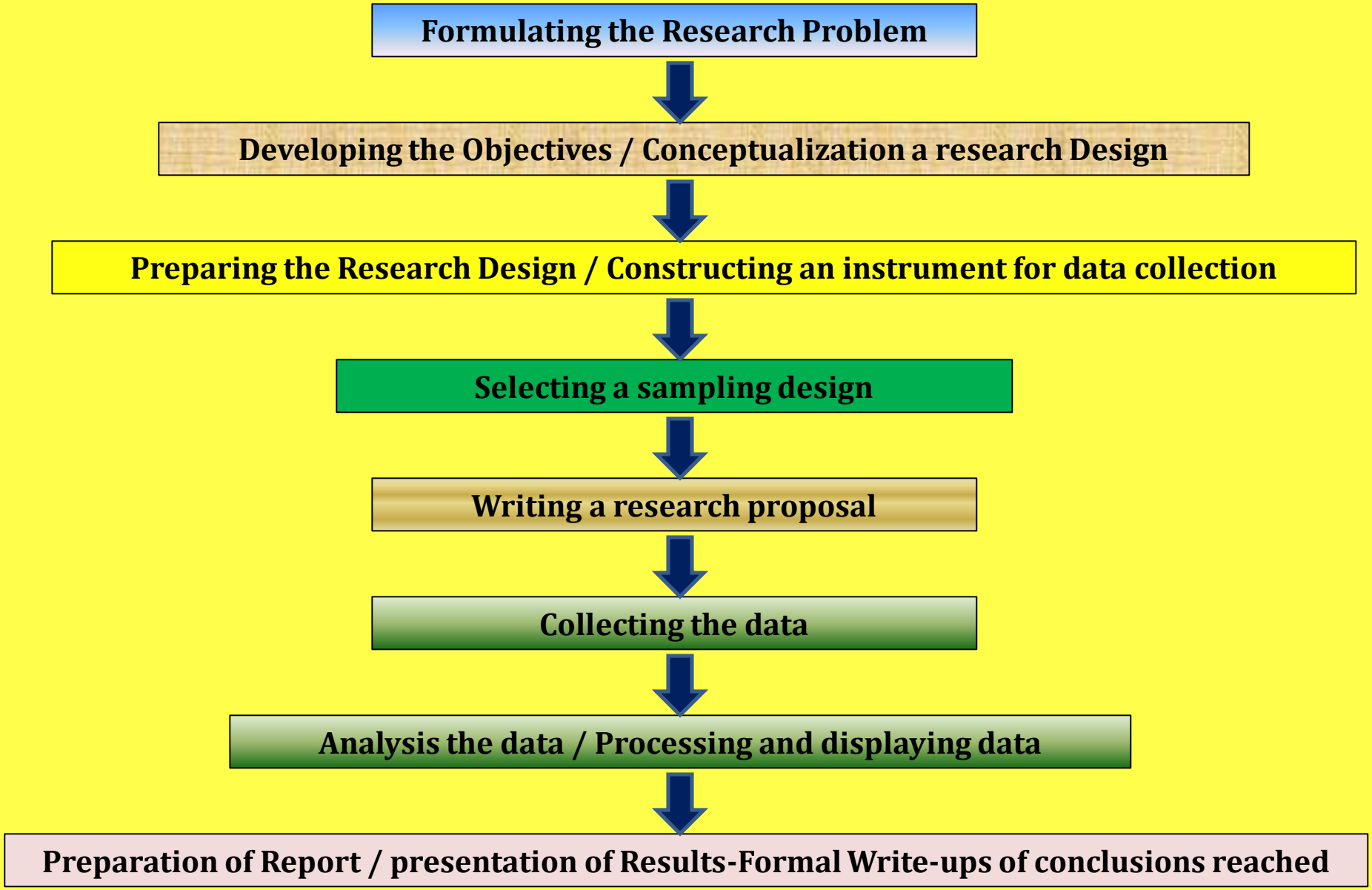
- Define Research Problem.
- Review of Literature.
- Formulate Hypotheses.
- Preparing the Research Design.
- Data Collection.
- Data Analysis.
- Interpretation and Report Writing



Steps in Research Process



Steps in Research Process



Main steps in Research - I

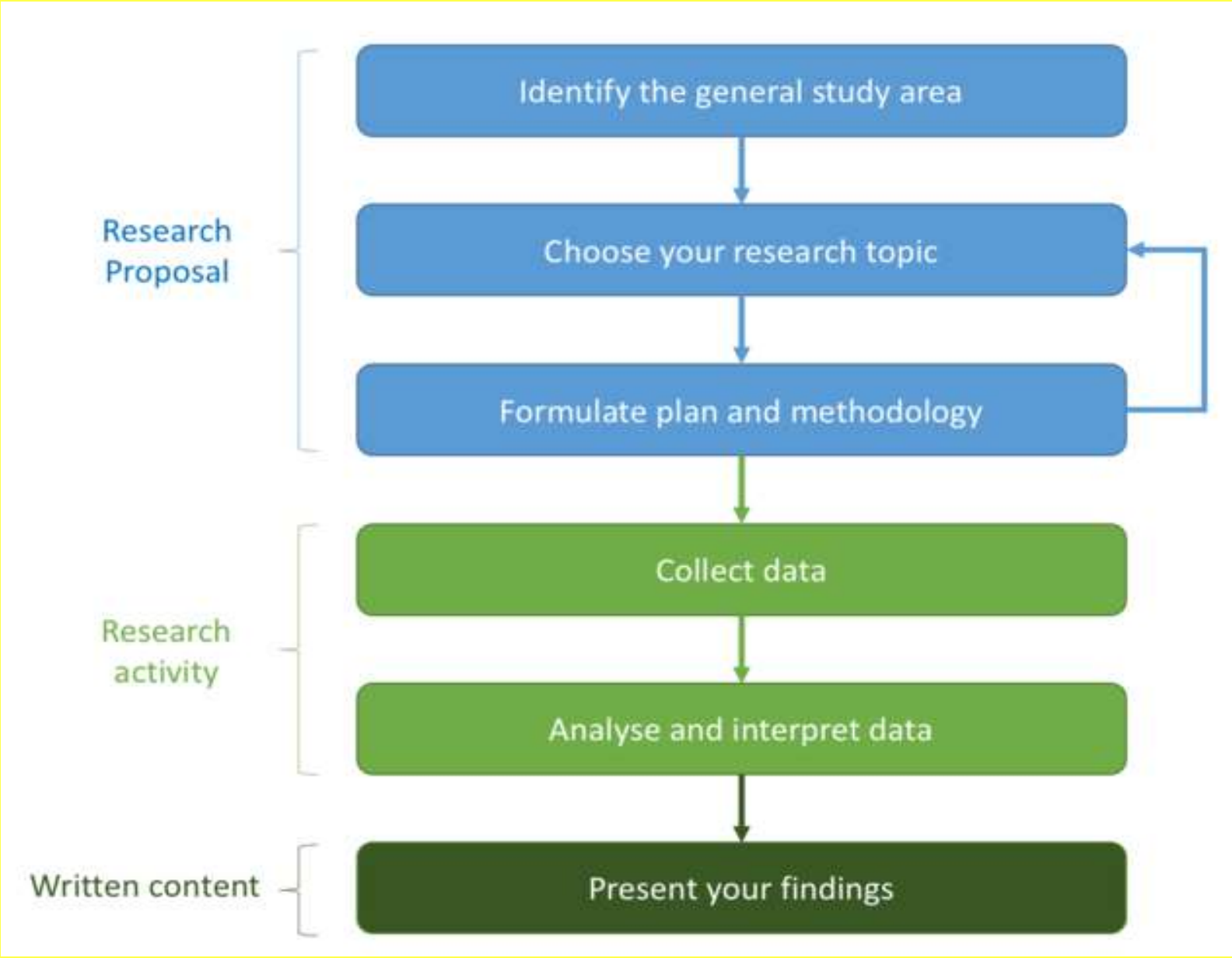
- Scientific research involves a systematic process that focuses on being objective and gathering a multitude of information for analysis so that the researcher can come to a conclusion.
- This process is used in all research and evaluation projects, regardless of the research method (scientific method of inquiry, evaluation research, or action research).
- The process focuses on testing hunches or ideas in a park and recreation setting through a systematic process. In this process, the study is documented in such a way that another individual can conduct the same study again.
- This is referred to as replicating the study. Any research done without documenting the study so that others can review the process and results is not an investigation using the scientific research process.
- The scientific research process is a multiple-step process where the steps are interlinked with the other steps in the process. If changes are made in one step of the process, the researcher must review all the other steps to ensure that the changes are reflected throughout the process.
- Parks and recreation professionals are often involved in conducting research or evaluation projects within the agency. These professionals need to understand the eight steps of the research process as they apply to conducting a study.

Main steps in Research - II

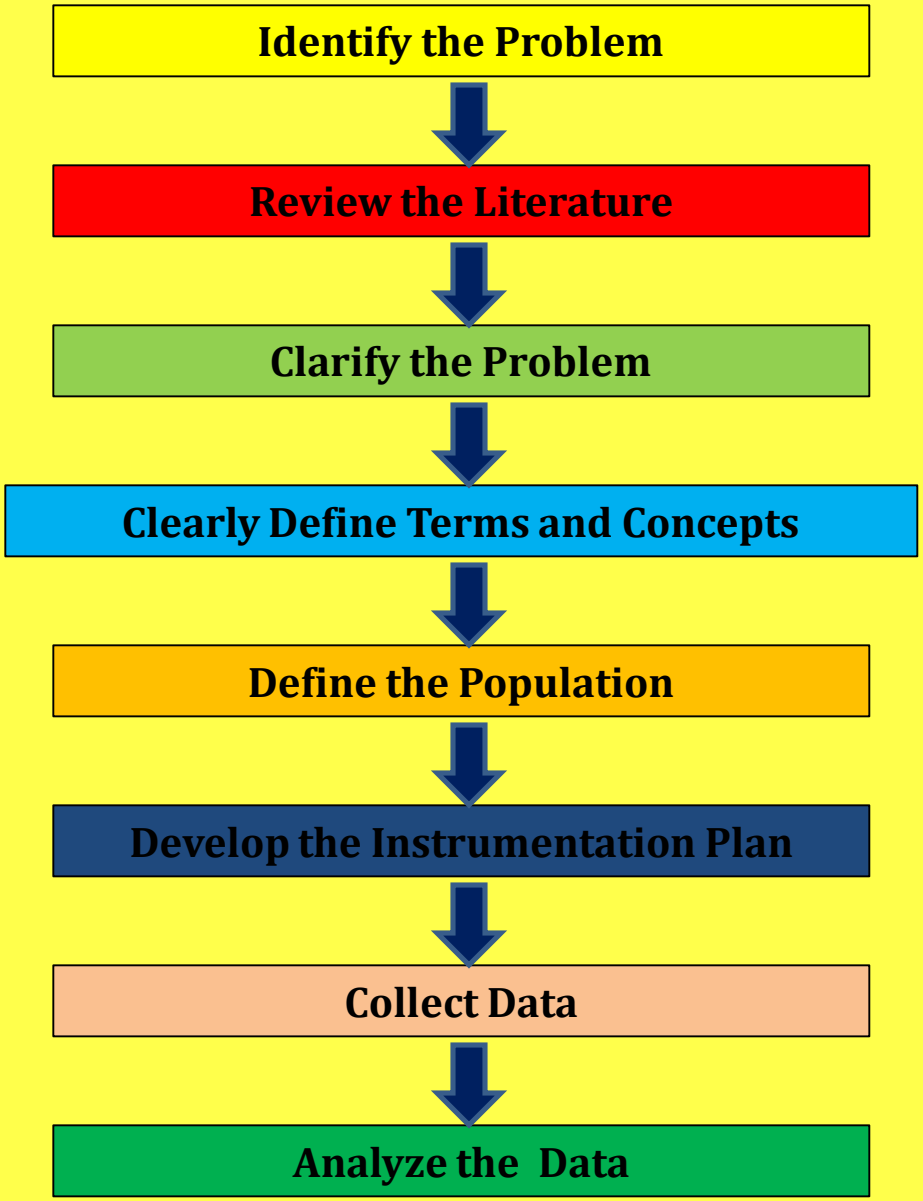
- **Identify and Develop your Topic.**
- **Find Background Information.**
- **Gather Data.**
  - ✓ Use catalogs to find books.
  - ✓ Use indexes to find periodical articles.
  - ✓ Find internet resources.,
- **Analyze/Evaluate what you find.**
- **What is your Thinking?**
- **Cite what you find**



Main steps in Research - III



Main steps in Research - IV



Identify the Problem

- The first step in the process is to identify a problem or develop a research question. The research problem may be something the agency identifies as a problem, some knowledge or information that is needed by the agency, or the desire to identify a recreation trend nationally.
- In the example in table 2.4, the problem that the agency has identified is childhood obesity, which is a local problem and concern within the community. This serves as the focus of the study.



Clarify the Problem

- Many times the initial problem identified in the first step of the process is too large or broad in scope. In step 3 of the process, the researcher clarifies the problem and narrows the scope of the study.
- This can only be done after the literature has been reviewed. The knowledge gained through the review of literature guides the researcher in clarifying and narrowing the research project.
- In the example, the programmer has identified childhood obesity as the problem and the purpose of the study.
- This topic is very broad and could be studied based on genetics, family environment, diet, exercise, self-confidence, leisure activities, or health issues.
- All of these areas cannot be investigated in a single study; therefore, the problem and purpose of the study must be more clearly defined.
- The programmer has decided that the purpose of the study is to determine if walking 10,000 steps a day for three days a week will improve the individual's health. This purpose is more narrowly focused and researchable than the original problem.

Clearly Define Terms and Concepts

- Terms and concepts are words or phrases used in the purpose statement of the study or the description of the study. These items need to be specifically defined as they apply to the study.
- Terms or concepts often have different definitions depending on who is reading the study. To minimize confusion about what the terms and phrases mean, the researcher must specifically define them for the study.
- In the obesity study, the concept of “individual's health” can be defined in hundreds of ways, such as physical, mental, emotional, or spiritual health. For this study, the individual's health is defined as physical health.
- The concept of physical health may also be defined and measured in many ways. In this case, the programmer decides to more narrowly define “individual health” to refer to the areas of weight, percentage of body fat, and cholesterol.
- By defining the terms or concepts more narrowly, the scope of the study is more manageable for the programmer, making it easier to collect the necessary data for the study. This also makes the concepts more understandable to the reader.

Formulation of Research Problem - I

- Any question that we want answered and any assumption or assertion that we want to challenge or investigate can become a research problem or a research topic for our study.
- However, it is important to remember that not all questions can be transformed into research problems and some may prove to be extremely difficult to study.
- According to Powers, Meenaghan and Twoomey 'Potential research questions may occur to us on a regular basis, but the process of formulating them in a meaningful way is not at all an easy task.'
- As a newcomer it might seem easy to formulate a problem but it requires considerable knowledge of both the subject area and research methodology.
- Once we examine a question more closely we will soon realize the complexity of formulating an idea into a problem which is researchable.
- It is essential for the problem we formulate to be able to withstand scrutiny in terms of the procedures required to be undertaken.
- Hence we should spend considerable time in thinking it through.

Formulation of Research Problem - II

- The first step in research is formulating a research problem, at this step, your goal should be clear that what you intend to achieve out of research.
- It is the most crucial step in the research process.
- The main function is to decide what you want to find out about.
- The way you formulate a problem determines almost every step that follows.
- All other steps in the research process will be influenced by research problem identification.
- So its extremely important that you should consider time, cost, expertise you have in the field of research you have chosen.
- The process of reviewing the literature helps you to understand the subject area better and thus helps you to conceptualize your research problem clearly and precisely.
- It also helps you to understand the relationship between your research problem and the body of knowledge in the area.
- Sometimes you should also consider your guide/ supervisor expertise knowledge.

Considerations in Selecting a Research Problem - I



- **Interest**
- **Magnitude**
- **Measurement of concepts**
- **Level of expertise**
- **Relevance**
- **Availability of Data**
- **Ethical Issues**
- **Interest**
- It should be the most important consideration.
- It keeps the researcher interested in it throughout the research process.
- **Magnitude**
- Not too broad or too big.
- Not too narrow.
- Manageable, specific and clear.



Considerations in Selecting a Research Problem - II

- **Measurement of concepts**
- If a concept is used in the study , be clear about its indicators and their measurement.
- **Level of expertise**
- It fits the level of researcher's level of research skills, needed resources, and time restrictions.
- **Relevance**
- It contributes to the improvement and understanding of educational theory and practice.
- **Availability of Data**
- If you want to collect information from secondary sources (office record, client records, already – published reports) , make sure that these data are available and in the format you want.
- **Ethical Issues**
- Do no harm to person /society involved.

Steps in Formulating a Research Problem

- The formulation of a research problem is the most crucial part of the research journey as the quality and relevance of your research project entirely depends upon it.
- The process of formulating a research problem consists of a number of steps. Working through these steps presupposes a reasonable level of knowledge in the broad subject area within which the study is to be undertaken and the research methodology itself.



Step-1: Identify a broad field or subject area of interest to you

- It is a good idea to think about the field in which we would like to work after graduation.

- This will help us to find an interesting topic, and one which may be of use to us in the future it is a good idea to think about the field in which we would like to work after graduation. This will help us to find an interesting topic, and one which may be of use to us in the future.

- **Example:** As a student of public health, intending to work with patients who have HIV/AIDS, you might like to conduct research on a subject area relating to HIV/AIDS.

Step-2: Dissect the broad areas into subareas

- We can select any subject area from the field such as community health or consumer research and go through the dissection process.
- In preparing this list of subareas we should also consult others who have some knowledge of the area and the literature in our subject area.
- Once we have developed an exhaustive list of the subareas from various sources, we proceed to the next stage where we select what will become the basis of your enquiry.

Step-3: Select what is of most interest to you

- It is neither advisable nor feasible to study all subareas. Out of this list, we should select issues or subareas about which we are passionate.
- This is because our interest should be the most important determinant for selection.
- One way to decide what interests us most is to start with the process of elimination. We should go through our list and delete all those subareas in which we are not very interested.
- We need to continue until we are left with something that is manageable considering the time available to us, our level of expertise and other resources needed to undertake the study.
- Once we are confident that we have selected an issue we are passionate about and can manage, we are ready to go to the next step.

Step - 4: Raise Research Questions

- At this step we ask ourselves, ‘What is it that I want to find out about in this subarea?’

- We should make a list of whatever questions come to our mind relating to our chosen subarea and if we think there are too many to be manageable, go through the process of elimination, as we did in Step 3.

Step - 5: Formulate Objectives

- Both our main objectives and our sub-objectives now need to be formulated, which grow out of our research questions. The main difference between objectives and research questions is the way in which they are written.
- Research questions are obviously that – questions. Objectives transform these questions into behavioral aims by using action-oriented words such as ‘to find out’, ‘to determine’, ‘to ascertain’ and ‘to examine’.
- Some researchers prefer to reverse the process; that is, they start from objectives and formulate research questions from them.
- Some researchers are satisfied only with research questions, and do not formulate objectives at all.
- If we prefer to have only research questions or only objectives, this is fine, but we should keep in mind the requirements of our institution for research proposals.

Step - 6: Assess your Objectives

- Now, we examine our objectives to ascertain the feasibility of achieving them through our research endeavor. Consider them in the light of the time, resources (financial and human) and technical expertise at your disposal.

Step – 7 : Double - Check

- We should go back and give final consideration to whether or not we are sufficiently interested in the study, and have adequate resources to undertake it.
- We should ask ourselves, ‘Am I really enthusiastic about this study?’ and ‘Do I really have enough resources to undertake it?’ Answer these questions thoughtfully and realistically.
- If we answer to one of them is ‘no’, reassess your objectives.

Formulation of Research Objectives

- Objectives are the goals we set out to attain in our study. Since these objectives inform us of what we want to achieve through the study, it is extremely important to word them clearly and specifically.
- **Objectives should be listed under two headings:**
  - Main objectives
  - Sub-objectives
- The **main objective** is an overall statement of the thrust of our study. It is also a statement of the main associations and relationships that we seek to discover or establish.
- The **sub-objectives** are the specific aspects of the topic that we want to investigate within the main framework of our study.

Importance of Formulating a Research Problem-I

- Formulation of research problem is the first and important step of research process. It is like the identification of a destination before undertaking a journey.

- A research problem is like the foundation of a building. The type and design of the building are dependent upon the foundation. If the foundation is well designed and strong you can expect the building to be also.

- The research problem serves as the foundation of a research study: if it is well formulated, you can expect a good study to follow.

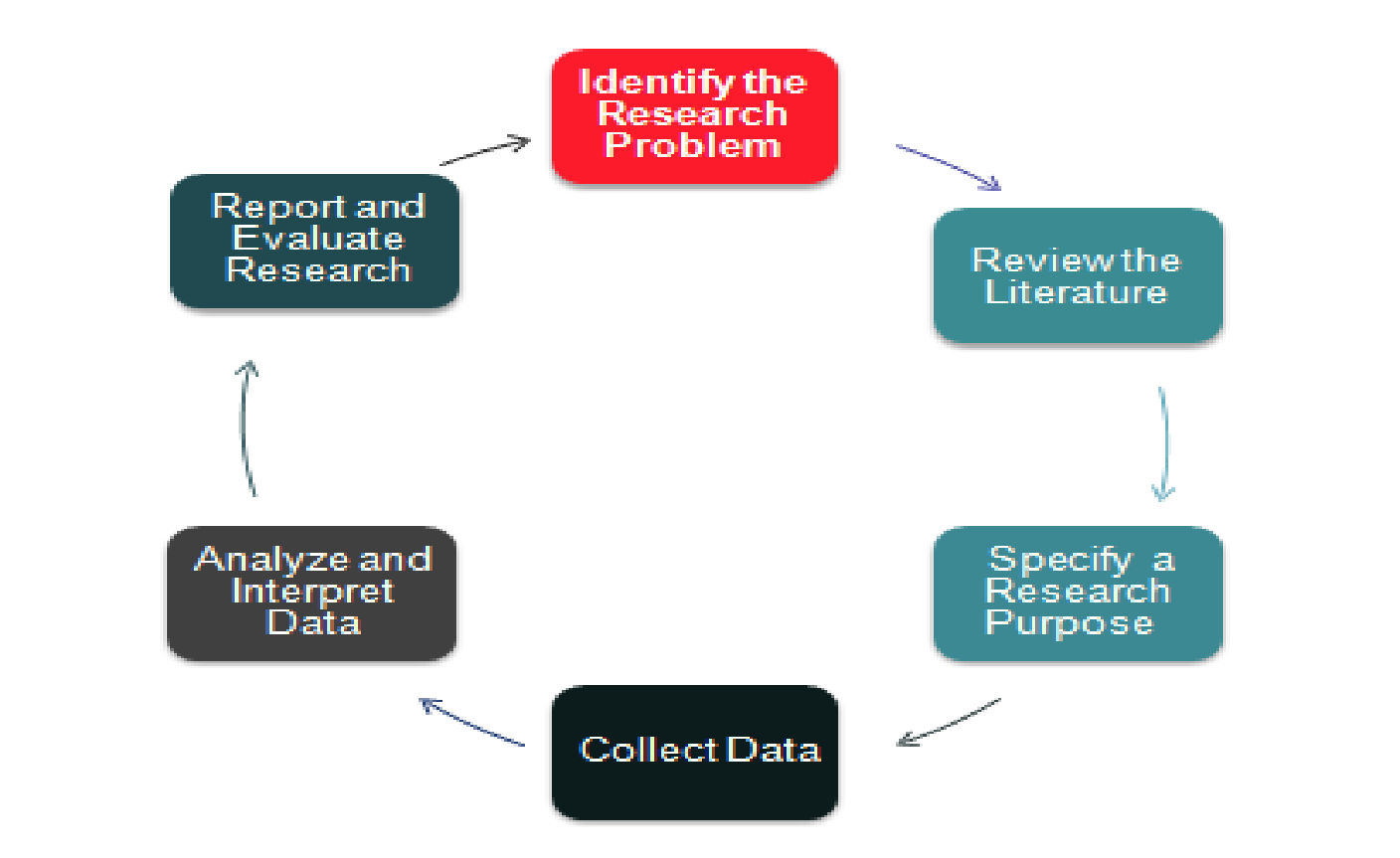
- **According to Kerlinger:**
- ***"If one wants to solve a problem, one must generally know what the problem is. It can be said that a large part of the problem lies in knowing what one is trying to do."***

Importance of Formulating a Research Problem-II

- **A research problem may take a number of forms, from the very simple to the very complex.**
- **The way you formulate a problem determines almost every step that follows:**
  - The type of study design that can be used.
  - The type of sampling strategy that can be employed.
  - The research instrument that can be used or developed.
  - The type of analysis that can be undertaken.
- **It is important for you to understand that the way you formulate a research problem determines all the subsequent steps that you have to follow during your research journey.**

Selection of a Research Problem

- Definition of a Problem.
- Selection of the Problem.
- Evaluation of the Problem.
- Refinement of the Research Problem.
- **Identification of Research Problem**



Process of Research - Selection of a Research Problem

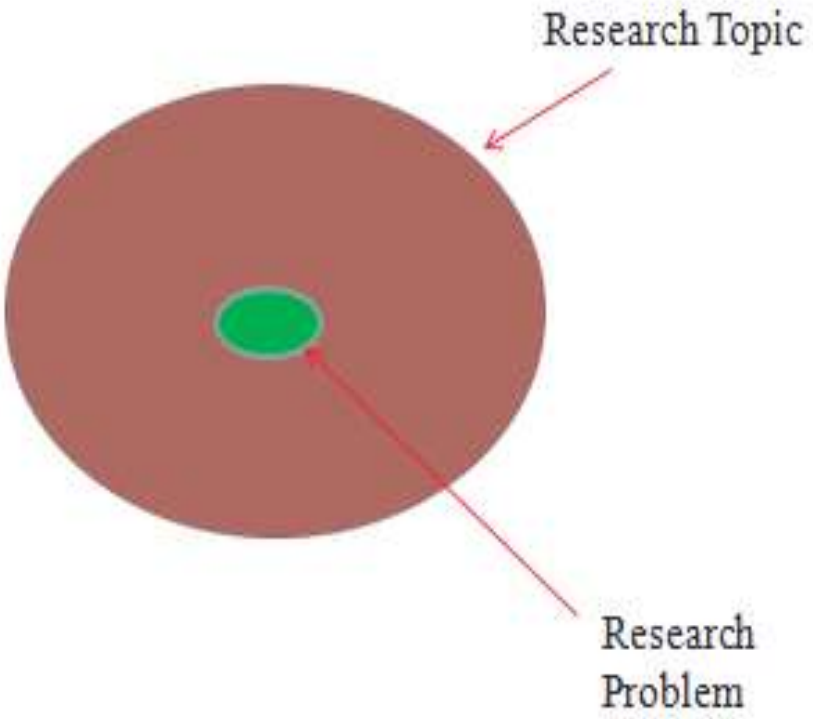


Research Topic to a Research Problem

- Idea

  ↓

  brainstorming
- Literature review
- Identify the variables for study
- Formulate research problems and questions/hypotheses



Sources of the Research Problems

- Previous researches/dissertations/studies/etc.
- Books/reports.
- Resource persons/professors.
- Brainstorming.
- Newspapers/TV/radios/medias.
- Seminars, workshops/conferences.
- Relations organizations/research centers/research institutes.
- Encyclopedia.
- Own critical observations.
- Discussions with fellows, colleagues, friends, etc.

Definition of a Research Problem

- A research problem is the situation that causes the researcher to feel apprehensive, confused and hard.
- Operationalization is then used to give some indication of the exact definitions of the variables, and the type of scientific measurement used.
- Operationalization is the process of strictly defining variables into measurable factors. The process defines ambiguous concepts and allows them to be measured, empirically and quantitatively.
- For example, intelligence may be measured with IQ and human responses could be measured with a questionnaire from '1- strongly disagree', to '5 - strongly agree'.

Selection of a Research Problem - I

- First step in research process is the choice of a suitable problem.
- Enlist the specific questions like WHO, WHAT, WHERE, WHEN and WHY of the problem.
- A problem well defined is half solved.
- The formulation of the problem is often more essential than its solution.
- The actual reasons (possible causes) for the discrepancy between what is (current situation) and what is desired (e.g. lack of motivation among students)

Selection of a Research Problem - II

- Criteria/Factors of selecting the problems.
- Academic background of the researcher.
- Researcher's experiences, aptitudes, ability , values , behavior etc.
- Researcher's training and orientation.
- Researcher's purpose and objectives.
- Originality and Utility of the research.
- Direction of the organization where the researcher works.
- Availability of data and information – primary and secondary Sources.
- Access to the study area.
- Resource availability – money, manpower.

Selection of a Research Problem - III

- Time factor.
- Supervisor's time, qualification, experiences, interest, etc.
- Nature of the problems – need oriented, problem solving, etc.
- Controversial topic.
- Nor too narrow/vague.
- Cooperation of the stakeholders.
- Avoidance of the troubles.





Selection of a Research Problem - III

- Time factor.
- Supervisor's time, qualification, experiences, interest, etc.
- Nature of the problems – need oriented, problem solving, etc.
- Controversial topic.
- Nor too narrow/vague.
- Cooperation of the stakeholders.
- Avoidance of the troubles.





Evaluation of a Research Problem - I

- Is the problem of current interest? Will the research results have social, educational or scientific value?
- Will it be possible to apply the results in practice?
- Does the research contribute to the science of education?
- Will the research opt new problems and lead to further research?
- Is there enough scope left within the area of research (field of research)?
- Will it be possible for another researcher to repeat the research?
- Will it have any value? and are you motivated to undertake the research?
- Do you have the necessary knowledge and skills to do the research? Are you qualified to undertake the research?
- Do you have the necessary funds for the research?
- Do you have access to the administrative, statistic and computer facilities the research necessitates?
- Will it be practically possible to undertake the research?
- Is the research free of any ethical problems and limitations?

Evaluation of a Research Problem - II

- **In short**
- Is the problem researchable?
- Is the problem new?
- Is the problem significant?
- Is the solution to this problem helpful for the development of further knowledge?
- Is the problem feasible for researcher?
- Research competencies.
- Interest and enthusiasm.
- Financial consideration.
- Time requirements.
- Administrative considerations.

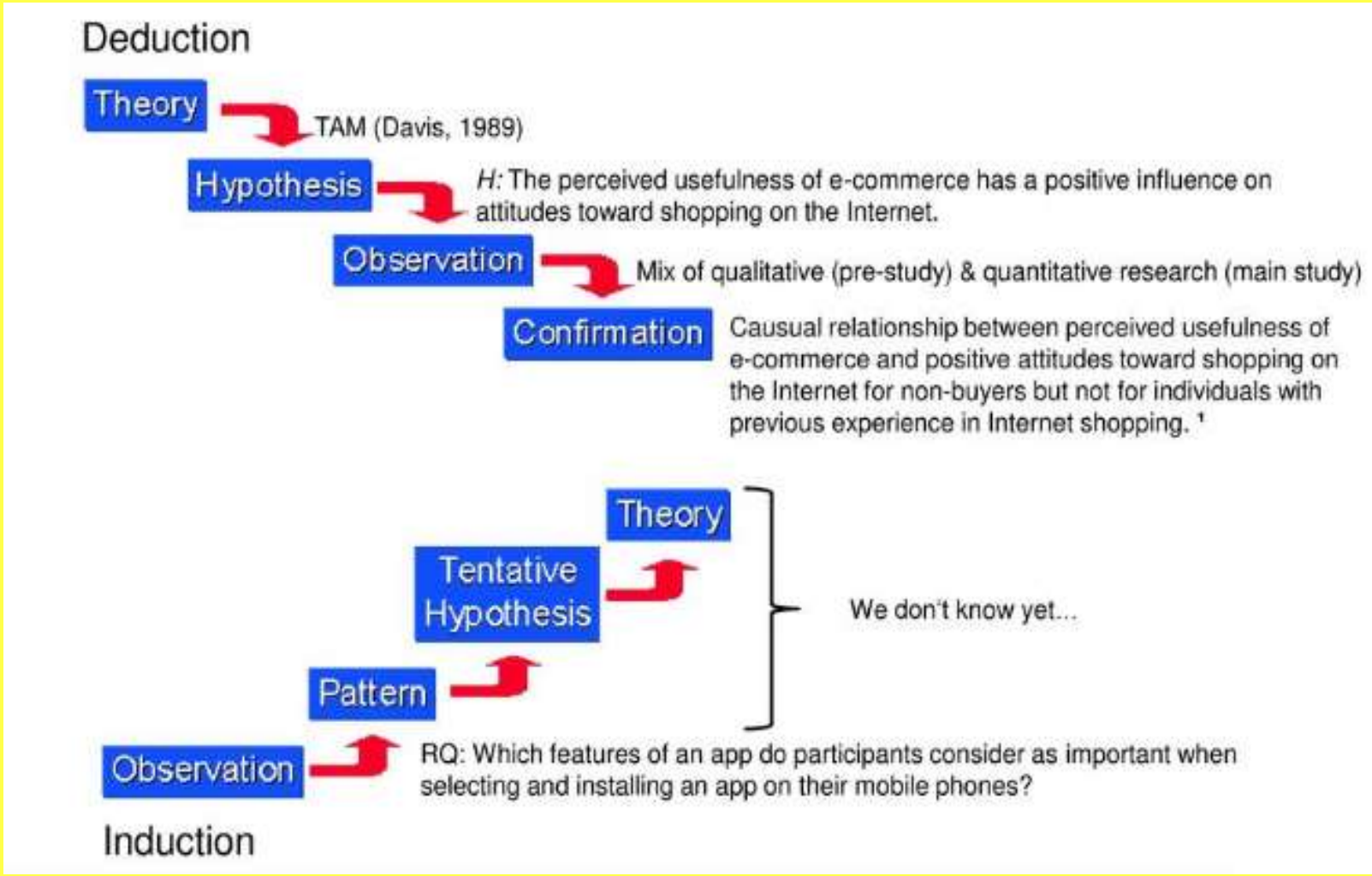


Consideration in Selecting a Research Problem

- When selecting a research problem there are a number of considerations to keep in mind which will help to ensure that your study will be manageable and that you remain motivated.
- **These considerations are:**
- **Interest**
- **Magnitude**
- **Measurement of concepts**
- **Level of expertise**
- **Relevance**
- **Availability of data**
- **Ethical issues**



Introduction to Social Science Research



Social Science Research-I

- Social sciences are not exact science like physical sciences.
- It deals with human beings.
- Human nature and mans environment are so complex, that it is more difficult to comprehend and predict human behavior than the physical phenomena.
- Social science research is a systematic method of exploring, analyzing and conceptualizing human life in order to extend, correct or verify knowledge of human behavior and social life.
- Social research seeks to find explanations to unexplained phenomena, to clarify the doubtful and correct the misconceived fact of social life.
- It involves the application of scientific method for understanding and analyzing of social life in order to correct and verify the existing knowledge as a system.

Social Science Research-II

- The main idea behind social research is to discover new inter relations, new knowledge, new facts and also to verify old ones.
- Human behavior may be involved by certain values and laws.
- The main purpose of social research is to discover those laws which can be proper guidelines for studying human contact and behavior.
- It is difficult see the underlying uniformities in the diversity of complex human behavior.
- Social research can be defined as the systematic and objective analysis and recording of controlled observations that may lead to the development of generalization, principles or theories resulting in prediction and possibly ultimate control of events in society.
- It attempts to answer or solve social problems

Definitions of Social Science Research

- **C. A. Moser :** “Social research is a systematized investigation to gain new knowledge about social phenomenon and problems.”
- **P.V. Young:** “Social research is a scientific undertaking which by means of logical methods, aim to discover new facts or old facts and to analyze their sequences, interrelationships, casual explanations and natural laws which govern them.”



Objectives of Social Science Research - I

- To facilitate the understanding of human behavior.
- To acquire knowledge about social phenomena, events, issue, problems etc.
- To identify functional relationship existing in the social phenomena.
- To find out the natural laws that regulates or directs social phenomena.
- To standardize the society concept, e.g. culture, struggle, generation gap, social distance etc.
- To formulate solution to social problems.
- To maintain social organization, remove social tension, misconception, etc.
- To develop social revival plan.

Objectives of Social Science Research - II

- There are five major objectives of social research,
- **Manipulation of Things, Concepts and Symbols,**
- **Generalization,**
- **Verification of Old Facts,**
- **Extension of Knowledge, and**
- **Knowledge May be Used for Theory Building or Practical Application.**





Characteristics of Social Science Research-I

- It is directed towards the solution of problems. The ultimate goal is to discover cause-and-effect relationship between social problems.
- It emphasis the development of generalizations, principles or theories that will be helpful in predicting future occurrences.
- It is based upon observable experience or empirical evidence.
- It demands accurate observations and description. Researchers may choose from a variety or non-qualitative description of their observations.
- It involves gathering new data from primary sources or using existence data for new purpose.
- Although social research activities may at time be somewhat random and unsystematic, it is more often characterized by carefully designed procedure that applies rigorous analysis.
- It requires expertise. The researcher knows what is already known about the problem and how others have investigated.

Characteristics of Social Science Research-II

- It strives to the objective and logical applying every possible test to validate the procedure employed, data collected and conclusion reached.
- It involves the guests for answer to unsolved problems.
- It is characterized by patient and unhurried activity. Researcher must expect disappointment and discouragement as they pursue the answer to difficult question.
- It is carefully recorded and reported. Each important term is defined, limiting factors are recognized, procedures are described in detail, reference are carefully documented, results are objectively recorded and conclusions are presented with scholarly caution and restraint.
- It is interdisciplinary in nature.
- It sometimes requires courage.
- It finds the cause of such issues and provides a solution to resolve the same. It also brings structural change in the life of people by studying more about the **social** situations. The **research** brings about a positive change in the society. The societal behavior is studied and analyzed to achieve broad goals.

Types of Social Science Research

- **Basic research:**

  It is also called fundamental research. It is undertaken to improve our understanding of certain problems that commonly occur in social setting and how to solve them. It undertaken for sole purpose of adding to our knowledge that is fundamental and generalisable. This type of research may have no immediate or planned application.

- **Applied research:**

  It is also called action or decisional research. It is undertaken in response to a social problem, which requires a solution. Its major purpose is to answer practical and useful question. The results are practically applied to solve immediate problems. It involves normative prescription. As applied research id concerned with knowledge that has immediate application. It is also called decisional research.

Functions of Social Science Research-I

- **Discovery of facts and their interpretation:** Social research provides answer to questions of what, when, how and why of man, social life and institutions. Discover of facts and their inter relationship help us to discard distortions and contribute to our understanding of social reality.

- **Diagnosis of problems and their analysis:** Our society has innumerable problems such as poverty, unemployment, economic inequality, social tension etc,. The nature and dimensions of such problems have to be diagnosed and analyzed. An analysis of problems leads to an identification of appropriate remedial actions.

- **Systematization of knowledge:** The facts discovered through research are systematized and the body of knowledge is developed. It contributes to the growth of theory building.

Functions of Social Science Research-II

- **Control over social phenomena:** Research in social science provides first hand information about the nature of social institutions. This knowledge helps us to control over the social phenomena.

- **Prediction:** Social research aims at finding an order among social fact and their casual relations. This affords a sound basis for prediction in several cases.

- **Development planning:** Systematic research can give us the required data base for planning and designing developmental schemes and programs.

- **Social welfare:** Social research can identify the causes of social evils and problems. It can thus help in taking appropriate remedial actions. It also provides guideline for social welfare..

Scope of Social Science Research

- The fields of social science research unlimited and the materials of research are endless. Every group of social phenomena, every phase of human life and every stages of past and present development are materials for the social scientist. The area of research in various social sciences provides vast scope for research in social sciences.
- **The main scope of social research are:**
- Social Science Research provides new insight in to the organized society and its social structure.
- It also provides new horizon in scientific explanation, advanced and tested principles of procedure and suggested new concepts.
- Another scope of social science research is that exemplified by studies and attempt to test or challenge existing theories and revise them the light of new evidence.
- It is helpful to establish new theory and established techniques of exploration.
- It is also provides contributions to existing stone of fruitful ideas, methodology and basic understanding of social life and control of its problems.

Importance of Social Science Research - I

- Social research makes social unity possible by providing solutions to the social groups and castes.
- It finds the cause of such issues and provides a solution to resolve the same.
- It also brings structural change in the life of people by studying more about the social situations.
- The research brings about a positive change in the society.
- The societal behavior is studied and analyzed to achieve broad goals.
- Social research also contributes to societal development by increasing creativity and innovation.
- The new idea leads to the up-gradation of society.
- Focusing on research enhances the knowledge thereby giving more power to the society.
- Creativity and innovation help in formulating new theories that help in the upliftment of the societal behavior.
- It changes the already accepted principles through its new research methodologies.

Importance of Social Science Research - II

- Social research studies the society and helps in making laws for the benefit of the government.
- It helps in smooth functioning of the society by bringing social order and control.
- The research studies the trends of the past and predicts the future.
- The prediction of the future trends many tasks, may it be businesses, government or the society.
- The government can make the rules by taking in consideration about the future trends and the businesses can design their strategies in the same way. There is never an end to the social science research.
- It helps diverse subject areas to achieve their motives. It plays an important role in all fields.
- Hence, the growth and development of a society are impossible without social science research.
- Social science research studies various theories on motivation, personality, leadership, team skills, etc. that are already accepted.

Research Approaches

- Research approaches are plans and the procedures for research that span the steps from broad assumptions to detailed methods of data collection, analysis, and interpretation.
- This plan involves several decisions, and they need not be taken in the order in which they make sense to me and the order of their presentation here.
- The overall decision involves which approach should be used to study a topic.
- Informing this decision should be the philosophical assumptions the researcher brings to the study; procedures of inquiry **(called research designs);** and specific research methods of data collection, analysis, and interpretation.
- The selection of a research approach is also based on the nature of the research problem or issue being addressed, the researchers' personal experiences, and the audiences for the study.
- Thus, in this book, research approaches, research designs, and research methods are three key terms that represent a perspective about research that presents information in a successive way from broad constructions of research to the narrow procedures of methods.

Research Approaches and Types - I

- There are three main research approaches:
- **Qualitative,**
- **Quantitative, and**
- **Mixed methods.**
- Unquestionably, the three approaches are not as discrete as they first appear. Qualitative and quantitative approaches should not be viewed as rigid, distinct categories, polar opposites, or dichotomies.
- Instead, they represent different ends on a continuum. A study tends to be more qualitative than quantitative or vice versa.
- Mixed methods research resides in the middle of this continuum because it incorporates elements of both qualitative and quantitative approaches.

Research Approaches and Types - II

- In a broad, interdisciplinary field such as Economics, Commerce, Management and Planning, research may be conducted in a number of ways. Five possible research approaches are suggested below. They are by no means mutually exclusive: a research project may include two or more of these approaches, or approaches other than those described.
- **Descriptive Study**
- **Explanatory Study**
- **Remedial Study**
- **Methodological Study**
- **Historical Study**

Research Approaches and Types - III

- **Qualitative research:** It is an approach for exploring and understanding the meaning individuals or groups ascribe to a social or human problem.
- The process of research involves emerging questions and procedures, data typically collected in the participant's setting, data analysis inductively building from particulars to general themes, and the researcher making interpretations of the meaning of the data.
- The final written report has a flexible structure.
- Those who engage in this form of inquiry support a way of looking at research that honors an inductive style, a focus on individual meaning, and the importance of rendering the complexity of a situation.

Research Approaches and Types - IV

- **Quantitative research:** It is an approach for testing objective theories by examining the relationship among variables.
- These variables, in turn, can be measured, typically on instruments, so that numbered data can be analyzed using statistical procedures.
- The final written report has a set structure consisting of introduction, literature and theory, methods, results, and discussion.
- Like qualitative researchers, those who engage in this form of inquiry have assumptions about testing theories deductively, building in protections against bias, controlling for alternative explanations, and being able to generalize and replicate the findings.

Research Approaches and Types - V

- **Mixed methods:** It is an approach to inquiry involving collecting both quantitative and qualitative data, integrating the two forms of data, and using distinct designs that may involve philosophical assumptions and theoretical frameworks.
- The core assumption of this form of inquiry is that the combination of qualitative and quantitative approaches provides a more complete understanding of a research problem than either approach alone.
- Overall these definitions have considerable information in each one of them.

Descriptive study

- This approach attempts to identify the characteristics of a problem through description. Because the subject cannot be described in all its detail, careful selection of facts must occur. Facts should be gathered according to pre-determined criteria and for the purpose of demonstrating relationships of interest. To the extent that the descriptive study of a particular problem provides one with a generalized understanding of a phenomenon that, in turn, can be employed to understand other specific problems, this approach is useful and acceptable.
- **Example:** This approach would be a description of an unusual planning program in operation in one Municipality, with the objective being to illustrate how the program differs from similar programs found in other Municipalities.

Explanatory study

- This approach attempts to find the answer to an enigmatic question.
- **Example:** why has an urban area acquired its particular shape, or why has a city council refrained from allowing mobile home parks in the community?
- The explanatory studies are designed to investigate origin – cause - effect relationships. The typical study includes the collection of empirical data for the formulation of hypotheses or less pretentious hunches and the subsequent test of these hypotheses by any one of a number of ways available to the researcher.

Remedial study

- The remedial study seeks to formulate plans to correct or improve undesirable social, economic, political, and environmental conditions.
- These undesirable conditions, causes and processes associated with them are analyzed. Plans, strategies, and policies are formulated to remedy the undesirable conditions.
- Implementation methods may be suggested. A remedial study may be directed to the solution of pollution of a lake ringed by cottages or to pedestrian and vehicular conflicts in the central business district.

Methodological study

- The methodological study attempts to devise, test or improve new research methods in Planning.
- The study may deal with the development of a specific technique for the discipline or may take a technique developed by another discipline and attempt to apply it to a Planning context. Possible examples of this approach are innumerable.
- **Example:** The application of new forecasting techniques developed in Management Science to a planning problem, or an attempt to develop methods for breaking down cross census information into smaller temporal or areal units.

Historical study

- If it is designed to facilitate a deeper understanding of historical processes and is not merely an attempt to fill gaps in our factual knowledge, the historical approach can be very useful.
- Studies detailing the transference of the modern town Planning movement from Europe to North America or the evolution of company towns in Ontario would be appropriate uses of the historical approach.



A Suggested Essay Format

- The following is a suggested outline for the essay. The structure may well vary, depending on the topic, the treatment given, and the advice of an advisor as how to most effectively present the material and research findings.
- **Abstract** (Mandatory)
- **Introduction**
  - Statement of problem
  - Significance of problem
  - Purpose of study
  - Assumptions and limitations
- **Review of preceding research and literature**
- **Survey of pertinent information**
  - Methodology and techniques employed
  - Sources of information, data
- **Analysis**
- **Summary and conclusions**
  - Recommendations for further research
- **References**
- **Appendices, if needed.**



Quantitative Research Approach – Data Analysis - I

- Statistical analysis is the usual method used in quantitative research approach. Quantitative data can be analyzed in several ways. Data collected has a certain level of measurements which initially influences the analysis. The identification of a particular level of measurement is the usually the first step in quantitative data analysis.
- **The four levels of measurements include:**
  - **Nominal data**: basic classification data; lack logical order - e.g. male or female
  - **Ordinal data**: has logical order but lack constant differences between values – e.g. Pizza size (large, medium, small)
  - **Interval data**: has logical order, is continuous, has standardized differences between values but lacks natural zero – e.g. Celsius degrees
  - **Ratio data**: has logical order, is continuous, has standardized differences between values, and has a natural zero – e.g. height, weight, age, length.

Quantitative Research Approach – Data Analysis - II

- After identifying a level of measurement, the next step is now to use a specific analysis technique in analyzing data. There are several procedures that can be used to analyze data.
- Main ones include :

- ***Data tabulation*** *(e.g. frequency distributions & percent distributions)*
- ***Data descriptives*** *(e.g. Mean, medium, mode, minimum and maximum values, etc.)*
- ***Data disaggregation*** *(tabulation of data across multiple categories)*
- ***Moderate and advanced analytical methods*** *(regression, correlation, variance analysis)*
- The above methods can be used invariably by all different true experimental, quasi-experimental and non-experimental quantitative research strategies.

Qualitative Research Approach – Data Analysis - I

- Textual data analysis is the usual method used in qualitative research approach. This involves identifying patterns and themes in data collected and then examining and interpreting these patterns and themes to draw meaning and answer research questions.
- The five strategies of qualitative research mentioned–ethnography, phenomenological, grounded theory, narrative and case studies–employ different analytical procedures. However, preliminary and some general steps in data analysis are common to all. These include:
- Immediate processing and recording of data.
- Commencement of data analysis soon after collection
- Reduction of data to meaningful information
- "Identification of meaningful patterns and themes" via
  - *Content analysis achieved by:*
    - *Coding the data for certain words or content*
    - *Identifying their patterns*
    - *Interpreting their meanings.*

Qualitative Research Approach – Data Analysis - II

- *Thematic analysis achieved by "grouping data into themes that answers research problem".*
- Display of data which include organizing data in forms of graphics, maps, tables, etc., to draw conclusions
- "Drawing of conclusion and verification"
- All the steps are should usually be accompanied by references to literature, triangulation, keeping the research purpose on focus and constant back-and-forth analysis between and among the steps.
- In most cases, analysis in qualitative research commences with **"emic" analysis** (analysis based on what a researchers knows or brings to a research study) and ends with **"etic" analysis** (analysis to derive meaning based on negotiations between what a researchers knows and what others or participants know).
- This means that subjectivity, training and personal experiences of a researcher, as well as other factors may influence the research process. For this reason, there is a need for innovative ways of research approaches to allow objectivity and comfortability.

Difference in Designing Qualitative and Quantitative Research

S. No.	Particulars	Qualitative Research	Quantitative Research
1.	Focus	Quality.	Quantity.
2.	Data	Words.	Numbers.
3.	Instruments	Researcher, Interview, Observation, Documents, Questionnaire.	Test, Observations.
4.	Sample	Small, Non-Random.	Large, Random.
5.	Design	Flexible.	Predetermined.
6.	Aim	Understand in depth.	Generalize.
7.	Findings	Comprehensive, Richly, Descriptive.	Precise, Numerical.

Unique Analysis Procedures and Strategy of Qualitative Research - I

- **Ethnography Research**
- Ethnography theory research approach focuses analysis on processes, patterns and trends that occur in individuals or across various groups regarding their cultural behavior but keeping in view the purpose of a research.
- While achieving the steps a research analysis adopts an iterative process in which cultural ideas that arise during active involvement during the research study are used to write a document while bearing in mind the context, norms, frequency and other factors.



Unique Analysis Procedures and Strategy of Qualitative Research - II

- **Phenomenological Research**
  Phenomenological theory research approach involves the process which allows the analysis method to follow the nature of data itself. This is called **emergent strategy.** Analysis focuses on the details that can be appreciated through an experience a subject lived.
- Some emergent strategies may include: Narratives *or interviews from artistic depictions, photo voice, etc., while paying attention to aspects such as:*
  - *Objects involved*
  - *Activities and their results*
  - *Time*
  - *Descriptive elements*
  - *Physical surroundings,*
  - *Characters or their aspects such as relationships*
  - *Social interactions*
  - *Results of activities .*

Unique Analysis Procedures and Strategy of Qualitative Research - III

- **Grounded Theory Research - 1**

  Grounded theory research approach involves the process of moving in and out of data collection and analysis processes, a process called **constant comparative analysis** or '**iteration**'. After this process, a researcher begins to develop a theory regarding the questions and data collected (Strauss & Corbin, 1994). This process of theory generating is also based on theoretical sampling which may incorporate background research. The process of theory generation is proceeded in three levels of data coding:

- ***LEVEL: 1 DATA CODING (OPEN CODING/ MEMOING/CATEGORY DEVELOPMENT)***
  - *Involves memoing– taking notes – giving names. A research may:*
    - ***Look for language such as action verbs or gerunds and points of dissonance which may denote an important experience or past time***
    - ***Employ biasness***
    - ***Focus on obtaining content and platform***
    - ***Use structured questions***
    - ***Read background after interview***
    - ***Collect and let data speaks then proceed to level***

Unique Analysis Procedures and Strategy of Qualitative Research - IV

- **Grounded Theory Research - 2**
- **LEVEL - 2:DATA CODING (AXIAL CODING/COLLAPSING/CATEGORIZING/RENAMING)**
  - **Involves renaming and forming categories**
  - **Ask semi-structured questions**
- **LEVEL - 3: DATA CODING (SELECTIVE CODING/THEME & THEORY CONSTRUCTION)**
  - Involves formation of themes and theories
  - Results in formation of grounded theory
- **The process of coding is usually a cycle**
  - Normally, different kinds (at least two) of interviews are used; first interview which may include L1 Coding; second interview which may include L2 Coding (using semi-structured questions "smart bombs"); third second interview which may include L3 Coding (using unstructured questions–smart bombs).

Unique Analysis Procedures and Strategy of Qualitative Research - V

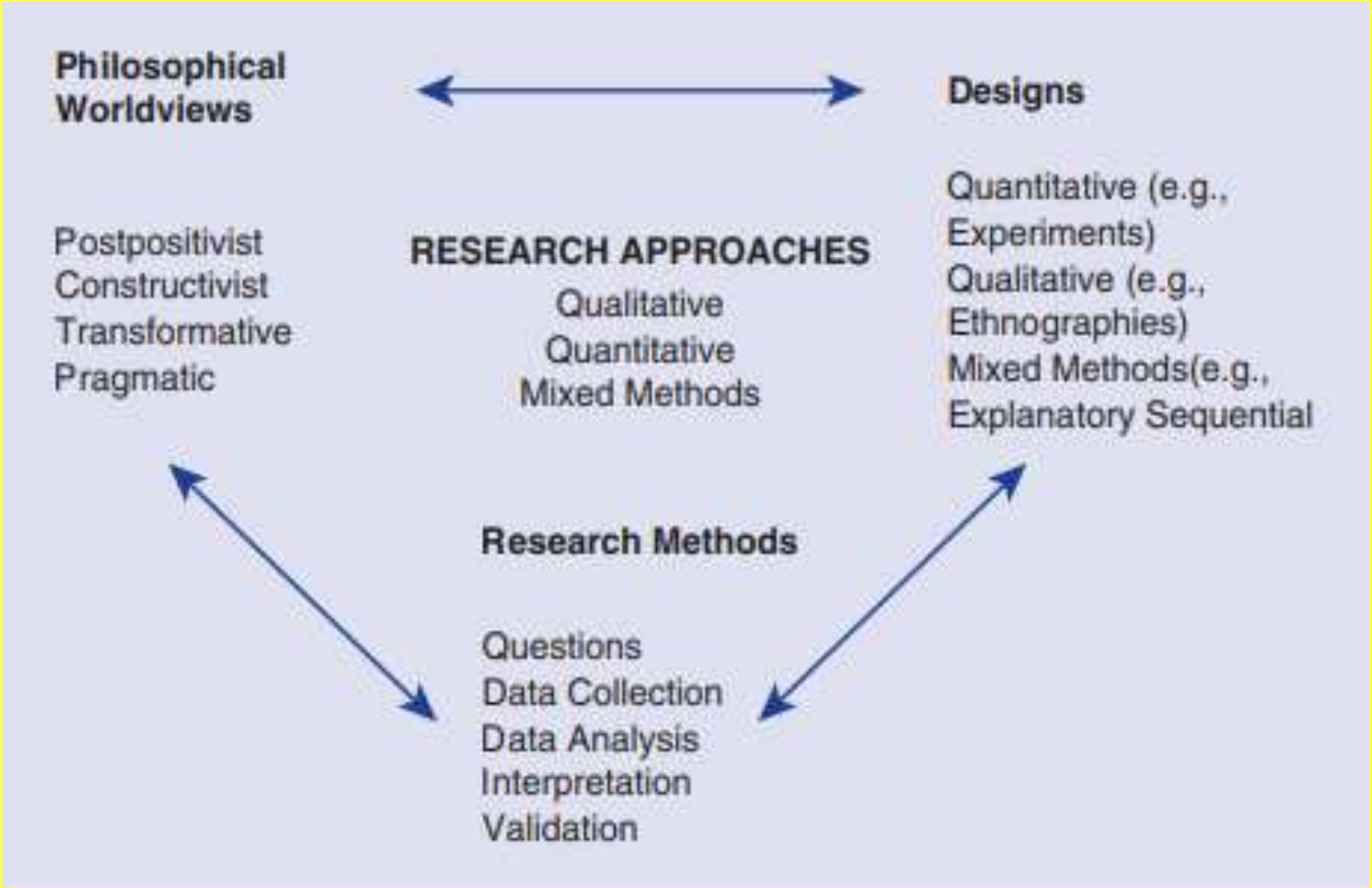
- **Narrative Research**

Narrative theory research analysis approach involves shaping, transforming, and uncovering an experience (usually human experience) in a story format–understood through a stirring linguistic depiction. Therefore it is characterized by in-depth discourses and may involve subjective intuitions. Analysis process converts experiences into words in form of writing, vernal or other communicable representation. Narrative analysis is aimed to derive a theme in an account of one's life.

- **Case Study Research**

Narrative theory research analysis approach involves in-depth examination of a case under study. Analysis is usually unusual and concurrent with data collection. It includes an iterative process in which initial analysis of data shapes subsequent data collection and analysis. 'The principle data analysis method for case studies is referred to as OTTR, which stands for "observe," "think," "test," and "revise".

A Framework for Research —The Interconnection of Worldviews, Design, and Research Methods



Problems of Social Science Research

- What are the major problems and challenges in social science research and let us proposed comprehensive solution accordingly. The considerable known challenges are development of conceptualizations, problem of maintaining objectivity, difficulty in the verification of the inferences, unpredictability and problem of Universalisation, etc.
- Besides above mentioned problems, what would be other challenges in respect to different societies, conditions and different nations?

Hypothesis - I

- A Hypothesis is a tentative relationship between two or more variables which direct the research activity to test it.
- A Hypothesis is a testable prediction which is expected to occur. It can be a false or a true statement that is tested in the research to check its authenticity.
- Sometimes, it is very difficult to start a research without having a valid foundation. Hence, the research builds a logical relationship between various phenomena to start working on the research. This logical relationship is relevant to the theme of the research.
- This logical relationship between various phenomena is called a hypothesis. This logical relationship or testable assumption gives a direction to the research, specifies the focus of the research and helps in framing research techniques.

Hypothesis - II

- Hypothesis is a proposition or a set of propositions, set forth as an explanation for the occurrence of some specified group of phenomena asserted merely as a provisional conjecture to guide some investigation or accepted as highly probable in the light of established facts.
- Simply, a mere assumption to be proved or disproved. But for a researcher, hypothesis is a formal question that he intends to resolve.
- Often hypothesis is a predictive statement capable of being tested by scientific methods, that relates an independent variable to some dependent variable.
- **Example:** Students who receive counseling will show greater increase in creativity than students not receiving counseling; (or) Car A is performing as well as Car B.
- In short, hypothesis is a proposition which can be put to test to determining its validity.

Characteristics of Hypothesis

- It should be clear and precise.
- It should be capable of being tested.
- It should be limited in scope and be specific.
- It should be stated in simple terms.
- It should be state the relationship between variables.
- It should be consistent with most known facts.
- It should be amenable to testing within a reasonable time.
- It must explains the fact that gave rise to the need for explanation.
- Empirically Testable.
- Simple and Clear.
- Specific and relevant.
- Predictable.
- Manageable.



Importance of Hypothesis

- It gives a direction to the research.
- It specifies the focus of the researcher.
- It helps in devising research techniques.
- It prevents from blind research.
- It ensures accuracy and precision.
- It saves resources – time, money and energy.





Types of Hypothesis - I

- Simple Hypothesis
- Complex Hypothesis
- Empirical Hypothesis
- Null Hypothesis
- Alternative Hypothesis
- Logical Hypothesis
- Statistical Hypothesis
- Question form of Hypothesis
- Working or Research Hypothesis

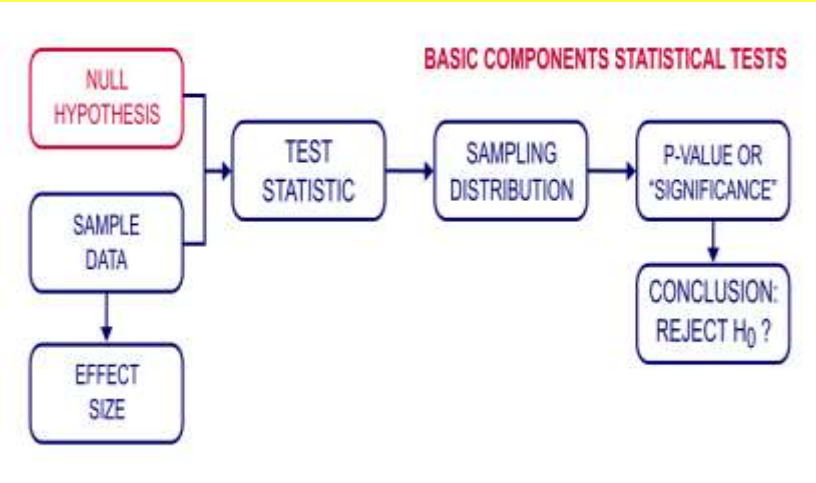


Types of Hypothesis - II

- **Simple Hypothesis**
- Simple hypothesis is that one in which there exists relationship between two variables one is called independent variable or cause and the other is dependent variable or effect.
- **Example:** Smoking leads to cancer and the higher ratio of unemployment leads to crimes.
- **Complex Hypothesis**
- Complex hypothesis is that one in which as relationship among variable exists. In this type dependent and independent variables are more than two.
- **Example:** Smoking and other drugs leads to cancer, tension, chest infection etc. and the higher ratio of unemployment poverty illiteracy leads to crimes like dacoit etc.

Types of Hypothesis - III

- **Empirical Hypothesis**
- Empirical which means it is based on evidences. In scientific method the word "Empirical" refers to the use of working hypothesis that can be tested using observation and experiment. Empirical data is produced by experiment and observation.
- **Example:** How is the ability of 9^{th} class students in learning moral values?
- **Null Hypothesis**
- Null the hypothesis that there is no significant difference between specified populations, any observed difference being due to sampling or experimental error. It is denoted by $\mathbf{H_0}$.
- **Example:** There is no correlation between the educational status with occupational pattern of the respondents.



Types of Hypothesis - IV

- **Alternative Hypothesis**
- The Alternative Hypothesis, denoted by **H_1 or H_a.** Is the hypothesis that sample observations are influenced by some non-random cause.
- The Alternative Hypothesis is the statement that the parameter has a value that somehow differs from the null hypothesis.
- The symbolic form of the alternative hypothesis must use one of these symbols: $\neq$ (Not Equal), $<$ (Less than), $>$ (Greater than).
- **Example:** There is a significant relationship between the educational status of the respondents as well as the occupational status.



Types of Hypothesis - V

- **Alternative Hypothesis (Clarification)**
- **Alternative Hypotheses can be statements of inequality:**

  There is a difference in the average score of ninth graders from Watson High School and the average score of ninth graders from Gregory High School on the Math's test.
- **Alternative Hypotheses can be statements of relationship .**

  There is a relationship between personality type and job success.
- **Alternative Hypotheses can be statements of casual relationship.**

  The social class of the consumer has an impact on alcoholic beverage consumption.
- **Key to Understanding of Alternative Hypothesis:**

  - Stating a Null Hypothesis and an Alternative Hypothesis is the first step in our 5-step method for hypothesis testing.
  - The alternative hypothesis is the opposite of the null hypothesis and vice versa.
  - Stating the alternative hypothesis as a mathematical comparison, rather than a word description, can make things easier to understand. The comparison must include an inequality, using one of these: ≠ (Not Equal), < (Less than), > (Greater than).
  - In a 1-tailed test, the alternative hypothesis points in the direction of the tail of the test.

Types of Hypothesis - VI

- **Statistical Hypothesis**
- A Hypothesis, that can be verified statistically, is known as a statistical hypothesis.
- It can be any hypothesis that has the quality of being verified statistically. It means using quantitative techniques, to generate statistical data, can easily verify it.
- It can also be said that the variables in a statistical hypothesis can be transformed into quantifiable sub-variable to test it statistically.
- **Question form of Hypothesis**
- It is the simplest form of empirical hypothesis.
- In simple case of investigation and research are adequately implemented by resuming a question.
- **Example:** How is the ability of 9^{th} class students in learning moral values?

5-Steps for Testing Hypothesis

Steps	Actions	Descriptions
Step - 1	State Null and Alternative Hypothesis.	Null Hypothesis : H0 = 0 Alternative Hypothesis : H1 = 0 Note: 1. **Two-tailed test:** If alternative hypothesis does not state direction (Greater or less). 2. **One-tailed test:** If alternative state direction.
Step – 2	Select level of Significance.	1. 0.01 level (1% level) – for consumer research 2. 0.05 level (5 % level) – for quality assurance 3. 0.10 level (10 % level) – for political pooling
Step – 3	Identify the test statistics.	**z** and **t** as test statistic, (Note: use **t** test when **n** is less than 30. If **n** is 30 and more use **z** test), and others **F** test (to test more than 2 means) and Chi-square for non-parameter statistic.
Step – 4	Formulate Decision Rule	Find the critical value of **z** from Normal Distribution table, or value **t** from **t** distribution table, or Chi-square, of **F** distribution table, where appropriate,
Step - 5	Arrive at Decision	Only one decision is possible in hypothesis testing do not reject null hypothesis or reject null hypothesis and accept alternative hypothesis.

Comparison of Null and Alternative Hypothesis

Basic for Comparison	Null Hypothesis	Alternative Hypothesis
Meaning	A null hypothesis is a statement, in which there is no relationship between two variables.	An alternative hypothesis is statement in which there is some statistical significance between two measured phenomenon.
Represents	No observed effect.	Some observed effect.
What is it?	It is what the researcher tries to disprove.	It is what the researcher tries to prove
Acceptance	No changes in opinions or actions.	Changes in opinions or actions
Testing	Indirect and implicit.	Direct and explicit.
Observations	Result of chance.	Result of real effect.
Denoted by	H_0	H_1
Mathematical formulation	= (Equal sign)	≠ (Unequal sign)

Logic of Testing Hypothesis

- A hypothesis test is a standard format for assessing statistical evidence. It is ubiquitous in scientific literature, most often appearing in the form of statements of statistical significance and notations like "$p < 0.01$" that pepper scientific journals.
- Hypothesis testing involves a substantial technical vocabulary: null hypotheses, alternative hypotheses, test statistics, significance, power, p-values, and so on. The last section of this chapter lists the terms and gives definitions.
- The technical aspects of hypothesis testing arise because it is a highly formal and quite artificial way of reasoning. This isn't a criticism. Hypothesis testing is this way because the "natural" forms of reasoning are inappropriate.
- **Example:** The stock market's ups and downs are reported each working day. Some people make money by investing in the market, some people lose. Is there reason to believe that there is a trend in the market that goes beyond the random-seeming daily ups and down

Contribution of Hypothesis

- It provides clarity to the research problem and research objectives.
- It describes, explains or predicts the expected results or outcome of the research.
- It indicates the type of research design.
- It directs the research study process.
- It identifies the population of the research study that is to be investigated or examined.
- It facilitates data collection, data analysis and data interpretation.Hypothesis
- It enables an investigator to start his research work.
- It may lead to formulations of another hypothesis.
- It leads to interpret results drawing conclusions related to original purpose.

Research Design - I

- Research design is the framework of research methods and techniques chosen by a researcher. The design allows researchers to hone in on research methods that are suitable for the subject matter and set up their studies up for success.
- The design of a research topic explains the type of research (experimental, survey and correlation, semi-experimental, review) and also its sub-type (experimental design, research problem, descriptive case-study).
- **3 Main types of Research Design:**
- There are three main types of research design,
- Data collection,
- measurement, and
- Analysis.

Research Design - II

- A Research Design is a framework or blueprint for conducting applied economic research and others social science researches. It details the procedures necessary for obtaining the information needed to structure or solve the research problems.
- **Definition:**
- **William Zikmund -** " Research design is a master plan specifying the methods and procedures for collection and analyzing the needed information".
- **Kerlinger –** "Research design is the plan, structure and strategy of investigation conceived so as to obtain answers to research questions and to control variance".



Purpose of Research Design

- **Research designs are used for the following purposes:**
- To minimize the expenditure.
- To facilitate the smooth scaling.
- To collect the relevant data and technique.
- To provide blue print for plans.
- To provide an overview to other experts.
- To provide a direction of a research.
- **Need of the Research Design:**
- It reduces inaccuracy.
- Helps to get maximum efficiency and reliability.
- Eliminates bias and marginal errors.
- Minimizes wastage of time.
- Helpful for collecting research materials.
- Helpful for testing of hypothesis.
- Provides an overview to other experts.
- Guides the research in the right direction.



Objects of Research Design

- To clearly define the direction of research study.
- To find out the solution to the research problem.
- To eliminate problems encountered in research.
- To increase the reliability of measurement.
- For a systematic and logical study.
- **Characteristics of Good Research Design:**
- Objectivity
- Reliability
- Generalizability
- Adequate information
- Flexibility
- Truthfulness
- Pre - causations
- Hypothesis
- Use of Library.



Research Designing for find out these Answer

- What is the study about ?
- Why is the study being made ?
- Where will the study be carried out ?
- What type of data is required ?
- Where can the required data be found ?
- What periods of time will the study include ?
- What techniques of data collection will be used ?
- How will the data be analyzed ?
- In what style will the report be prepared ?
- **Phases in Research Designing:**



Essential Elements of the Research Design

- An impactful research design usually creates a minimum bias in data and increases trust in the accuracy of collected data. A design that produces the least margin of error in experimental research is generally considered the desired outcome.
- The essential elements of the research design are,
- Accurate purpose statement.
- Techniques to be implemented for collecting and analyzing research.
- The method applied for analyzing collected details.
- Type of research methodology.
- Probable objections for research.
- Settings for the research study.
- Timeline.
- Measurement of analysis.

Key Characteristics of Research Design -I

- Proper research design sets your study up for success. Successful research studies provide insights that are accurate and unbiased. You'll need to create a survey that meets all of the main characteristics of a design. There are four key characteristics of research design:
- Neutrality
- Reliability
- Validity
- Generalization
- **Neutrality:** When you set up your study, you may have to make assumptions about the data you expect to collect. The results projected in the research design should be free from bias and neutral. Understand opinions about the final evaluated scores and conclusion from multiple individuals and consider those who agree with the derived results.

Key Characteristics of Research Design - II

- **Reliability:** With regularly conducted research, the researcher involved expects similar results every time. Your design should indicate how to form research questions to ensure the standard of results. You'll only be able to reach the expected results if your design is reliable.
- **Validity:** There are multiple measuring tools available. However, the only correct measuring tools are those which help a researcher in gauging results according to the objective of the research. The questionnaire developed from this design will then be valid.
- **Generalization:** The outcome of your design should apply to a population and not just a restricted sample. A generalized design implies that your survey can be conducted on any part of a population with similar accuracy.
- The above factors affect the way respondents answer the research questions and so all the above characteristics should be balanced in a good design.

Broad classification of the Research Design

- A researcher must have a clear understanding of the various types of research design to select which model to implement for a study. Like research itself, the design of the study can be broadly classified into quantitative and qualitative.
- **Qualitative research design:** Qualitative Research determines relationships between collected data and observations based on mathematical calculations. Theories related to a naturally existing phenomenon can be proved or disproved using statistical methods. Researchers rely on qualitative research design methods that conclude "why" a particular theory exists along with "what" respondents have to say about it.
- **Quantitative research design:** Quantitative Research is for cases where statistical conclusions to collect actionable insights are essential. Numbers provide a better perspective to make critical business decisions. Quantitative research design methods are necessary for the growth of any organization. Insights drawn from hard numerical data and analysis prove to be highly effective when making decisions related to the future of the business.

Types of Research Design - I

- Descriptive research design
- Experimental research design
- Corelational research design
- Diagnostic research design
- Explanatory research design
- **Descriptive research design:**
- In a descriptive design, a researcher is solely interested in describing the situation or case under their research study. It is a theory-based design method which is created by gathering, analyzing, and presenting collected data.
- This allows a researcher to provide insights into the why and how of research. Descriptive design helps others better understand the need for the research. If the problem statement is not clear, you can conduct exploratory research.

Types of Research Design -II

- **Experimental research design:** Experimental Research design establishes a relationship between the cause and effect of a situation. It is a causal design where one observes the impact caused by the independent variable on the dependent variable.
- For **example,** one monitors the influence of an independent variable such as a price on a dependent variable such as customer satisfaction or brand loyalty. It is a highly practical research design method as it contributes to solving a problem at hand.
- The independent variables are manipulated to monitor the change it has on the dependent variable. It is often used in social sciences to observe human behavior by analyzing two groups.
- Researchers can have participants change their actions and study how the people around them react to gain a better understanding of social psychology.

Types of Research Design -III

- **Co-relational research design:** Co-relational Research is a non-experimental research design technique that helps researchers establish a relationship between two closely connected variables.
- This type of research requires two different groups. There is no assumption while evaluating a relationship between two different variables, and statistical analysis techniques calculate the relationship between them.
- A correlation coefficient determines the correlation between two variables, whose value ranges between -1 and +1. If the correlation coefficient is towards +1, it indicates a positive relationship between the variables and -1 means a negative relationship between the two variables.

Types of Research Design -IV

- **Diagnostic research design:** In diagnostic design, the researcher is looking to evaluate the underlying cause of a specific topic or phenomenon.
- This method helps one learn more about the factors that create troublesome situations.
- This design has three parts of the research:
- Inception of the issue
- Diagnosis of the issue
- Solution for the issue
- **Explanatory research design:** Explanatory design uses a researcher's ideas and thoughts on a subject to further explore their theories.
- The research explains unexplored aspects of a subject and details about what, how, and why of research questions.

Major Steps in Preparing Research Design

- **Statement of research problem is an essential step in preparing a model:**
- Review of Earlier Literature.
- Sources of Information to be tapped.
- Development of Bibliography.
- Objectives of study.
- Socio – Economic, Cultural and Political factors.
- Geographical areas to be covered.
- Periods of Time to be covered or Time Dimension of the study.
- Dimensions of the study.
- The basis for selecting for selecting the data.
- Techniques of study.
- The control of Error.
- Establish the reliability and validity of test instruments.
- Chapter scheme.

Unit – II
METHODS AND TECHNIQUES OF DATA COLLECTION

Dr. G. YOGANANDHAM
Professor and Head,
Department of Economics,
Thiruvalluvar University.

Data Collection

- Data collection is a methodical process of gathering and analyzing specific information to proffer solutions to relevant questions and evaluate the results. It focuses on finding out all there is to a particular subject matter. Data is collected to be further subjected to hypothesis testing which seeks to explain a phenomenon.

- For collectors of data, there is a range of outcomes for which the data is collected. But the key purpose for which data is collected is to put a researcher in a vantage position to make predictions about future probabilities and trends.

- The core forms in which data can be collected are primary and secondary data. While the former is collected by a researcher through first-hand sources, the latter is collected by an individual other than the user.

Methods of Data Collection

- Data collection method is impacted by the method of research you choose. It is usually done through:
  - Data collection by the individual researcher.
  - Collection through hired researchers.
  - Collection through firms.
- **Data Collection Formats**
- Format of data is influenced by the method of research, as it could be Printed questionnaires.
- Interview sheets (in-person or telephonic).
- Focus group notes.
- Observation notes.
- Email or web responses.
- Content analysis notes, pictures, documentaries.
- Printed, e-records, scanned data.
- Literature review.

Data Preparation

- Preparation of data file.
- It is important to convert raw data into a usable data for analysis.
- The analysis and results will surely depend on the quality of data.
- There are possibilities of errors in handling instruments, raw data, transcribing, data entry, assigning codes, values, value labels.
- Data need to be cleaned to fulfill the analysis conations.

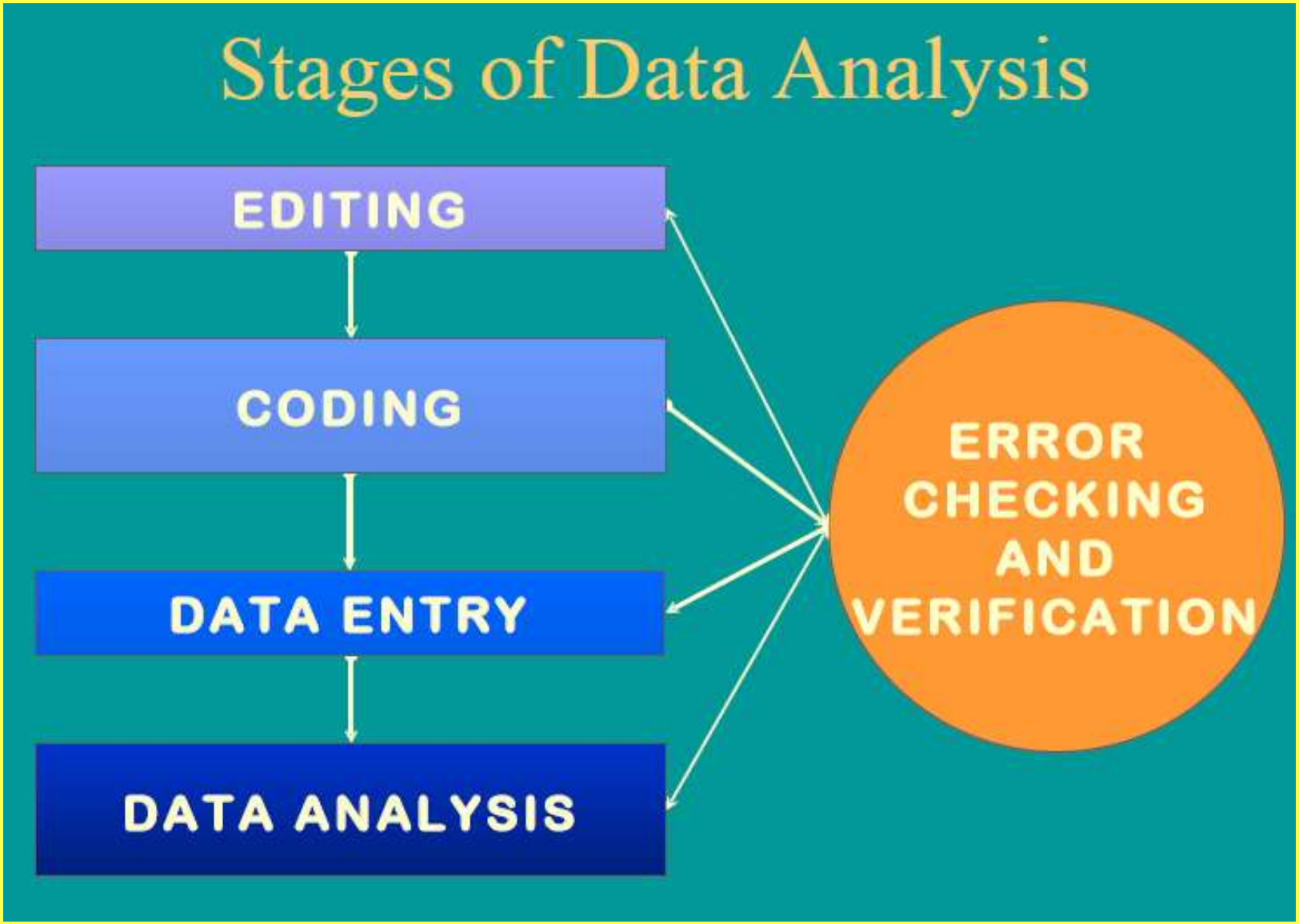


Data Analysis

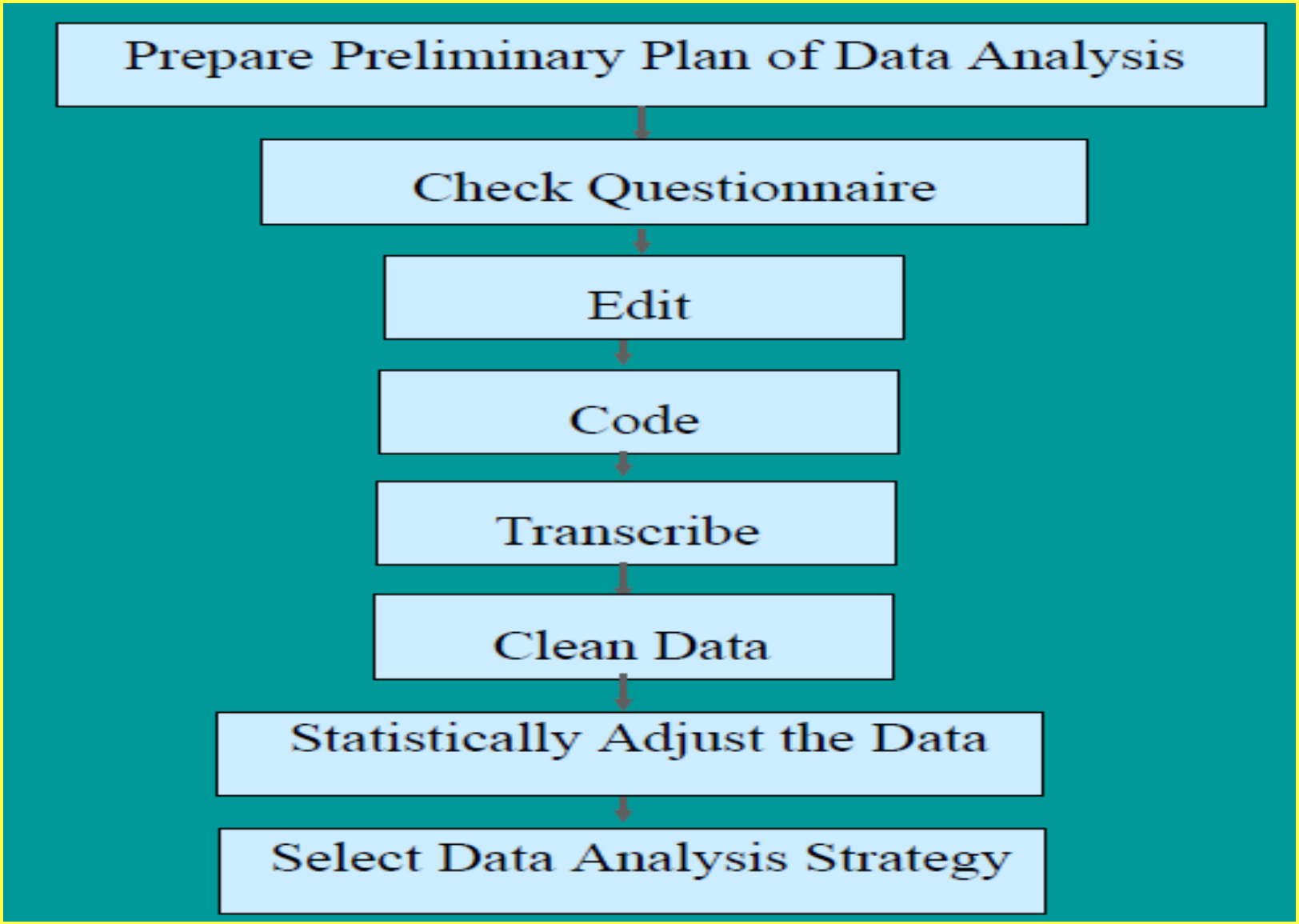
- **Analysis of data is influenced by a number of factors. They are but not limited to:**
- The purpose of research.
- The type research questions and hypothesis.
- The method of research and format of data.
- Use of software for management, manipulation and analysis of data
- Researchers skills and capabilities.
- Techniques used for data.
- The quality of the data.



Stages of Data Analysis



Data Preparation Process



Questionnaire Checking

- A questionnaire returned from the field may be unacceptable for several reasons.
- Parts of the questionnaire may be incomplete.
- The pattern of responses may indicate that the respondent did not understand or follow the instructions.
- The responses show little variance.
- One or more pages are missing.
- The questionnaire is received after the pre-established cutoff date.
- The questionnaire is answered by someone who does not qualify for participation.
- We need to find valid questionnaires for data analysis.
- Each questionnaire/response need allotment of a case number for future reference.
- Questionnaire/response need filing in an order for retrieval and verification.

Types of Data Collection - I

- Before broaching the subject of the various types of data collection. It is pertinent to note that data collection in itself falls under two broad categories; Primary data collection and secondary data collection.
- Primary Data Collection
  - ✓ Qualitative Research Method
  - ✓ Quantitative Research Method
- Secondary Data Collection
- **Primary Data Collection**

Primary data collection by definition is the gathering of raw data collected at the source. It is a process of collecting the original data collected by a researcher for a specific research purpose. It could be further analyzed into two segments; qualitative research and quantitative data collection methods.

Types of Data Collection - II

- **Qualitative Research Method**

The qualitative research methods of data collection does not involve the collection of data that involves numbers or a need to be deduced through a mathematical calculation, rather it is based on the non-quantifiable elements like the feeling or emotion of the researcher. An example of such a method is an open-ended questionnaire.



Types of Data Collection - III

- **Quantitative Research Method**

Quantitative methods are presented in numbers and require a mathematical calculation to deduce. An example would be the use of a questionnaire with close-ended questions to arrive at figures to be calculated Mathematically. Also, methods of correlation and regression, mean, mode and median.



Types of Data Collection - IV

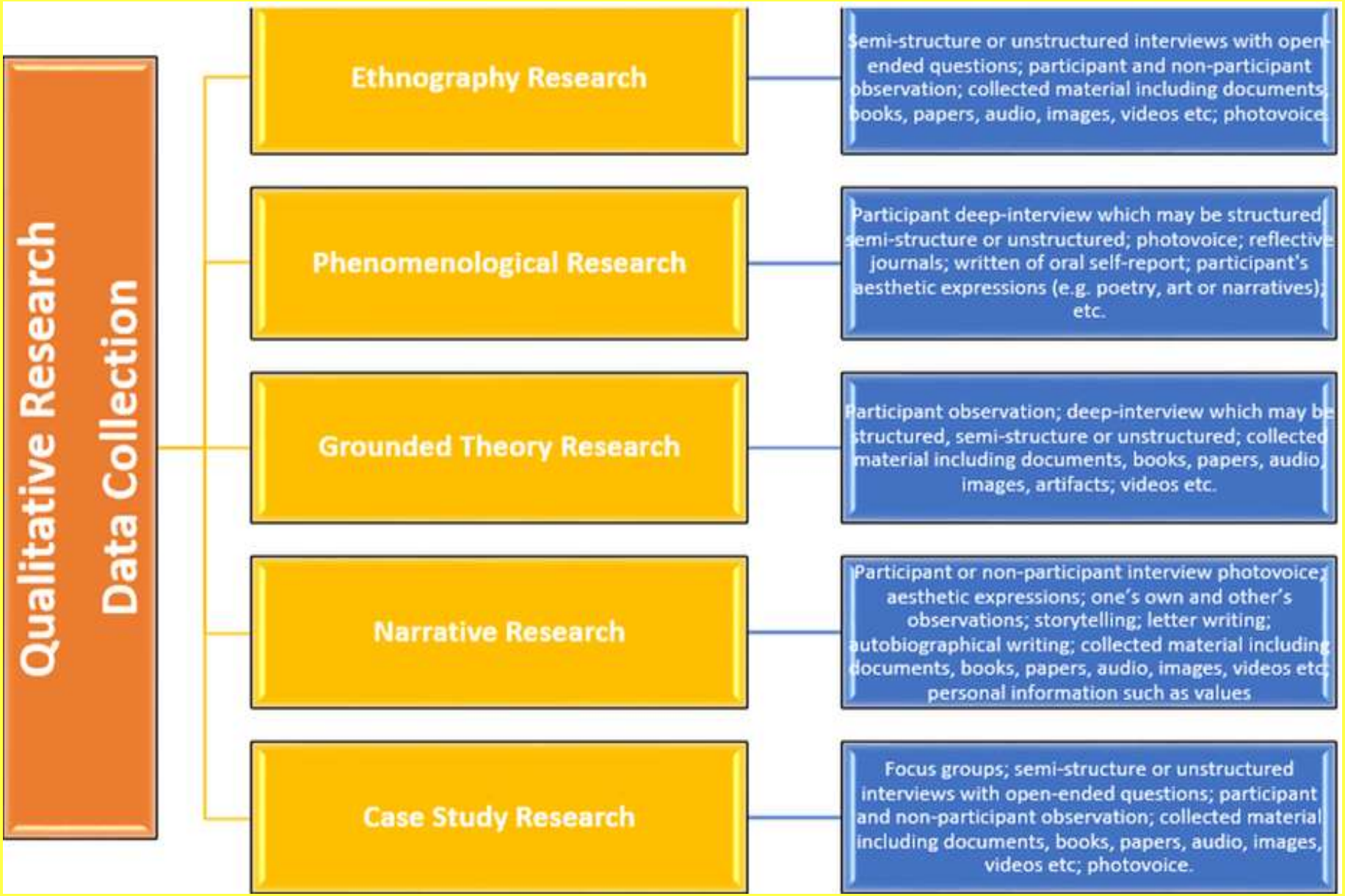
- **Secondary Data Collection**

  Secondary data collection, on the other hand, is referred to as the gathering of second-hand data collected by an individual who is not the original user. It is the process of collecting data that is already existing, be it already published books, journals and/or online portals. In terms of ease, it is much less expensive and easier to collect.

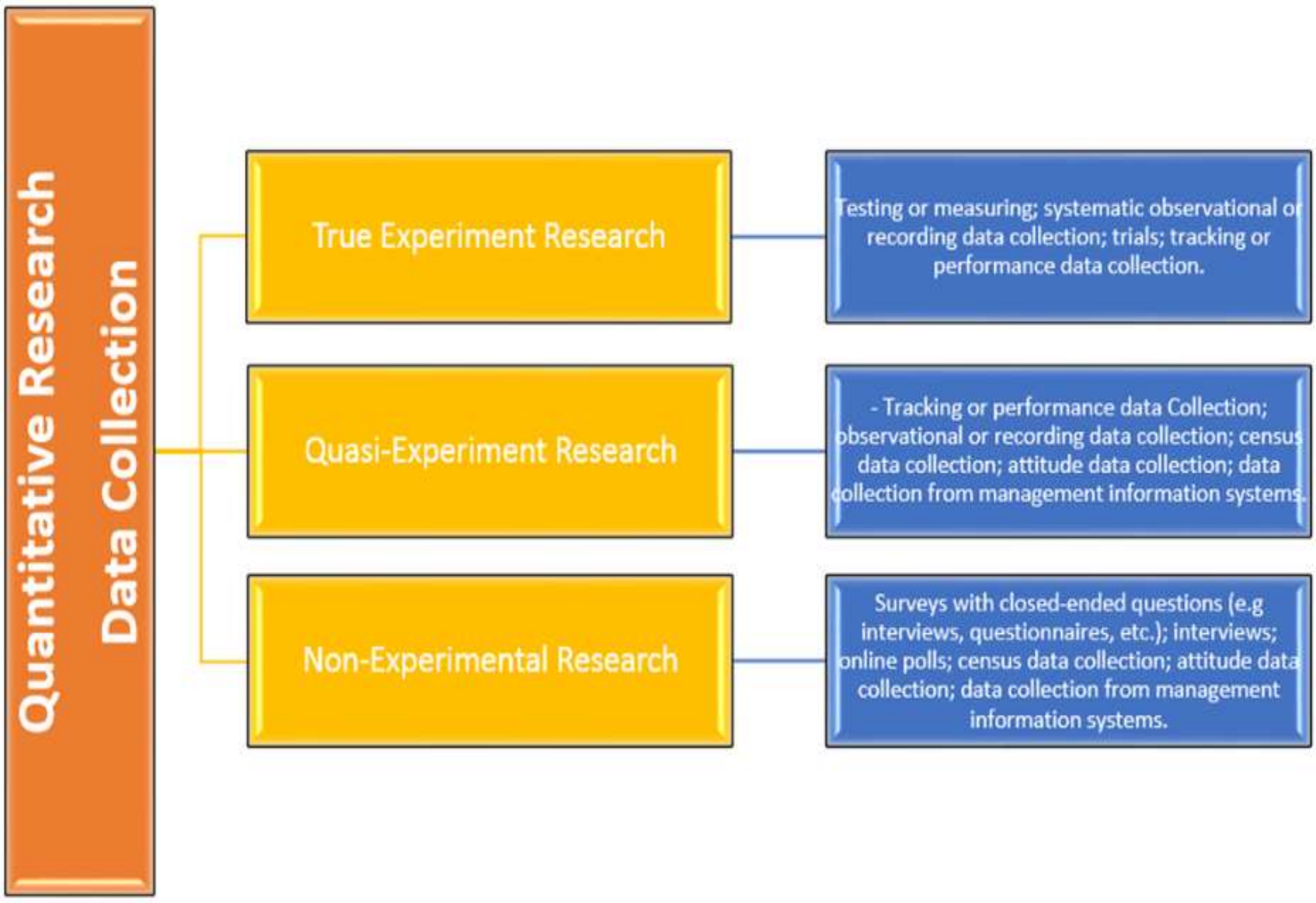
- Your choice between Primary data collection and secondary data collection depend on the nature, scope and area of your research as well as its aims and objectives.



Qualitative Research Data Collection



Quantitative Research Data Collection



Difference between Quantitative & Qualitative Data Collection Methods

Quantitative Data Collection Methods	Qualitative Data Collection Methods
Fixed and more structured techniques	Semi-structured or unstructured techniques
Instrument or test based	Not instrument or test based
Number-based to collect information	Text-based to collect information
Usually measurable	Not usually measureable
Used for statistical tests	Not used for statistical tests
Usually large sample size	Usually small sample size
Employs strong scientific control	Lacks strong scientific control
Data type is usually rational or interval	Data is usually ordinal or nominal
Interviews forms include closed-ended questions	Interview forms include open-ended questions

Importance of Data Collection

- There are a bunch of underlying reasons for collecting data, especially for a researcher. Some of them are,
- **Integrity of The Research:** A key reason for collecting data, be it through quantitative or qualitative methods is to ensure that the integrity of the research question is indeed maintained.
- **Reduce the likelihood of errors:** The correct use of appropriate data collection of methods reduces the likelihood of errors consistent with the results.
- **Decision Making:** To minimize the risk of errors in decision making, it is important that accurate data is collected so that the researcher doesn't make uninformed decisions.
- **Save Cost and Time:** Data collection saves the researcher time and funds that would otherwise be misspent without a deeper understanding of the topic or subject matter.
- **To support a need for a new idea, change and/or innovation:** To prove the need for a change in the norm or the introduction of new information that will be widely accepted, it is important to collect data as evidence to support these claims.

Data Collection Tool

- Data collection tools refer to the devices/instruments used to collect data, such as a paper questionnaire or computer-assisted interviewing system. Case Studies, Checklists, Interviews, Observation sometimes, and Surveys or Questionnaires are all tools used to collect data.
- It is important to decide the tools for data collection because research is carried out in different ways and for different purposes. The objective behind data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the questions that have been posed.
- The objective behind data collection is to capture quality evidence that allows analysis to lead to the formulation of convincing and credible answers to the questions that have been posed.
- The Form plus' online data collection tool is perfect for gathering primary data, i.e. raw data collected from the source. You can easily get data with at least three data collection methods with our online and offline data gathering tool. i.e online questionnaires, Focus Groups and Reporting.
- In our previous articles, we've explained why quantitative research methods are more effective than qualitative methods. However, with Form plus data collection tool, you can gather all types of primary data for academic, opinion or product research.

Seven top data collection methods and tools for Academic, Opinion or Product Research

- Interview method
- Questionnaires
- Reporting
- Existing Data
- Observation
- Focus Groups
- Combination Research



TELEPHONIC INTERVIEWS



- Contacting samples on telephone
- Uncommon method may be used in developed regions

MERITS

- Flexible compare to mailing method
- Faster than other methods
- Cheaper than personal interview method
- Callbacks are simple and economical also
- High response than mailing method.



when it is not possible to contact the respondent directly, then interview is conducted through – Telephone.

Interview method - I

- An interview is a face-to-face conversation between two individuals with the sole purpose of collecting relevant information to satisfy a research purpose. Interviews are of different types namely; **Structured, Semi-structured and Unstructured** with each having a slight variation from the other.
- **Structured Interview:** Simply put, it is a verbally administered questionnaire. In terms of depth, it is surface level and is usually completed within a short period. For speed and efficiency, it is highly recommendable, but it lacks depth.
- **Semi-structured Interview:** In this method, there subsist several key questions which cover the scope of the areas to be explored. It allows a little more leeway for the researcher to explore the subject matter.
- **Unstructured Interview:** It is an in-depth interview that allows the researcher to collect a wide range of information with a purpose. An advantage of this method is the freedom it gives a researcher to combine structure with flexibility even though it is more time-consuming.

Interview method - II

- **Pros**
- In-depth information.
- Freedom of flexibility.
- Accurate data.
- **Cons**
- Time-consuming.
- Expensive to collect.



- **What are the best Data Collection Tools for Interviews?**

For collecting data through interviews, here are a few tools you can use to easily collect data.

- **Audio Recorder:** An audio recorder is used for recording sound on disc, tape, or film. Audio information can meet the needs of a wide range of people, as well as provide alternatives to print data collection tools.

Interview method - III

- **Digital Camera:** An advantage of a digital camera is that it can be used for transmitting those images to a monitor screen when the need arises.
- **Camcorder:** A camcorder is used for collecting data through interviews. It provides a combination of both an audio recorder and a video camera. The data provided is qualitative in nature and allows the respondents to answer questions asked exhaustively. If you need to collect sensitive information during an interview, a camcorder might not work for you as you would need to maintain your subject's privacy.
- Want to conduct an interview for qualitative data research or special report? Use this online interview consent form template to allow the interviewee to give their consent before you use the interview data for research or report. With premium features like e-signature, upload fields, form security, etc., Formplus Builder is the perfect tool to create your preferred online consent forms without coding experience.

Questionnaires - I

- This is the process of collecting data through an instrument consisting of a series of questions and prompts to receive a response from individuals it is administered to. Questionnaires are designed to collect data from a group.
- For clarity, it is important to note that a questionnaire isn't a survey, rather it forms a part of it. A survey is a process of data gathering involving a variety of data collection methods, including a questionnaire.
- On a questionnaire, there are three kinds of questions used. They are; fixed-alternative, scale, and open-ended. With each of the questions tailored to the nature and scope of the research.
- **Pros**
- Can be administered in large numbers and is cost-effective.
- It can be used to compare and contrast previous research to measure change.
- Easy to visualize and analyze.
- Questionnaires offer actionable data.
- Respondent identity is protected.
- Questionnaires can cover all areas of a topic.
- Relatively inexpensive.



Questionnaires - II

- **Cons**
- Answers may be dishonest or the respondents lose interest midway.
- Questionnaires can't produce qualitative data.
- Questions might be left unanswered.
- Respondents may have a hidden agenda.
- Not all questions can be analyzed easily.
- **What are the best Data Collection Tools for Questionnaire?**
- **Form plus online questionnaire**

  Form plus lets you create powerful forms to help you collect the information you need. Form plus helps you create the online forms that you like. The Form plus online questionnaire form template to get actionable trends and measurable responses. Conduct research, optimize knowledge of your brand or just get to know an audience with this form template. The form template is fast, free and fully customizable.

- **Paper Questionnaire**

  A paper questionnaire is a data collection tool consisting of a series of questions and/or prompts for the purpose of gathering information from respondents. Mostly designed for statistical analysis of the responses, they can also be used as a form of data collection.

Difference between Interview Schedule and Questionnaire - I

S. No.	Description	Interview Schedule	Questionnaire
1.	Non-Response from Respondents	Very low because they are filled by enumerators who are able to get answers to all questions.	High as many people do not responded many return it without answering all questions.
2.	Contact with respondents	Direct contact exist.	Direct contact may or may not exist as it can be sent by post also.
3.	Coverage of area	Limited geographical area as difficulty lies in sending enumerators over a relatively wider area.	Wider geographical area coverage is possible.
4.	Reliability of data collection	High. The information collected is generally complete and accurate as enumerators can remove difficulties.	Less. Risk of collection incomplete and wrong information is high, particularly when people are unable to understand questions properly.

Continue….

Difference between Interview Schedule and Questionnaire - II

S. No.	Description	Interview Schedule	Questionnaire
5.	Interviewer Bias and Cheating	Relatively High.	Relatively Low.
6.	Success	Here the success mainly depends on the honestly and competence of enumerators not on the attractive looks as they are filled by enumerators not by respondents.	Here the success lies on the quality, physical appearance of the questionnaire.
7.	Education Level of the respondents	Information can be gathered even when the respondents are illiterate.	Can be used only when respondents are literate and cooperative,
8.	Cost	Relatively Expensive since considerable amount of money is spent in appointing enumerators and preparing schedules.	Relatively cheaper since we have to spend money only in preparing the questionnaire and sending the same to the respondents.

Researchers have prepared guidelines for fieldworkers in asking questions. The guidelines include:

- Be thoroughly familiar with the questionnaire.
- Ask the questions in the order in which they appear in the questionnaire.
- Use the exact wording given in the questionnaire.
- Read each question slowly.
- Repeat questions that are not understood.
- Ask every applicable question.
- Follow instructions and skip patterns, probing carefully.

Probing techniques

- Repeating the question
- Repeating the respondents' reply
- Boosting or reassuring the respondent
- Eliciting clarification
- Using a pause (silent probe)
- Using objective/neutral questions or comments



Data entry

- Once the questionnaire is coded appropriately, researchers input the data into statistical software package. This process is called data entry.
- **Data Cleaning**
- Data cleaning focuses on error detection and consistency checks as well as treatment of missing responses. The first step in the data cleaning process is to check each variable for data that are out of the range or as otherwise called logically inconsistent data.
- Such data must be corrected as they can hamper the overall analysis process. Most advance statistical packages provide an output relating to such inconsistent data. Inconsistent data must be closely examined as sometimes they might not be inconsistent and be representing legitimate response.

Reporting - I

- **Definition:** Data reporting is the process of gathering and submitting data to be further subjected to analysis. The key aspect of data reporting is reporting accurate data because of inaccurate data reporting leads to uninformed decision making.
- **Pros**
- Informed decision making.
- Easily accessible.
- **Cons**
- Self-reported answers may be exaggerated.
- The results may be affected by bias.
- Respondents may be too shy to give out all the details.
- Inaccurate reports will lead to uninformed decisions.



Reporting - II

- **What are the best Data Collection Tools for Reporting?**
- Reporting tools enable you to extract and present data in charts, tables, and other visualizations so users can find useful information. You could source data for reporting from Non-Governmental Organizations (NGO) reports, newspapers, website articles, hospital records.
- **NGO Reports:** Contained in NGO reports is an in-depth and comprehensive report on the activities carried out by the NGO, covering areas such as business and human rights. The information contained in these reports are research-specific and forms an acceptable academic base towards collecting data. NGOs often focus on development projects which are organized to promote particular causes.
- **Newspapers:** Newspaper data are relatively easy to collect and are sometimes the only continuously available source of event data. Even though there is a problem of bias in newspaper data, it is still a valid tool in collecting data for Reporting.

Reporting - III

- **Website Articles:** Gathering and using data contained in website articles is also another tool for data collection. Collecting data from web articles is a quicker and less expensive data collection Two major disadvantages of using this data reporting method are biases inherent in the data collection process and possible security/confidentiality concerns.
- **Hospital Care records:** Health care involves a diverse set of public and private data collection systems, including health surveys, administrative enrollment and billing records, and medical records, used by various entities, including hospitals, CHCs, physicians, and health plans. The data provided is clear, unbiased and accurate, but must be obtained under the legal means as medical data is kept with the strictest regulations.

Existing Data - I

- This is the introduction of new investigative questions in addition to/other than the ones originally used when the data was initially gathered. It involves adding measurement to a study or research. An example would be sourcing data from an archive.
- **Pros**
- Accuracy is very high.
- Easily accessible information.
- **Cons**
- Problems with evaluation.
- Difficulty in understanding.
- **What are the Best Data Collection Tools for Existing Data?**
- The concept of Existing data means that data is collected from existing sources to investigate research questions other than those for which the data were originally gathered. Tools to collect existing data include:

Existing Data - II

- **Research Journals** - Unlike newspapers and magazines, research journals are intended for an academic or technical audience, not general readers. A journal is a scholarly publication containing articles written by researchers, professors, and other experts.
- **Surveys** - A survey is a data collection tool for gathering information from a sample population, with the intention of generalizing the results to a larger population. Surveys have a variety of purposes and can be carried out in many ways depending on the objectives to be achieved.

Observation - I

- This is a data collection method by which information on a phenomenon is gathered through observation. The nature of the observation could be accomplished either as a complete observer, an observer as a participant, a participant as an observer or as a complete participant. This method is a key base of formulating a hypothesis.
- **Pros**
- Easy to administer.
- There subsists a greater accuracy with results.
- It is a universally accepted practice.
- It diffuses the situation of an unwillingness of respondents to administer a report.
- It is appropriate for certain situations.
- **Cons**
- Some phenomena aren’t open to observation.
- It cannot be relied upon.
- Bias may arise.
- It is expensive to administer.
- Its validity cannot be predicted accurately.

Observation - II

- **What are the best Data Collection Tools for Observation?**

- Observation involves the active acquisition of information from a primary source. Observation can also involve the perception and recording of data via the use of scientific instruments. The best tools for Observation are:

- **Checklists -** state specific criteria, allow users to gather information and make judgments about what they should know in relation to the outcomes. They offer systematic ways of collecting data about specific behaviors, knowledge, and skills.
- **Direct observation -** This is an observational study method of collecting evaluative information. The evaluator watches the subject in his or her usual environment without altering that environment.

Focus Group - I

- The opposite of quantitative research which involves numerical based data, this data collection method focuses more on qualitative research. It falls under the primary category for data based on the feelings and opinions of the respondents. This research involves asking open-ended questions to a group of individuals usually ranging from 6-10 people, to provide feedback.
- **Pros**
- Information obtained is usually very detailed.
- Cost-effective when compared to one-on-one interviews.
- It reflects speed and efficiency in the supply of results.
- **Cons**
- Lacking depth in covering the nitty-gritty of a subject matter.
- Bias might still be evident.
- Requires interviewer training
- The researcher has very little control over the outcome.
- A few vocal voices can drown out the rest.
- Difficulty in assembling an all-inclusive group.

Focus Group - II

- **What are the best Data Collection Tools for Focus Groups?**
- A focus group is a data collection method that is tightly facilitated and structured around a set of questions. The purpose of the meeting is to extract from the participants' detailed responses to these questions. The best tools for tackling Focus groups are:

- **Two-Way -** One group watches another group answer the questions posed by the moderator. After listening to what the other group has to offer, the group that listens are able to facilitate more discussion and could potentially draw different conclusions**.**

- **Dueling-Moderator -** There are two moderators who play the devil's advocate. The main positive of the dueling-moderator focus group is to facilitate new ideas by introducing new ways of thinking and varying viewpoints.

Combination Research - I

- This method of data collection encompasses the use of innovative methods to enhance participation to both individuals and groups. Also under the primary category, it is a combination of Interviews and Focus Groups while collecting qualitative data. This method is key when addressing sensitive subjects.
- **Pros**
- Encourage participants to give responses.
- It stimulates a deeper connection between participants.
- The relative anonymity of respondents increases participation.
- It improves the richness of the data collected.
- **Cons**
- It costs the most out of all the top 7.
- It's the most time-consuming.

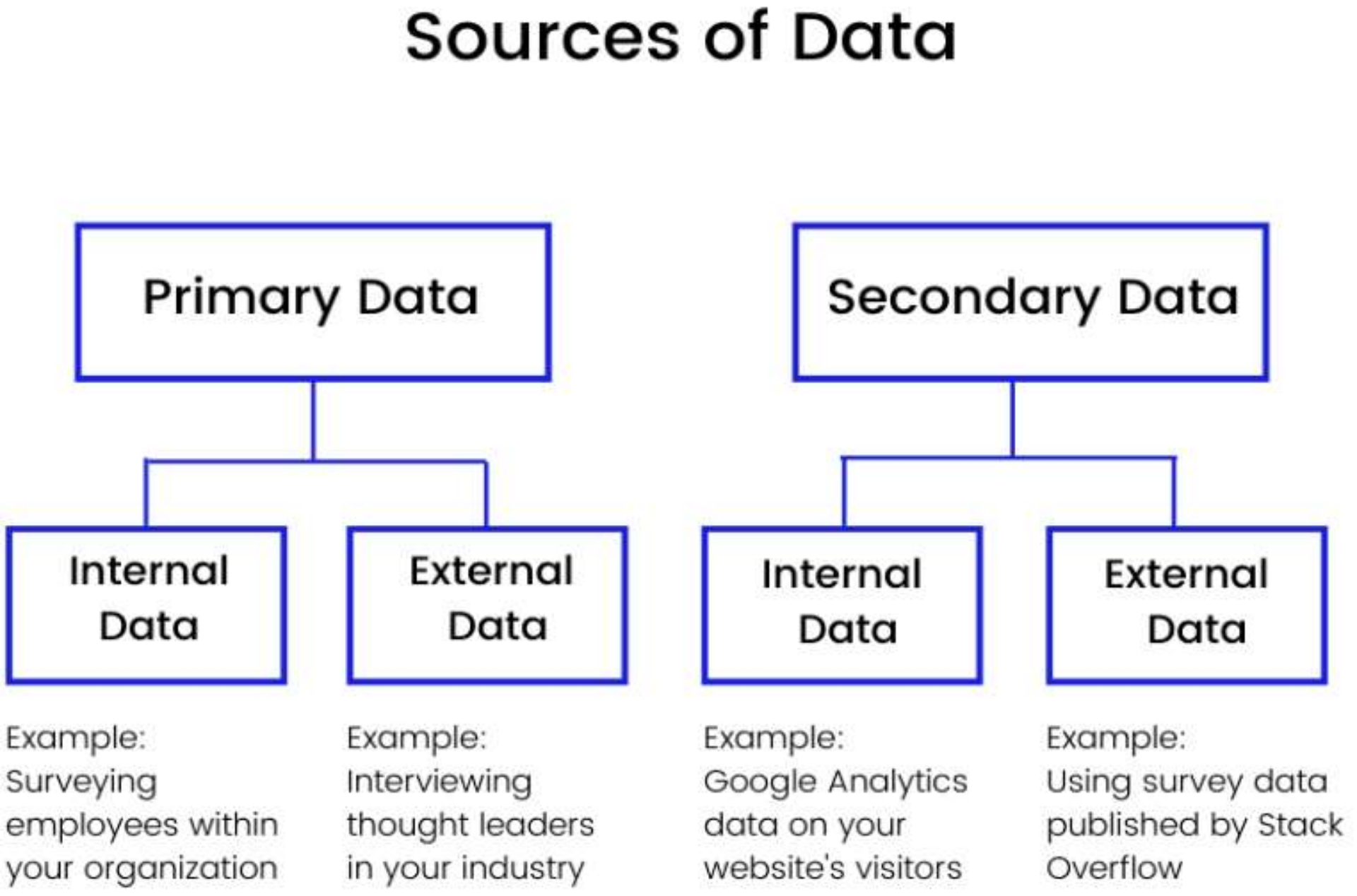
Combination Research - II

- **What are the best Data Collection Tools for Combination Research?**
- The Combination Research method involves two or more data collection methods, for instance, interviews as well as questionnaires or a combination of semi-structured telephone interviews and focus groups. The best tools for combination research are:
- **Online Survey -** The two tools combined here are online interviews and the use of questionnaires. This is a questionnaire that the target audience can complete over the Internet. It is timely, effective and efficient. Especially since the data to be collected is quantitative in nature.
- **Dual-Moderator -** The two tools combined here are focus groups and structured questionnaires. The structured questioners give a direction as to where the research is headed while two moderators take charge of proceedings. Whilst one ensures the focus group session progresses smoothly, the other makes sure that the topics in question are all covered. Dual-moderator focus groups typically result in a more productive session and essentially leads to an optimum collection of data.

Sources of Data - I

- Sources of Data can be classified into 2 types. Statistical sources refer to data that are gathered for some official purposes and incorporate censuses and officially administered surveys. Non-statistical sources refer to the collection of data for other administrative purposes or for the private sector.
- What are the different sources of data?
- Following are the two sources of data:
- **Internal Source:** When data are collected from reports and records of the organization itself, it is known as the internal source.
- For **example,** a company publishes its 'Annual Report' on Profit and Loss, Total Sales, Loans, Wages etc.
- **External Source:** When data are collected from outside the organization, it is known as the external source.
- For **example,** if a Tour and Travels Company obtains information on 'Karnataka Tourism' from Karnataka Transport Corporation, it would be known as external sources of data.

Sources of Data - II



Constructing an Instrument for data collection

- Construction of an instrument for research total depends upon how you are going to collect data and various method around it.
- All that is needed is what sort of information is required for the qualitative and quantitative study. This first practical step in research study also requires some validation and reliability for research instrument you have constructed.



Method of Collecting Primary Data - I

- **Direct personal Investigation (i.e. Interview Method)**
- Under this method the investigator collects the data personally. He approaches the objects, conducts the enquiry on the spot, collects information. For the successful collection of data under this method, it's necessary that the investigator should be polite, unbiased and aware of the local conditions, environment and languages.
- **Indirect oral investigation (i.e. through enumerators)**
- Under this the investigator doesn't collects the information directly, instead he gets them indirectly through those persons who know the information and who are ready to part away with the information they posses. This method is used in case where direct contact is not possible.
- **Investigation through Local reporters Questionnaire**
- In this method data are not collected through specially appointed agents/ correspondents/investigators/enumerators but local agents or correspondents are requested to collect the information. They collect the information in their own fashion and according to their liking's and disliking.

Method of Collecting Primary Data - II

- Investigation through mailed Questionnaire
- Under this method the investigator may either appoint local agents/ investigators/ enumerators and mail them the questionnaire for collecting information or collecting by his own using mail or telephone.
- Investigation through Observation
- In this watching behaviour of object and most commonly used method in behavioural science. One major problem in this method is that Natural response is not obtained when respondent know he is observed.



Method of Collecting Secondary Data

- **Published Sources**
  - **International Publication:** UNO publishes a number of journals and periodicals which are best sources of secondary data for national as well as international facts.
  - **Government Publications:** The Govt. of India as well as other state Govt. publish data for their use as well as for the use of researcher and others. These data are very much accurate and reliable for various types of investigations.
  - **Publication:** Municipal corporations, boards, etc. publish data pertaining to birth rate, death rate, literacy, etc. These can be effectively put to use by investigators.
  - **Commercials Research, Educational Institute, Unions, Organizations:** These institutions also conduct enquiry and publish their reports through which data can be had for use of investigators enquiring on the related subjects.
- **Unpublished Sources**
- In some of the cases, enquiry are conducted, data are classified and analysed but due to some reasons or other findings are not published though they are preserved either by the individual or a society. They can also be used by the investigators as Secondary Data.

Difference between Primary and Secondary Data - I

S. No.	Primary Data	Secondary Data
1.	Real time data.	Past data.
2.	Sure about sources of data.	Not sure about sources of data.
3.	Costly and Time consuming process.	Cheap and No time consuming process.
4.	Avoid biasness of response data.	Can not know if data biasness or not.
5.	More flexible.	Less flexible.

Difference between Primary and Secondary Data - II

S. No.	Criteria	Primary Data	Secondary Data
1.	Accuracy	A direct approach that starts from scratch. Data is highly accurate and zooms in on your research problem.	An Indirect approach that hinges on existing data. May not be specific to your research problem.
2.	Control	Researchers have a high degree of control.	Less control over data.
3.	Relevancy	Very relevant to your research.	Negligible or less relevant to a new research.
4.	Ownership	Researcher has ownership of the data.	No ownership of the data.
5.	Cost and Time	Expensive and time consuming task.	Saves you money and time,
6.	Accessibility	Not freely or easily available to the public.	Available to the public and your competitors.
7.	Bias	The data does not favor the researcher who collected the data.	The data may lean toward the researcher wh collected the data.
8.	Sampling Errors	Errors in sampling arise by selecting the wrong sample size.	Errors arise by selecting outdated, irrelevant data that is no longer pertinent to the research problem.
9.	Sources	Websites, journals, dissertations, literature reviews, balance sheets, and more.	Case studies, surveys, interviews, questionnaires, field , observations, experiments, etc.

Important Sources of Secondary Data - I

- Sources of secondary data includes books, personal sources, journal, newspaper, website, government record etc. Secondary data are known to be readily available compared to that of primary data. It requires very little research and need for manpower to use these sources.
- With the advent of electronic media and the internet, secondary data sources have become more easily accessible. Some of these sources are highlighted below.
- Books
- Published sources
- Unpublished Personal Sources
- Journal
- Newspapers
- Websites
- Blogs
- Diaries
- Government Records
- Podcasts



Important Sources of Secondary Data - II

- **Books:** Books are one of the most traditional ways of collecting data. Today, there are books available for all topics you can think of. When carrying out research, all you have to do is look for a book on the topic being researched on, then select from the available repository of books in that area.
- Books, when carefully chosen are an authentic source of authentic data and can be useful in preparing a literature review.
- **Published Sources:** There are a variety of published sources available for different research topics. The authenticity of the data generated from these sources depends majorly on the writer and publishing company.
- Published sources may be printed or electronic as the case may be. They may be paid or free depending on the writer and publishing company's decision.

Important Sources of Secondary Data - III

- **Unpublished Personal Sources:** This may not be readily available and easily accessible compared to the published sources. They only become accessible if the researcher shares with another researcher who is not allowed to share it with a third party.
- **Example:** the product management team of an organization may need data on customer feedback to assess what customers think about their product and improvement suggestions. They will need to collect the data from the customer service department, which primarily collected the data to improve customer service.
- **Journal:** Journals are gradually becoming more important than books these days when data collection is concerned. This is because journals are updated regularly with new publications on a periodic basis, therefore giving to date information.
- Also, journals are usually more specific when it comes to research. For example, we can have a journal on, "Secondary data collection for quantitative data" while a book will simply be titled, "Secondary data collection".

Important Sources of Secondary Data - IV

- **Newspapers:** In most cases, the information passed through a newspaper is usually very reliable. Hence, making it one of the most authentic sources of collecting secondary data.

- The kind of data commonly shared in newspapers is usually more political, economic, and educational than scientific. Therefore, newspapers may not be the best source for scientific data collection.
- **Websites:** The information shared on websites are mostly not regulated and as such may not be trusted compared to other sources. However, there are some regulated websites that only share authentic data and can be trusted by researchers.
- Most of these websites are usually government websites or private organizations that are paid, data collectors.

Important Sources of Secondary Data - V

- **Blogs:** Blogs are one of the most common online sources for data and may even be less authentic than websites. These days, practically everyone owns a blog and a lot of people use these blogs to drive traffic to their website or make money through paid ads.
- Therefore, they cannot always be trusted. For example, a blogger may write good things about a product because he or she was paid to do so by the manufacturer even though these things are not true.
- **Diaries:** They are personal records and as such rarely used for data collection by researchers. Also, diaries are usually personal, except for these days when people now share public diaries containing specific events in their life.
- A common example of this is Anne Frank's diary which contained an accurate record of the Nazi wars.

Important Sources of Secondary Data - VI

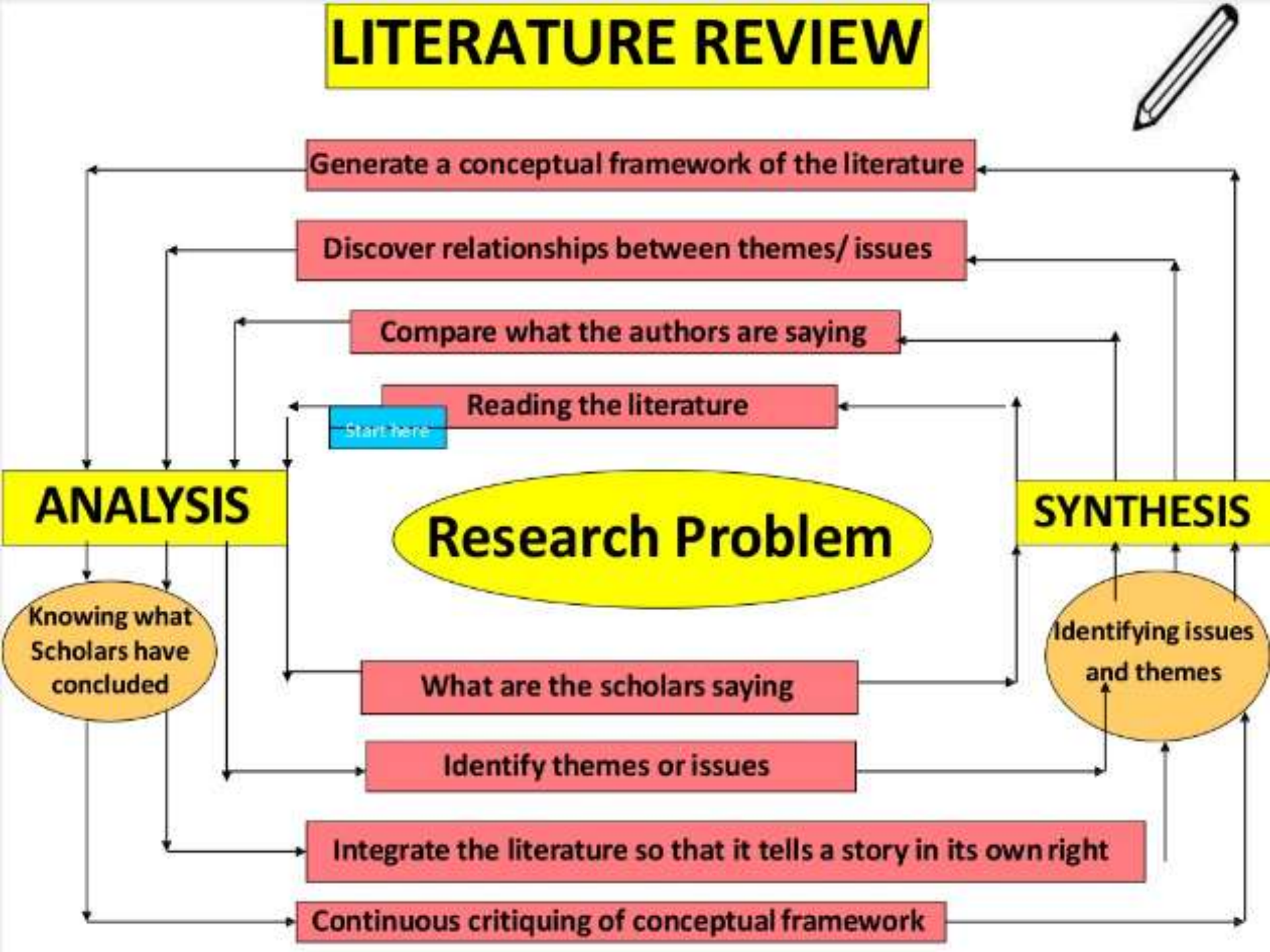
- **Government Records:** Government records are a very important and authentic source of secondary data. They contain information useful in marketing, management, humanities, and social science research.
- Some of these records include; census data, health records, education institute records, etc. They are usually collected to aid proper planning, allocation of funds, and prioritizing of projects.
- **Podcasts:** Podcasts are gradually becoming very common these days, and a lot of people listen to them as an alternative to radio. They are more or less like online radio stations and are generating increasing popularity.
- Information is usually shared during podcasts, and listeners can use it as a source of data collection.
- **Some other sources of data collection include:**
- Letters
- Radio stations
- Public sector records.

Review of Literature

- A Literature Review is a written summary of journals, articles, books and other documents that describe the past and current state of information on the topic of your research study.
- Review of literature is one of the most important steps in the research process.
- It is an account of what is already known about a particular phenomenon.
- The main purpose of literature review is to convey to the readers about the work already done & the knowledge & ideas that have been already established on a particular topic of research.
- Literature review is a laborious task, but it is essential if the research process is to be successful.

Meaning of Literature Review

- A review of literature is description and analysis of the literature relevant to a particular field or topic. It provides an overview of what world already had been carried out, who are the key researchers who did that work.
- It is actually the reading of the works of others before commencing on our own research work.
- Literature review can pave the way for better research.
- It can help in identifying the relevance of the research.
- A review tells the reader that the researcher know the research in the area. A good review increases a reader's confidence in the researcher's professional competence, ability and background.
- A good review places a research project in a context and demonstrates its relevance by making connections to a body of knowledge.


LITERATURE REVIEW
Generate a conceptual framework of the literature
Discover relationships between themes/ issues
Compare what the authors are saying
Reading the literature
Start here
ANALYSIS
Research Problem
SYNTHESIS
Knowing what Scholars have concluded
What are the scholars saying
Identifying issues and themes
Identify themes or issues
Integrate the literature so that it tells a story in its own right
Continuous critiquing of conceptual framework

Importance of Literature Review - I

- Identification of a research problem & Development or refinement of research questions.
- Generation of useful research questions or projects / activities for the discipline.
- Orientation to what is known & not known about an area of inquiry to ascertain what research can best contribute to knowledge.
- Determination of any gaps or inconsistencies in a body of knowledge.
- Discovery of unanswered questions about subjects, concepts or problems.
- Determination of a need to replicate a prior study in different study settings or different samples or size or different study populations.

Importance of Literature Review - II

- Identification of relevant theoretical or conceptual framework for research problems.
- Identification or development of new or refined clinical interventions to test through empirical research.
- Description of the strengths & weaknesses of design / methods of inquiry & instruments used in earlier research work.
- Development of hypothesis to be tested in a research study.
- Helps in planning the methodology of the present research study.
- It also helps in development of research instruments.
- Identification of suitable design & data collection methods for a research study.

Sources of Review of Related Literature

- **Primary Sources:**
- In the primary sources of information, the author reports his own work directly in the form of research articles books, monographs, dissertation and theses.
- Research published by educational journals are an example of this type of source.
- **Secondary Sources:**
- The author compiles and summarizes the finding of the work done by others and gives interpretation of these findings.
- **Example:** Encyclopedia of Education, education indexes, abstracts, bibliographies, bibliographical references and quotations etc.

Resources of Literature Review

The main sources from where literature can be searched are as ...



www.drjayeshpatidar.blogspot.com

PRIMARY AND SECONDARY SOURCES OF REVIEW OF RELATED LITERATURE/SCOPE

1. **Text Book**

2. **Reference material**

e.g.- Albert J. Walford, Guide to Reference Material.

This is a two volume work which covers:

(1) Science and technology (1966) and

(2)philosophy and psychology, religion, social sciences, geography and history (1968)

A. Encyclopedias :- These serve a store house of information and usually contain well rounded discussion and selected bibliographies that are prepared by specialist.

Sample of encyclopedias that researcher in the field of education might use

(a) Encyclopedia of Modern Education, Henry D. Rivlin and H. Schueller ed.,(New York: Philosophical Library, 1943). This comprehensive work of about 200 authorities has been edited by Henry D.Rivlin and H. Schueller. It stresses present-day problems, trends, theories and practices. The articles are accompanied by brief bibliographies and there is a system of cross references.

B. Dictionaries:-

They serve as constant guides to the researcher. A few known dictionaries are:

(a)Dictionary of Education, (New York: Mc Graw Hill Book Co.,1973) This dictionary covers 33,000 technical and professional terms. It also includes educational terms used in various countries.

(b)Dictionary of Sociology, Totowa, N.J.,(Littlefield,Adams and Co.).In this dictionary sociological terms are defined in non-technical language.



C. Thesaurus:-

A Thesaurus is the opposite of a dictionary. One turns to the Thesaurus when one has an idea, but does not yet have appropriate word to convey it. Thesaurus lists together the synonyms and antonyms of words. A researcher should use these reference in conjunction with a good dictionary to ensure precision of expression. Roget's International Thesaurus of Words and Phrases, (New York: Crowell, Collier and Macmillan) is a good example of Thesaurus.

- **D. Yearbook,Almanacs and Handbooks:-**

A large amount of current information on educational problems , thought and practices may be found in yearbooks, almanacs and handbooks. Some yearbooks cover a new topic of current interest each year and some others give more general reviews of events . A few examples of yearbooks , almanacs and handbooks are:

- **(a)** The Handbook of Research on Teaching, N.L.Gage(ed.), (Chicago: Rand McNally & Co.,1963)This handbook presents a comprehensive research information on teaching with extensive bibliographies.

- **(b)** Education Yearbook. (New York: Macmillion Co.,1972-date). It includes statistical data on major educational issues and movements with a comprehensive bibliography and reference guide.

- **(c)** World Almanac-Book of facts, (New York: Newspaper Enterprise Association, 1968-date) It provides up-to-date statistics and data concerning events, progress and conditions in social, educational, political, religious, geographical, commercial, financial and economic fields.

- **E. Directories and bibliographies:**

Directories are used by a researcher to locate the names and addresses of persons , periodicals , publishers or organizations when he wants to obtain information , about financial assistance or research material and equipments . A few important directories in the USA and the UK are:

- **(a)** Guide to American Educational Directories. It lists one volume over 12000 educational and allied directories . The directories are listed alphabetically and are arranged under subject headings.

- (b) Educator's World, (Englewood, Colo.: Fisher Publishing Co.,1972-date)d . This is an annual guide to more than 1600 education associations publications research centres and foundations.
- (c) American Library Directory, (New York: R . R. Bowker Co.,1923-date). This directory provides a biannual guide to private , state, municipal , institutional, and collegiate libraries in the USA and Canada. It includes information on special collections, number of holdings , staff salaries , budgets and affiliations.
- (d) Mental Health Directory,(Washington , D .C.: National Institute of Mental Health ,Govt. Printing Office, 1964-date). This annual publication lists national , state and local mental health agencies in USA.

- (e) NEA Handbook for Local , State and National Associations, (Washington, D. C. : National Associations 1945 – date) . This is an annual publication and contains listing and comprehensive reports of state and national officers of affiliated associations and departments.

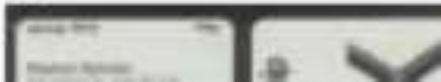


(f) The Education Directory,(Washington :US Office of Education , Superintendent of Documents, 1912 –date). This directory is published annually in five parts ., It deals with names , educational agencies , officials , institutions and other relevant data.

3. The Index : A Periodical index serves the same purpose as the index of a book or the card file of a library . It identifies the source of the article or of the book cited by listing the titles alphabetically, under author , and the readers should read all such directions before trying to locate the references.

A list of some important educational indexes is:

(a) Education Index, (New York: H.W.Wilson Co.,1929-date).It is published monthly . It indexes more than 250 educational periodicals , many yearbooks , bulletins and monographs published in the United States , Canada and Great Britain.The material on adult education , business education , curriculum , educational administration , educational psychology , educational research , exceptional children ,



higher education ,guidance , health and physical education , international education and teacher education are included in this index.

(b)ERIC Educational Document Index, (Washington, D.C.: National Institute of Education, Government Printing Office,1966-date).

(c)International Guide to Educational Documentation, (Paris: UNESCO).

(d) British Education Index .

(e)Index to Selected Educational Periodicals,(Leeds : Librarians of Institute of Education, 1945 – date)

4.Research Periodicals

Periodicals give much more up to date treatment to current question in education than books possibly can. They also publish articles of temporary, local or limited interest that never appear in the book form. There are approx 2100 journals that are related with education.

Example: Ulrich's Classified Guide to a Periodicals, Foreign and Domestic, (New York: Bowker),provides a comprehensive list of periodicals related to education. Periodicals are grouped in a subject classification and are alphabetically arranged.

Each entry includes title; sub-title, date of the origin , frequency of the publication , annual index , cumulative indexes , of each periodicals.

- In India , many periodicals are published by some associations . They provide a medium for dissemination of educational research and exchange of experience among research workers , teachers , scholars and other interested in educational research and related fields and professions.

5.Abstract:

Abstract includes the brief summaries of the contents of the research study or article. They serve the most useful reference guides to the researcher.

In America the most useful of these references are following :

The Review of Educational Research : It gives an excellent overview of the work that has been done in the field and about the recent developments. This publication ,between 1931 and 1969 ,reviewed about every three years each of the following eleven major areas of education: (i) administration (ii) curriculum (iii) educational measurement (iv) educational psychology (v) educational sociology

(vi)guidance and counselling (vii) language art ,fine art, natural sciences mathematics (viii) research methods (ix) special programmes (x) mental and physical development and (xi) teaching personnel.

Research in Education : RIE is published monthly since 1966 by the Educational Resource Information Centre (ERIC) and indexed annually. Each monthly issue of it is divided into three sections:

(i) Document Section

(ii) Project Section

(iii) Accession Number Section.

- Psychological abstract : This is published by the American Psychological Association since 1927. It is published bimonthly and contains abstracts of articles appearing in all over the 530 journals ,mostly educational periodicals.

- Education Abstract: This is a publication of UNESCO ,which began in1949 and has been published monthly except July and August. Each introductory essay devoted to a particular aspect of education ,is followed by abstract of books and documents, selected from various countries dealing with the topic under consideration.a

In addition to the above periodicals ,a researcher may also consult the following publications :

(i) Annual Review of Psychology (1950-date);

(ii)Child Development Abstract and Bibliography (1927-date);

(iii) Psychological Bulletin(1904-date);

(iv) Sociological Abstracts(1952-date);

(v)Educational Administration Abstracts (1966-date);

(vi) Sociology of Education Abstracts (1965-date);

(vii) Mental Retardation Abstracts (1964-date)

(viii) Dissertation Abstracts International(1952-date)

6.Theses and Dissertations: Theses and Dissertations are usually preserved by the universities that award the author their doctoral and masters degree. Sometimes these are published in whole or in part in various educational journal or periodicals.

In the USA ,reference of doctoral dissertations in all the fields , including education , can be found in sources compiled by various agencies.

. For the period1912 to 1938 , the Library of Congress issued the annual List of American Doctoral Dissertations for published studies.

. The Association of Research Libraries published the list of Doctoral Dissertations accepted by the American Universities from 1933-1934 to1954-1955.

7.Dissertation Abstract:

In India only few universities publish abstract of dissertations and theses have been completed at the institution.

- Kurukshetra University, Kurukshetra (Haryana) published : Abstract of M.Ed. Dissertations Vol.I, 1966;

  Abstract of M.Ed. Dissertations Vol.II,1967;

  Abstract of M.Ed. Dissertations Vol.III ,1968;

- Dissertation Abstracts International ,May1970, abstracts dissertations in the humanities, social sciences , physical sciences and engineering . It is published monthly.

8. Survey

M. B Buch ed. A Survey of Research in Education, (Centre of Advanced Study in Education , Baroda :M.S. University,1973).This publication contains all research studies in education completed in Indian Universities up to 1972.The abstract of all the studies have been classified into 17 meaningful areas of education . They are :

(i) Philosophy of Education (ii) History of Education (iii) Sociology of Education (iv) Economics of Education (v) Comparative Education (vi) Personality , Learning and Motivation (vii) Guidance and Counselling (viii) Test and Measurement (ix) Curriculum ,Methods and Textbook (x) Educational Technology



(xi) Correlates of Achievement (xii) Educational Evaluation and Examination (xiii) Teaching and teaching Behaviour (xiv) Teacher Education (xv) Educational Administration (xvi)Higher Education and (xvii) Non Formal Education.

- M. B Buch ed. Second Survey of Research in Education (1971-1978) (Baroda :Society for Educational Research and Development,1979).
- This publication incorporates 839 research studies completed during the period of 1972-1978.
- It follows the same pattern of organization of 17 research areas as A Survey of Research in Education (1973).

- The abstracts are arranged alphabetically for each area and continuously numbered throughout the volume.

- Each abstract contains the title of the study , the objective and hypotheses examined , methodolgy including the sample , tools of the research , the statistical techniques used and the findings.

- The trend reports are based not on the research completed during the period 1972-1978 ,but on the total research activities during the period 1940 to 1978.

- M. B Buch ed. Third Survey of Research in Education (1978-1983) New Delhi :NCERT 1987.
- The publication comprises of 20 chapters beginning with a comprehensive review for the general trend of research in education in India .
- In all ,1481 research abstracts have been presented after being classified under the 17 areas.
- A special feature of the volume is the chapter on 'Research on Indian Education Abroad' which presents a review of 192 doctoral dissertations submitted to American and British universities, covering a period of around two decades.

9 .Newspapers and Pamphlets

- Many articles of particular interest to a researcher may locate through pamphlets and newspapers. Current newspaper provides up-to-date information on speeches . Seminars , conferences ,new trends and number of other topics. Old newspapers, which preserve a record of past events ,movements and ideas are particularly useful in historical inquiries.

10. Government Documents

- Govt. documents are rich source of information. They include statistical data ,research studies , official reports laws and other material that are not available elsewhere. These are available in national ,state as well as local level govt. offices.

11. Monographs



- Monographs are also major source of information on ongoing research.

- In USA , universities and teacher's colleges publish many research studies in education in the form of monographs.

- A few examples of these are Supplementary Educational Monographs, Educational Research Monographs and Lincoln School Monographs.

12.Computer Generated Reference Material



- School Research Information Service (SRIS), Direct Access to Reference Information (DATRIX) ,and Psychological Abstract and Retrieval Service (PASAR) in the United States provide a number of computer generated reference sources that may have a great deal of time and effort of the researcher.
- SRIS provides a computer printout of abstracts for a moderate fee.
- DATRIX provides computerised retrieval for Dissertation Abstract , from 1928 to date.

- ***Online Journals:*** Following are the website addresses for journals & magazines that are available online:

- http://www.nursefriendly.com/nursing/linksections/nursingjournal.html
- http://www.nsna.org
- http://www.healthweb.org
- http://www.ispub.com/ostia/index.php
- http://www.healthweb.org.browse.cfm?catergy=1727
- http://www.juns.nursing.arizona.edu
- http://www.medbioworld.com
- http://www.nursingworld.org.ojin
- http://www.eaa-knowledge.com.ojni/#
- http://www.nursingweek.com

- ***Other online databases:*** Many other online database can be searched for free by nurses from the following websites:
- http://www.aidsinfo.nih.gov (HIV/AIDS information)
- http://www.hazmap.nlm.nih.gov (information on hazardous agents)
- http://www.child.nih.gov (combined health information database)
- http://www.toxinet.nlm.nih.gov (toxicology database network)
- ***Cancer Lit (Cancer literature)***
- ***EMBASE (exerpta Medica Database)***
- ***ETOH (Alcohol & alcohol problems science database)***
- ***Health STAR (Health services technology administration & research)***
- ***Radix (Nursing managed care databse)***
- ***CD-ROM (Compact disc-read only memory) with research***

www.drjayeshpatidar.blogspot.com

ORGANIZING THE RELATED LITERATURE (STEPS)





- Recording reference material
- Recording content of references
- Preparation of the related research report



STEPS IN CONDUCTING A LITERATURE REVIEW

1. **Identify Key Terms**

By narrowing your topic to a few key terms.



-Write a preliminary "working title" for a project and select two to three key words.

-Pose a short, general question that you would like answered in the study.



-Look in a catalog of terms

-Visit online database .e.g :ERIC database.

STEPS IN CONDUCTING A LITERATURE REVIEW

2 Locate literature



--Use academic libraries

--Use on line libraries e.g. :University of Michigan, University of California and Library of Congress.



--Use both primary and secondary sources.

STEPS IN CONDUCTING A LITERATURE REVIEW

3. Critically evaluate and select the literature.

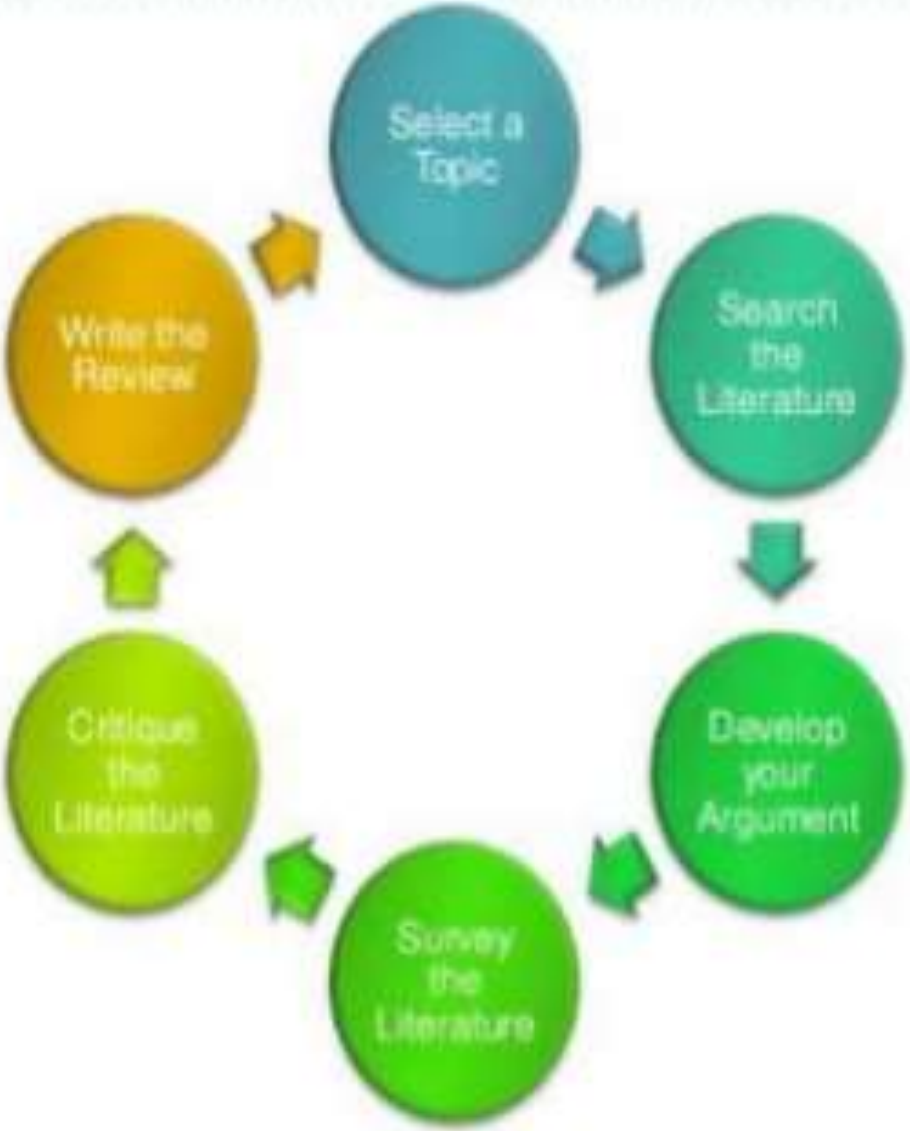


4. Organize the literature.

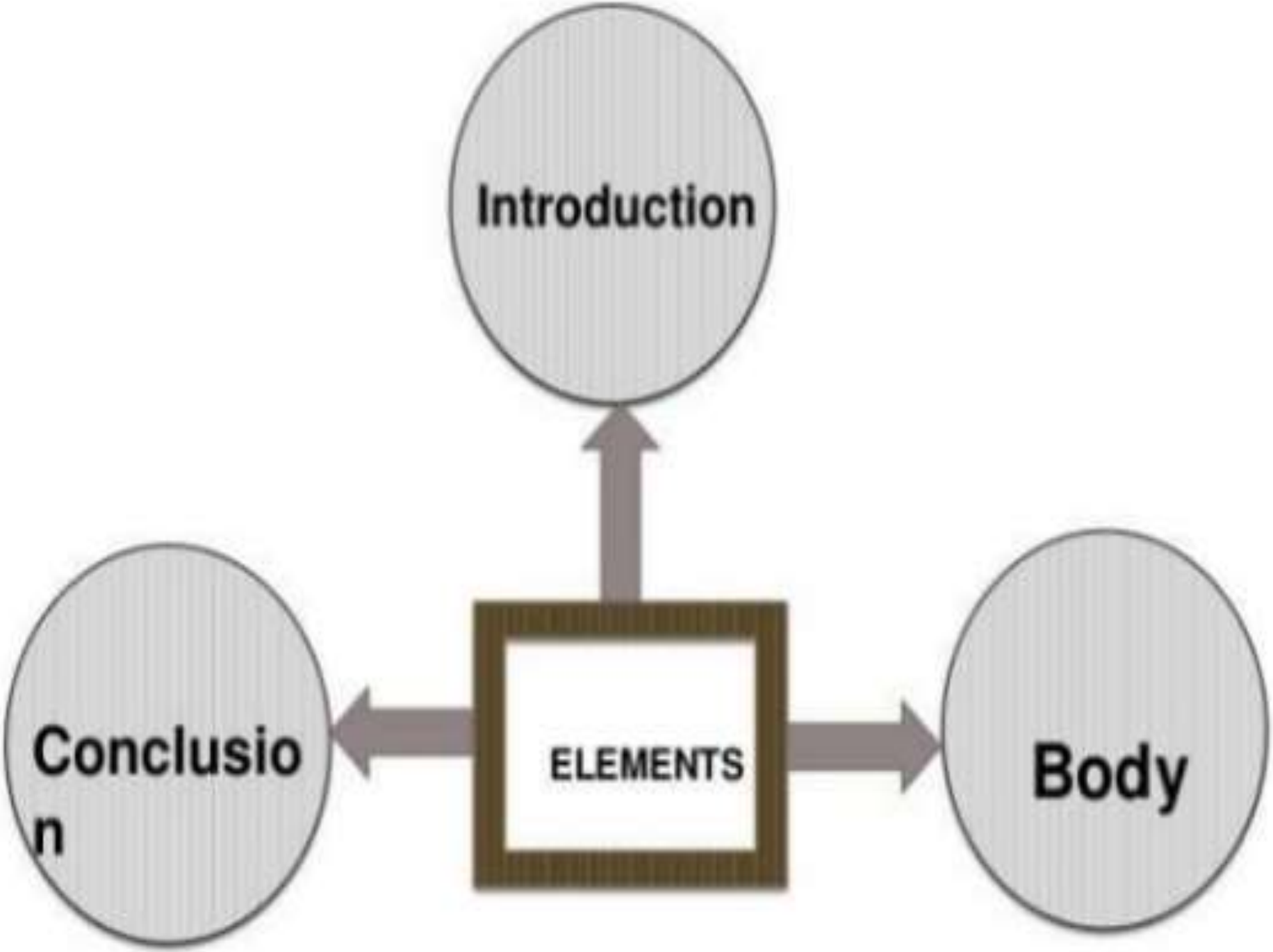
5. Write a literature review that reports summaries of the literature for inclusion in your research report.



The Literature Review Process



From *The Literature Review* (2009) by Machi and McEvoy


Introduction
Conclusio
n
ELEMENTS
Body

Literature Review Structure

- **Introduction:** Introduce the topic/problem and the context within which it is found.
- **Body:** Examine past research in the area highlighting methodological and/ or theoretical developments, areas of agreement, contentious areas, important studies and so forth. Keep the focus on your area of interest and identify gaps in the research that your research/investigation will attempt to fill. State clearly how your work builds on or responds to earlier work.
- **Conclusion:** Summarize what has emerged from the review of literature and reiterate conclusions.

Writing the Introduction…

While writing the introduction, following steps should be taken care of:

- Define or identify the general topic, issue, or area of concern, thus, providing appropriate context for reviewing the literature.
- Point out overall trends in what has been published about the topic or conflicts in theory, methodology, evidence, & conclusion or gaps in research & scholarship, or a single problem or new perspective of immediate interest.
- Establish the writer's point of view for reviewing the literature, explain the criteria to be used in analyzing & comparing literature & organization or review

POINTS TO BE CONSIDERED FOR LITERATURE REVIEW



- ***Be specific & be succinct:***



Briefly state specific findings listed in an article, specific methodologies used in a study, or other important points. Literature reviews are not the place for long quotes or in-depth analysis of each point.

- ***Be selective:***

Researcher should narrow down a lot of information into a small space for literature review. Just the most important points (i.e. those most relevant to the review's focus) must be mentioned in each work of review.

- ***Focus of current topics:***

Researcher needs to analyse points such as if it is a current article, & if not, how old it is: has its claims, evidence, or arguments been superseded by more recent work; if it is not current, then if it is important for historical background ; etc.

- ***Ensure evidence for claims:***

Researcher should focus on what support is given for claims made in literature. What evidence & what type (experimental, statistical, anecdotal, etc.) of evidences are offered? Is the evidence relevant & sufficient? What arguments are given? What assumptions are made, & are they warranted?

- ***Focus on sources of evidences:***

Researchers should ensure the reliability of the sources of the evidence or other information – if they are from author's own experiments, surveys, historical records, government documents, etc. He should check how reliable those sources are.

- ***Account of contrary evidences:***

  Does the author take into account contrary or conflicting evidence & arguments? How does the author address disagreements with other researchers?

- ***Reference citation:***

  Any references cited in the literature review must be included in the bibliography. The common practice is that the reviewer does not list references in the bibliography that are not directly cited in the literature review or elsewhere in the paper /thesis.

- ***Avoid abbreviations:***

  Avoid technical terms, jargons & abbreviations.

- ***Simple & accurate sentence structure:***

  A researcher should use simple sentences & must avoid errors of grammar & punctuation

- ***Organization of literature review:***

A literature review is organized by subtopic, not by individual references. In a typical literature review, the writers may cite several references in the same paragraph & may cite the same reference in more than one paragraph, if that source address more than one of the subtopics in the literature review. Typically, discussion of each sources is quite brief. The contribution the present reviewers make is organizing the ideas from the sources into a cogent argument or narrative that includes their perspectives.

- ***Referring original source:***

The reviewer should focus on citing the material that originates with each reference. This may require a careful reading of the reference. If the reference author refers to another source whose ideas are relevant or interesting, it is better to track & use that original reference.

Review the Literature

- Now that the problem has been identified, the researcher must learn more about the topic under investigation.
- To do this, the researcher must review the literature related to the research problem. This step provides foundational knowledge about the problem area.
- The review of literature also educates the researcher about what studies have been conducted in the past, how these studies were conducted, and the conclusions in the problem area.
- In the obesity study, the review of literature enables the programmer to discover horrifying statistics related to the long-term effects of childhood obesity in terms of health issues, death rates, and projected medical costs.
- In addition, the programmer finds several articles and information from the Centers for Disease Control and Prevention that describe the benefits of walking 10,000 steps a day.
- The information discovered during this step helps the programmer fully understand the magnitude of the problem, recognize the future consequences of obesity, and identify a strategy to combat obesity (i.e., walking).



Steps to Writing A Literature Review

❖ Find a working Topic

Look at your specific area of study. Think about what interests you, and what is fertile ground for study. Talk to your professor, brainstorm, and read lecture notes and recent issues of periodicals in the field.

❖ Review the literature

Using keywords, search a computer database. It is best to use at least two databases relevant to your discipline. Remember that the reference lists of recent articles and reviews can lead to valuable papers. Make certain that you also include any studies contrary to your point of view.

❖ Focus your topic narrowly and select papers accordingly

Consider the following:

- ✓ **What interests you?**
- ✓ **What interests others?**
- ✓ **What time span of research will you consider?**

❖ Read the selected articles thoroughly and evaluate them

- ✓ **What assumptions do most/ some researchers seem to be making ?**
- ✓ **What methodologies do they use? What testing procedures, subjects, material tested?**
- ✓ **Evaluate and synthesize the research findings and conclusions drawn**
- ✓ **Note experts in the field: names/ labs that are frequently referenced**
- ✓ **Note conflicting theories, results, methodologies**

❖ Developing subtopics in the review of literature process

Note things such as:

- **Findings that are common/ contested**
- **Two or three important trends in the research**
- **The most influential theories**

❖ Develop a Working Thesis

Write a one or two sentence statement summarizing the conclusion you have reached about the major trends and developments you see in the research that has been done on your subject.



- **Organize your own paper based on the Findings From steps 4&5**

Develop headings / subheadings. If your literature review is extensive, find a large table surface, and on it place post-it notes or filing cards to organize all your finding into categories . Move them around if you decide that (a) they fir better under different headings, or (b) you need to establish new topic headings.

- **Write the body of the paper**

Follow the plan you have developed above, making certain that each section links logically to the one before and after, and that you have divided your sections by themes or subtopics, not by reporting the work of individual theorists or researchers.

❖ Look at what you have written; focus on analysis, not description

So if your paper still does not appear to be defined by a central, guiding concept, or if it does not critically analyse the literature selected, then you should make a new outline based on what you have said in each section and paragraph of the paper, and decide whether you need to add information, to delete off - topic information, or to restructure the paper entirely.



Review the Key Factors Involved - I

- As a researcher, you must work closely with your team of researchers to define and test the influencing factors and the wider context involved in your study.
- These might include demographic and economic trends or the business environment affecting the question at hand.
- To do this you have to identify the factors that will affect the research project and begin formulating different methods to control for them.
- You also need to consider the relationships between factors and the degree of control you have over them.
- **Example:** you may be able to control the loading speed of your website but you can't control the fluctuations of the stock market. Doing this will help you determine whether the findings of your project will produce enough information to be worth the cost.

Review the Key Factors Involved - II

- You need to determine:
  - ✓ Which factors affect the solution to the research problem.
  - ✓ Which ones can be controlled and used for the purposes of the company, and to what extent.
  - ✓ The functional relationships between the factors
  - ✓ Which ones are critical to the solution of the research problem.
- **Example:** The research team at the running shoe company is hard at work. They explore the factors involved and the context of why YoY sales are down for trail shoes, including things like what the company's competitors are doing, what the weather has been like – affecting outdoor exercise – and the relative spend for the brand from year to year.
- The final factor is within the company's control, although the first two are not. They check the figures and determine marketing spend has a significant impact on the company.

Prioritize - I

- Once you and your research team have a few observations with promise, prioritize them based on their business impact and importance.
- It may be that you can answer more than one question with a single study, but don't do it at the risk of losing focus on your overarching research question.
- **Questions to ask:**
- Who? Who are the people with the problem? Are they end-users, stakeholders, teams within your business? Have you validated the information to see what the scale of the problem is?
- What? What is its nature and what is the supporting evidence?
- Why? What is the business case for solving the problem? How will it help?
- Where? How does the problem manifest and where is it observed?
- To help you understand all dimensions, you might want to consider focus groups or preliminary interviews with external (including consumers and existing customers) and internal (salespeople, managers and other stakeholders) parties to provide what is sometimes much – needed insight into a particular set of questions or problems.

Prioritize - II

- **Example:** After observing and investigating, the running shoe researchers come up with a few candidate research questions, including:
- What is the relationship between US average temperatures and sales of our products year on year?
- At the present time, how does our customer base rank Competitor X and Competitor Y's trail running shoe compared to our brand?
- What is the relationship between marketing spend and trail shoe product sales over the last 12 months?
- They opt for the final question, because the variables involved are fully within the company's control, and based on their initial research and stakeholder input, seem the most likely cause of the dive in sales.
- The research question is specific enough to keep the work on course towards an actionable result, but it allows for a few different avenues to be explored, such as the different budget allocations of offline and online marketing and the kinds of messaging used.

Align

- Get feedback from the key teams within your business to make sure everyone is aligned and has the same understanding of the research question and the actions you hope to take based on the results.
- Now is also a good time to demonstrate the ROI of your research and lay out its potential benefits to your stakeholders.
- Different groups may have different goals and perspectives on the issue.
- This step is vital for getting the necessary buy-in and pushing the project forward.
- **Example:** The running shoe company researchers now have everything they need to begin. They call a meeting with the sales manager and consult with the product team, marketing team, and C-suite to make sure everyone is aligned and bought in to the direction of the research.
- They identify and agree that the likely course of action will be a rethink of how marketing resources are allocated, and potentially testing out some new channels and messaging strategies.

Collect Data

- Once the instrumentation plan is completed, the actual study begins with the collection of data. The collection of data is a critical step in providing the information needed to answer the research question.
- Every study includes the collection of some type of data—whether it is from the literature or from subjects—to answer the research question.
- Data can be collected in the form of words on a survey, with a questionnaire, through observations, or from the literature.
- In the obesity study, the programmers will be collecting data on the defined variables: weight, percentage of body fat, cholesterol levels, and the number of days the person walked a total of 10,000 steps during the class.
- The researcher collects these data at the first session and at the last session of the program.
- These two sets of data are necessary to determine the effect of the walking program on weight, body fat, and cholesterol level.
- Once the data are collected on the variables, the researcher is ready to move to the final step of the process, which is the data analysis.

The process of data collection can be defined in four stages

- Selection of fieldworkers;
- Training of fieldworkers;
- Supervision of fieldworkers;
- Evaluation of fieldwork and fieldworkers.



Validation of Data

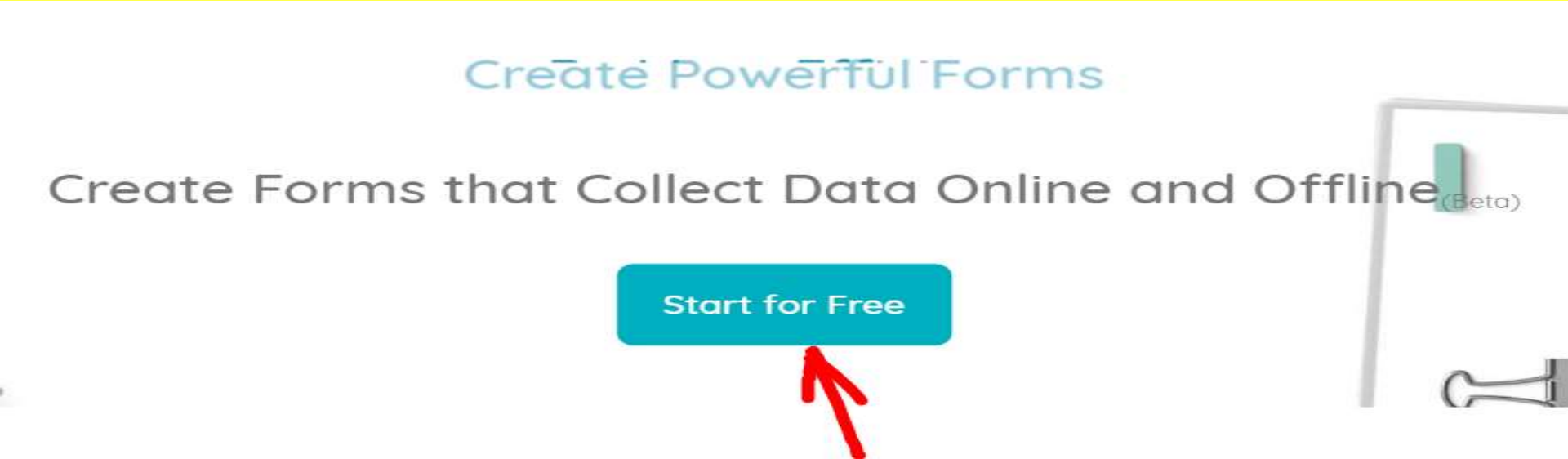
- After the data is coded, it is validated for data entry errors. The data is then used for further analysis. The purpose of validating the data is that it has been collected as per the specifications in the prescribed format or questionnaire.
- **Example:** If the respondent is asked to rate a particular aspect on 1 to 7, then the obvious responses should be 1 or 2….,7. Any other inputted number is not considered as valid.
- In validation of the data, the above data will be restricted to the integers between 1 and 7. This minimizes the errors.
- The other validations are age within a number like 100, dates such as birth dates, joining dates, etc should not be future dates, etc.

Classification

- Data having a common characteristic are placed in one class and in this way the entire data divided into a number of groups or classes. Classification can be one of the following two types, depending upon the nature of the phenomenon involved:
- Classification according to attributes: As stated above, data are classified on the basis of common characteristics which can either be descriptive (such as literacy, sex, honesty, etc.) or numerical (such as weight, height, income, etc.).
- Classification according to class-intervals: Data relating to income, production, age, weight, etc. come under this category. Such data are known as statistics of variables and are classified on the basis of class intervals.
- For instance, persons whose incomes, say are within Rs.201 to Rs.400 can form one group, those whose incomes are within Rs.401 can form another group.

Steps to Collect Data with Surveys or Questionnaires - I

- As a business owner who wants to regularly track the number of sales made in your business, you need to know how to collect data. Follow these 4 easy steps to collect real-time sales data for your business using Formplus.
- **Step 1 - Register on Formplus**
  - ✓ Visit formplus on your PC or mobile device.
  - ✓ Click on the Start for Free button to start collecting data for your business.
  - ✓ Register with your email address or with Google in 2 seconds

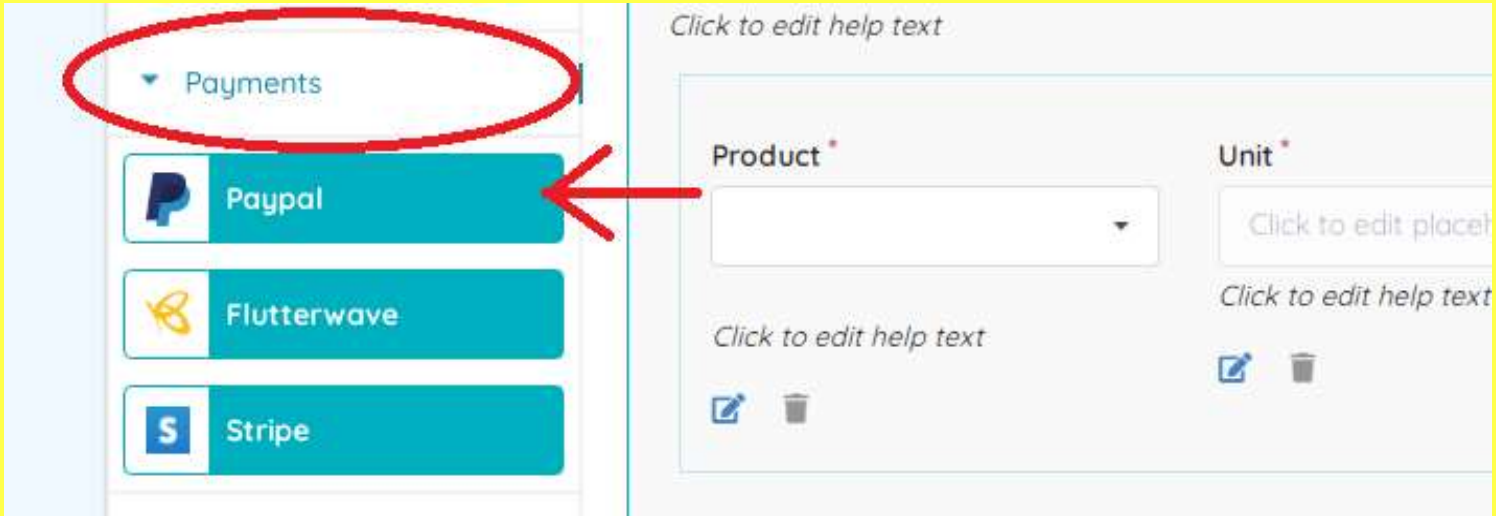


Steps to Collect Data with Surveys or Questionnaires - II

- **Step 2 - Start Creating Surveys For Free**
  - Go to the Forms tab beside your Dashboard in the Formplus menu.
  - Click on Create Form to start creating your survey.



  - Take advantage of the dynamic form fields to add questions to your survey.
  - You can also add payment options that allow you to receive payments using Paypal, Flutterwave, and Stripe.



Steps to Collect Data with Surveys or Questionnaires - III

Step 3 - Customize Your Survey and Start Collecting Data

- Go to the Customize tab to beautify your survey by adding colors, background images, fonts, or even a custom CSS.
- You can also add your brand logo, color and other things to define your brand identity.
- Preview your form, share, and start collecting data.



Step 4 - Track Responses Real-time

- Track your sales data real-time in the **Analytics** section.

Steps to Collect Data with Surveys or Questionnaires - III

- After identifying a research topic, doing a literature background research, establishing philosophical assumptions and focus problem, deciding on an appropriate research paradigm and methodology with specific purpose, designing a research plan and collecting sufficient data, the next step in the research process is data analysis and interpretation, which precedes reporting of research.
- Data analysis is, therefore, a process that involves examining, and molding collected data for interpretation to discover relevant information, draw or propose conclusions and support decision-making to solve a research problem.
- This involves interpreting data to answer research questions and making research findings be ready for dissemination. Data analysis also serves as a reference for future data collection and other research activities.
- **During data analysis:** D*ata collected is transformed into information and knowledge about a research performed*
- *relationships between variables are explored*
- *meanings are identified and information is interpreted*.
- Like other research methods, data analysis procedures in quantitative research approach are different from those in qualitative research approach.

Conceptualizing a Research Design

- Once the research problem is identified the next step is to identify the knowledge gap, selection of appropriate methods, sampling strategy and time frame.
- A faulty design may lead to misleading findings and therefore it is important to ensure that your selection of research design permits the accurate prediction of outcome under the given set of limitation.
- While selecting a particular research design you must have strong reason and justification for your selection; and you should be aware of its strengths, weaknesses and limitations.

Selecting a sample design

- Sampling design selection is another crucial step in the research process, the research result will largely depend on the category of sampling design you have chosen for your research study.
- The three categories of sampling design random/probability sampling designs, non-random/non-probability sampling designs and ‘mixed’ sampling design will affect your strategies and you must be able to select most appropriate for your study by considering the strength and weakness of each along with their use case.

Writing a research proposal

- Having done all the preparatory work, the next step is to put everything together in a way that provides adequate information about your research study, for you research supervisor and others.
- Universities and other institutions may have different requirements regarding the style and content of a research proposal, but the majority of institutions would require most of what is set out here.
- This also varies within an institution, from discipline to discipline or from supervisor to supervisor. Therefore it should contain the following information about your study:
- Objectives of the study.
- List of hypotheses if you are testing any.
- Study design you are proposing to use.
- The research instruments you are planning to use.
- Information on sample size and sampling design.
- Information on data processing procedures.
- The study's problems and limitations.
- The proposed time-frame.

Processing and Displaying Data

- Processing of data you have collected through various methods requires different ways of analysis; analysis will be quantitative for descriptive study. If you want quantitative analysis, it is also necessary to decide upon the type of analysis required and how it should be presented.
- **Writing a Research Report:**
- Universities and other institutions may have different requirements regarding the style and content of a research proposal but there are two broad categories of reports:
  - ✓ Quantitative
  - ✓ Qualitative
- Writing the report is the last and, for many, the most difficult step of the research process. This report informs the world what you have done, what you have discovered and what conclusions you have drawn from your findings. If you are clear about the whole process. Your report should be written in an academic style and be divided into different chapters and/or sections based upon the main themes of your study.

Research Question

- The terms “research question” and “research problem” are often used interchangeably.
- Some researchers think in terms of a single research problem and a number of research questions that arise from it. The questions are lines of enquiry to explore in trying to solve the overarching research problem.
- It may be useful to think of problems as coming out of your business data – that’s the O - data (otherwise known as operational data) like sales figures and website metrics.
- For **example,** why do sales peak at certain times of the day, or why are customers abandoning their online carts at the point of sale. Research questions are a tool you use to solve those business problems.
- **Steps to Defining a Research Question**
- Observe and Identity
- Review the key factors involved
- Prioritize
- Align

Sampling

- **Definition:** Sampling is a technique of selecting individual members or a subset of the population to make statistical inferences from them and estimate characteristics of the whole population.
- Different sampling methods are widely used by researchers in market research so that they do not need to research the entire population to collect actionable insights.
- It is also a time-convenient and a cost-effective method and hence forms the basis of any research design. Sampling techniques can be used in a research survey software for optimum derivation.
- **Example:** If a drug manufacturer would like to research the adverse side effects of a drug on the country's population, it is almost impossible to conduct a research study that involves everyone.
- In this case, the researcher decides a sample of people from each demographic and then researches them, giving him/her indicative feedback on the drug's behavior.

Sampling

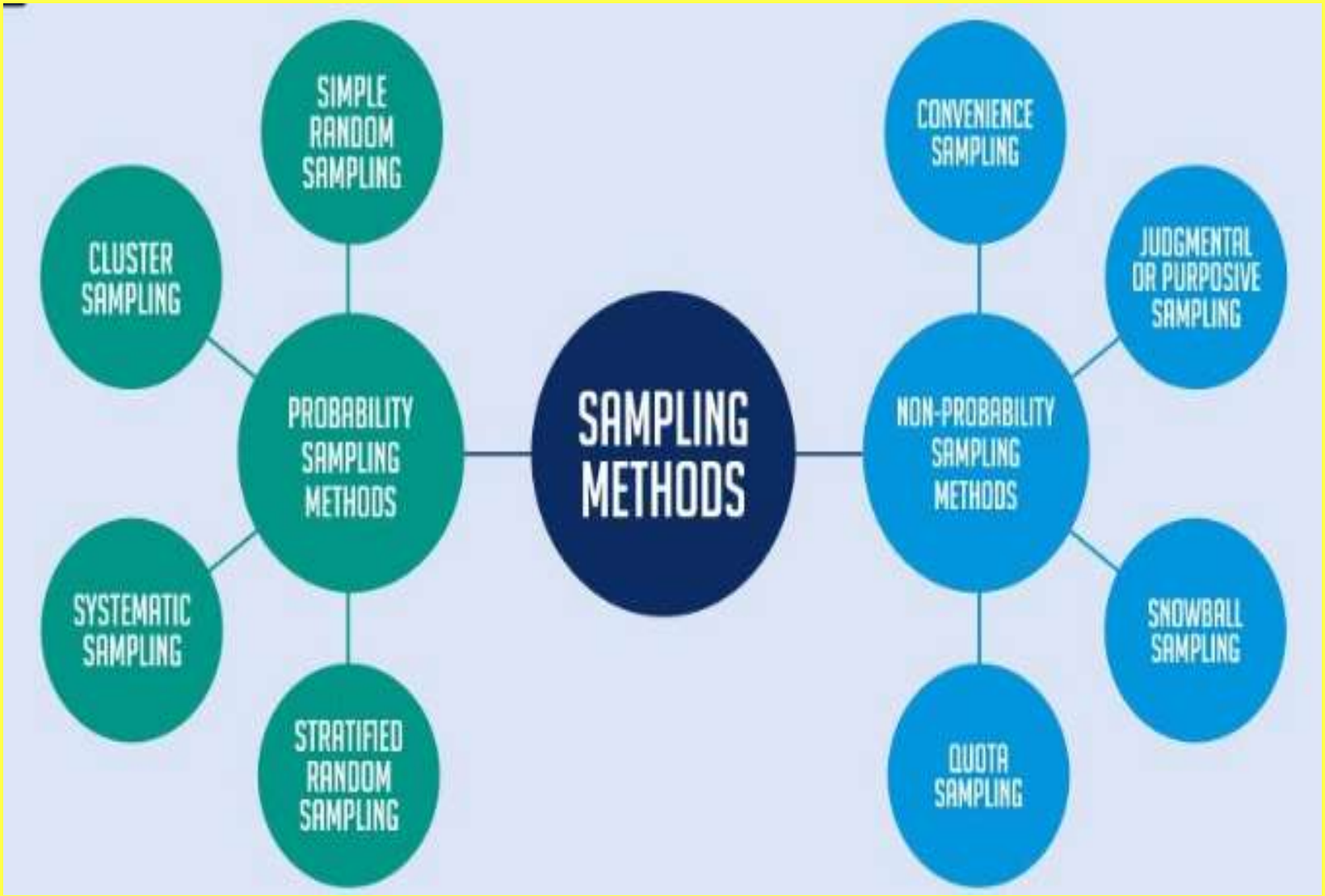
- When you conduct research about a group of people, it's rarely possible to collect data from every person in that group. Instead, you select a sample.
- The sample is the group of individuals who will actually participate in the research.
- To draw valid conclusions from your results, you have to carefully decide how you will select a sample that is representative of the group as a whole. There are two types of sampling methods:
- **Probability Sampling** involves random selection, allowing you to make statistical inferences about the whole group.
- **Non-Probability** Sampling involves non-random selection based on convenience or other criteria, allowing you to easily collect initial data.
- You should clearly explain how you selected your sample in the methodology section of your paper or thesis.

Classification of Sampling Techniques



- Sampling Techniques
  - Nonprobability Sampling Techniques
    - Convenience Sampling
    - Judgmental Sampling
    - Quota Sampling
    - Snowball Sampling
  - Probability Sampling Techniques
    - Simple Random Sampling
    - Systematic Sampling
    - Stratified Sampling
    - Cluster Sampling
    - Other Sampling Techniques

Sampling Methods



Types of Sampling Methods



Probability Sampling

Probability sampling utilizes random sampling techniques to create a sample. This group of sampling methods give all the members of a population equal chances of being selected.

Non-probability Sampling

It is a group of sampling techniques where the samples are collected in a way that does not give all the units in the population equal chances of being selected. Probability sampling does not involve random selection at all.

Types of Probability Sampling:

SIMPLE RANDOM SAMPLING

It is a fully random method of selecting subjects. All you need to do is ensure that all the subjects of a population are on the list and then randomly select the desired number of subjects.



Types of Non-probability Sampling:

CONVENIENCE SAMPLING

As the name suggests, this method involves collecting units that are the easiest to access: your local school, the mall, your nearest church and etc. It forms an accidental sample.

STRATIFIED RANDOM SAMPLING

Involves the division of a population into smaller groups, called 'strata'. Then the researcher randomly selects the final items proportionally from the different strata.



QUOTA SAMPLING

It aims to create a sample where the groups (e.g. males vs. females workers) are proportional to the population.The population is divided into groups and the samples are gathered from each group to meet a quota.

SYSTEMATIC SAMPLING

It is appropriate if we have a list of subjects arranged in a systematic order such as geographical order. The process includes first selecting a starting point in the population and then performing subsequent observations by using a constant interval between samples taken.



JUDGMENT SAMPLING

It is a sampling methodology where the researcher selects the units of the sample based on their knowledge. The selection includes nothing but purposely hand picking units from the population based on the researcher's knowledge.

CLUSTER RANDOM SAMPLING

It randomly selects members from a list that is too large. Cluster sampling design is used when natural groups occur in a population. The entire population is subdivided into clusters (groups) and random samples are then gathered from each group.

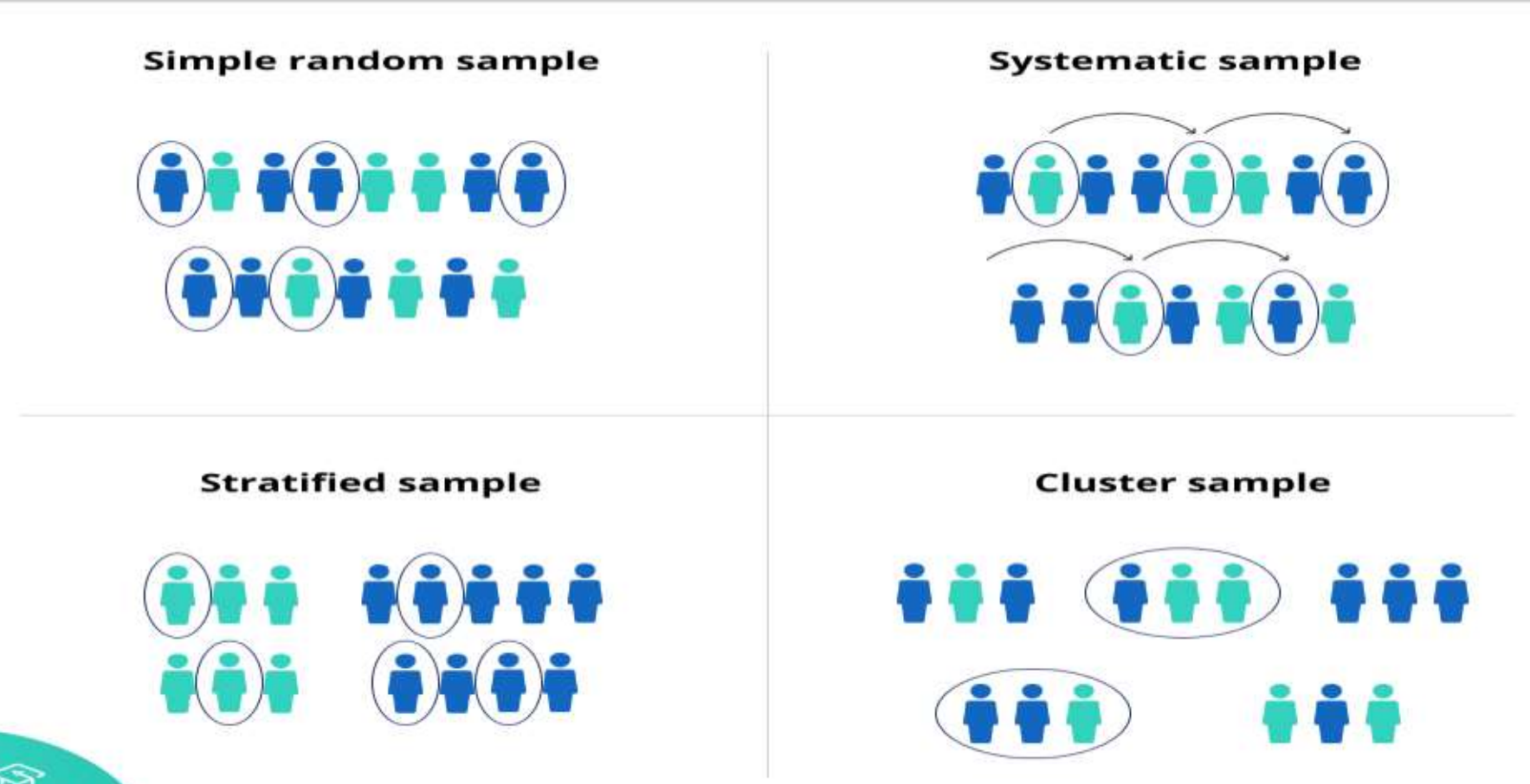


SNOWBALL SAMPLING

It is a method where researcher recruits other individuals for the study. This method is used only when the population is very hard-to-reach. For example, these include populations such as working prostitutes and current heroin users.

Probability Sampling Methods

- Probability sampling means that every member of the population has a chance of being selected. It is mainly used in quantitative research. If you want to produce results that are representative of the whole population, you need to use a probability sampling technique.
- There are four main types of probability sample.



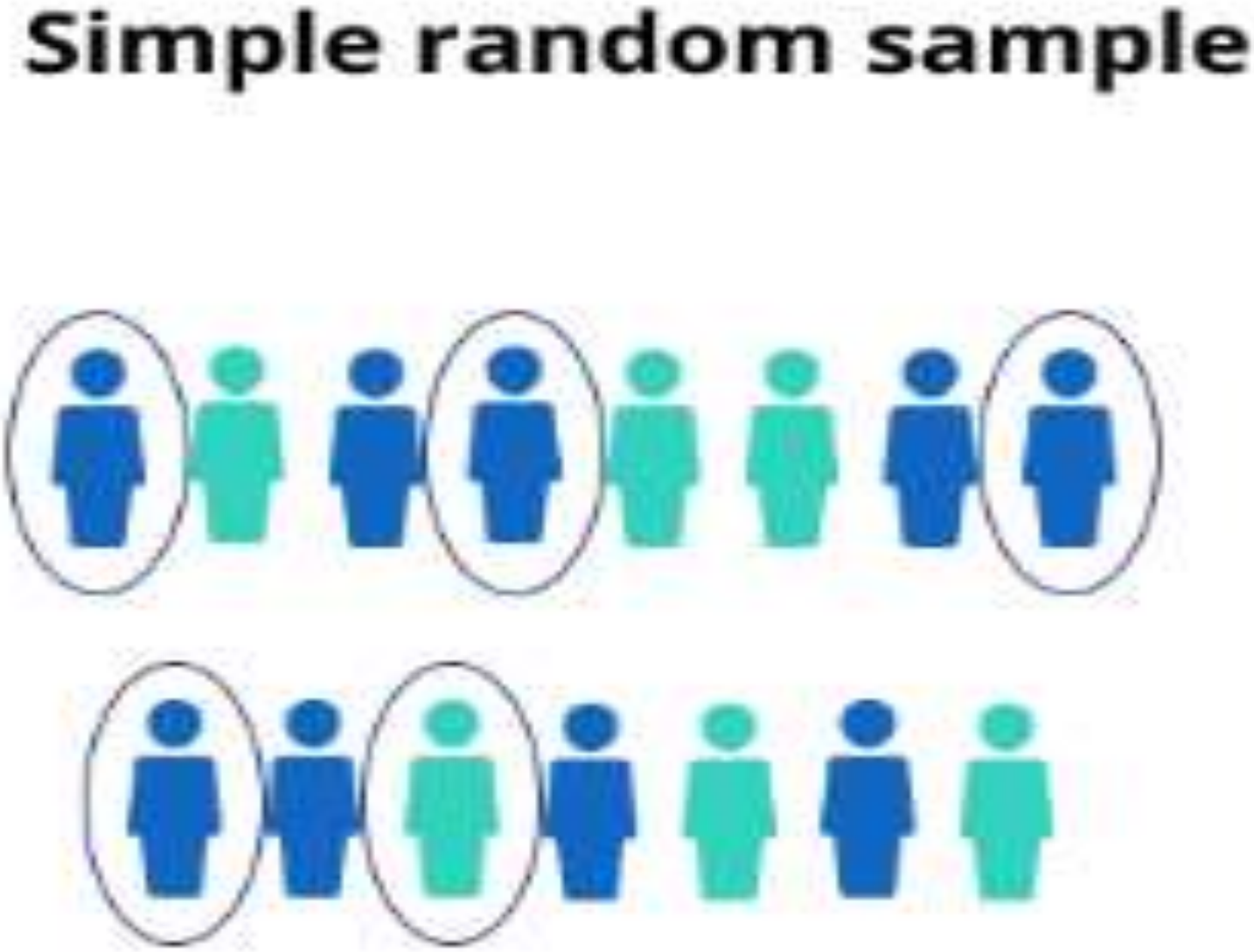
Scribbr

Simple Random Sampling

- In a simple random sample, every member of the population has an equal chance of being selected.
- Your sampling frame should include the whole population.
- To conduct this type of sampling, you can use tools like random number generators or other techniques that are based entirely on chance.
- **Example:** You want to select a simple random sample of 100 employees of Company X. You assign a number to every employee in the company database from 1 to 1000, and use a random number generator to select 100 numbers.



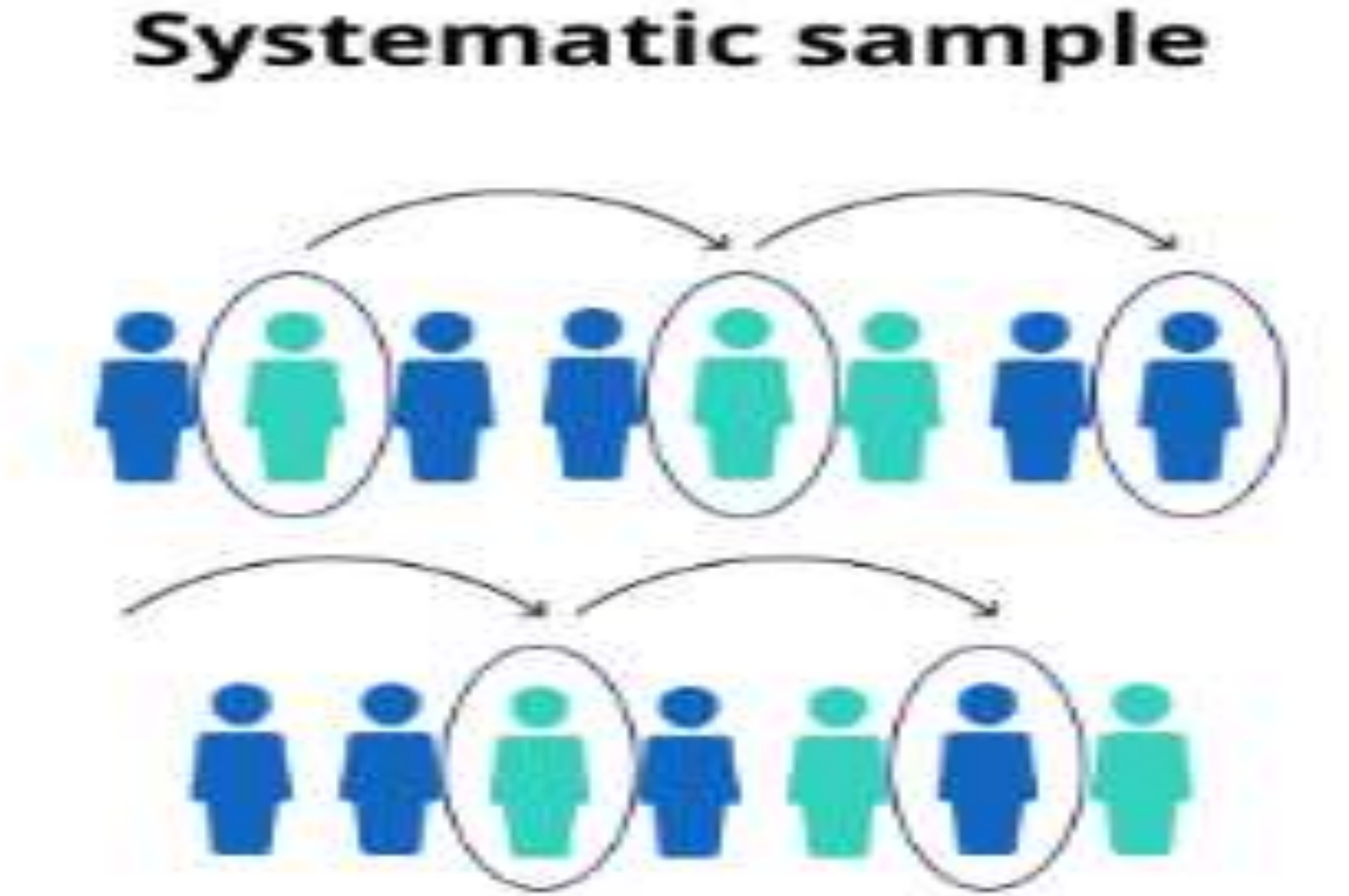
Simple Random Sampling



Systematic Sampling

- Systematic sampling is similar to simple random sampling, but it is usually slightly easier to conduct.
- Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.
- **Example:** All employees of the company are listed in alphabetical order. From the first 10 numbers, you randomly select a starting point: number 6. From number 6 onwards, every 10th person on the list is selected (6, 16, 26, 36, and so on), and you end up with a sample of 100 people.
- If you use this technique, it is important to make sure that there is no hidden pattern in the list that might skew the sample. For example, if the HR database groups employees by team, and team members are listed in order of seniority, there is a risk that your interval might skip over people in junior roles, resulting in a sample that is skewed towards senior employees.

Systematic Sampling



Stratified Sampling

- This sampling method is appropriate when the population has mixed characteristics, and you want to ensure that every characteristic is proportionally represented in the sample.
- You divide the population into subgroups (called strata) based on the relevant characteristic (e.g. gender, age range, income bracket, job role).
- From the overall proportions of the population, you calculate how many people should be sampled from each subgroup. Then you use random or systematic sampling to select a sample from each subgroup.
- **Example:** The company has 800 female employees and 200 male employees. You want to ensure that the sample reflects the gender balance of the company, so you sort the population into two strata based on gender. Then you use random sampling on each group, selecting 80 women and 20 men, which gives you a representative sample of 100 people.

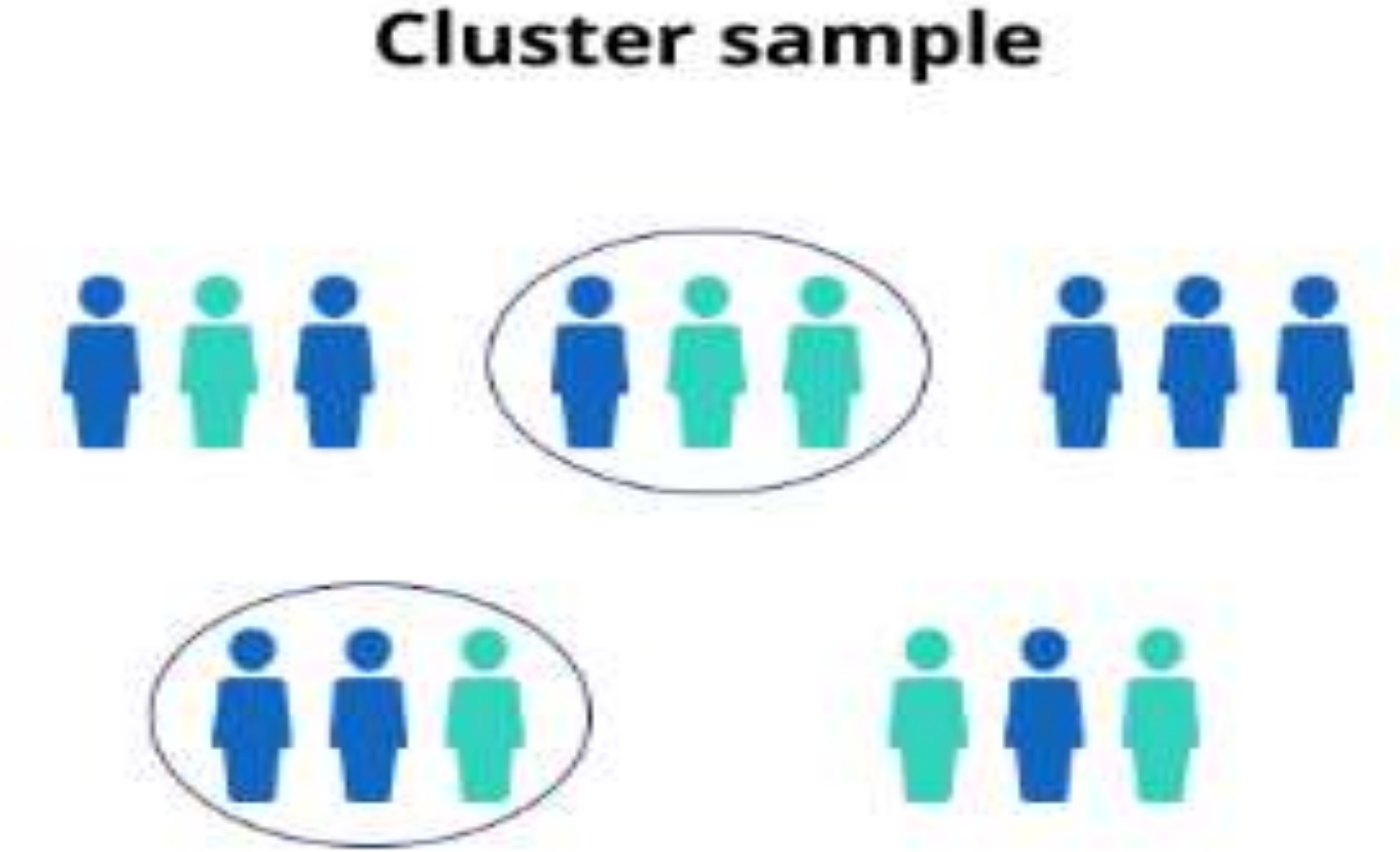
Stratified Sampling



Cluster Sampling

- Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, you randomly select entire subgroups.
- If it is practically possible, you might include every individual from each sampled cluster.
- If the clusters themselves are large, you can also sample individuals from within each cluster using one of the techniques above.
- This method is good for dealing with large and dispersed populations, but there is more risk of error in the sample, as there could be substantial differences between clusters. It's difficult to guarantee that the sampled clusters are really representative of the whole population.
- **Example:** The company has offices in 10 cities across the country (all with roughly the same number of employees in similar roles). You don't have the capacity to travel to every office to collect your data, so you use random sampling to select 3 offices – these are your clusters.

Cluster Sampling



Multi - Stage Sampling

- With this sampling process the respondents are chosen through a process of defined stages.
- For example residents within Islington (London) may have been chosen for a survey through the following process:
- Throughout the UK the south east may have been selected at random, (stage 1),
- Within the South East London is selected again at random (stage 2),
- Islington is selected as the borough (stage 3),
- Then polling districts from Islington (stage 4) and
- Then individuals from the electoral register (stage 5).
- As demonstrated five stages were gone through before the final selection of respondents were selected from the electoral register.

Probability Sampling Methods



Multi-Stage Sample

Systematic Sample

Simple Random Sample



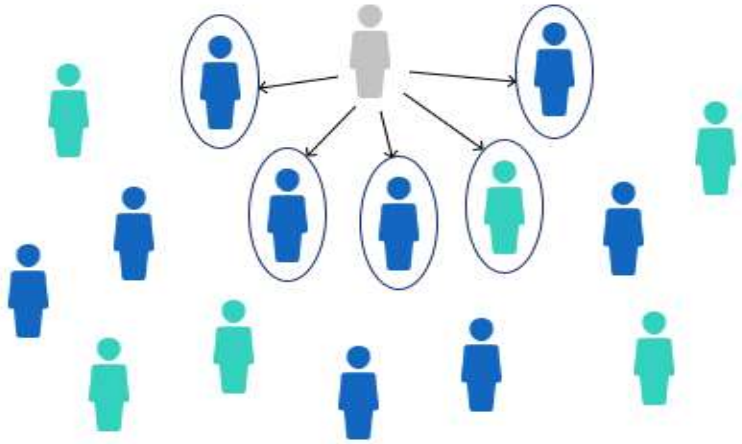
Non-Probability Sampling Methods

- In a non-probability sample, individuals are selected based on non-random criteria, and not every individual has a chance of being included.
- This type of sample is easier and cheaper to access, but it has a higher risk of sampling bias, and you can't use it to make valid statistical inferences about the whole population.
- Non-probability sampling techniques are often appropriate for exploratory and qualitative research. In these types of research, the aim is not to test a hypothesis about a broad population, but to develop an initial understanding of a small or under-researched population.

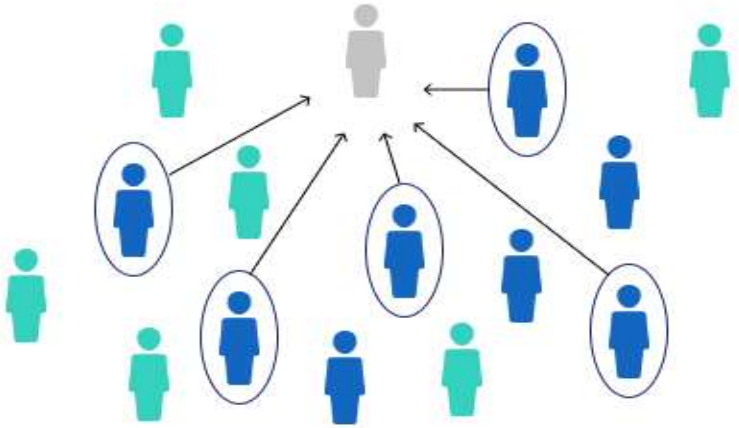


Types of Non-Probability Sampling Methods

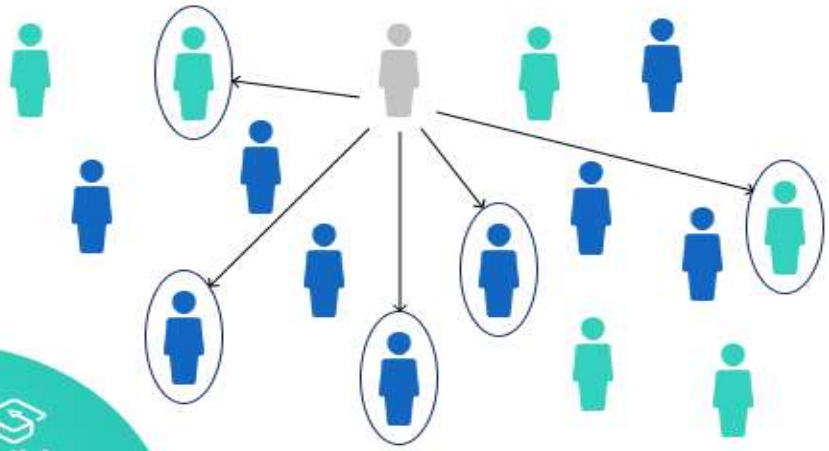
Convenience sample



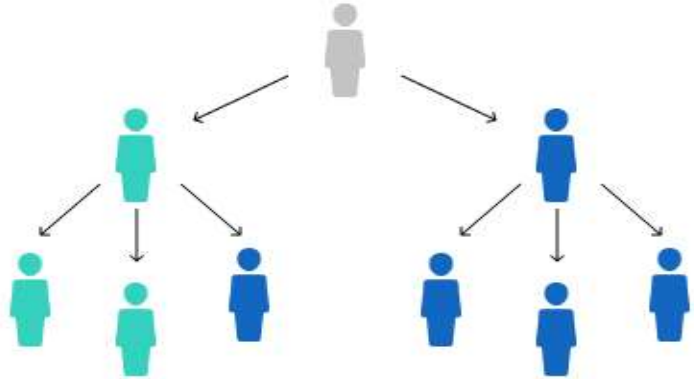
Voluntary response sample



Purposive sample



Snowball sample



Scribbr

Non-Probability Sampling Methods



Convenience Sample

Quota Sample

Dimensional Sample



Difference between Probability and Non-Probability Sampling

Particulars	Probability Sampling Methods	Non-Probability Sampling Methods
Definition	Probability Sampling is a sampling technique in which samples from a larger population are chosen using a method based on the theory of probability.	Non-probability sampling is a sampling technique in which the researcher selects samples based on the researcher's subjective judgment rather than random selection.
Alternatively Known as	Random sampling method.	Non-random sampling method.
Population selection	The population is selected randomly.	The population is selected arbitrarily.
Nature	The research is conclusive.	The research is exploratory.
Sample	Since there is a method for deciding the sample, the population demographics are conclusively represented.	Since the sampling method is arbitrary, the population demographics representation is almost always skewed.

Continue….

Difference between Probability and Non-Probability Sampling

Particulars	Probability Sampling Methods	Non-Probability Sampling Methods
Time Taken	Takes longer to conduct since the research design defines the selection parameters before the market research study begins.	This type of sampling method is quick since neither the sample or selection criteria of the sample are undefined.
Results	This type of sampling is entirely unbiased and hence the results are unbiased too and conclusive.	This type of sampling is entirely biased and hence the results are biased too, rendering the research speculative.
Hypothesis	In probability sampling, there is an underlying hypothesis before the study begins and the objective of this method is to prove the hypothesis.	In non-probability sampling, the hypothesis is derived after conducting the research study.
Requirement of Resources	More Resources in terms of time, cost and efforts.	Less Resources are required.

Continue….

Difference between Probability and Non-Probability Sampling

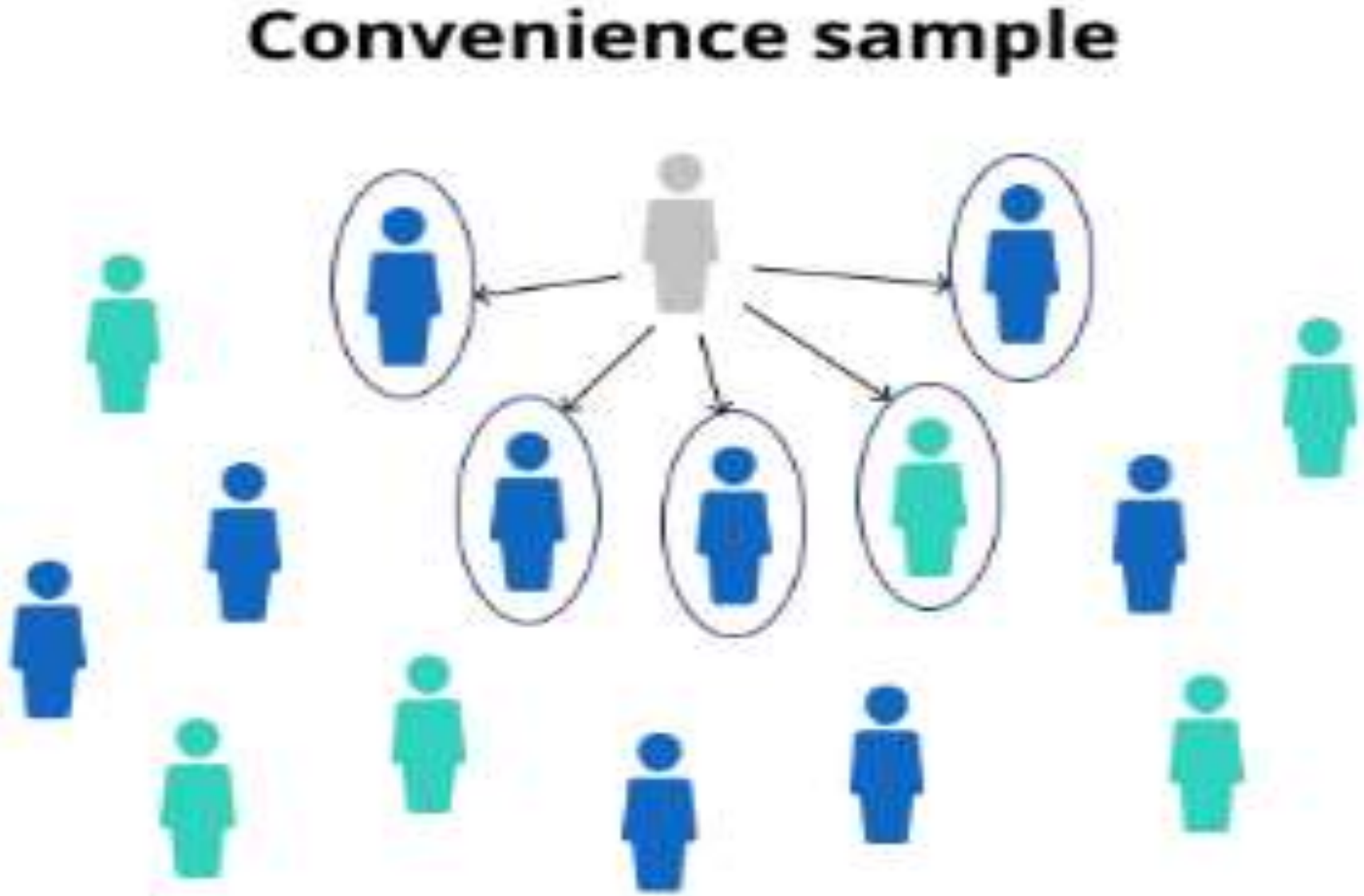
Particulars	Probability Sampling Methods	Non-Probability Sampling Methods
Selection of Sample	Random: No subjective judgment of researcher is involved.	Non Random: Subjective judgments of researcher is involved.
Quality of Inferences drawn	Generalize to the population.	Non Generalizable to the population.
Best suited for researchers	To Understand a population	To understand or develop a concept or idea.
Applicable to the kind of population	Elements which are finite in number which precisely defined and a specific category.	Elements which are infinite, too general category, nor precisely defined.
Chances of error and biases	Sampling errors and systematic biases are less likely to occur.	Prone to encounter with systematic errors and sampling biases.
Types	Simple random sampling, systematic random sampling, stratified random sampling, cluster sampling, Multistage sampling.	Volunteer sampling, Convenient Sampling, Purposive Sampling, Quota Sampling, Snowball Sampling, Matched Sampling, Genealogy based Sampling.

Convenience Sampling

- A convenience sample simply includes the individuals who happen to be most accessible to the researcher.
- This is an easy and inexpensive way to gather initial data, but there is no way to tell if the sample is representative of the population, so it can't produce generalizable results.
- **Example:** You are researching opinions about student support services in your university, so after each of your classes, you ask your fellow students to complete a survey on the topic.
- This is a convenient way to gather data, but as you only surveyed students taking the same classes as you at the same level, the sample is not representative of all the students at your university.



Convenience Sampling

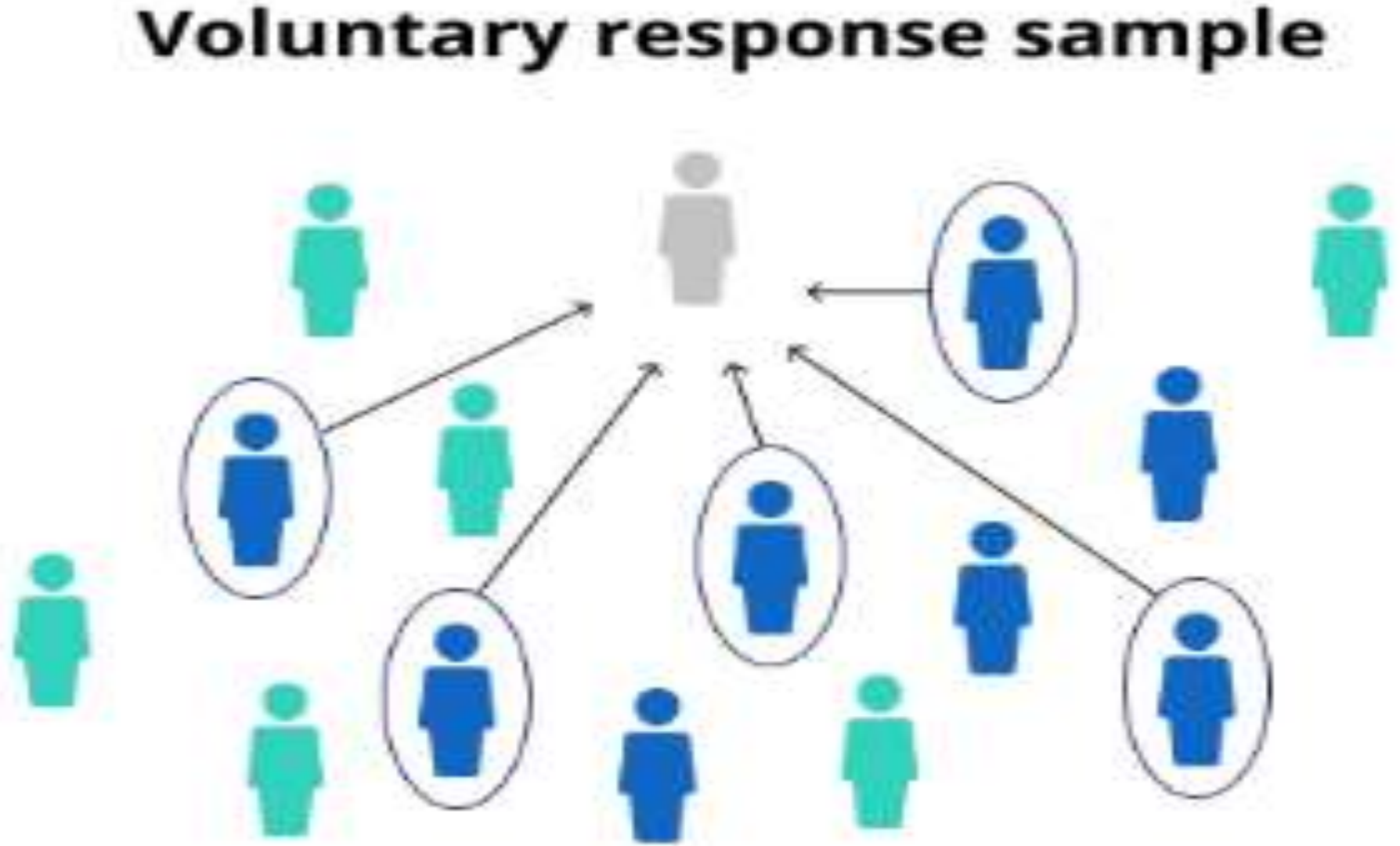


Voluntary Response Sampling

- Similar to a convenience sample, a voluntary response sample is mainly based on ease of access. Instead of the researcher choosing participants and directly contacting them, people volunteer themselves (e.g. by responding to a public online survey).
- Voluntary response samples are always at least somewhat biased, as some people will inherently be more likely to volunteer than others.
- **Example:** You send out the survey to all students at your university and a lot of students decide to complete it.
- This can certainly give you some insight into the topic, but the people who responded are more likely to be those who have strong opinions about the student support services, so you can't be sure that their opinions are representative of all students.



Voluntary Response Sampling

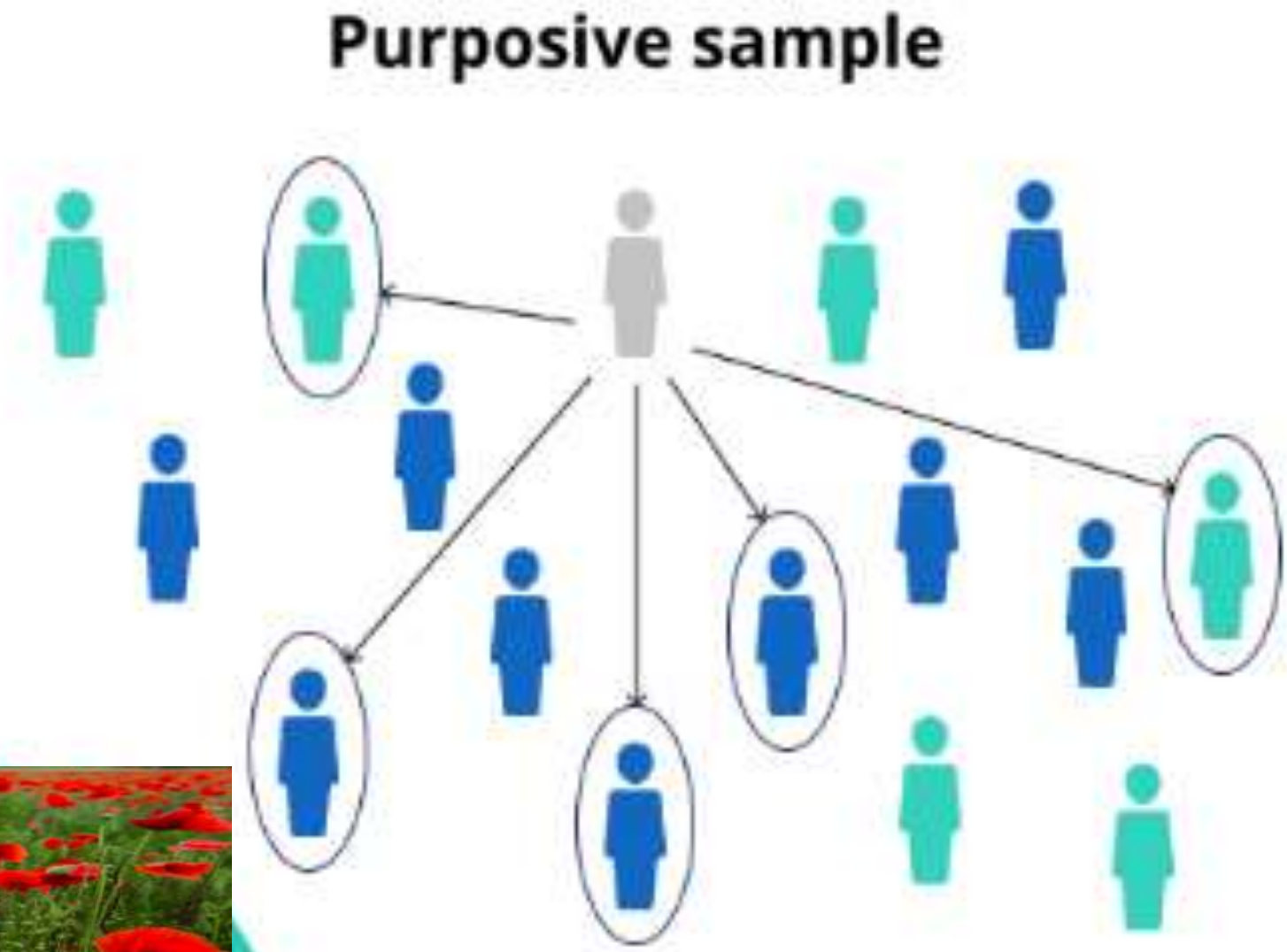


Purposive Sampling

- This type of sampling involves the researcher using their judgement to select a sample that is most useful to the purposes of the research.
- It is often used in qualitative research, where the researcher wants to gain detailed knowledge about a specific phenomenon rather than make statistical inferences.
- An effective purposive sample must have clear criteria and rationale for inclusion.
- **Example:** You want to know more about the opinions and experiences of disabled students at your university, so you purposefully select a number of students with different support needs in order to gather a varied range of data on their experiences with student services.



Purposive Sampling



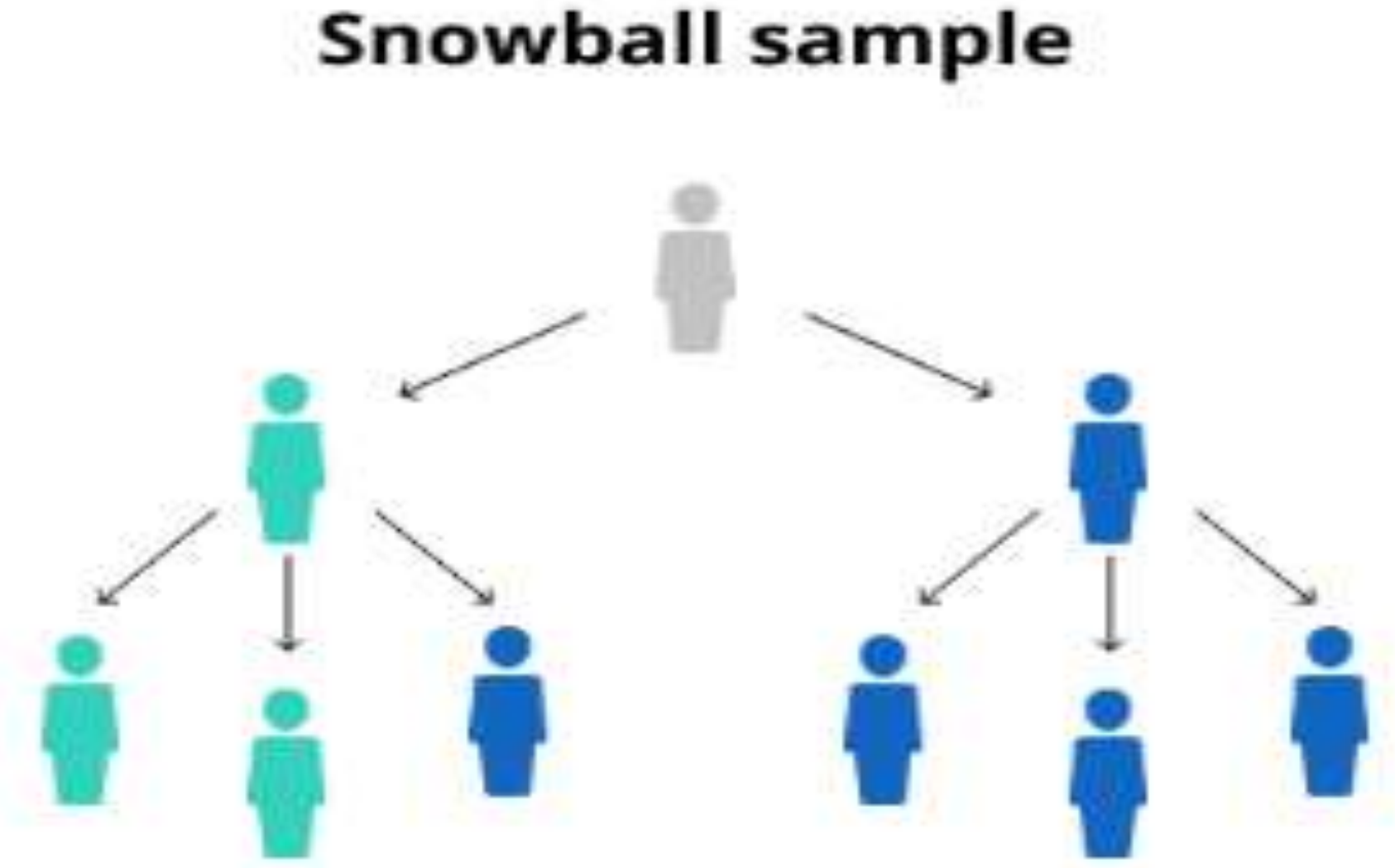


Snowball Sampling

- If the population is hard to access, snowball sampling can be used to recruit participants via other participants.
- The number of people you have access to "snowballs" as you get in contact with more people.
- **Example:** You are researching experiences of homelessness in your city. Since there is no list of all homeless people in the city, probability sampling isn't possible.
- You meet one person who agrees to participate in the research, and she puts you in contact with other homeless people that she knows in the area.

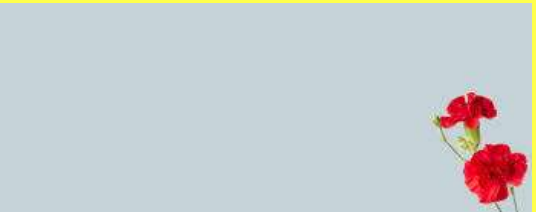


Snowball Sampling



Quota and Dimensional Sampling

- **Quota Sampling:** Using this method the sample respondents are made up of potential purchasers of your product or the market that you would like to research.
- The sample will contain people who meet certain criteria e.g. age, social group, gender. Splitting the sample into age is a popular way to apply quota sampling, the researcher will be asked to interview a set number (quota) of people from different age groups e.g. 16-25, 25-40, 40-55, 55 and above.
- Quota sampling ensures that the sample contains people satisfying all of the characteristics in the market being researched.
- **Dimensional Sampling:** Dimensional sampling is an extension to quota sampling. The researcher takes into account several characteristics e.g. gender, age, income, residence and education.
- The researcher must ensure that there is at least one person in the study representing each of the chosen characteristics.
- For **example** out of 10 people the researcher ensures they have interviewed 2 people that are a certain gender, 2 a certain age group and 2 who have an income between £25000 and £30000.



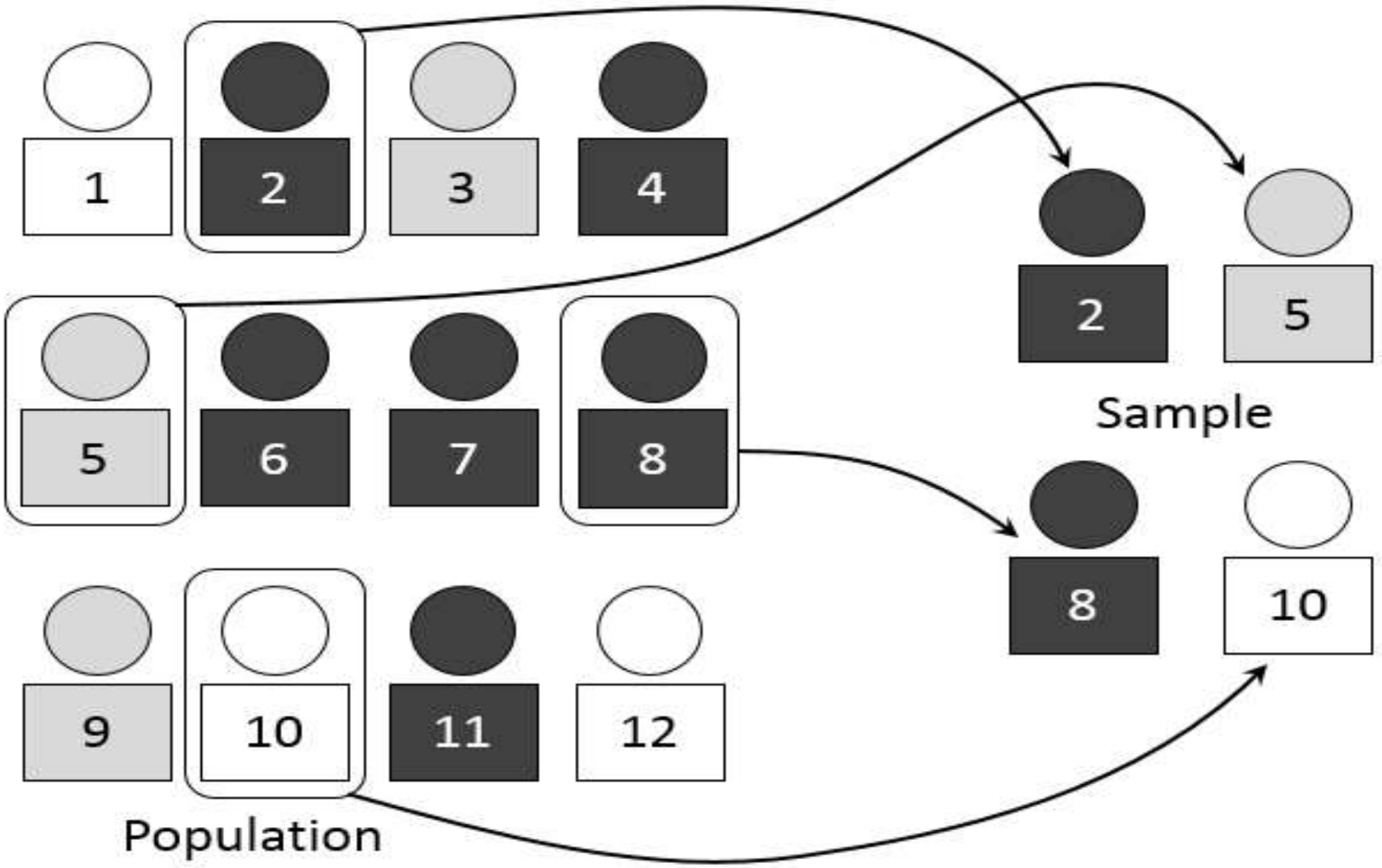
Judgmental Sampling

- In this method, the subjects are selected by the choice of the investigators.
- The researcher assumes specific characteristics for the sample (e.g. male/female ratio = 2/1) and therefore, they judge the sample to be suitable for representing the population.
- This method is widely criticized due to the likelihood of bias by investigator judgment [5].
- Judgment sampling (a type of purposive sampling) occurs when units are selected for inclusion in a study based on the professional judgment of the researcher. This is in contrast to probability sampling techniques in which units are drawn with some probability (e.g., randomly) from the population of interest.
- Judgment sampling, also referred to as judgmental sampling or authoritative sampling, is a non-probability sampling technique where the researcher selects units to be sampled based on his own existing knowledge, or his professional judgment.
- **Example:** A researcher may decide to draw the entire sample from one "representative" city, even though the population includes all cities.

Expert Sampling

- Expert sampling (or judgment sampling) is where you draw your sample from experts in the field you're studying. It's used when you need the opinions or assessment of people with a high degree of knowledge about the study area. When used in this way, expert sampling is a simple sub-type of purposive sampling.
- A second reason to use experts is to validate another sampling method (Singh, 2007). For example, let's say you want to use snowball sampling to identify addicts in your area. You are concerned that using this non-random sampling method will adversely affect your results and the way your results are perceived by others. You can ask a panel of experts their opinion on whether snowball sampling is the most appropriate sampling method.
- "Expert" doesn't necessarily have to mean highly education and skilled in a field (although it often does). Sometimes it can refer to a more abstract construct of the "perfect" embodiment of a certain trait or set of characteristics. For example, Kitamaya & Cohen (2010, p.212) the term *cultural expert* does not have anything to do with education level, but rather it means "…the people most immersed in, most competent in, or who most embody a culture in it's "pure" form."

Expert Sampling



Examples, Advantages and Limitations of the Four Sampling Techniques

Sampling Technique	Example	Advantages	Limitations
Simple random sampling	The names of all 1,000 children are placed into a computer database. The computer is then instructed to randomly select 100 names. These children and their parents are then contacted.	Representative of the population	May be difficult to obtain the list May be more expensive
Stratified random sampling	The names of all 1,000 children are placed into a computer database and organized by grade (sixth, seventh, eighth). The computer is then instructed to randomly select 35 names from each of the three grades. These children and their parents are then contacted.	Representative of the population	May be difficult to obtain the list May be more expensive
Convenience sampling	The researcher knows one of the middle-school teachers, and the teacher volunteers her 35 students for the study. These children and their parents are then contacted.	Simple Easy Convenient No complete member list needed	May not be representative of the population
Quota sampling	Using the middle-school directory, the researcher selects the first 20 sixth-grade boys, the first 20 sixth-grade girls, the first 20 seventh-grade boys, the first 20 seventh-grade girls, the first 20 eighth-grade boys, and the first 20 eighth-grade girls. These children and their parents are then contacted.	Simple Easy Convenient No complete member list needed	May not be representative of the population

Criteria for Sample Size Determination - 1

- Suppose that we are to conduct an investigation comparing populations, $\prod_A, \prod_B$.
- Sample A comprises n_A units of observation from $\prod_A$
- Sample B comprises n_B units of observation from $\prod_B$
- Suppose that $n_A = n_B$ and that $n = n_A + n_B$
- The responses will be quantitative, and the analysis will use a t-test.
- How should we choose n?

Let,

μ_A = Mean response for $\prod_A$

μ_B = Mean response for $\prod_B$



Criteria for Sample Size Determination - 2

- Null Hypothesis is $H_0 : \mu_A = \mu_B$
- From the data, we will obtain the sample means $\overline{X}_A$ and $\overline{X}_B$ and sample standard deviations S_A and S_B for groups A and B.
- Once we have the data, we can:
- Reject H_0 and say that $\mu_A > \mu_B$
- Reject H_0 and say that $\mu_A < \mu_B$
- Not reject H_0.

When $n_A = n_B = n/2$, the t-statistic is

$$t = \frac{\bar{x}_A - \bar{x}_B}{S\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{(\bar{x}_A - \bar{x}_B)\sqrt{n}}{2S}$$

where

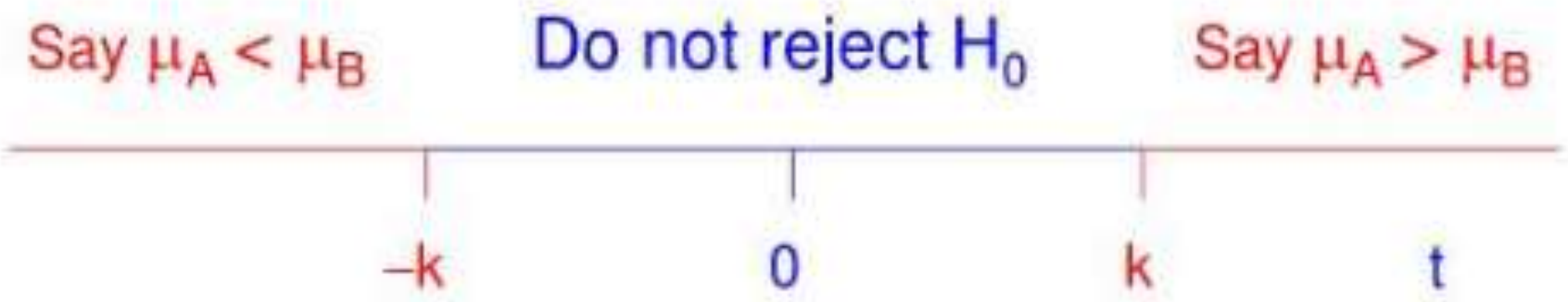
$$S = \sqrt{\frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}} = \sqrt{\frac{S_A^2 + S_B^2}{2}}$$

t will tend to be positive if $\mu_A > \mu_B$, negative if $\mu_A < \mu_B$ and close to zero if $\mu_A = \mu_B$

Criteria for Sample Size Determination - 3

We will:

- Reject H_0 and say that $\mu_A > \mu_B$ if $t \geq k$
- Reject H_0 and say that $\mu_A < \mu_B$ if $t \leq -k$
- Not reject H_0 if $-k < t < k$



Now we need to find both n and k

Criteria for Sample Size Determination - 4

Suppose that, in truth, $\mu_A = \mu_B$

This does not mean that we will observe $\bar{x}_A = \bar{x}_B$ nor $t = 0$

In fact, we may observe $t \geq k$ or $t \leq -k$, just by chance

This means that we might reject H_0 when H_0 is true

This is called ***type I error***

Criteria for Sample Size Determination - 5

Suppose that, in truth, $\mu_A = \mu_B + \delta$

where $\delta > 0$, and is of a magnitude that would be scientifically worth detecting

We may still observe $t \leq k$ by chance

This means that we might fail to reject H_0 when H_0 is false

This is called ***type II error***

Criteria for Sample Size Determination - 6

The probability that $t \geq k$ or $t \leq -k$, when $\mu_A = \mu_B$, is called the *risk of type I error,* and is denoted by α

(This is for a two-sided alternative: the probability that $t \geq k$, when $\mu_A = \mu_B$, is the *risk of type I error for a one-sided alternative* and is equal to $\alpha/2$)

The probability that $t \leq k$, when $\mu_A = \mu_B + \delta$ is called the *risk of type II error,* and is denoted by β

The probability that $t \geq k$, when $\mu_A = \mu_B + \delta$ is called the *power,* and is equal to $1 - \beta$

Criteria for Sample Size Determination - 7

Reducing type I error

Increase k – make it difficult to reject H_0

Increasing power

Decrease k – make it easy to reject H_0

Reducing type I error and increasing power simultaneously

Increase n – this will make the study more informative, but it will cost more

Finding the Sample Size - 1

Suppose that the true standard deviation within each of the populations Π_A and Π_B is σ

Then $t \approx Z$ where

$$Z = \frac{(\bar{x}_A - \bar{x}_B)\sqrt{n}}{2\sigma}$$

Z follows the normal distribution, with standard deviation 1

When $\mu_A = \mu_B$, Z has mean 0

When $\mu_A = \mu_B + \delta$, Z has mean $\delta\sqrt{n}/(2\sigma)$

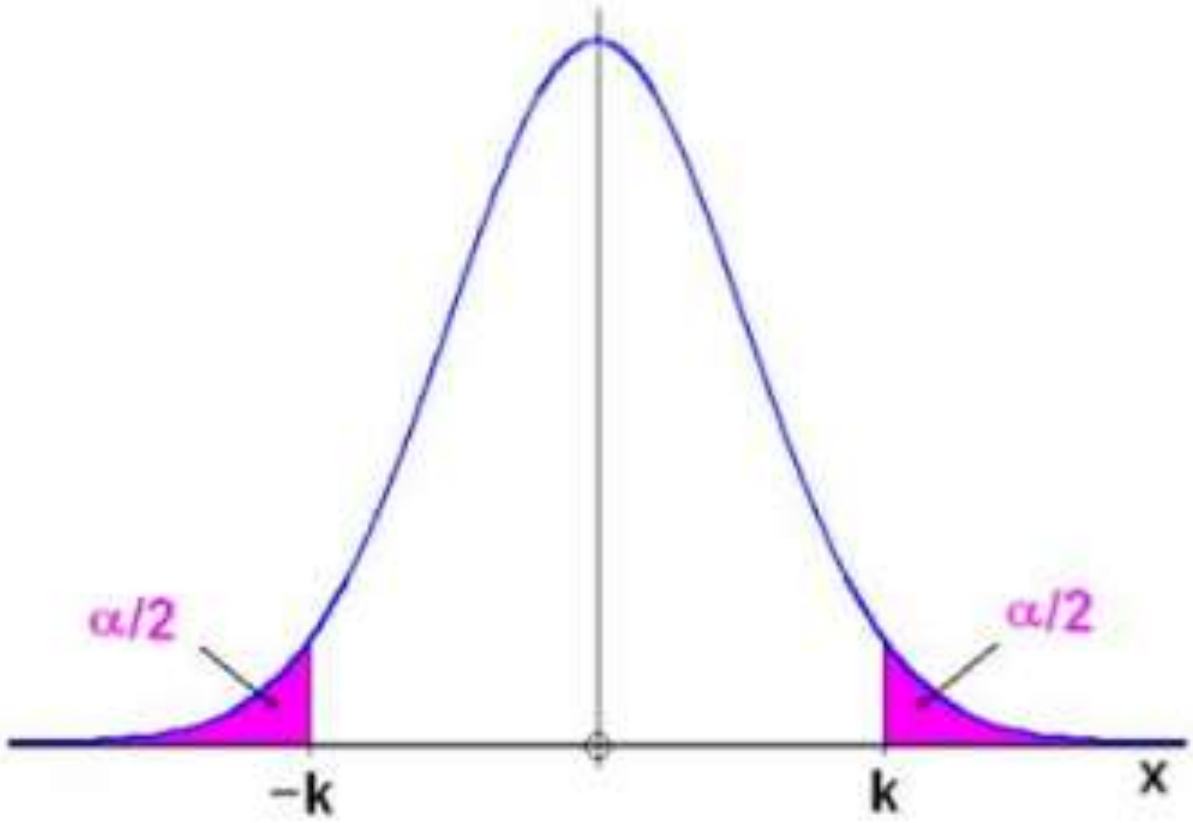
Finding the Sample Size - 2

Specify that the type I risk of error (two-sided) should be α:

$$P(Z \geq k \text{ or } Z \leq -k : \mu_A = \mu_B) = \alpha \qquad (1)$$

Under H_0, Z is normally distributed with mean 0 and st dev 1

k is the value exceeded by a normal (0, 1) random variable with prob $\alpha/2$



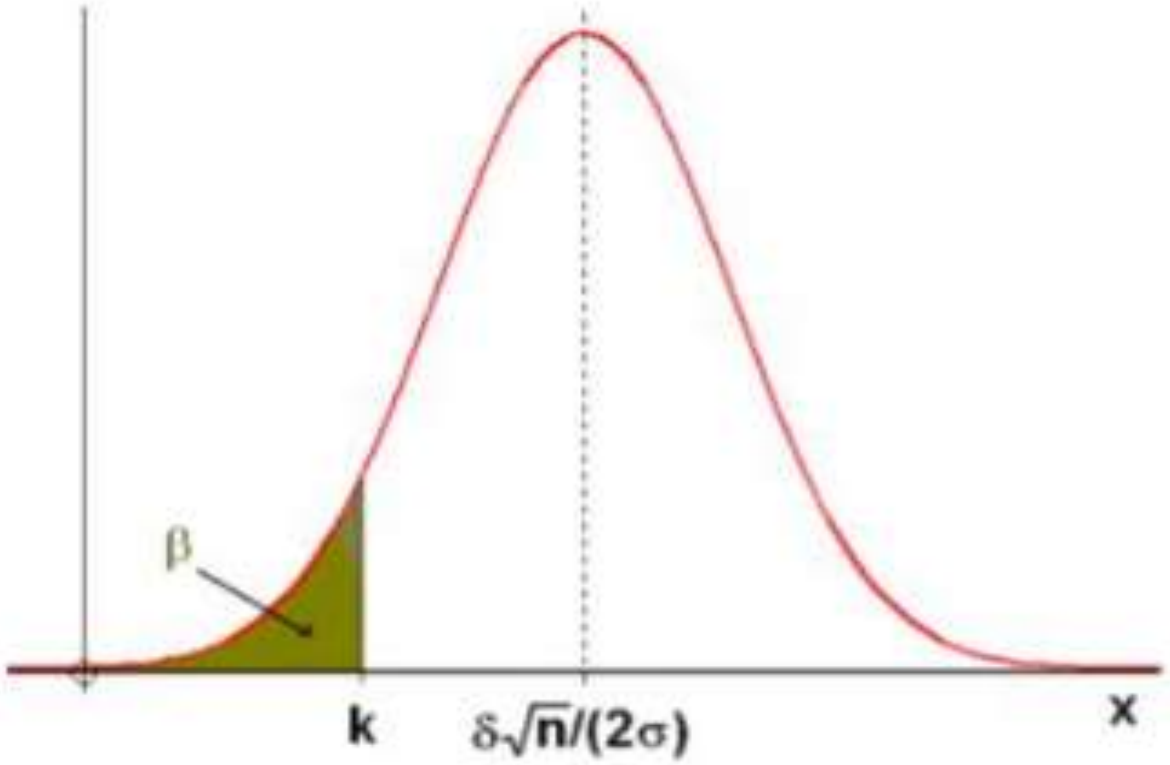
Finding the Sample Size - 3

Specify that the type II risk of error should be β:

$$P(Z \le k : \mu_A = \mu_B + \delta) = \beta \qquad (2)$$

Under H_0, Z is normally distributed with mean $\delta\sqrt{n}/(2\sigma)$ and st dev 1

$k - \delta\sqrt{n}/(2\sigma)$ is the value exceeded by a normal (0, 1) random variable with prob $1 - \beta$



Finding the Sample Size - 4

For $\alpha = 0.05$ and $1 - \beta = 0.90$, we have

$$k = 1.960 \quad \text{and} \quad k - \delta\sqrt{n}/(2\sigma) = -1.282$$

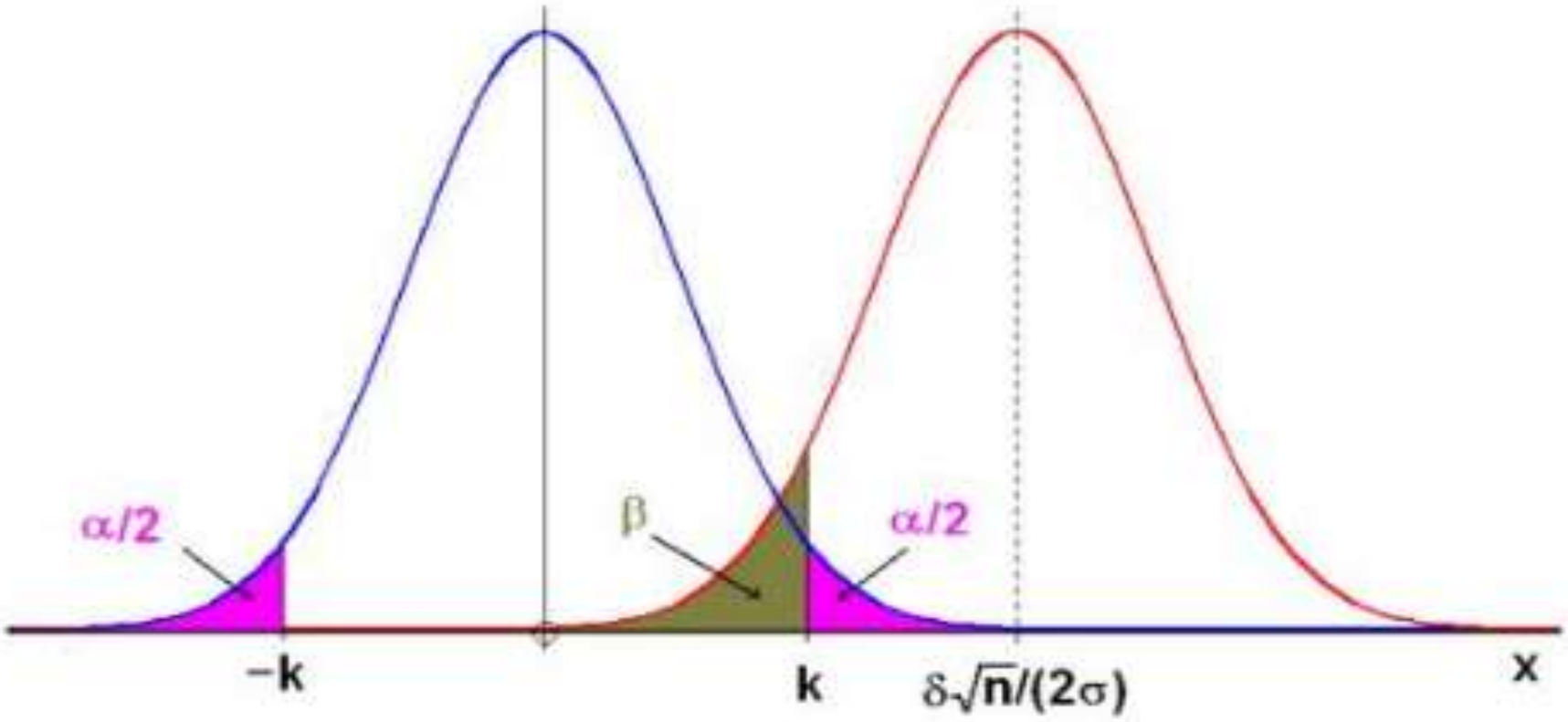
Thus

$$n = 4\sigma^2 \frac{(1.960 + 1.282)^2}{\delta^2} = 42.030\left(\frac{\sigma}{\delta}\right)^2$$

$n/\left(\frac{\sigma}{\delta}\right)^2$		Power: $1 - \beta$		
		0.8	0.9	0.95
Type I error: α	0.1	24.730	34.255	43.289
	0.05	31.396	42.030	51.979
	0.01	46.716	59.518	71.257

Finding the Sample Size - 5

Sample size increases: – *as* σ *increases*
– *as* δ *decreases*
– *as* α *decreases*
– *as* $1 - \beta$ *increases*



Some Simple Variations - 1

Unequal randomisation

The power of a study depends on $\frac{n_E n_C}{n}$

which, for equal sample sizes is equal to $\frac{(n/2)(n/2)}{n} = \frac{n}{4}$

For $n_E = Rn_C$, $n = Rn_C + n_C$ and so

$$\frac{n_E n_C}{n} = \frac{\{Rn/(R+1)\}\{n/(R+1)\}}{n} = \frac{Rn}{(R+1)^2}$$

Some Simple Variations - 2

Unequal randomisation

So, the overall sample size is multiplied by the factor

$$F = n \Big/ \frac{4Rn}{(R+1)^2} = \frac{(R+1)^2}{4R}$$

and n_E by F_E and n_C by F_C, where

$$F_E = \frac{R+1}{2} \quad \text{and} \quad F_C = \frac{R+1}{2R}$$

R	1	2	3	5	10
F	1	1.125	1.333	1.800	3.025
F_E	1	1.500	2.000	3.000	5.500
F_C	1	0.750	0.667	0.600	0.550

Some Simple Variations - 3

Unknown standard deviation

The sample size formula depends on guessing σ

If this guess is smaller than the truth, the sample size will be too small and the study underpowered

If this guess is larger than the truth, the sample size will be too large and the sample size unnecessarily large

A more accurate calculation can be based on the t-distribution rather than the normal, but this makes little difference and does not overcome the dependence on σ

Some Simple Variations - 4

Unknown standard deviation

Often, the final analysis will be based on a linear model, not just a t-test

The formulae given can still be used, but σ is now the residual standard deviation (the SD about the fitted model)

Fitting the right factors will reduce the residual standard deviation, and so the sample size will also be reduced

– but you have to guess what σ will be in advance!

Some Simple Variations - 5

Sample size for estimation

The sample size can be determined to give a confidence interval of specified width W

The 95% confidence interval for $\delta = \mu_A - \mu_B$ is of the form

$$(\bar{x}_A - \bar{x}_C) - 1.96\,\sigma\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)} \le \delta \le (\bar{x}_A - \bar{x}_C) + 1.96\,\sigma\sqrt{\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}$$

when sample sizes are large (Lecture 1, Slide 24)

When $n_A = n_B = n/2$, this has length

$$2 \times 1.96\,\sigma\sqrt{\left(\frac{4}{n}\right)} = 7.84\frac{\sigma}{\sqrt{n}}$$

Some Simple Variations - 6

Sample size for estimation

We need to set

$$7.84\frac{\sigma}{\sqrt{n}} = W$$

which means that

$$n = 61.47\left(\frac{\sigma}{W}\right)^2$$

Some Simple Variations - 7

Binary data

For $R = 1$, $\alpha = 0.05$ and $1 - \beta = 0.90$, we have

$$n = \frac{4(1.960+1.282)^2}{\theta^2 \bar{p}(1-\bar{p})} = \frac{42.030}{\theta^2 \bar{p}(1-\bar{p})}$$

where

$$\theta = \log_e\left(\frac{p_E}{1-p_E}\right) - \log_e\left(\frac{p_C}{1-p_C}\right) \quad \text{and} \quad \bar{p} = \frac{Rp_E + p_C}{R+1}$$

p_C is the anticipated success rate in Π_C, and p_E the improved rate in Π_E to be detected with power $1 - \beta$

Some Simple Variations - 8

Examples for binary data: $R = 1$, $\alpha = 0.05$ and $1 - \beta = 0.90$

p_C	p_E	θ	$\bar{p}(1-\bar{p})$	n
0.1	0.2	0.811	0.1275	502
0.1	0.3	1.350	0.1600	144
0.1	0.5	2.197	0.2100	42
0.3	0.4	0.442	0.2275	946
0.3	0.5	0.847	0.2400	244
0.3	0.7	1.695	0.2500	60
0.4	0.5	0.405	0.2475	1034
0.4	0.6	0.811	0.2500	256
0.4	0.8	1.792	0.2400	56
0.5	0.6	0.405	0.2475	1034
0.5	0.7	0.847	0.2400	244
0.5	0.9	2.197	0.2100	42

Some Simple Variations - 9

Binary data

This approach is based on the *log-odds ratio*

$$\theta = \log_e\left(\frac{p_E}{1-p_E}\right) - \log_e\left(\frac{p_C}{1-p_C}\right)$$

Many other approximate formulae exist

All give similar answers when sample sizes are large: exact calculations can be made for small sample sizes

Further Considerations - 1

Setting the values for α and β

The standard scientific convention is to ensure that α will be small, and allow any risks to be taken with β

For example, if an SD or a control success rate is underestimated at the design stage, the study will be underpowered – the analysis maintains the type I error α at the cost of losing power

α is the community's risk of being given a false conclusion
β is the scientist's risk of not proving his/her point

Further Considerations - 2

Exceptions

If the scientist wishes to prove the null hypothesis (equivalence testing)

- *then β should be kept small, while α can be inflated if necessary*

In a pilot study, preliminary to a larger confirmatory study

- *type I errors can be rectified in the next study, but type II errors will mean that the next study is not conducted at all*

Further Considerations - 3

Finally:

- Many more sample size formulae exist – see Machin et al. (1997)
- Software also exists: nQuery advisor, PASS
- Ensure that the sample size formula used matches the intended final analysis
- In complicated situations, the whole study can be simulated on the computer in advance to determine its power

Choice of Sample Size - 1

- Suppose that the null hypothesis is false and that the true value of the mean is $\mu = \delta + \mu_0$, where $\delta > 0$
- Formulas that determine the appropriate sample size to obtain a particular value of β for a given δ and α are
- Two sided-test alternative hypotheses

$$n \simeq \frac{(z_{\alpha/2} + z_\beta)^2 \sigma^2}{\delta^2}$$

- For either of the one-sided alternative hypotheses

$$n = \frac{(z_\alpha + z_\beta)^2 \sigma^2}{\delta^2}$$

- Where $\delta = \mu - \mu_0$

Choice of Sample Size - 2

- Two-sided Sample Size
  - Sample size $n=n_1=n_2$ required to detect a true difference in means Δ of with power at least $1-\beta$

$$n \cong \frac{(z_{\alpha/2} + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

  - Where Δ is the true difference in means of interest
- One-sided Sample Size

$$n = \frac{(z_\alpha + z_\beta)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta - \Delta_0)^2}$$

Factors Influencing the Choice of Sample Size

- Three factors are used in the sample size calculation and thus, determine the sample size for simple random samples. These factors are:
- 1) the margin of error,
- 2) the **confidence** level, and
- 3) the proportion (or percentage) of the sample that will chose a given answer to a survey question.
- **Example for Choice of Sample size:**

If the standard deviation of the mean for the sampling distribution of random samples of size 36 from a large or infinite population is 2, how large must the sample size become if the standard deviation is to be reduced to 1.2?

Answer: $n = 100$.

Unit – III
Data Presentation and Analysis

Dr. G. YOGANANDHAM
Professor and Head,
Department of Economics,
Thiruvalluvar University.

Data Analysis

- **Data analysis** is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making.
- The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis.
- Whenever we take any decision in our day-to-day life is by thinking about what happened last time or what will happen by choosing that particular decision.
- This is nothing but analyzing our past or future and making decisions based on it.
- For that, we gather memories of our past or dreams of our future. So that is nothing but data analysis.
- Now same thing analyst does for business purposes, is called Data Analysis.

Planning Your Analysis

- Leave enough time for data entry and data formatting.
  - ✓ Can take much longer than you expect.
- In your codebook – note the type of variable for each measurement / question.
- This will allow you to plan the proper levels and types of analysis.
- If your research question requires a level of analysis your variables will not allow, you will need to transform them.
  - ✓ Create ‘dummy’ variables.
  - ✓ Collapse categories.
- Determine the level of significance acceptable & apply proper tests.
- Proper planning will make things easier later.
- Take good notes on any transformations, etc. that you do.
- Save all the elements of your analysis programs.

Analyze the Data - I

- All the time, effort, and resources dedicated to steps 1 through 7 of the research process culminate in this final step.
- The researcher finally has data to analyze so that the research question can be answered.
- In the instrumentation plan, the researcher specified how the data will be analyzed.
- The researcher now analyzes the data according to the plan. The results of this analysis are then reviewed and summarized in a manner directly related to the research questions.
- In the obesity study, the researcher compares the measurements of weight, percentage of body fat, and cholesterol that were taken at the first meeting of the subjects to the measurements of the same variables at the final program session.
- These two sets of data will be analyzed to determine if there was a difference between the first measurement and the second measurement for each individual in the program.
- Then, the data will be analyzed to determine if the differences are statistically significant.

Analyze the Data - II

- If the differences are statistically significant, the study validates the theory that was the focus of the study.
- The results of the study also provide valuable information about one strategy to combat childhood obesity in the community.
- As you have probably concluded, conducting studies using the eight steps of the scientific research process requires you to dedicate time and effort to the planning process.
- You cannot conduct a study using the scientific research process when time is limited or the study is done at the last minute.
- Researchers who do this conduct studies that result in either false conclusions or conclusions that are not of any value to the organization.
- This is an excerpt from *Applied Research and Evaluation Methods in Recreation*.

Types of Data and Analysis - I

- **Quantitative Data**
  - ✓ Counts, Frequencies, Tallies.
  - ✓ Statistical Analyses (as appropriate)
- **Qualitative Data**
  - ✓ Coding
  - ✓ Patterns, themes, theory building
- **Combination**
  - ✓ Process each element as appropriate
- **Types of Data and Formatting Technique**
- **Quantitative Data**
  - ✓ Must “quantify” the data.
  - ✓ Convert (data reduce) from collection format into numeric database.
- **Qualitative Data**
  - ✓ Must process the data (type/enter/describe)
  - ✓ Convert from audio/video to text.
- **Combination**
  - ✓ Process each element as appropriate.

Types of Data Analysis - II

- There are several **types of Data Analysis** techniques that exist based on business and technology. However, the major types of data analysis are:
- Text Analysis
- Statistical Analysis
  - ✓ Descriptive Analysis
  - ✓ Inferential Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis
- **Text Analysis**

  Text Analysis is also referred to as Data Mining. It is a method to discover a pattern in large data sets using databases or data mining tools. It used to transform raw data into business information. Business Intelligence tools are present in the market which is used to take strategic business decisions. Overall it offers a way to extract and examine data and deriving patterns and finally interpretation of the data.

Types of Data Analysis - III

- **Statistical Analysis**

  Statistical Analysis shows "What happen?" by using past data in the form of dashboards. Statistical Analysis includes collection, Analysis, interpretation, presentation, and modeling of data. It analyses a set of data or a sample of data. There are two categories of this type of Analysis - Descriptive Analysis and Inferential Analysis.

- **Descriptive Analysis**

  Analyses complete data or a sample of summarized numerical data. It shows mean and deviation for continuous data whereas percentage and frequency for categorical data.

- **Inferential Analysis**

  Analyses sample from complete data. In this type of Analysis, you can find different conclusions from the same data by selecting different samples.

Types of Data Analysis - IV

- **Diagnostic Analysis**

  Diagnostic Analysis shows "Why did it happen?" by finding the cause from the insight found in Statistical Analysis. This Analysis is useful to identify behavior patterns of data. If a new problem arrives in your business process, then you can look into this Analysis to find similar patterns of that problem. And it may have chances to use similar prescriptions for the new problems.

- **Predictive Analysis**

  Predictive Analysis shows "what is likely to happen" by using previous data. The simplest example is like if last year I bought two dresses based on my savings and if this year my salary is increasing double then I can buy four dresses. But of course it's not easy like this because you have to think about other circumstances like chances of prices of clothes is increased this year or maybe instead of dresses you want to buy a new bike, or you need to buy a house.

- So here, this Analysis makes predictions about future outcomes based on current or past data. Forecasting is just an estimate. Its accuracy is based on how much detailed information you have and how much you dig in it.

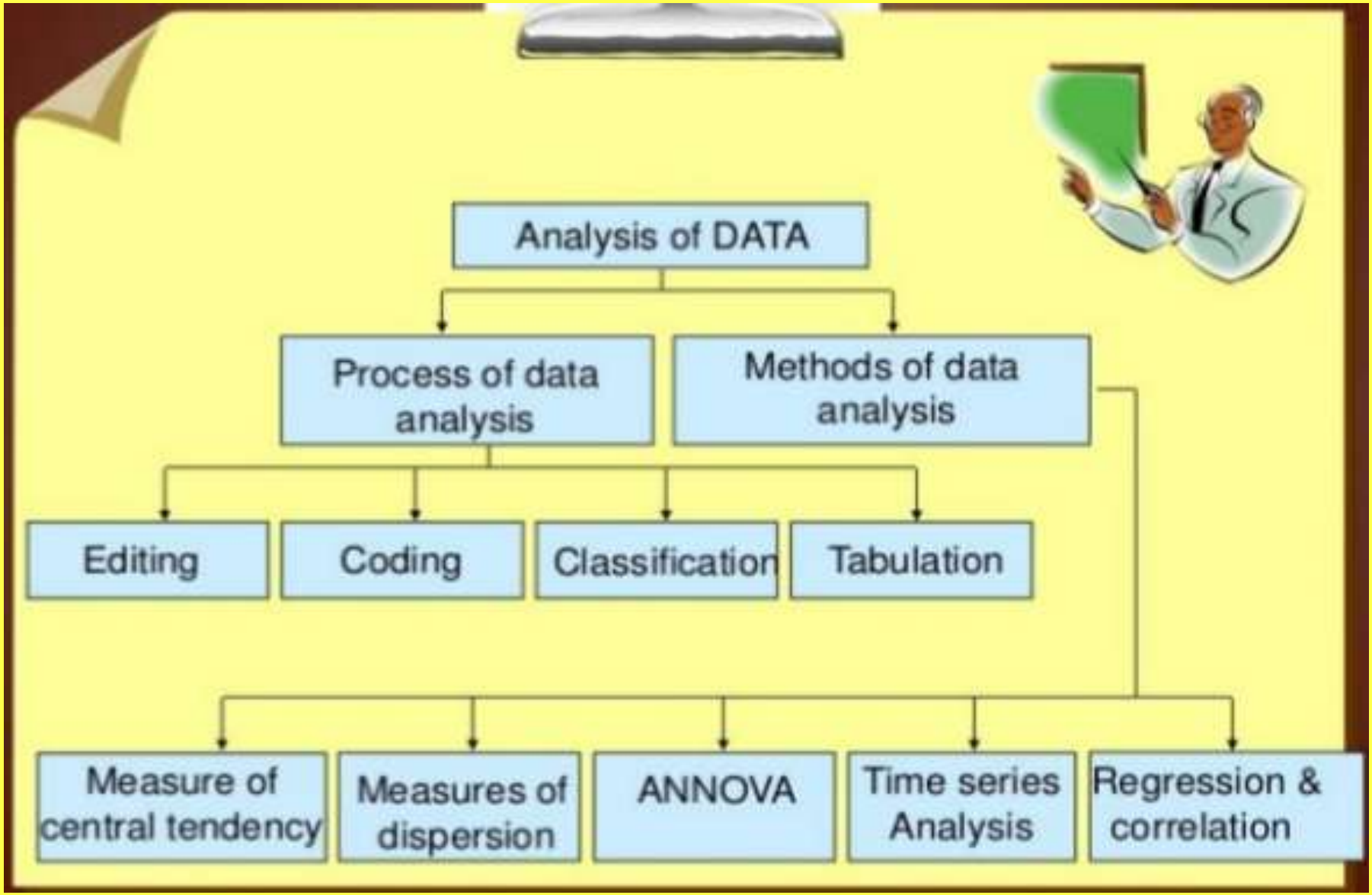
Types of Data Analysis - V

- **Prescriptive Analysis**

  Prescriptive Analysis combines the insight from all previous Analysis to determine which action to take in a current problem or decision. Most data-driven companies are utilizing Prescriptive Analysis because predictive and descriptive Analysis are not enough to improve data performance. Based on current situations and problems, they analyze the data and make decisions.

- Prescriptive analytics suggests various courses of action and outlines what the potential implications could be reached after predictive analysis. Prescriptive analysis generating automated decisions or recommendations requires specific and unique algorithmic and clear direction from those utilizing the analytical techniques.

Classification of Analysis of Data



Processing the Data



Processing the data

- Editing
  - Field Editing
  - Central Editing
- Coding
- Classification
- Tabulation
- Graphing

Process of Data Analysis - I

- The **Data Analysis Process** is nothing but gathering information by using a proper application or tool which allows you to explore the data and find a pattern in it. Based on that information and data, you can make decisions, or you can get ultimate conclusions.
  - ✓ Data Analysis consists of the following phases:
  - ✓ Data Requirement Gathering
  - ✓ Data Collection
  - ✓ Data Cleaning
  - ✓ Data Analysis
  - ✓ Data Interpretation
  - ✓ Data Visualization

Process of Data Analysis - II

- **Data Analysis**

  Once the data is collected, cleaned, and processed, it is ready for Analysis. As you manipulate data, you may find you have the exact information you need, or you might need to collect more data. During this phase, you can use data analysis tools and software which will help you to understand, interpret, and derive conclusions based on the requirements.

- **Data Interpretation**

  After analyzing your data, it's finally time to interpret your results. You can choose the way to express or communicate your data analysis either you can use simply in words or maybe a table or chart. Then use the results of your data analysis process to decide your best course of action.

- **Data Visualization**

  Data visualization is very common in your day to day life; they often appear in the form of charts and graphs. In other words, data shown graphically so that it will be easier for the human brain to understand and process it. Data visualization often used to discover unknown facts and trends. By observing relationships and comparing datasets, you can find a way to find out meaningful information.

- **Summary:**
  - ✓ **Data analysis means a process of cleaning, transforming and modeling data to discover useful information for business decision-making**
  - ✓ **Types of Data Analysis are Text, Statistical, Diagnostic, Predictive, Prescriptive Analysis**
  - ✓ **Data Analysis consists of Data Requirement Gathering, Data Collection, Data Cleaning, Data Analysis, Data Interpretation, Data Visualization**

Process of Data Analysis - III

- **Data Requirement Gathering**

  First of all, you have to think about why do you want to do this data analysis? All you need to find out the purpose or aim of doing the Analysis. You have to decide which type of data analysis you wanted to do! In this phase, you have to decide what to analyze and how to measure it, you have to understand why you are investigating and what measures you have to use to do this Analysis.

- **Data Collection**

  After requirement gathering, you will get a clear idea about what things you have to measure and what should be your findings. Now it's time to collect your data based on requirements. Once you collect your data, remember that the collected data must be processed or organized for Analysis. As you collected data from various sources, you must have to keep a log with a collection date and source of the data.

- **Data Cleaning**

  Now whatever data is collected may not be useful or irrelevant to your aim of Analysis, hence it should be cleaned. The data which is collected may contain duplicate records, white spaces or errors. The data should be cleaned and error free. This phase must be done before Analysis because based on data cleaning, your output of Analysis will be closer to your expected outcome.

Quantifying Data

- Before we can do any kind of analysis, we need to quantify our data.
- "Quantification" is the process of converting data to a numeric format.
- Convert social science data into a " machine – readable" form, a form that can be read & manipulated by computer programs.
- Some transformations are simple:
- **Assign numeric representations to nominal or ordinal variables:**
  - ✓ Turning male into "1" and female into "2".
  - ✓ Assigning "3" to very interested, "2" to somewhat interested, "1" to not interested.
- **Assign numeric values to continuous variables**
  - ✓ Turning born in 1973 to "35".
  - ✓ Number of children = "02".

Developing Code Categories

- Some data are more challenging. Open-ended responses must be coded.
- Two basic approaches:
  - ✓ Begin with a coding scheme derived from the research purpose.
  - ✓ Generate codes from the data.
- **Coding Quantitative Data**
- **Goal – Reduce a wide variety of information to a more limited set of variable attributes:**
- "What is your occupation?"
  - ✓ Use pre-established scheme: Professional, Managerial, Clerical, Semi- skilled, etc.
  - ✓ Create a scheme after reviewing the data.
  - ✓ Assign value to each category in the scheme: professional = 1, Managerial = 2, etc.
  - ✓ Classify the response: "Secretary" is "Clerical" and is coded as "3".

Coding Quantitative Data

- **Points to Remember:**
  - ✓ If the data are coded to maintain a good amount of detail, they can always be combined (reduced) later.
  - ✓ However, if you start off with too little detail, you can't get it back.
  - ✓ If you are using a survey / questionnaire, it is a good idea to do your coding on the form so that it can be entered properly (i.e. create a "codebook").
- **Codebook Construction**
- **Purposes:**
  - ➢ Primary guide used in the coding process.
    - ✓ Should note the value assigned to each variable attribute (response).
  - ➢ Guide for locating variables and interpreting cods in the data file during analysis.
  - ➢ If you are doing your own input, this will also guide data set construction.

Hands – on Exercise 1

- **Create a mini-codebook by coding the survey instrument**
  - ✓ Note column spaces / locations.
  - ✓ Note variable attribute values.
  - ✓ Pay attention to the box at the bottom, special instructions.
- **Entering Data**
- Optical scan sheets (usually ASCII output).
  - ✓ Limits possible responses.
- CATI system / on-line: entered while collected.
- Data entry specialists enter the data into SPSS data matrix, Excel spreadsheet or ASCII file.
  - ✓ Typically, work off a coded Questionnaire.
- In Excel or Access, follow procedures from class:
  - ✓ Format tables with proper variable columns.
  - ✓ Enter data for each case.

Entering Data

- **In SPSS**
  - ✓ Import an ASCII file and name variables / column headings.
  - ✓ Or, create variables / column heading & enter each case.
- ASCII files are useful because they can be transformed or used in almost all analysis programs.
- Upload to SPSS, Excel, or use directly with SAS.
- **Into an ASCII file**
  - ✓ Using Notepad.
  - ✓ Use your coded survey to show you the proper entry order.
- **Into an ASCII file**
  - ✓ Use the Command prompt (Accessories → Command Prompt)
  - ✓ Type "Edit".
- If you open an ASCII file in Excel, you will get a wizard to convert the data.
- Delimited or Fixed width.
- If fixed width, add column breaks.
- Opens as Excel workbook.

Hands – on Exercise - 2

- Complete the survey (fill – in your answers).
- Create a ‘dataset’.
- Enter the data from your survey using either Notepad or the Edit program from the command Prompt.
- **Quantitative Analysis**
- You should choose a level of analysis that is appropriate for your research question.
- You should choose the type of statistical analysis appropriate for the variables you have.
  - ✓ Nominal / Categorical, Ordinal, or Continuous.
- **Quantitative Level Analysis**
- Univariate – Simplest form, describe a case in term of a single variable.
- Bivariate – Subgroup comparisons, describe a case in terms of two variables simultaneously.
- Multivariate – Analysis of two or more variables simultaneously.

Univariate Analysis

- Describing a case in terms of the distribution of attributes that comprise it.
- **Example:** Gender – Number of Women, Number of Men.
- You should always begin your analysis by running the basic univariate frequencies and checking to be sure data were entered properly.
- Frequency distributions.
- Measures of central tendency
  - ✓ Mean, Median, Mode.
- **Presenting Univariate Data**
- **Goals:** Provide Reader with the fullest degree of detail regarding the data.
- Present data in manageable form.
- Simple and Straightforward.

Uni-variate Analysis

- Describing a case in terms of the distribution of attributes that comprise it.
- **Example:** Gender – Number of Women, Number of Men.
- You should always begin your analysis by running the basic univariate frequencies and checking to be sure data were entered properly.
- Frequency distributions.
- Measures of central tendency
  - ✓ Mean, Median, Mode.
- **Presenting Uni-variate Data**
  - **Goals:** Provide Reader with the fullest degree of detail regarding the data.
  - Present data in manageable form.
  - Simple and Straightforward.
- **Subgroup Comparison**
  - Describe subsets of cases, subjects or respondets.
- **Examples:**
  - "Collapsing" response categories: Age categories, open responses, etc.
  - Handling "Don't Knows" : code separately, make missing if appropriate.

Bi-Variate Analysis

- Describe a case in terms of two variables simultaneously.
- **Example:**
  - ✓ Gender
  - ✓ Attitudes toward equality for men and women.
  - ✓ How does a respondents gender affect his or her attitude toward equality for men and women?
- Cross tabulations / Correlations
- **Constructing Bi-variate Tables**
- Divide cases into groups according to the attributes of the independent variable.
- Describe each subgroup in terms of attributes of the dependent variable.
- Read the table by comparing the independent variable subgroups in terms of a given attribute of the dependent variable.
- DV goes in the rows, IV goes in the columns.

Bivariate Analysis

- Bivariate Tables / Crosstabs are appropriate for all types of variables, but the proper inferential statistic will vary by variable type.
- Continuous variables are typically made into categorical variables for this type of analysis.
- Recode variables.
- **Example:** Create "Age" (18-34, 35-50, 51-65, 66+)
- **Appropriate Types of Analysis**

		Independent Variable		
		Nominal	Ordinal	Interval/Ratio
Dependent Variable	Nominal	*Crosstabs* Chi-square Lambda	*Crosstabs* Chi-square Lambda	
	Ordinal	*Crosstabs* Chi-square Lambda	*Crosstabs* Chi-square Lambda Gamma Kendall's tau Somers' *d*	
	Interval/Ratio	*Means* *t*-test ANOVA	*Means* *t*-test ANOVA	*Correlate* Pearson *r* Regression (*R*)

Bivariate Analysis: Correlations

- Bivariate Correlation analysis is appropriate for continuous variables (interval, ratio).
- Other types of variables are often recoded into 'Dummy' variables (value 0 or 1) for these purposes.
- **Examples:** Gender becomes two variables 'Male' (1 = Yes) & 'Female' (1 = Yes).
- Present in Correlation Matric.
- **Multivariate Analysis**
- Analysis of more than two variables simultaneously.
- Can be used to understand the relationship between multiple variables more fully.
- Most typical: Regression Analysis.
- Ordinal (technically inappropriate but it happens), continuous, dummy variables.
- Type of regression analysis will depend on the tupe of variables.
  - ✓ OLS (continuous)
  - ✓ Logistic (Other types)

Coding - I

- The rules for interpreting, classifying, and recording data in the coding process.
- The actual numerical or other character symbols.
- Coding means assigning a code, usually a number, to each possible response to each question. The code includes an indication of the column position (field) and data record it will occupy.
- **Coding Questions**
- Fixed field codes, which mean that the number of records for each respondent is the same and the same data appear in the same column(s) for all respondents, are highly desirable.
- If possible, standard codes should be used for missing data. Coding of structured questions is relatively simple, since the response options are predetermined.
- In questions that permit a large number of responses, each possible response option should be assigned a separate column.

Coding - II

- **Guidelines for coding unstructured questions:**
- Category codes should be mutually exclusive and collectively exhaustive.
- Only a few (10% or less) of the responses should fall into the “other” category.
- Category codes should be assigned for critical issues even if no one has mentioned them.
- Data should be coded to retain as much detail as possible.
- **Code Book**
- A codebook contains coding instructions and the necessary information about variables in the data set.
- A codebook generally contains the following information:
- Column number
- Record number
- Variable number
- Variable name
- Question number
- Instructions for coding

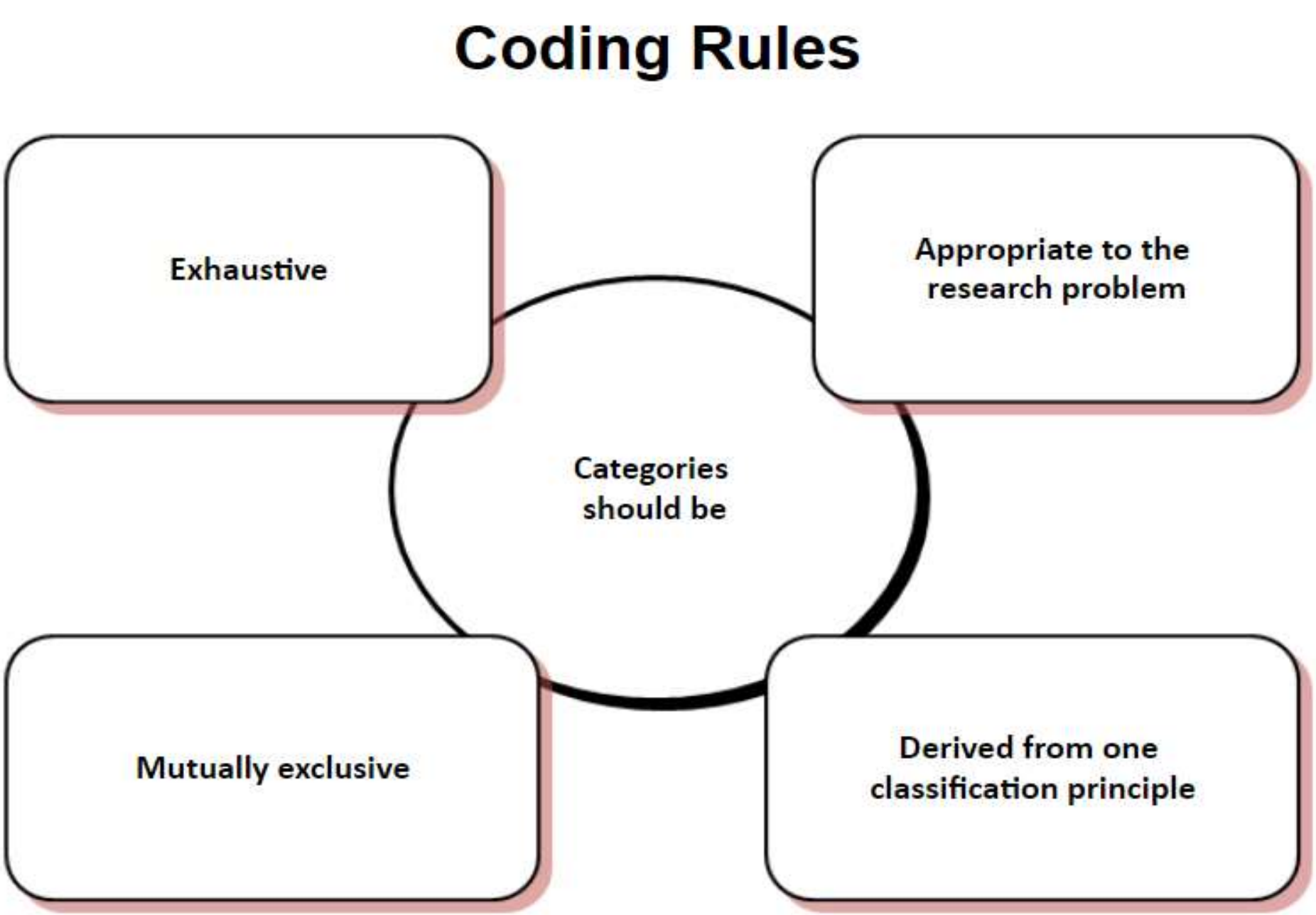
Coding - III

- Coding is a skill where you take instructions (the steps in a task) and translate it into a language the computer understands since computers do not communicate like humans.
- They communicate in a language called BINARY and it is uses 0's and 1's. Coders write the instructions using a programming language.
- Simply put, coding is used for communicating with computers. People use coding to give computers and other machines instructions on what actions to perform.
- Further, we use coding to program the websites, apps, and other technologies we interact with every day.
- Coding is what makes it possible for us to create computer software, apps and websites.
- Many coding tutorials use that command as their very first example, because it's one of the simplest examples of code you can have - it 'prints' (displays) the text' Hello, world! 'onto the screen.

Coding - IV

- Coding, in the simplest of terms, is telling a computer what you want it to do, which involves typing in step-by-step commands for the computer to follow.
- Computers are not clever things, however they are very obedient. They will do exactly what you want them to do, so long as you tell them how to do it correctly.
- Learning to code has been likened to learning a foreign language, or perhaps more specifically a family of foreign languages.
- There are many different coding languages, each one designed with certain things in mind.
- **Examples:** Include C, a ‘low level’ but fast programming language that is good for anything graphically intensive like games; Javascript, which was specifically designed for dealing with web content; and Perl, a multi-functional language that is often referred to as the ‘swiss army knife’ of programming.

Coding Rules



Data Coding Example

- Sample Record: Excel Sheet for two-wheeler owners

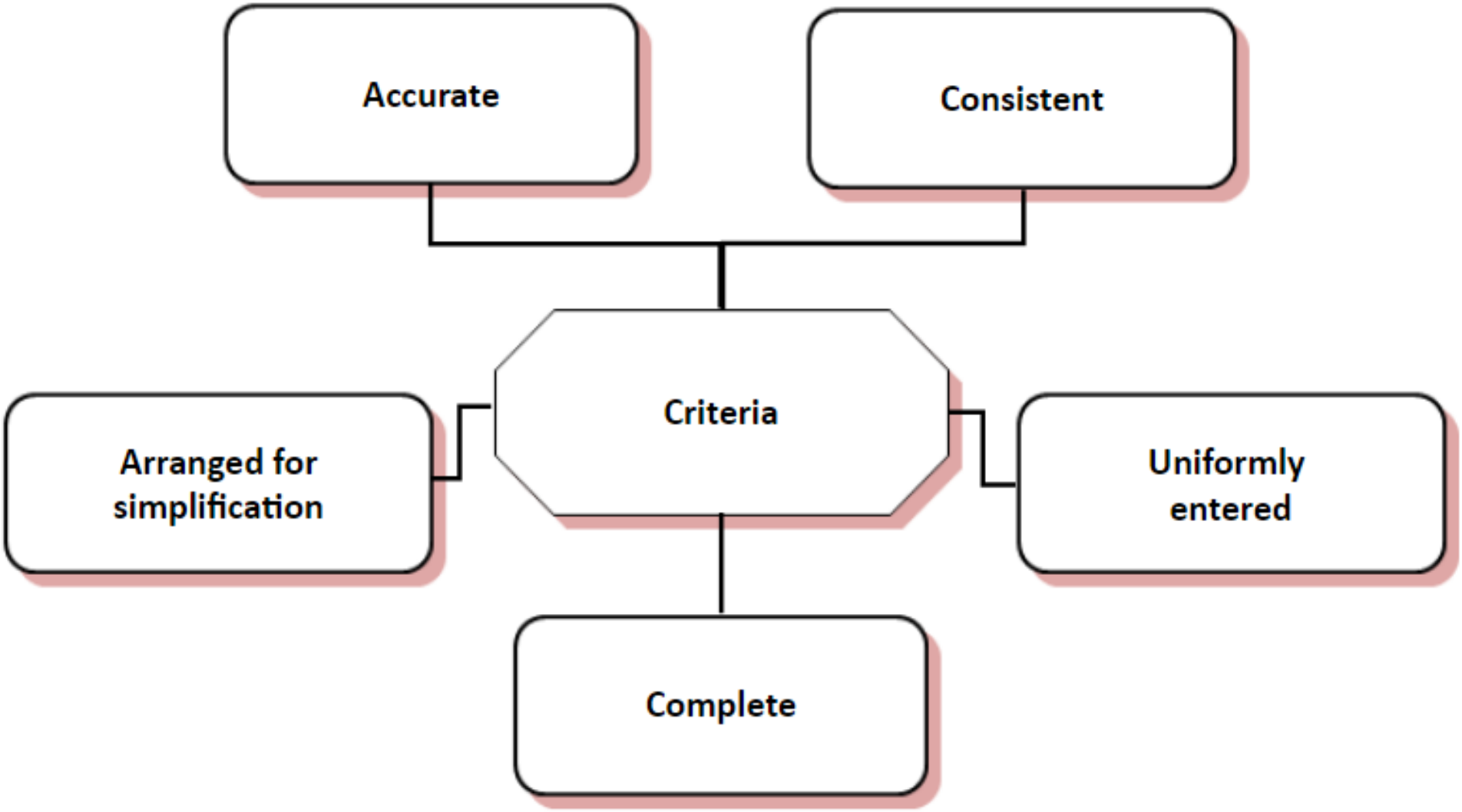
Unit	Occupation	Vehicle	Km / day	Marital Status	Family Size
1	4	1	20	1	3
2	3	2	25	2	1
3	5	1	25	1	4
4	2	1	15	2	2
5	4	2	20	2	4
6	5	2	35	2	6
7	1	1	40	1	3
8	5	2	20	2	4

Editing - I

- The process of checking and adjusting the data.
  - ✓ For omissions
  - ✓ For legibility
  - ✓ For consistency
- And readying them for coding and storage.
- Editing Responses.
- Treatment of Unsatisfactory Results.
- **Returning to the Field – The questionnaires with** unsatisfactory responses may be returned to the field, where the interviewers re-contact the respondents.
- **Assigning Missing Values – If returning the** uestionnaires to the field is not feasible, the editor may assign missing values to unsatisfactory responses.
- **Discarding Unsatisfactory Respondents – In this** approach, the respondents with unsatisfactory responses are simply discarded.

Editing - II

Editing Data Cont.



Reasons for Editing



Editing and Need of Editing

- Information gathered during data collection may lack uniformity. An research checks for completeness, accuracy, consistency of data, collection, and arranging the data in a systematic manner this is know as Editing.
- **Example:** Data collected through questionnaire and schedules may have answers which may not be ticked at proper places, or some questions may ne left unanswered. The research has to take a decision as to how to edit it.
- **Need of Editing**
- **Clarify Responses:** Bringing clarity is important otherwise the researcher can draw wrong inferences from the data. Sometimes the respondents make some spelling and grammatical mistakes the editor needs to correct them.
- **Make Omissions:** The researcher may also to make some omissions in the responses, by chance or by some mistake some responses are left incomplete, the editor has to see what has been an oversight by the respondent (ex. – Illiterate People)

Need of Editing

- **Avoid Biased Editing:** The researcher needs to be very objective and should not try to hide or remove any information. He should not add anything in the responses without any sound reason.
- **Make Judgements:** Sometimes the respondents leave something incomplete, to complete the sentence or a phase the editor has to make a judgement. He should have to have good judgement to do so. He should do it so well that his personal bias do not involve in the responses.
- **Check Handwriting:** Handwriting issues needs also be resolved by the editor. Some people write very fast and in this way they write so that comprehension of the text becomes difficult. In electronically sent questionnaires this problem never arises.
- **Re-contact the respondent:** If some information is least comprehendible and no logical meaning can be taken, interviewees can be re-contacted to know what they meant by that. Incase, the data in the questionnaire is not correct and the editor cannot take any meaning from it. The editor should ask the respondents. Re-contact with them and get their help.

Types of Editing

- **Field Editing:**
- This is a type of editing that relates to abbreviated or illegible written form of gathered data. Such editing is more effective when done on same day or the very next day after the interview. The investigator must not jump to conclusion while doing field editing.
- **Central Editing:**
- Such type of editing relates to the time when all data collection process has been completed. Here a single or common editor corrects the errors like entry in the wrong place, entry in wrong unit etc. As a rule all the wrong answers should be dropped from the final results

Benefits of Editing

- The data obtained is complete in all respects.
- It is accurate in terms of information recorded and responses sought.
- The response format is in the form that was instructed.
- The data is structured in a manner that entering the information will not be a problem.
- **Checklist for Research Paper Editing**
- **Your Name or Contact details as needed.**
- **Title**
- **Page Numbers**
- **Formatting**
- **Font and Size**
- **Line Spacing**
- **Margins**



Tabulation - I

- Tabulation is a systematic & logical presentation of numeric data in rows and columns, to facilitate comparison and statistical analysis.
- It facilitates comparison by bringing related information close to each other and helps in further statistical analysis and interpretation.
- The large mass of confusing data is easily reduced to reasonable form, that is understandable to kind. The data once arranged in a suitable form, gives the condition of the situation at a glance, or gives a bird eye view.
- **Objective of Data Tabulation:**
- To carry out investigations.
- To do comparisons.
- To locate omissions and errors in the data.
- To use space economically.
- To study the trends.
- To simplify data.
- To use it as future references.



Tabulations - II

- Tabulation is the process of summarizing raw data and displaying the same in compact form (i.e., in the form of statistical tables) for further analysis. In a broader sense, tabulation is an orderly arrangement of data in columns and rows.
- Tabulation is essential because of the following reasons:
  - It conserves space and reduces explanatory and descriptive statement to a minimum.
  - It facilitates the process of comparison.
  - It facilitates the summation of items and the detection of errors and omissions.
  - It provides a basis for various statistical computations.

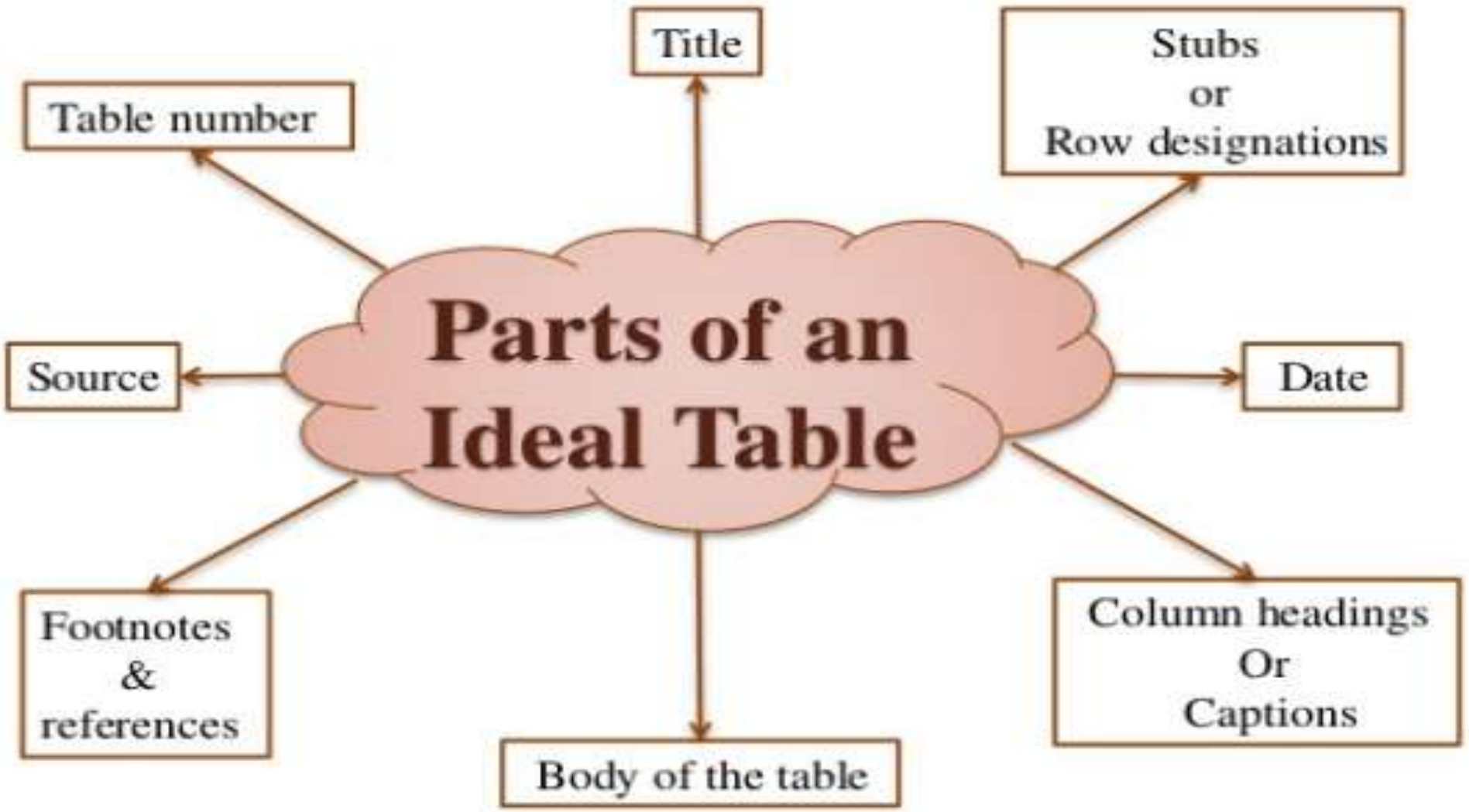
Rules of Tabulation - I

- There are no hard and fast rules for the tabulation of data but for constructing good table, following general rules should be observed while tabulating statistical data.
- The table should suit the size of the paper and, therefore, the width of the column should be decided before hand.
- Number of columns and rows should neither be too large nor too small.
- As far as possible figures should be approximated before tabulation. This would reduce unnecessary details.
- Items should be arranged either in alphabetical., chronological or geographical order or according to size.

Rules of Tabulation - II

- The sub-total and total of the items of the table must be written.
- Percentages are given in the tables if necessary.
- Ditto marks should not be used in a table because sometimes it create confusion.
- Table should be simple and attractive.
- A table should be logical, well-balanced in length and breadth and the comparable columns should be placed side by side.
- Light / heavy / thick or double rulings may be used to distinguish sub columns, main columns and totals.
- For large data more than one table may be used.

Parts of an Ideal Table - I



Parts of an Ideal Table - II

- **Table Number:**
- A Number must be allotted to the table for identification, particularly when there are many tables in a study.
- **Title:**
- the title should explain what is contained in the table. It should be clear, brief and set in bold type on top of the table. It should also indicate the time and place to which the data refer
- **Data:**
- The date of preparation of the table should be given.
- **Stubs or Row designations:**
- Each row of the table should be given brief heading. Such designations of rows are called ”stubs” or “stub items” and the entire column is called “stub column”.
- **Footnotes and References:**
- If any item has not been explained properly, a separate explanatory note should be added at the bottom of the table.

Parts of an Ideal Table - III

- **Column Heading or Captions:**
- Column designation is given on top of each column to explain to what the figures in the column refer. It should be clear and precise. This is called a “caption” or “heading”. Columns should be numbered if there are four or more columns.
- **Body of the table:**
- The data should be arranged in such a way that any figure can be located easily. Various types of numerical variables should be arranged in an ascending order, i.e., from left to right in rows and from top to bottom in columns. Column and row totals should be given.
- **Source:**
- At the bottom of the table note should be added indicating the primary and secondary sources from which data have been collected.

Importance of Tabulation

- Under tabulation, data is divided into various parts and for each part there are totals and sub totals. Therefore, relationship between different parts can be easily known.
- Since data are arranged in a table with a title and a number so these can be easily identified and used for the required purpose.
- Tabulation makes the data brief. Therefore, it can be easily presented in the form of graphs.
- Tabulation presents the numerical figures in an attractive form.
- Tabulation makes complex data simple and as a result of this, it becomes easy to understand the data.
- This form of the presentation of data helpful in finding mistakes.
- Tabulation is useful in condensing the collected data.
- Tabulation makes it easy to analyze the data from tables.
- Tabulation is very cheap mode to present the data. It saves time as well as space.
- Tabulation is a device to summaries the large scattered data. So, the maximum information may becollected from the tables.

Limitations of Tabulation

- Tables contain only numerical data. They do not contain details.
- Qualitative expression is not possible through tables.
- Tables can be used by experts only to draw conclusions. Common men do not understand them properly.
- **Methods of Tabulation**
- **Simple Tabulation**
- **Double Tabulation**
- **Complex Tabulation**
- **Simple Tabulation:** Simple Tabulation is when the data are tabulated to one characteristic. For example, the survey that determined the frequency or number of employees of a firm owning different brands of mobile phones like blackberry, Nokia, I Phone, etc.

Methods of Tabulation

- **Double Tabulation:** Double Tabulation is when two characteristic of data are tabulated. For example, frequency or number of male and female employees in the firm owning different brand of mobile phones like Blackberry, I Phone, etc.
- **Complex Tabulation**: Complex Tabulation of data that includes more than two characteristics. For example, frequency or number of male and female and the total employees owning different brand of mobile phones like Blackberry, Nokia, I phone, etc.
- Cross Tabulations is also a sub-type of complex tabulation that includes cross-classifying factors to build a contingency table of counts of frequencies at each combination of factor levels.
- A contingency table is a display format used to analyze and record the possible relationship between two or more categorical variables.

Data Interpretation – I

- Data interpretation is the process of reviewing data through some predefined processes which will help assign some meaning to the data and arrive at a relevant conclusion.
- It involves taking the result of data analysis, making inferences on the relations studied, and using them to conclude.
- Therefore, before one can talk about interpreting data, they need to be analyzed first. What then, is data analysis?
- Data analysis is the process of ordering, categorizing, manipulating, and summarizing data to obtain answers to research questions.
- It is usually the first step taken towards data interpretation.
- It is evident that the interpretation of data is very important, and as such needs to be done properly.
- Therefore, researchers have identified some data interpretation methods to aid this process.

Data Interpretation - II

- Data interpretation methods are how analysts help people make sense of numerical data that has been collected, analyzed and presented.
- Data, when collected in raw form, may be difficult for the layman to understand, which is why analysts need to break down the information gathered so that others can make sense of it.
- For example, when founders are pitching to potential investors, they must interpret data (e.g. market size, growth rate, etc.) for better understanding.
- There are 2 main methods in which this can be done, namely:
- Quantitative methods and
- Qualitative methods.



Qualitative Data Interpretation Method

- The qualitative data interpretation method is used to analyze qualitative data, which is also known as categorical data. This method uses texts, rather than numbers or patterns to describe data.
- Qualitative data is usually gathered using a wide variety of person-to-person techniques, which may be difficult to analyze compared to the quantitative research method.
- Unlike the quantitative data which can be analyzed directly after it has been collected and sorted, qualitative data needs to first be coded into numbers before it can be analyzed. This is because texts are usually cumbersome, and will take more time and result in a lot of errors if analyzed in its original state. Coding done by the analyst should also be documented so that it can be reused by others and also analyzed.
- There are 2 main types of qualitative data, namely; nominal and ordinal data. These 2 data types are both interpreted using the same method, but ordinal data interpretation is quite easier than that of nominal data.
- In most cases, ordinal data is usually labelled with numbers during the process of data collection, and coding may not be required. This is different from nominal data that still needs to be coded for proper interpretation.

Quantitative Data Interpretation Method

- The quantitative data interpretation method is used to analyze quantitative data, which is also known as numerical data.
- This data type contains numbers and is therefore analyzed with the use of numbers and not texts.
- Quantitative data are of 2 main types, namely; discrete and continuous data.
- Continuous data is further divided into interval data and ratio data, with all the data types being numeric.
- Due to its natural existence as a number, analysts do not need to employ the coding technique on quantitative data before it is analyzed.
- The process of analyzing quantitative data involves statistical modeling techniques such as standard deviation, mean and median.

Statistical Methods Used in Analyzing Quantitative Data

- **Mean**
- The mean is a numerical average for a set of data and is calculated by dividing the sum of the values by the number of values in a dataset. It is used to get an estimate of a large population from the dataset obtained from a sample of the population.
- Example: Online job boards in the US use the data collected from a group of registered users to estimate the salary paid to people of a particular profession.
- The estimate is usually made using the average salary submitted on their platform for each profession.
- **Frequency distribution**
- This technique is used to assess the demography of the respondents or the number of times a particular response appears in research.
- It is extremely keen on determining the degree of intersection between data points.

Statistical Methods Used in Analyzing Quantitative Data

- **Standard deviation**
- This technique is used to measure how well the responses align with or deviates from the mean.
- It describes the degree of consistency within the responses; together with the mean, it provides insight into data sets.
- In the job board example highlighted above, if the average salary of writers in the US is $20,000 per annum, and the standard deviation is 5.0, we can easily deduce that the salaries for the professionals are far away from each other.
- This will birth other questions like why the salaries deviate from each other that much.
- With this question, we may conclude that the sample contains people with few years of experience, which translates to a lower salary and people with many years of experience, translating to a higher salary. However, it does not contain people with mid-level experience.
- **Some other interpretation processes of quantitative data include:**
  - **Regression analysis**
  - **Cohort analysis**
  - **Predictive and prescriptive analysis.**

Tips for Collecting Accurate data for Interpretation - I

- **Identify the Required Data Type**
  - Researchers need to identify the type of data required for particular research.
  - Is it nominal, ordinal, interval or ratio data?
  - The key to collecting the required data to conduct research is to properly understand the research question.
  - If the researcher can understand the research question, then he can identify the kind of data that is required to carry out the research.
  - **Example:** when collecting customer feedback, the best data type to use is the ordinal data type.
  - Ordinal data can be used to access a customer's feelings about a brand and is also easy to interpret.

Tips for Collecting Accurate data for Interpretation - II

- **Avoid Biases**
- There are different kinds of biases a researcher might encounter when collecting data for analysis.
- Although biases sometimes come from the researcher, most of the biases encountered during the data collection process is caused by the respondent.
- There are 2 main biases, that can be caused by the President, namely:
  - ✓ Response bias and
  - ✓ Non-response bias.
- Researchers may not be able to eliminate these biases, but there are ways in which they can be avoided and reduced to a minimum.
- Response biases are biases that are caused by respondents intentionally giving wrong answers to responses, while non-response bias occurs when the respondents don't give answers to questions at all.
- Biases are capable of affecting the process of data interpretation.

Tips for Collecting Accurate data for Interpretation - III

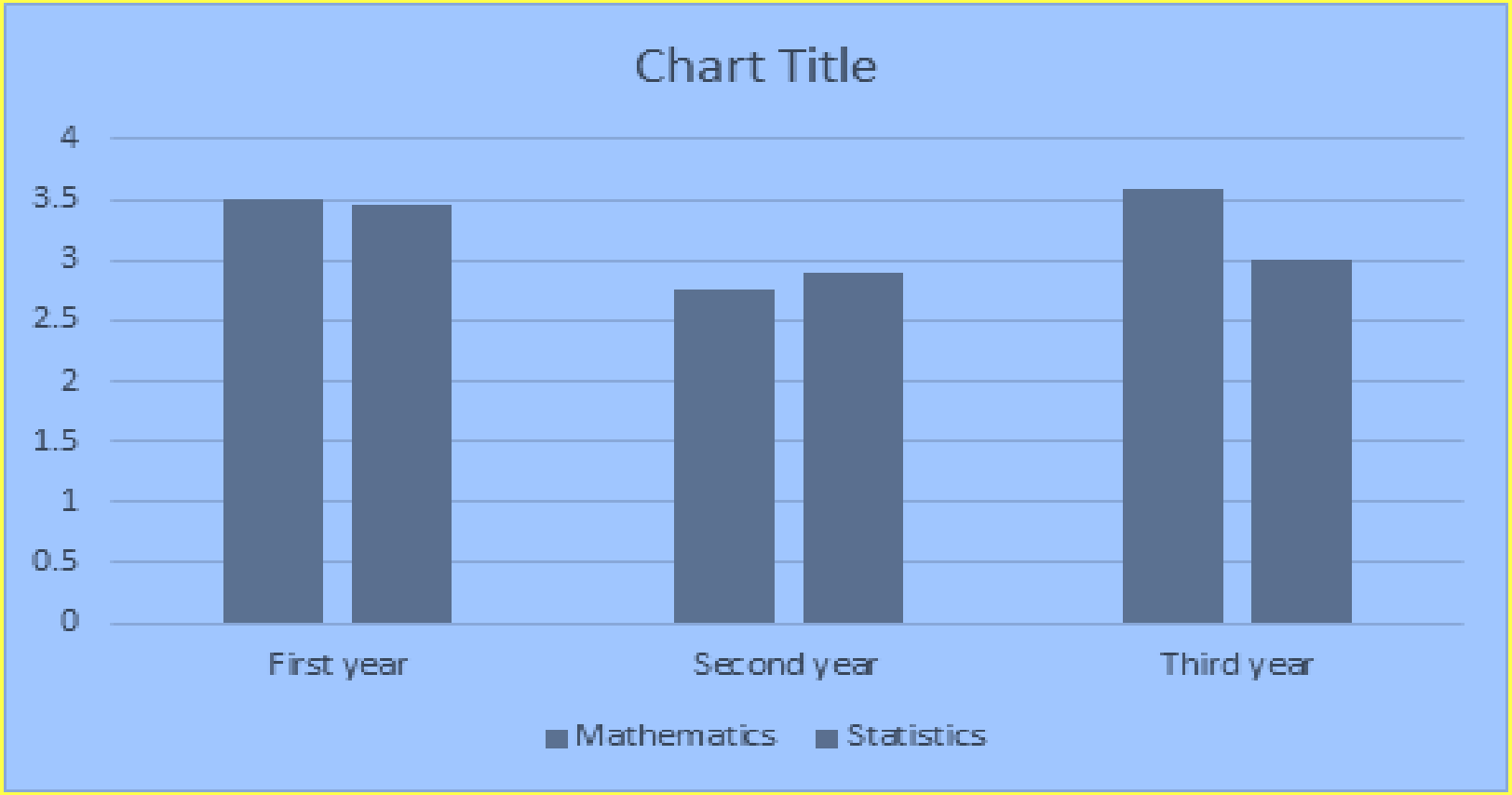
- **Use Close Ended Surveys**
  - Although open-ended surveys are capable of giving detailed information about the questions and allow respondents to fully express themselves, it is not the best kind of survey for data interpretation.
  - It requires a lot of coding before the data can be analyzed.
  - Close-ended surveys, on the other hand, restrict the respondents' answer to some predefined options, while simultaneously eliminating irrelevant data.
  - This way, researchers can easily analyze and interpret data.
  - However, close-ended surveys may not be applicable in some cases, like when collecting respondent's personal information like name, credit card details, phone number, etc.

Visualization Techniques in Data Analysis

- One of the best practices of data interpretation is the visualization of the dataset. Visualization makes it easy for a layman to understand the data, and also encourages people to view the data, as it provides a visually appealing summary of the data.
- **Different Techniques of Data Visualization:**
- **Bar Graphs**
- Bar graphs are graphs that interpret the relationship between 2 or more variables using rectangular bars.
- These rectangular bars can be drawn either vertically or horizontally, but they are mostly drawn vertically.
- The graph contains the horizontal axis (x) and the vertical axis (y), with the former representing the independent variable while the latter is the dependent variable.
- Bar graphs can be grouped into different types, depending on how the rectangular bars are placed on the graph.

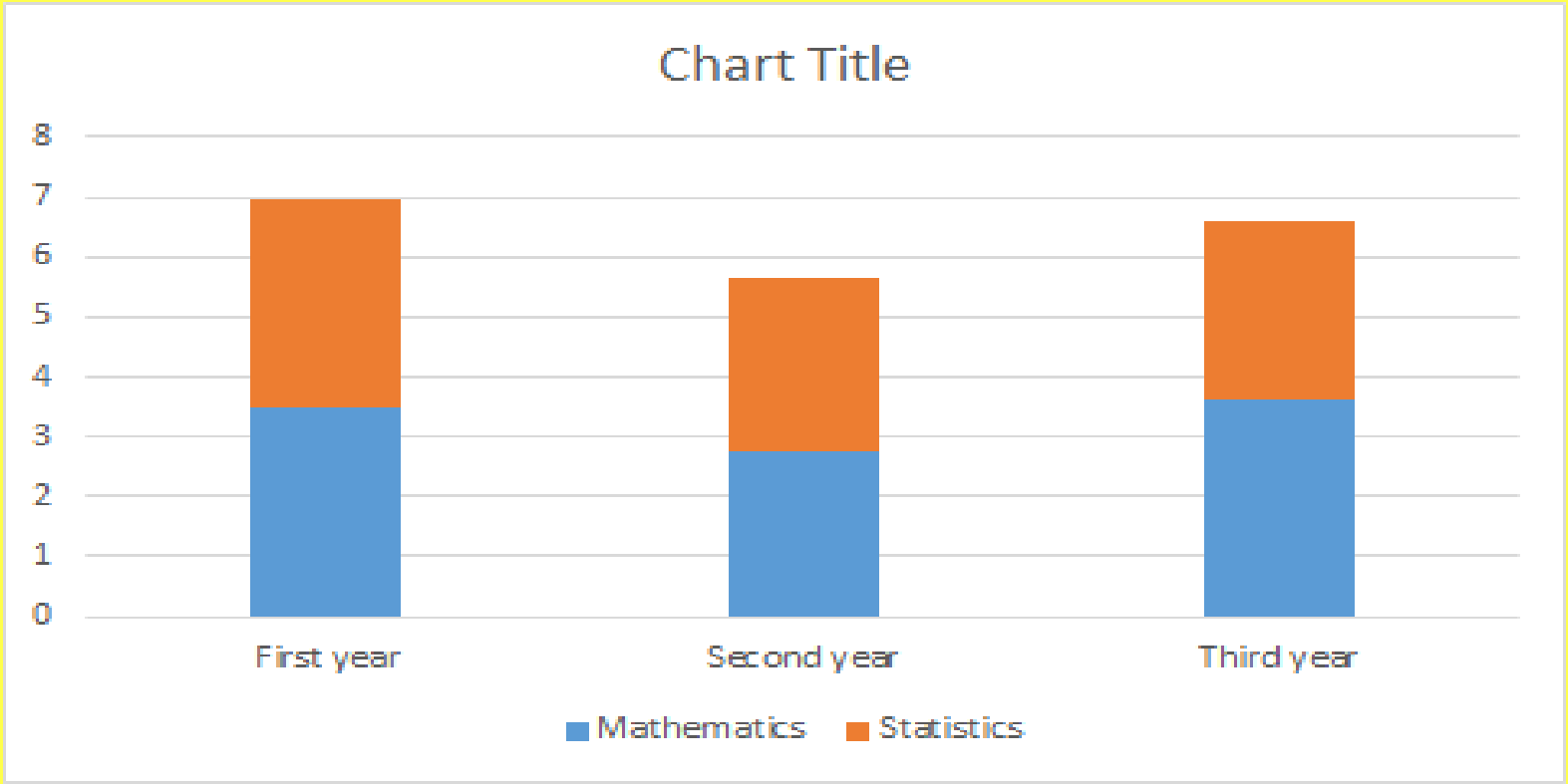
Types of Bar Graphs - I

- **Grouped Bar Graph**
- The grouped bar graph is used to show more information about variables that are subgroups of the same group with each subgroup bar placed side-by-side like in a histogram. For example, let us consider the subgroups of the average CGPA of students in the mathematics and statistics department of an institution.



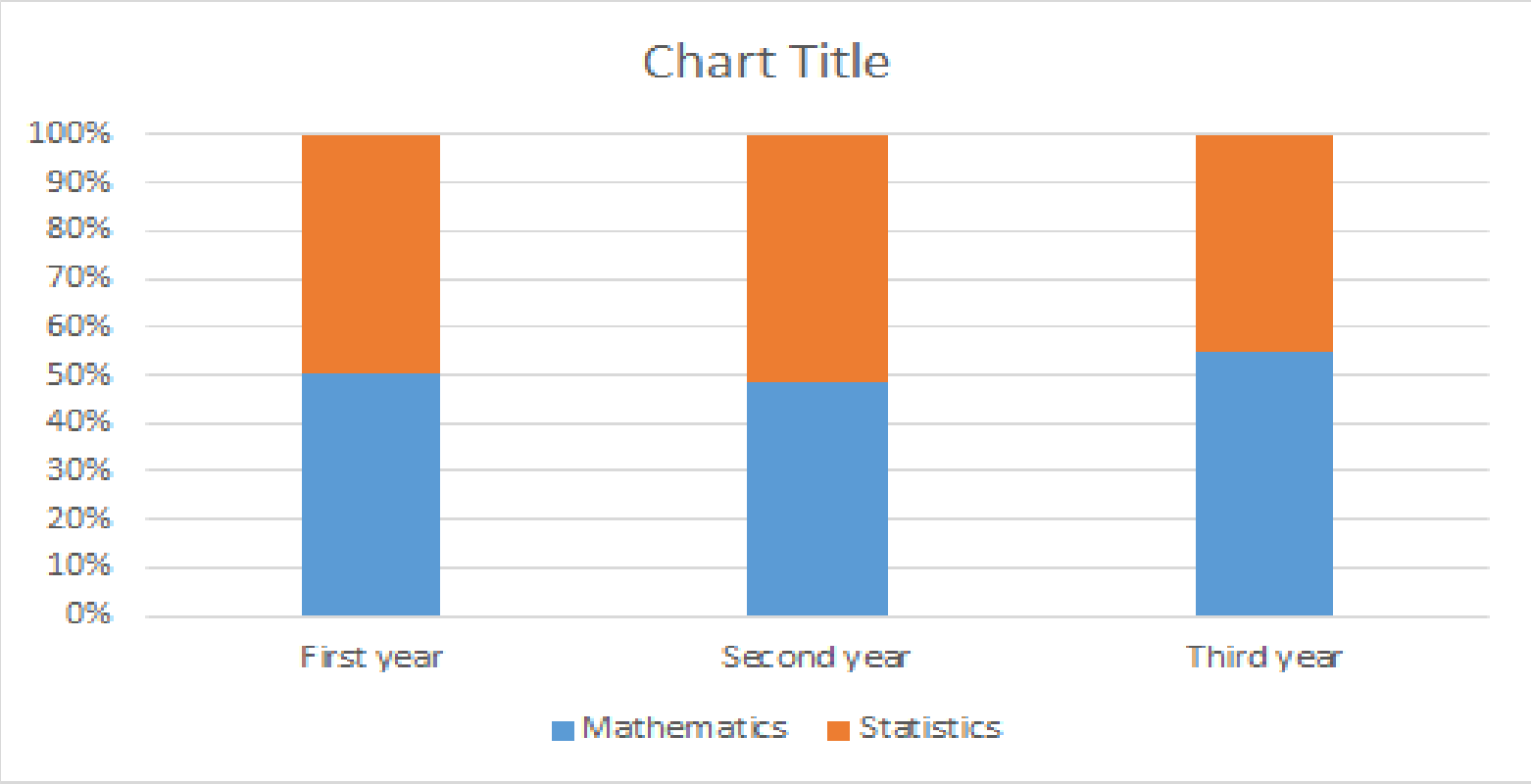
Types of Bar Graphs - II

- **Stacked Bar Graph**
- A stacked bar graph is a grouped bar graph with its rectangular bars stacked on top of each other rather than placed side by side. By representing the CGPA example highlighted above using a stacked bar graph, we get the graph below.
- In this graph, the rectangular bars representing the mathematics and statistics department are stacked on top of each other with each group representing, first year, second year, and third-year students.



Types of Bar Graphs - III

- **Segmented Bar Graph**
- Segmented bar graphs are stacked bar graphs where each rectangular bar shows 100% of the dependent variable. It is mostly used when there is an intersection between the variable categories.
- The segmented bar graph shows the percentage of the average CGPA accumulated by the mathematics and statistics department of the institution.



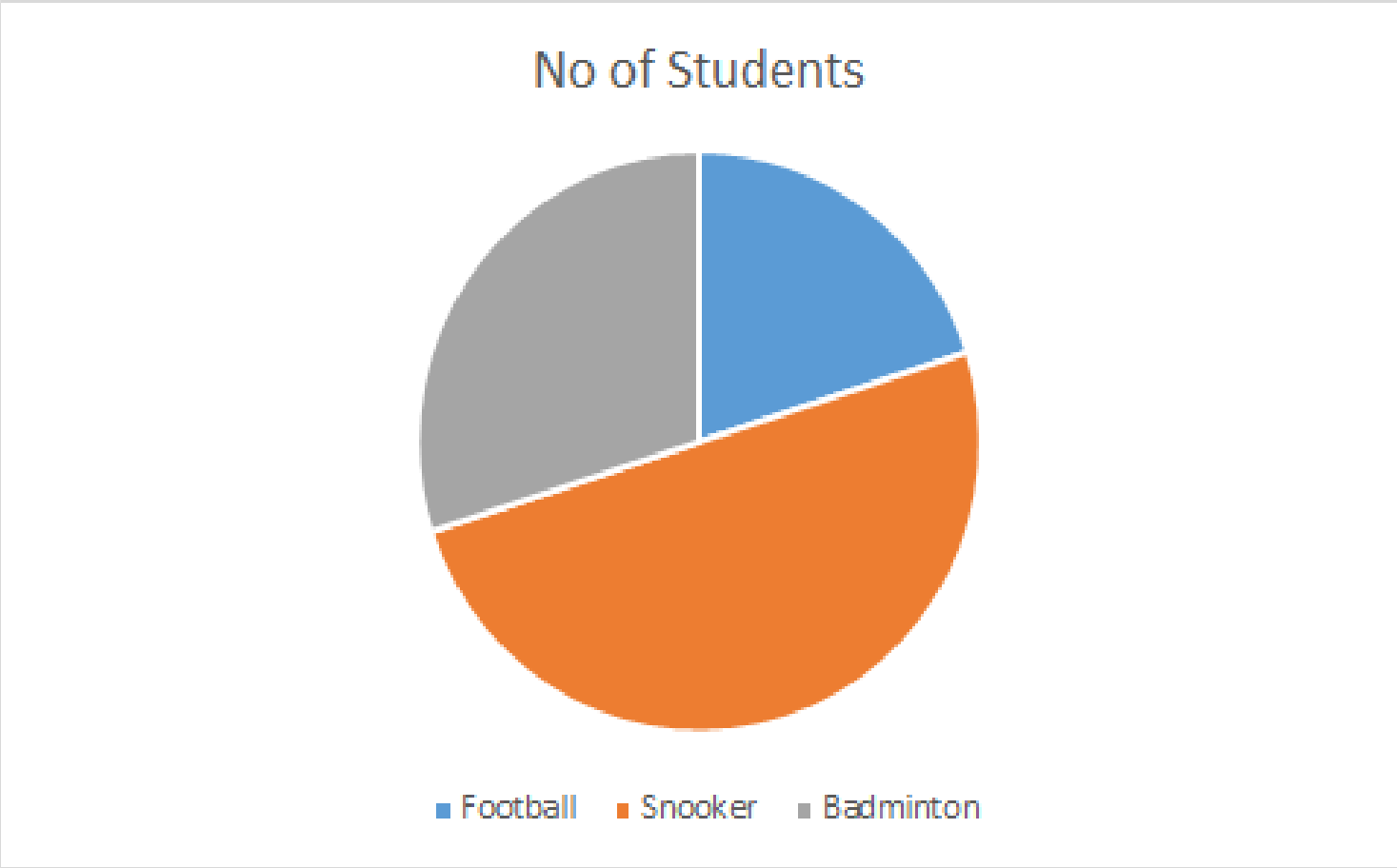
Pie Chart

- A pie chart is a circular graph used to represent the percentage of occurrence of a variable using sectors.
- The size of each sector is dependent on the frequency or percentage of the corresponding variables.
- There are different variants of the pie charts, but for the sake of this article, we will be restricting ourselves to only 3. For better illustration of these types, let us consider the following examples.
- **Pie Chart Example:** There are a total of 50 students in a class, and out of them, 10 students like Football, 25 students like snooker, and 15 students like Badminton.

Types of Pie Chart - I

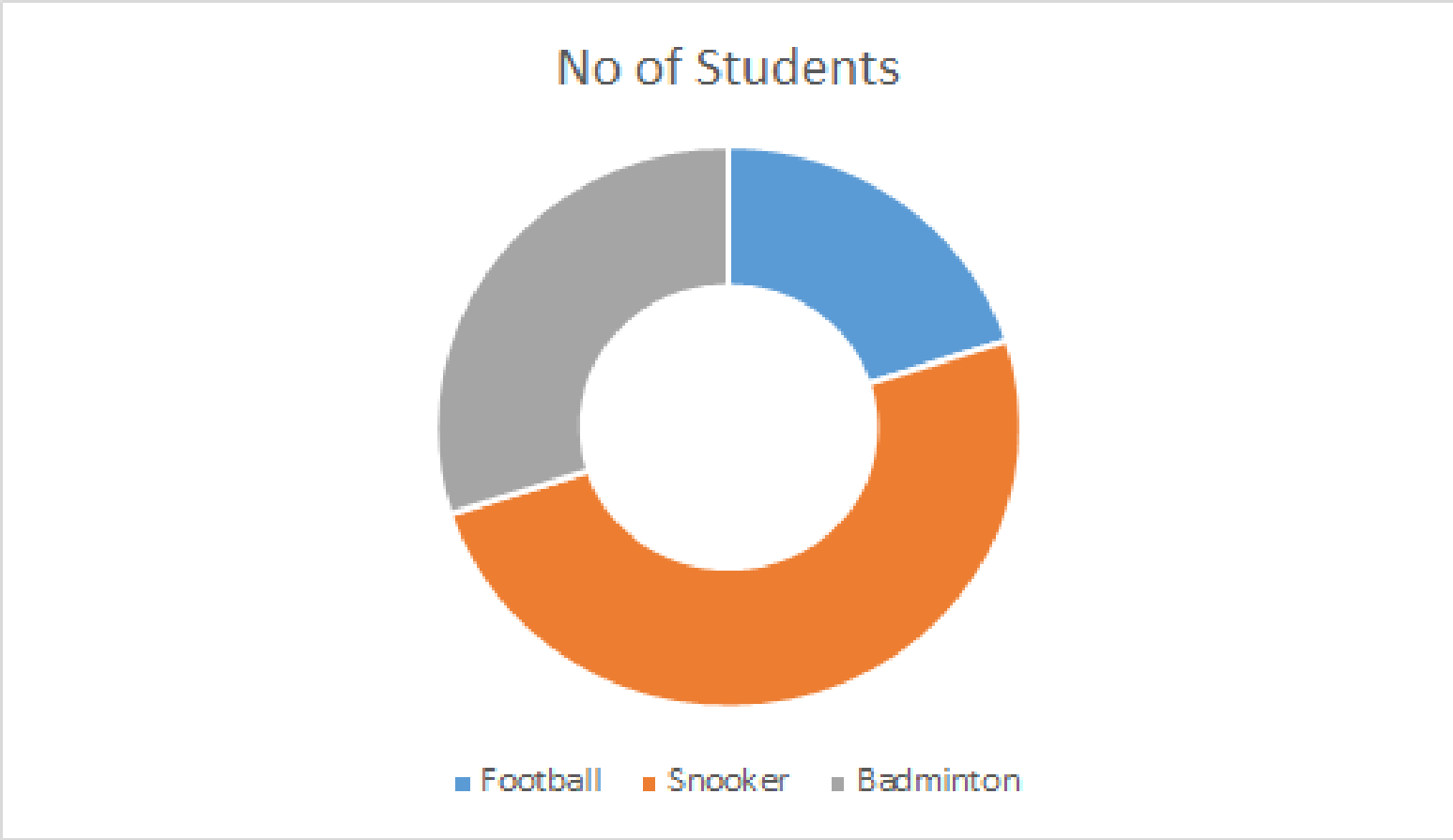
- **Simple Pie Chart**

The simple pie chart is the most basic type of pie chart, which is used to depict the general representation of a bar chart.



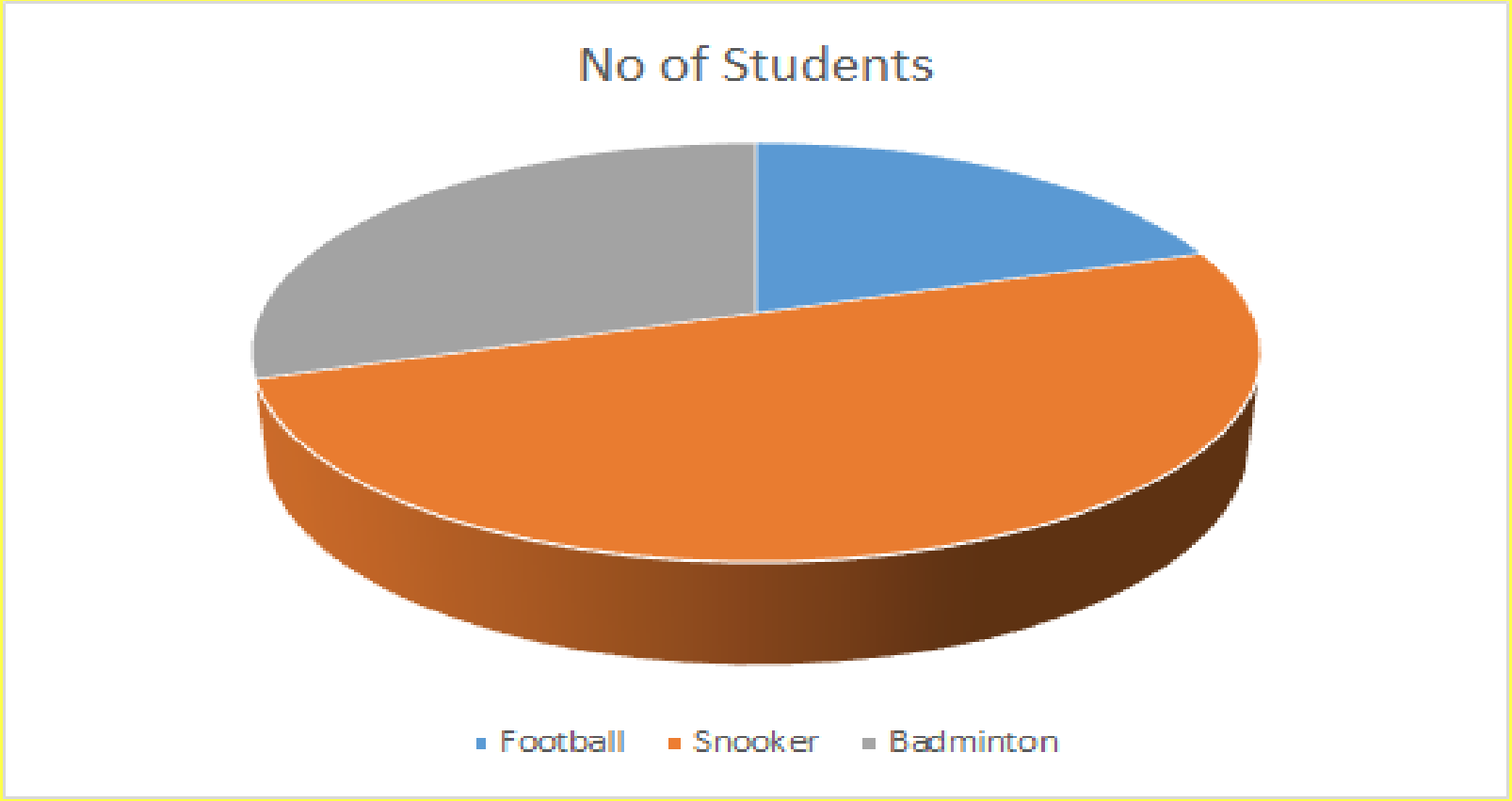
Types of Pie Chart - II

- **Doughnut Pie Chart**
- Doughnut pie is a variant of the pie chart, with a blank centre allowing for additional information about the data as a whole to be included.



Types of Pie Chart - III

- **3D Pie Chart**
- 3D pie chart is used to give the chart a 3D look and is often used for aesthetic purposes. It is usually difficult to reach because of the distortion of perspective due to the third dimension.



Tables - I

- Tables are used to represent statistical data by placing them in rows and columns.
- They are one of the most common statistical visualization techniques and are of 2 main types, namely:
- Simple and
- Complex tables.
- **Simple Tables**

Simple tables summarize information on a single characteristic and may also be called a univariate table. An example of a simple table showing the number of employed people in a community concerning their age group.

Age Group	16-25	25-35	36-50	51-64
No. of Employed	15	30	35	25

Tables - II

- **Complex Tables**

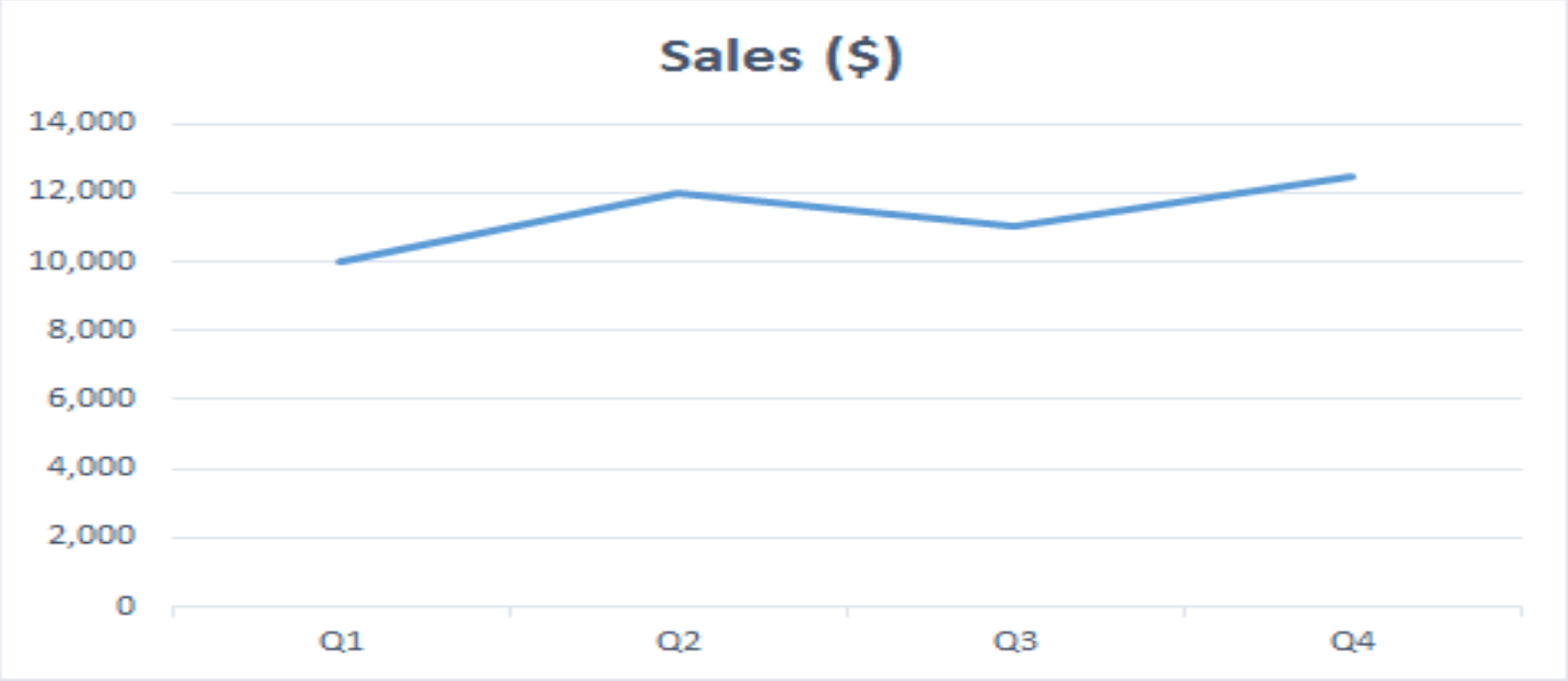
As its name suggests, complex tables summarize complex information and present them in two or more intersecting categories. A complex table example is a table showing the number of employed people in a population concerning their age group and sex as shown in the table below.

Age Group	No. of Employed		Total
	Male	Female	
16-25	10	5	15
25-35	16	14	30
36-50	25	10	35
51-64	15	10	25

Line Graph - I

- Line graphs or charts are a type of graph that displays information as a series of points, usually connected by a straight line. Some of the types of line graphs are highlighted below.
- **Simple Line Graphs**

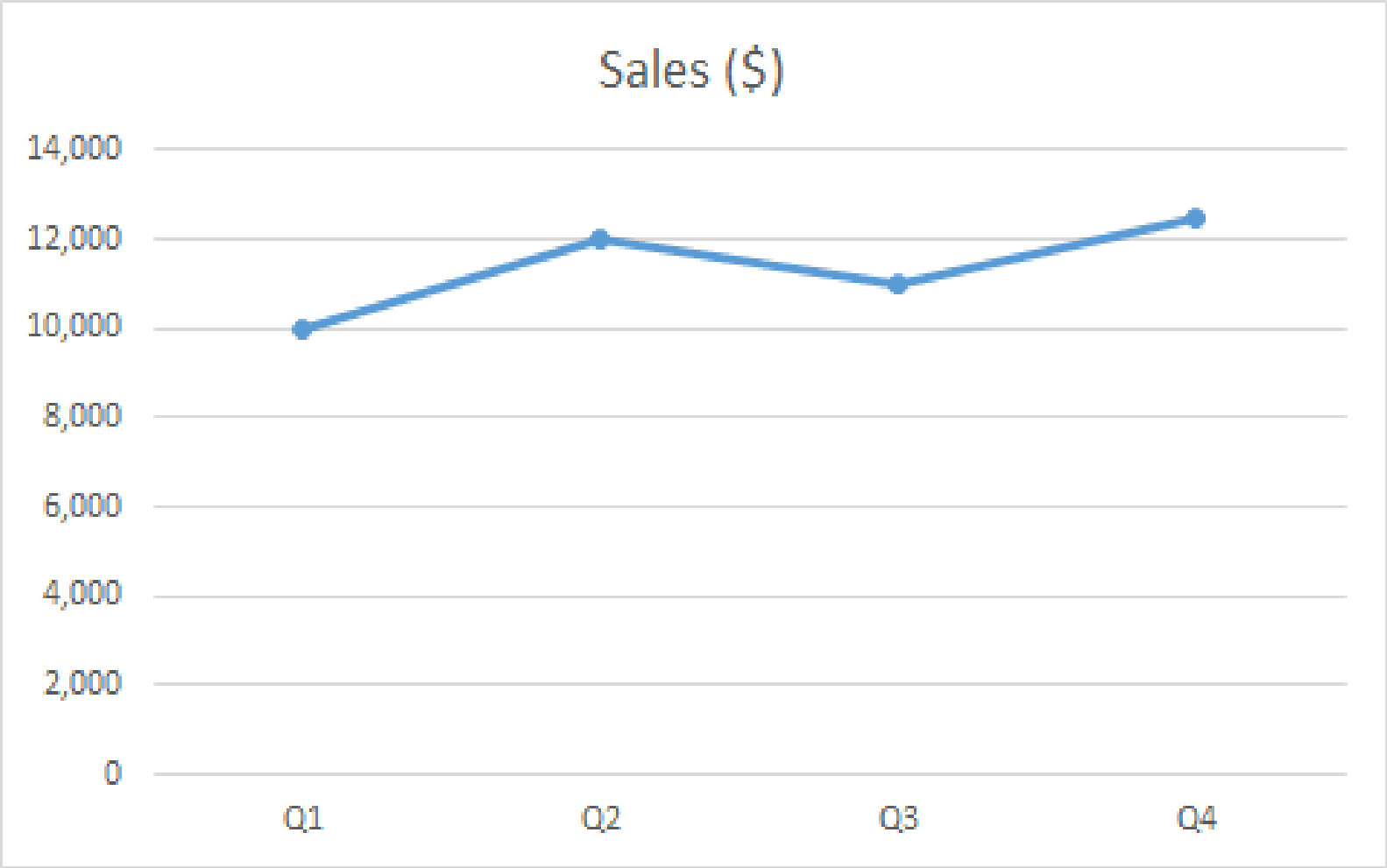
  Simple line graphs show the trend of data over time, and may also be used to compare categories. Let us assume we got the sales data of a firm for each quarter and are to visualize it using a line graph to estimate sales for the next year.



Line Graph - II

- **Line Graphs with Markers**

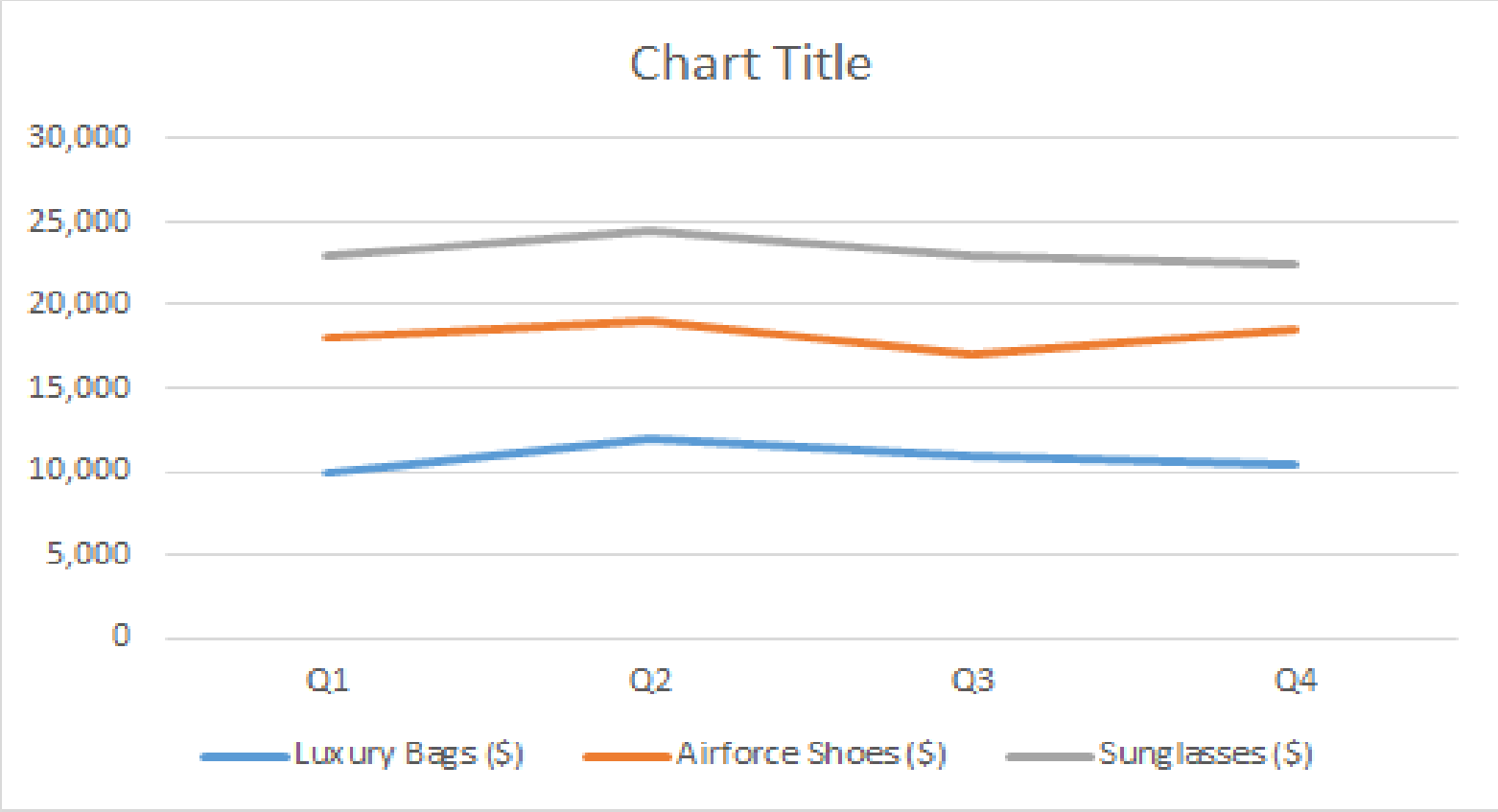
These are similar to line graphs but have visible markers illustrating the data points.



Line Graph - III

- **Stacked Line Graphs**

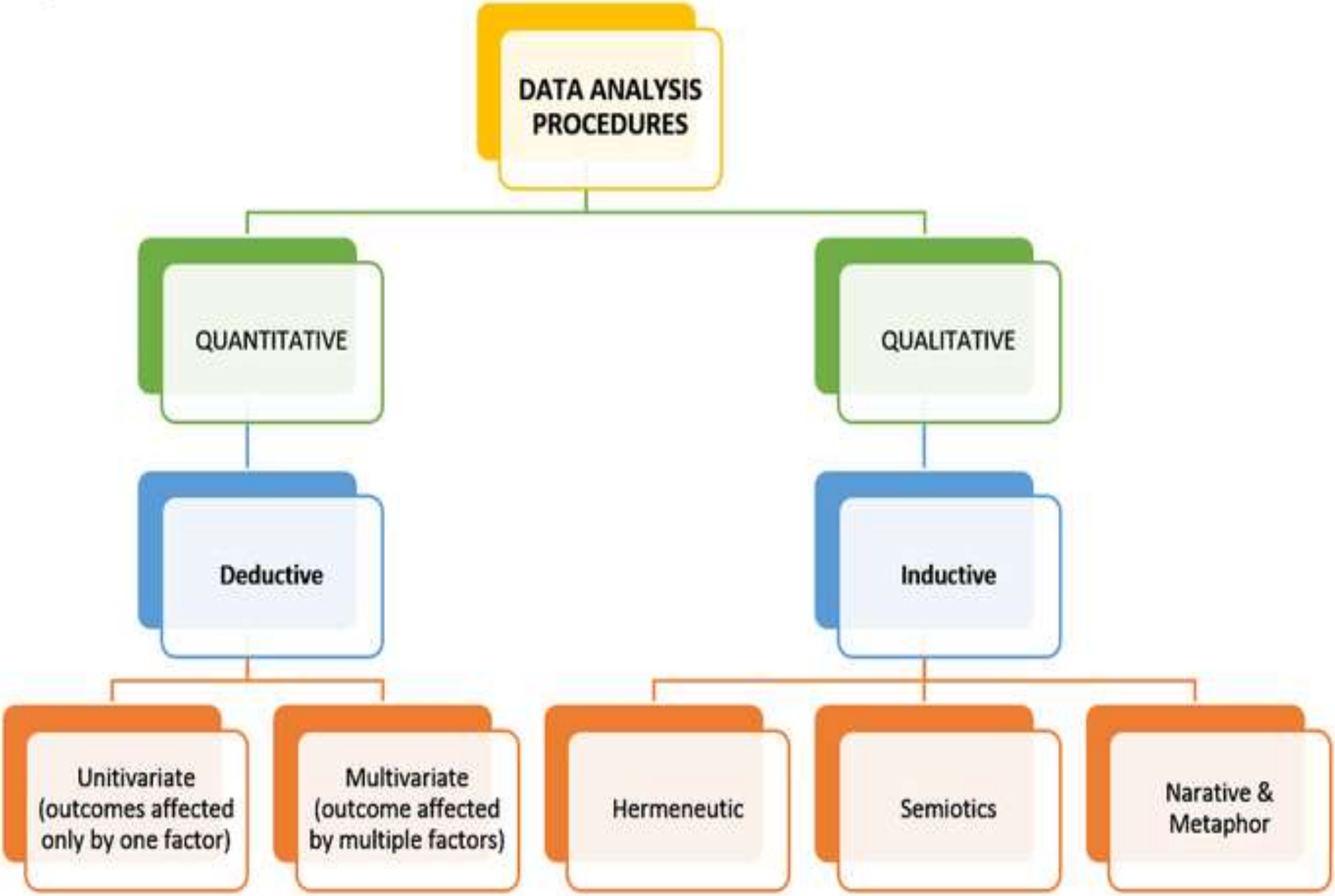
Stacked line graphs are line graphs where the points do not overlap, and the graphs are therefore placed on top of each other. Consider that we got the quarterly sales data for each product sold by the company and are to visualize it to predict company sales for the next year.



Advantages and Disadvantages of Visualization Techniques

S. No.	Visualization Techniques	Advantages	Disadvantages
1.	Bar Graph	✓ It helps to summarize a large data. ✓ Estimations of key values can be made at a glance. ✓ Can be easily understood.	✓It may require additional explanation. ✓It can be easily manipulated. ✓It doesn't properly describe the dataset.
2.	Pie Chart	✓It is visually appealing. ✓Best for comparing small data samples.	✓It can only compare small sample sizes. ✓Unhelpful with observing trends over time.
3.	Tables	✓Can Contain large data sets. ✓Helpful in comparing 2 or more similar things.	✓They do not give detailed information. ✓Maybe time-consuming.
4.	Line Graph	✓Great for visualizing trends and changes over time. ✓It is simple to construct and read.	✓It can not compare different variables at a single place or time.

General differences of Data Analysis Procedures

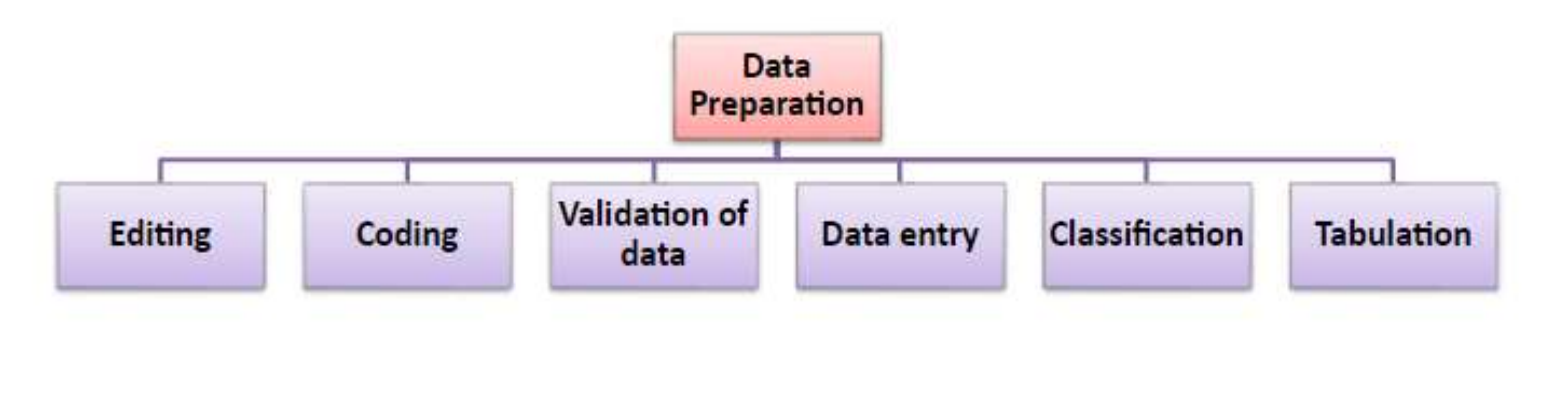


Quantitative and Qualitative Data Analysis Procedures

Quantitative Data Analysis Methods	Qualitative Data Analysis Methods
Involve statistical analysis (number-analysis)	Involve non–statistical analysis (text-analysis)
Seek deductive inferences	Seek inductive inferences
Focus on quantifiable phenomena (comparisons differences, trends, variable relationships; etc.)	Focus on meanings (descriptions of themes)
Involve data clustering analysis for relationships in non-hypothesis testing	Involve data structuring and coding into groups and themes
Involve systematic, predetermined and categorical analysis	Involve in-depth, un-predetermined, fluid, and non-categorical analysis
Assume possibility of value-free inquiry	Consider attention of researcher's impact on others' values
Adopt an objective approach	May adopt a subjective approach
Are usually narrow, specific and seek uniform generalizations	Are general, broad and is context sensitive

Data Preparation

- The data collected from the respondents is generally not in the form to be analyzed directly.
- After the responses are recorded or received, the next stage is that of preparation of data i.e. to make the data amenable for appropriate analysis.
- Data preparation includes, editing, coding, and data entry and is the activity that ensures the accuracy of the data and their conversion from raw form to reduced and classified forms that are more appropriate for analysis.
- Preparing a descriptive statistical summary is another preliminary step leading to an understanding of the collected data.



Data Preparation and Preprocessing

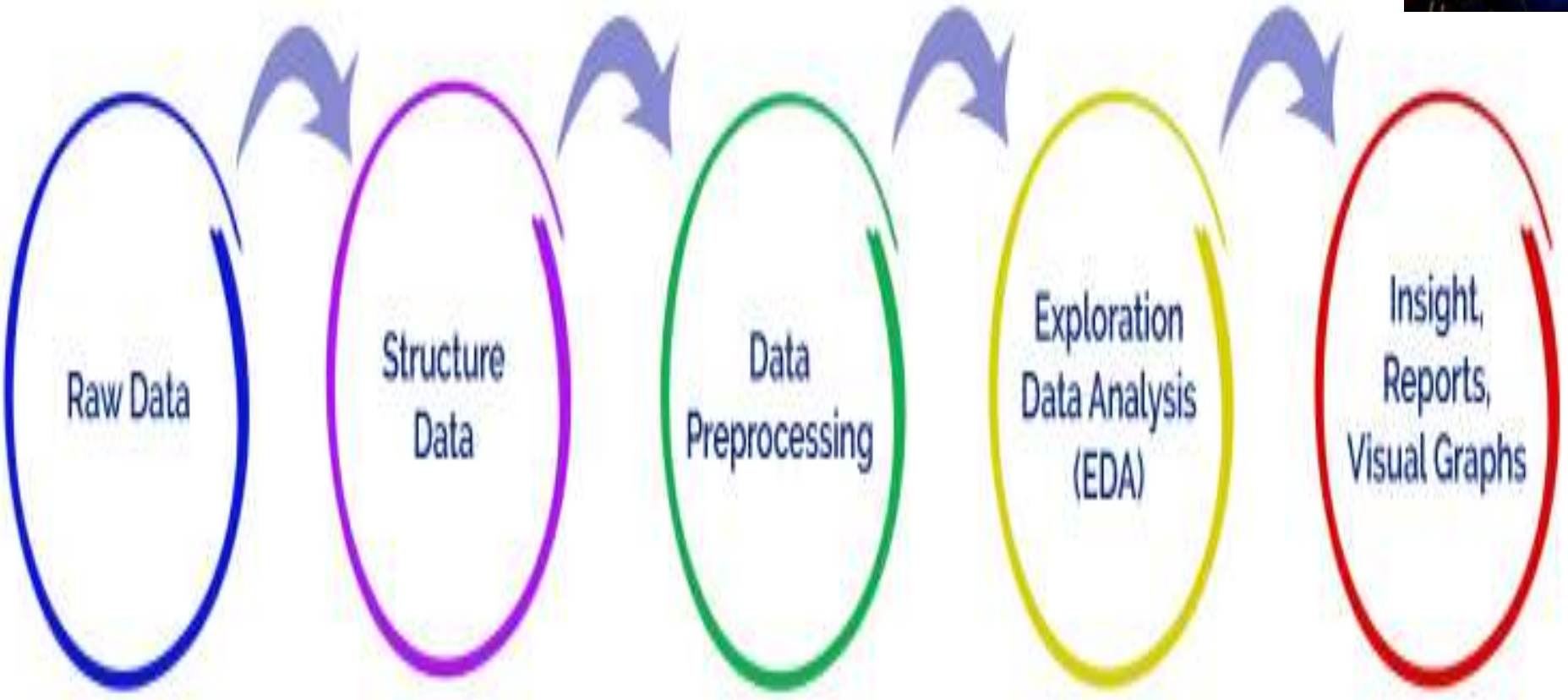
- Deep learning and Machine learning are becoming more and more important in today's ERP (Enterprise Resource Planning).
- During the process of building the analytical model using Deep Learning or Machine Learning the data set is collected from various sources such as a file, database, sensors, and much more.
- But, the collected data cannot be used directly for performing the analysis process.
- Therefore, to solve this problem Data Preparation is done. It includes two techniques that are listed below :
  - Data Preprocessing
  - Data Wrangling
- **Data Preparation Architecture**
- Data Preparation process s an important part of data science. It includes two concepts such as Data Cleaning and Feature Engineering.
- These two are compulsory for achieving better accuracy and performance in the Machine Learning and Deep Learning projects.

Data Preparation and Description

- Data Preparation includes editing, coding and data entry.
- It is the activity that ensures the accuracy of the data and their conversion from raw form to reduced and classified forms that are more appropriate for analysis.
- Preparing a descriptive statistic summary is another preliminary step that allows data entry errors to be identified and corrected.

Architecture of Data Preparation





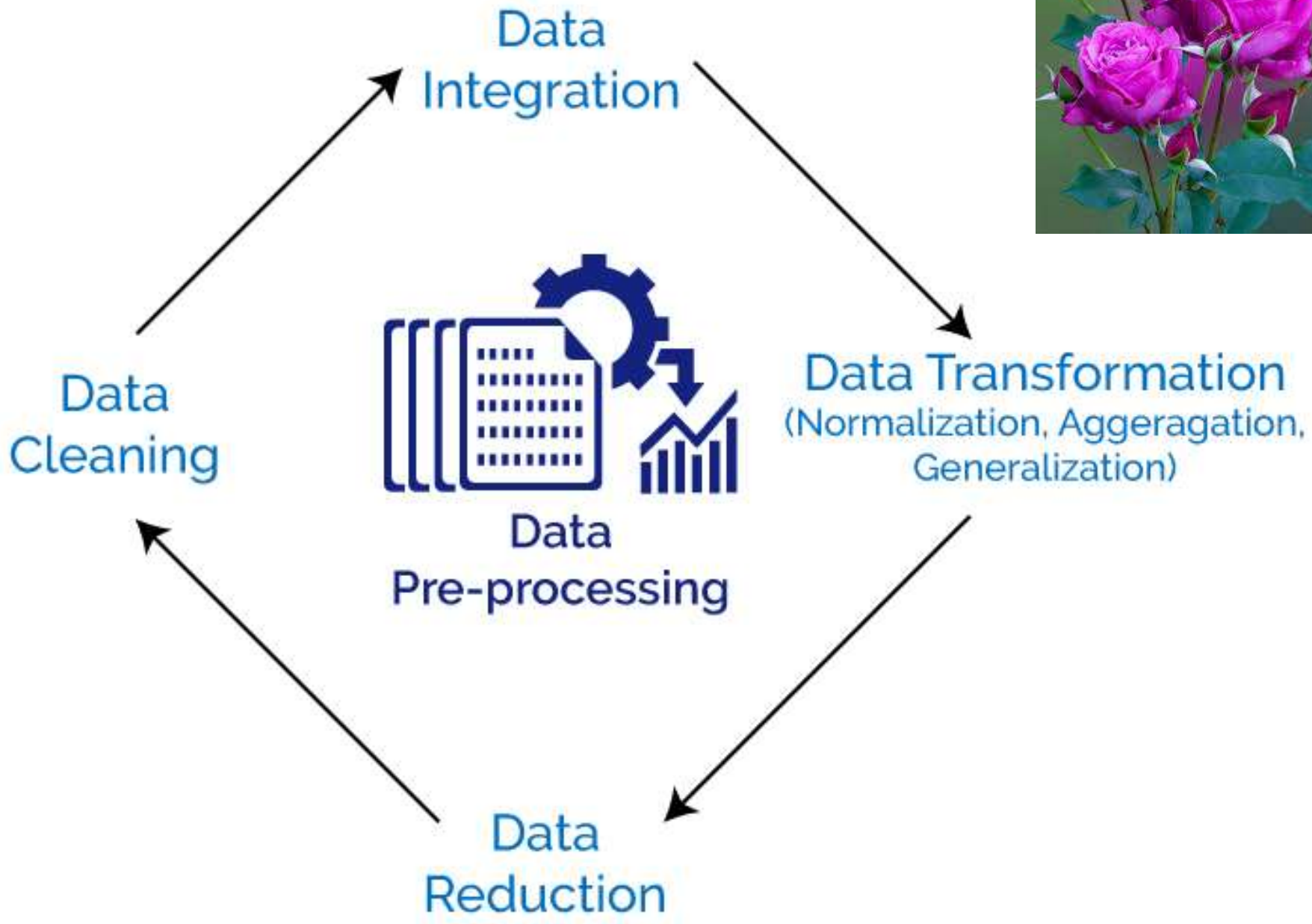
Data Preprocessing

- Data Preprocessing is a technique that is used to convert the raw data into a clean data set.
- In other words, whenever the data is gathered from different sources it is collected in raw format which is not feasible for the analysis.
- Therefore, certain steps are executed to convert the data into a small clean data set.
- This technique is performed before the execution of the Iterative Analysis. The set of steps is known as Data Preprocessing. It includes :
  - Data Cleaning
  - Data Integration
  - Data Transformation
  - Data Reduction



Data Preprocessing





Need of Data Preparation Process & Preprocessing

- For achieving better results from the applied model in Machine Learning and Deep Learning projects the format of the data has to be in a proper manner, this is where term Data Preparation is used.
- Some specified Machine Learning and Deep Learning model need information in a specified format, for example, Random Forest algorithm does not support null values, therefore to execute random forest algorithm null values has to be managed from the original raw data set.
- Another aspect of Data Preparation and analysis is that the data set should be formatted in such a way that more than one Machine Learning and Deep Learning algorithms are executed in one data set, and the best out of them is chosen.

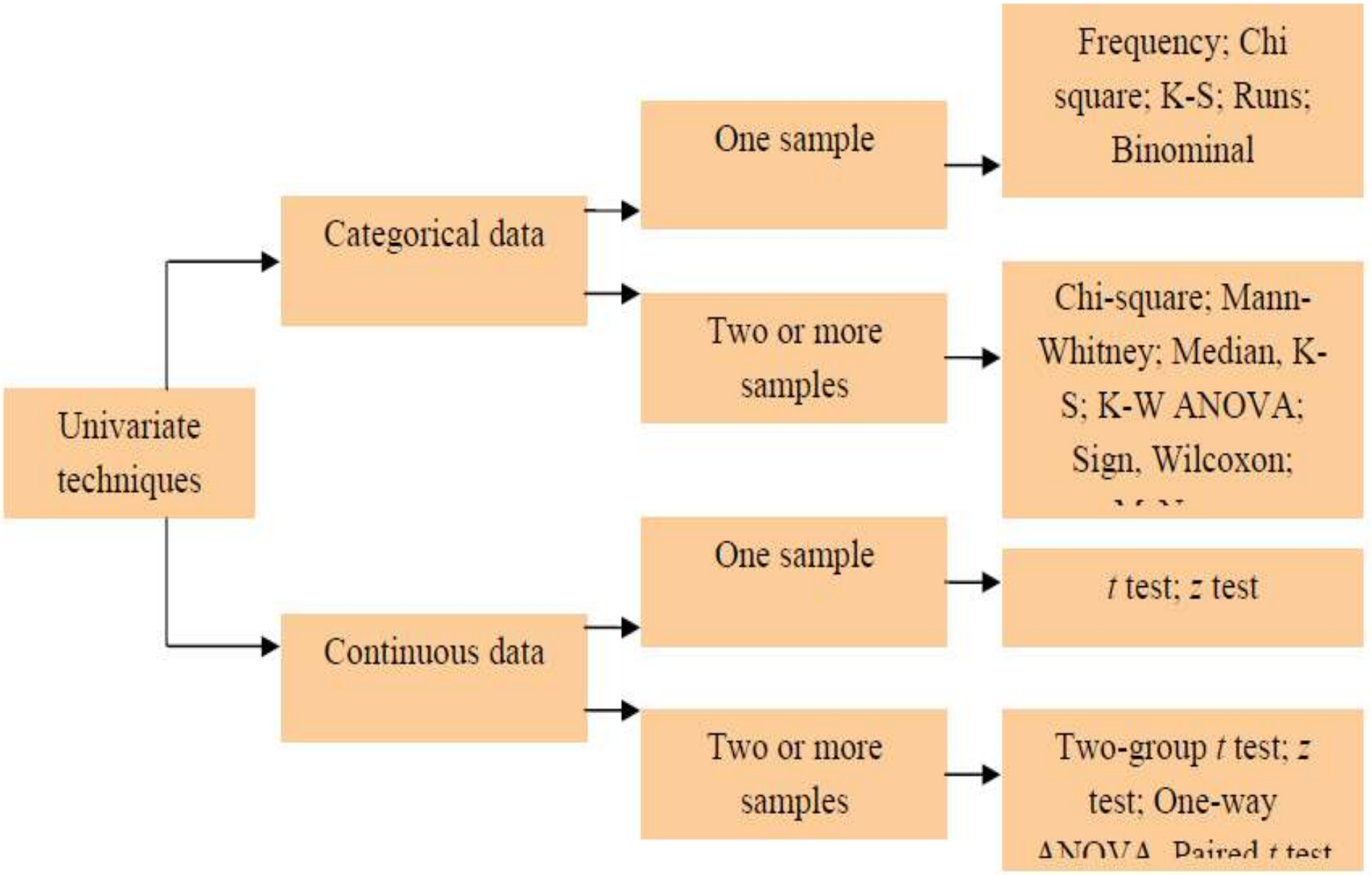
Importance of Data Preparation

- Data Preprocessing is necessary because of the presence of unformatted real-world data. Mostly real-world data is composed .
- **Inaccurate data (missing data) –** There are many reasons for missing data such as data is not continuously collected, a mistake in data entry, technical problems with biometrics, and much more, which requires proper Data Preparation.
- **The presence of noisy data (erroneous data and outliers) –** The reasons for the existence of noisy data could be a technological problem of gadget that gathers data, a human mistake during data entry and much more.
- **Inconsistent data –** The presence of inconsistencies are due to the reasons such that existence of duplication within data, human data entry, containing mistakes in codes or names, i.e., violation of data constraints and much more necessitate Data Preparation and analysis.
- Therefore, to handle raw data, Data Preprocessing is performed.

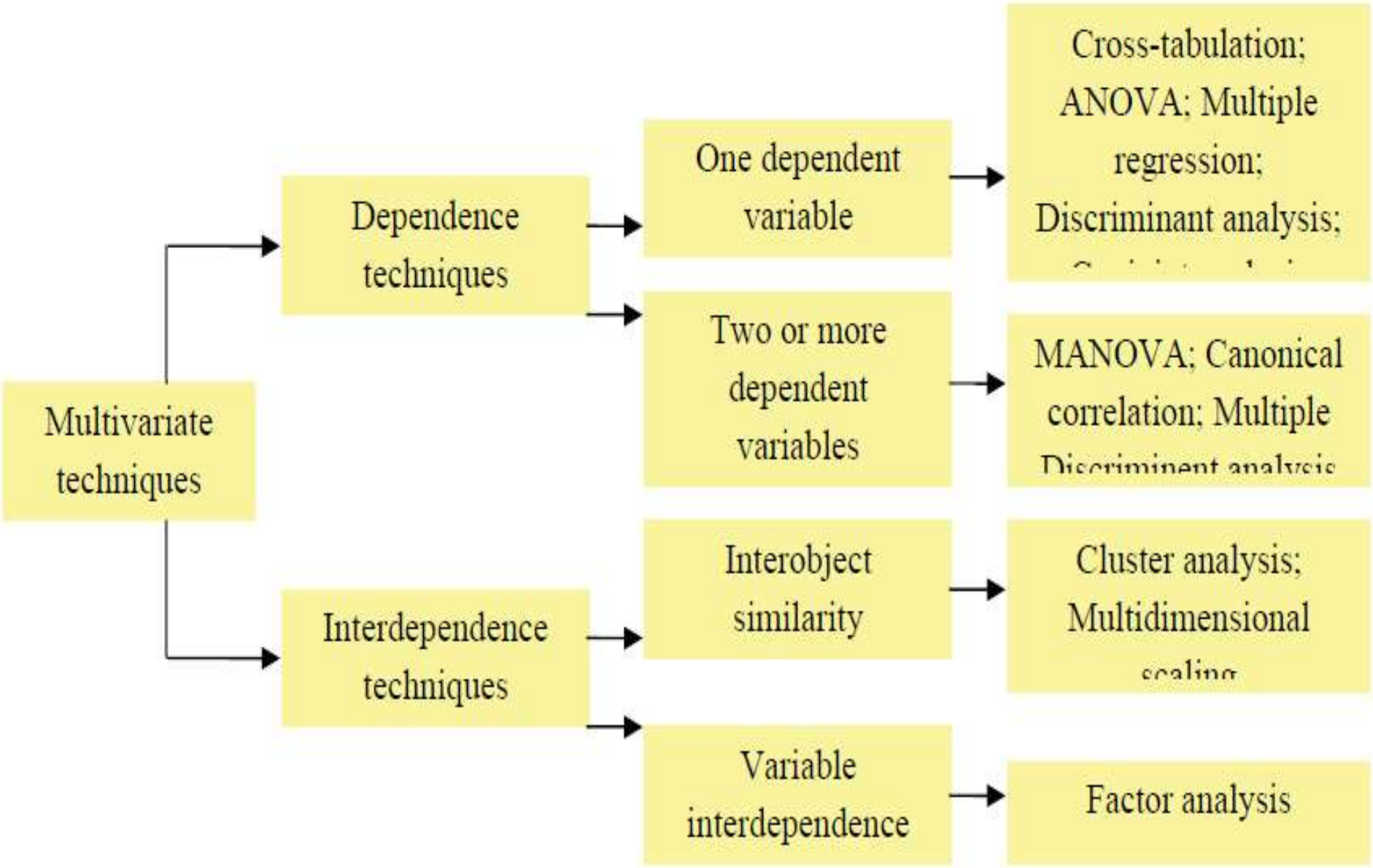
Preliminary Analysis

- Preliminary Analysis (My Preference) (lots of other terms: Initial feasibility, Initial study phase, Problem statements and Analysis, others; gathering information and needs.
- My definition of Preliminary Analysis includes all the items due for your first deliverable:
  - Project Description
    - ✓ Domain Modeling; Glossary
  - Use Case Collection
    - ✓ Brief Description of Action/Use Cases
  - Project Scope – Preliminary
  - Project Plan - Preliminary

Univariate Techniques



Multivariate Techniques



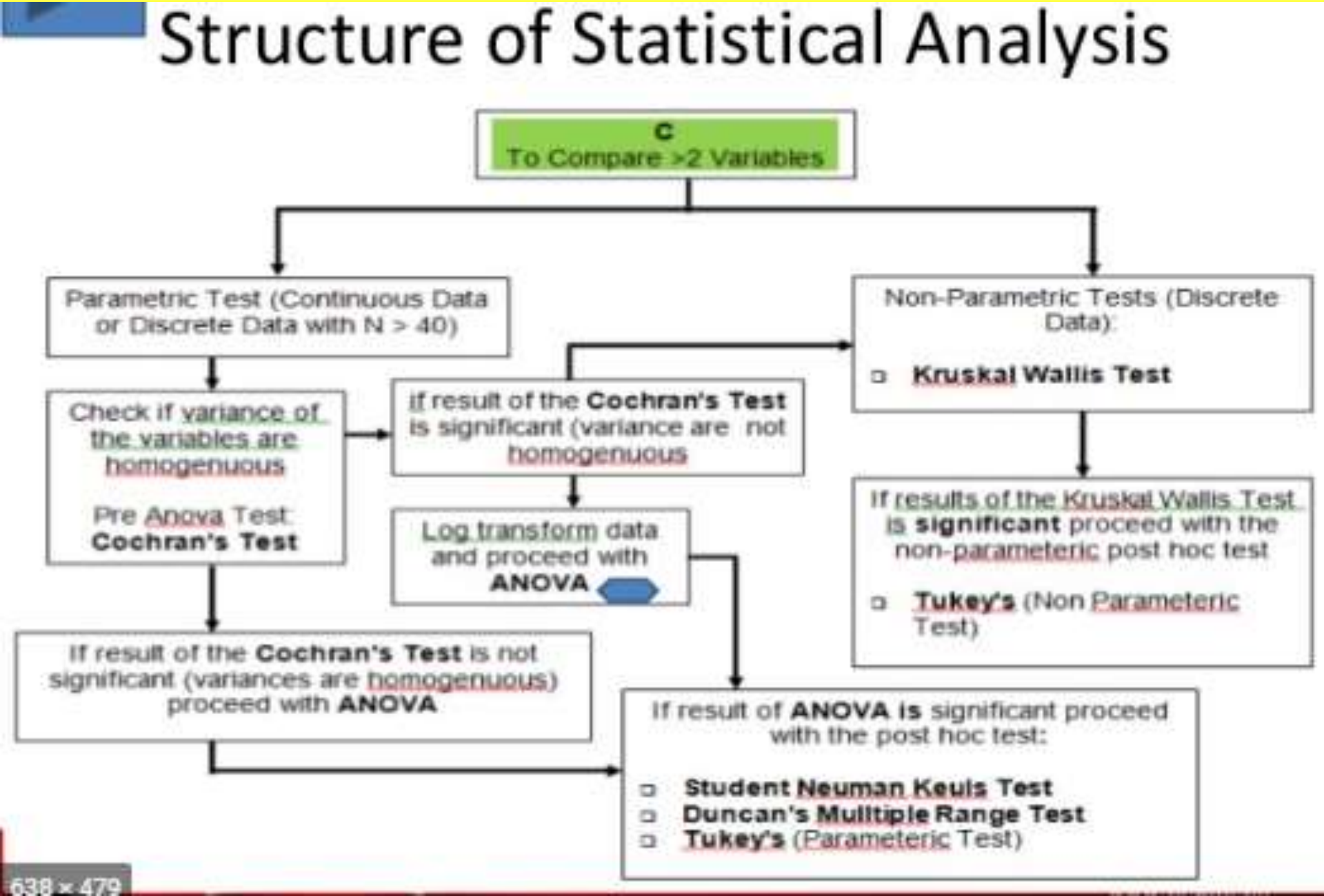
Statistical Analysis - I

- Statistical analysis is the science of collecting data and uncovering patterns and trends. It's really just another way of saying “statistics.” After collecting data you can analyze it to: Summarize the data. For example, make a pie chart.
- Statistics is basically a science that involves data collection, data interpretation and finally, data validation. Statistical data analysis is a procedure of performing various statistical operations. It is a kind of quantitative research, which seeks to quantify the data, and typically, applies some form of statistical analysis. Quantitative data basically involves descriptive data, such as survey data and observational data.
- Statistical data analysis generally involves some form of statistical tools, which a layman cannot perform without having any statistical knowledge. There are various software packages to perform statistical data analysis. This software includes Statistical Analysis System (SAS), Statistical Package for the Social Sciences (SPSS), Stat soft, etc.

Statistical Analysis - II

- Small relationships are detectable between any data sets.
  - They may be caused by chance
  - They may be caused by something else.
- Statistical tests help us decide, for any given sets of data, which of these possibilities is more likely.
- If we find that it is highly probable that the relationship was caused by chance, then we assume it was caused by chance. i.e. we accept the null hypothesis.
- If we find that it is highly improbable that the relationship was caused by chance, then we assume it was caused by something else. i.e. we reject the null hypothesis.

Structure of Statistical Analysis



638 × 479

Statistical Cycle

The statistical cycle



7 MAIN TYPES OF STATISTICAL ANALYSIS



DESCRIPTIVE TYPE

As the name suggests, the descriptive statistic is used to describe! It describ
the basic features of data and shows or summarizes data in a rational wa

However, descriptive statistics do not allow making conclusions. You can not g
conclusions and make generalizations that extend beyond the data at han

INFERENTIAL TYPE



Inferential statistics is a result of more complicated mathematical estimations, and allow us to infer trends about a larger population based on samples of "subjects" taken from it.

This type of statistical analysis is used to study the relationships between variables within a sample, and you can make conclusions, generalizations or predictions about a bigger population.



PREDICTIVE ANALYTICS

Predictive analytics uses techniques such as statistical algorithms an
machine learning to define the likelihood of future results, behavio
and trends based on both new and historical dat

It is important to note that no statistical method can "predict" the future with 100%
surety. Businesses use these statistics to answer the question "What might happen?

PRESCRIPTIVE ANALYTICS



Prescriptive analytics is a study which examines data to answer the question "What should be done?"

Prescriptive analytics aim to find the optimal recommendations for a decision making process. It is all about providing advice.



CAUSAL ANALYSIS

When you would like to understand and identify the reasons wh
things are as they are, causal analysis comes to help. This type
analysis answers the question "Why

The business world is full of events that lead to failure. The causal seeks
identify the reasons why? It is better to find causes and to treat them instea
of treating symptom

EXPLORATORY DATA ANALYSIS (EDA)



EDA is an analysis approach that focuses on identifying general patterns in the data and to find previously unknown relationships.

EDA alone should not be used for generalizing or predicting. EDA is used for taking a bird's eye view of the data and trying to make some feeling or sense of it. Commonly, It is the first step in data analysis, performed before other formal statistical techniques.



MECHANISTIC ANALYSIS

The mechanistic analysis is about understanding the exact changes
given variables that lead to changes in other variables. Howeve
mechanistic does not consider external influence

Types of Statistical Analysis

- Descriptive Type of Statistical Analysis
- Inferential Type of Statistical Analysis
- Prescriptive Analysis
- Predictive Analysis
- Casual Analysis
- Exploratory Data Analysis
- Mechanistic Analysis





Descriptive Statistical Analysis - I

- Descriptive statistical analysis as the name suggests helps in describing the data.
- It gets the summary of data in a way that meaningful information can be interpreted from it.
- Using descriptive analysis, we do not get to a conclusion however we get to know what in the data is i.e. we get to know the quantitative description of the data.
- For instance, consider a simple example in which you must determine how well the student performed throughout the semester by calculating the average.
- This average is nothing but the sum of the score in all the subjects in the semester by the total number of subjects.
- This single number is describing the general performance of the student across a potentially wide range of subject experiences.
- Whenever we try to describe a large set of observations with a single value, we run into the risk of either distorting the original data or losing any important information. The student average won’t determine the strong subject of the student.

Descriptive Statistical Analysis - II

- It won't tell you the specialty of the student or you won't come to know which subject was easy or strong. In spite of these limitations, Descriptive statistics can provide a powerful summary which may be helpful in comparisons across the various unit.
- **There are two types of statistics that are used to describe data:**
- **Measures of central tendency:** In this, a single value attempts to describe the data by using its central position with the given set. They are also classified as a summary set. In order to get the central value, they use averaging(mean), median or mode.
- **The measure of spread:** In this, the data is summarized by describing how well the data is spread out. For example, if the mean score of 100 students is 55 then there will be students whose score will be less than 55 or more than 55. Which means their score will be spread out in a way that their mean is 55. To describe the spread, we can use either of the statistical technique i.e. range, quartiles, variation, standard deviation, and absolute deviation.

Inferential Statistics

- The group of data that contains the information we are interested in is known as population.
- Inferential Statistics is used to make a generalization of the population using the samples. Where the sample is drawn from the population itself.
- It is necessary that the samples properly demonstrate the population and should not be biased. The process of achieving these kinds of samples is termed as sampling.
- Inferential Statistics comes from the fact that the sampling naturally incurs sampling errors and is thus not expected to perfectly represent the population.
- There are two types of Inferential Statistics method used for generalizing the data:
  - Estimating Parameters
  - Testing of Statistical Hypothesis
- The above two are the main types of statistical analysis.

Prescriptive Analysis

- “What should be done?” Prescriptive Analysis work on the data by asking this question.
- It is the common area of business analysis to identify the best possible action for a situation.
- Its whole idea is to provide advice that aims to find the optimal recommendation for a decision-making process.
- It is related to descriptive and predictive analysis.
- The descriptive analysis describes the data i.e. what has happened, and predictive analytics predicts what might happen prescriptive analysis find the best option among the available choice.
- Techniques used in the prescriptive analysis are simulation, graph analysis, business rules, algorithms, complex event processing, and machine learning.

Predictive Analysis

- “What might happen?” Predictive analysis is used to make a prediction of future events. It is based upon the current and historical facts.
- It uses statistical algorithm and machine learning techniques to determine the likelihood of future results, trends based upon historical and new data and behavior.
- Business is implementing predictive analytics to increase the competitive advantage and reduce the risk related to an unpredictable future.
- The main users of predictive analysis are marketing, financial service, online service providers and insurance companies. Techniques used in Predictive analysis are data mining, modeling, A.I., etc.

Casual Analysis

- “Why?” Casual Analysis helps in determining why things are the way they are.
- Since the current business world is full of events that might lead to failure, Casual Analysis seeks to identify the reason for it.
- It tries to get the root cause, i.e. the basic reason why something can happen.
- This is a common technique used in the IT industry for the quality assurance of the software. And industries that address major disasters.



Exploratory Data Analysis

- It is an Exponential to the inferential statistics and is mostly used by the data scientists.
- It is an analytical approach that focuses on identifying patterns in the data and figure out the unknown relationships.
- The purpose of Exploratory Data Analysis is to get check the missing data, find unknown relationships and check hypotheses and assumptions.
- It shouldn’t be used alone as it only provides a birds-eye view of the data and gets some insight into it.
- It is the first step in data analysis that should be performed before the other formal statistical techniques.



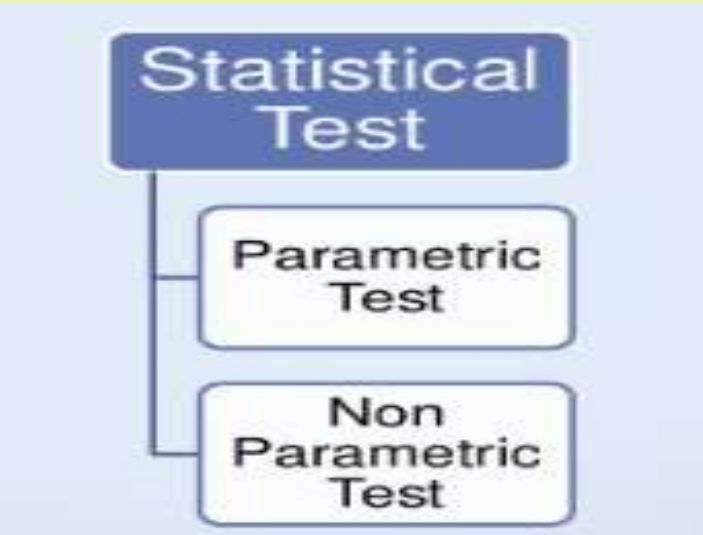
Mechanistic Analysis

- Mechanistic Analysis plays an important role in big industries.
- Though it is not among the common type of statistical analysis method still it's worth discussing.
- It is used for understanding the exact changes in the given variable that leads to the other variables.
- It works on the assumption that the given system gets affected by the interaction of its internal component.
- It does not consider external influence. It is useful in a system containing clear definitions like biological science.



Statistical Test

- These are intended to decide whether a hypothesis about distribution of one or more populations should be rejected or accepted.
- These may be:

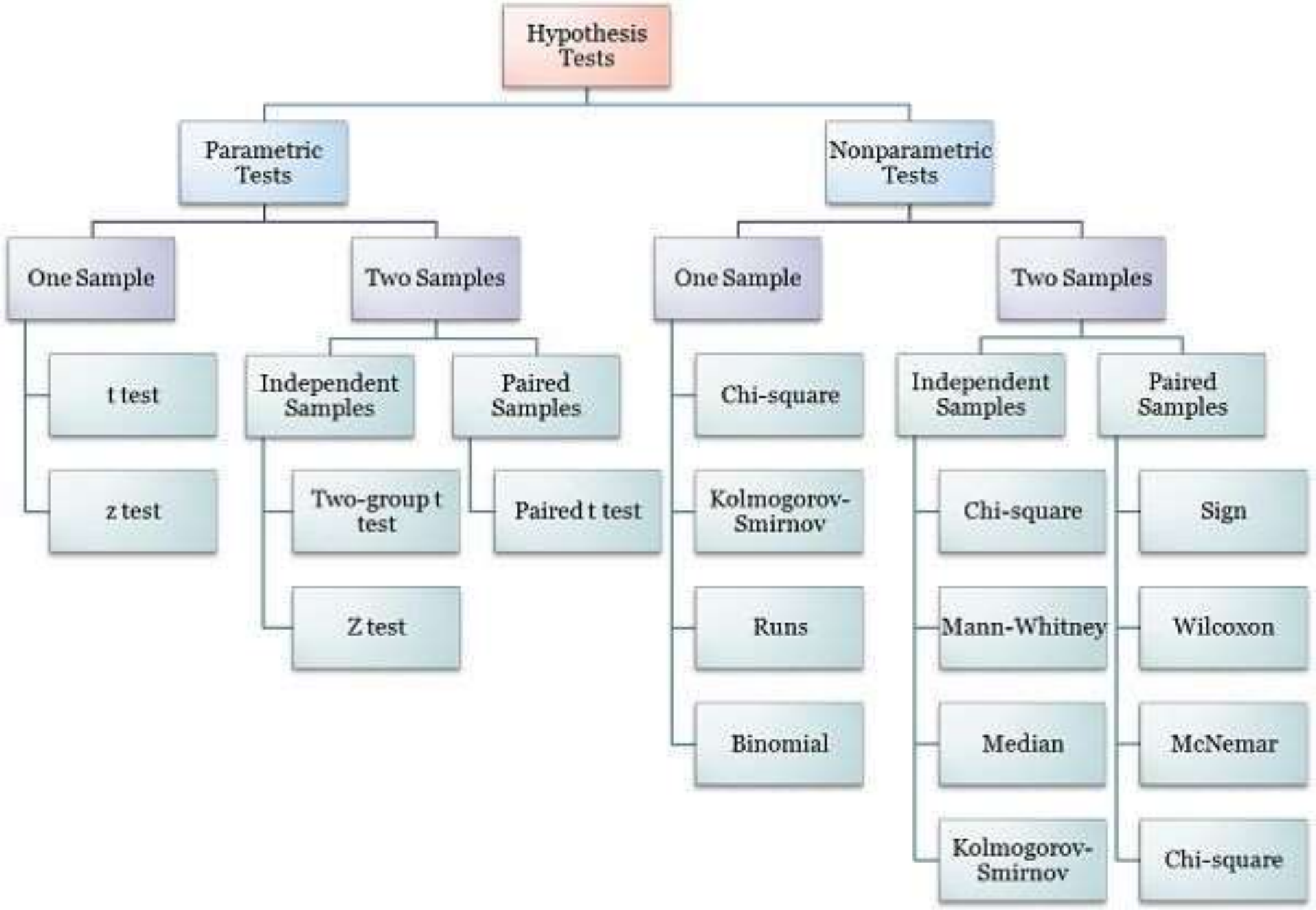


- These tests the statistical significance of the:
  - Difference in sample and population means.
  - Difference in two sample means.
  - Several population means.
  - Difference in proportions between sample and population.
  - Difference in proportions between two independent populations.
  - Significance of association between two variables.

System for Statistical Analysis

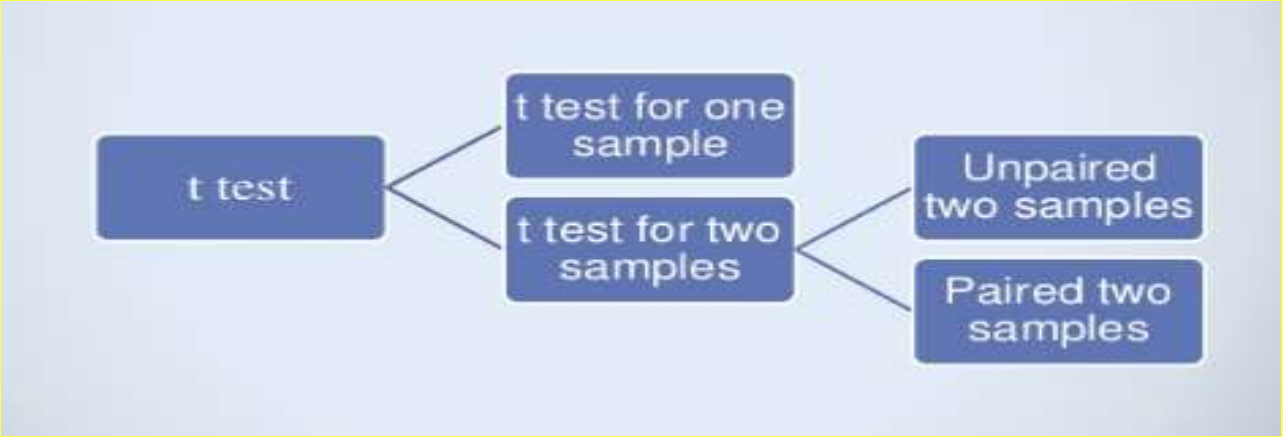
- State the Research Hypothesis.
- State the Level of Significance.
- Calculate the test statistic.
- Compare the calculated test statistic with the tabulated values.
- Decision.
- Statement of Result.
- **Parametric Tests:**
- Used for Quantitative Data.
- Use for continuous variables.
- Used when data are measured on approximate interval or ratio scales of measurement.
- Data should follow normal distribution.

Classification of Hypothesis Tests

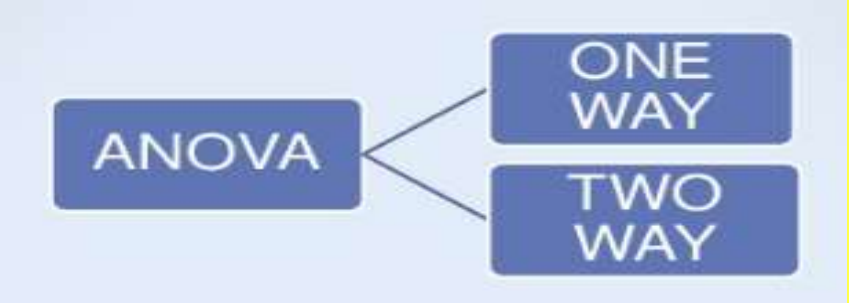


Parametric Tests

- t test (n<30)

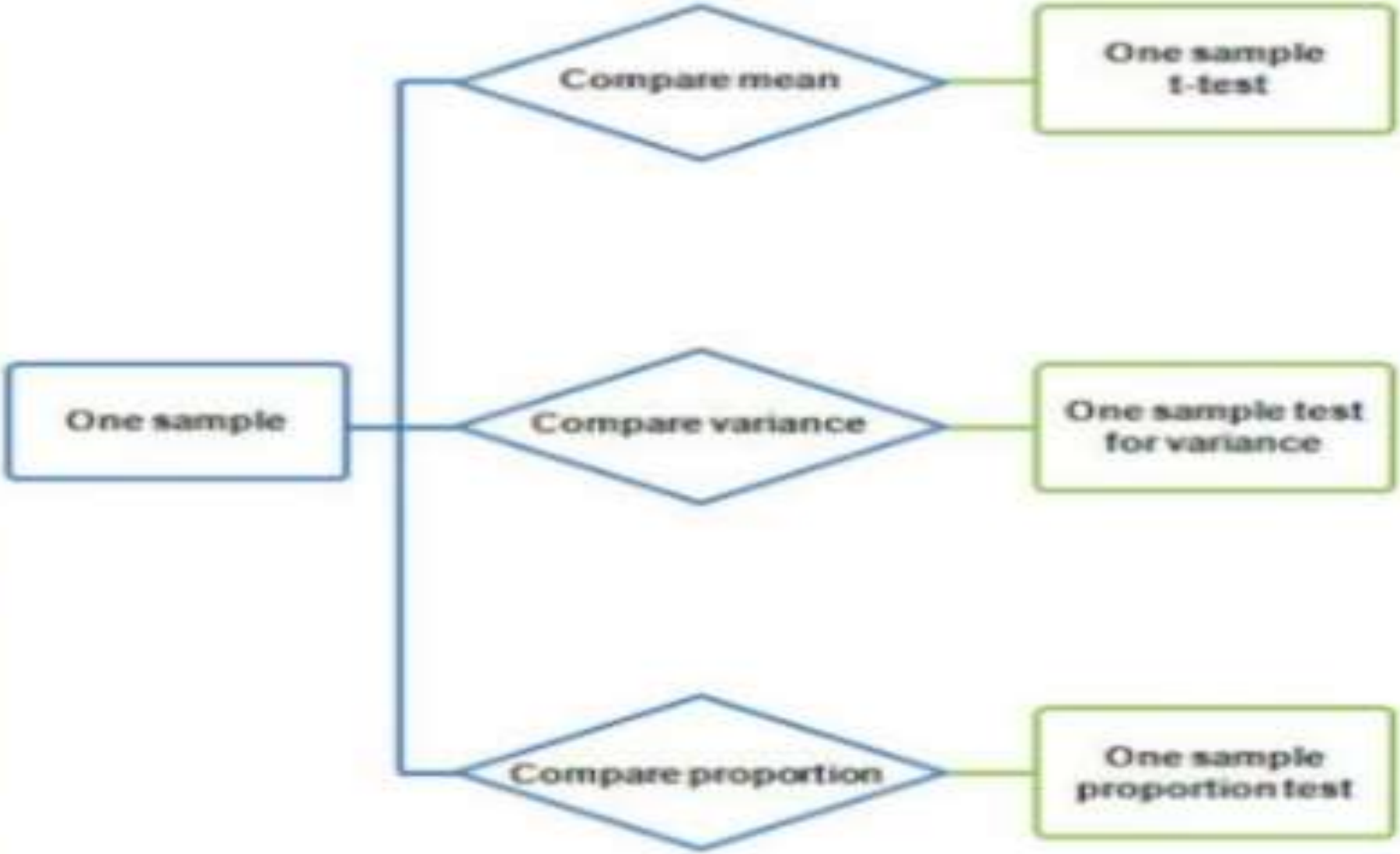


- ANOVA (Analysis of Variance)

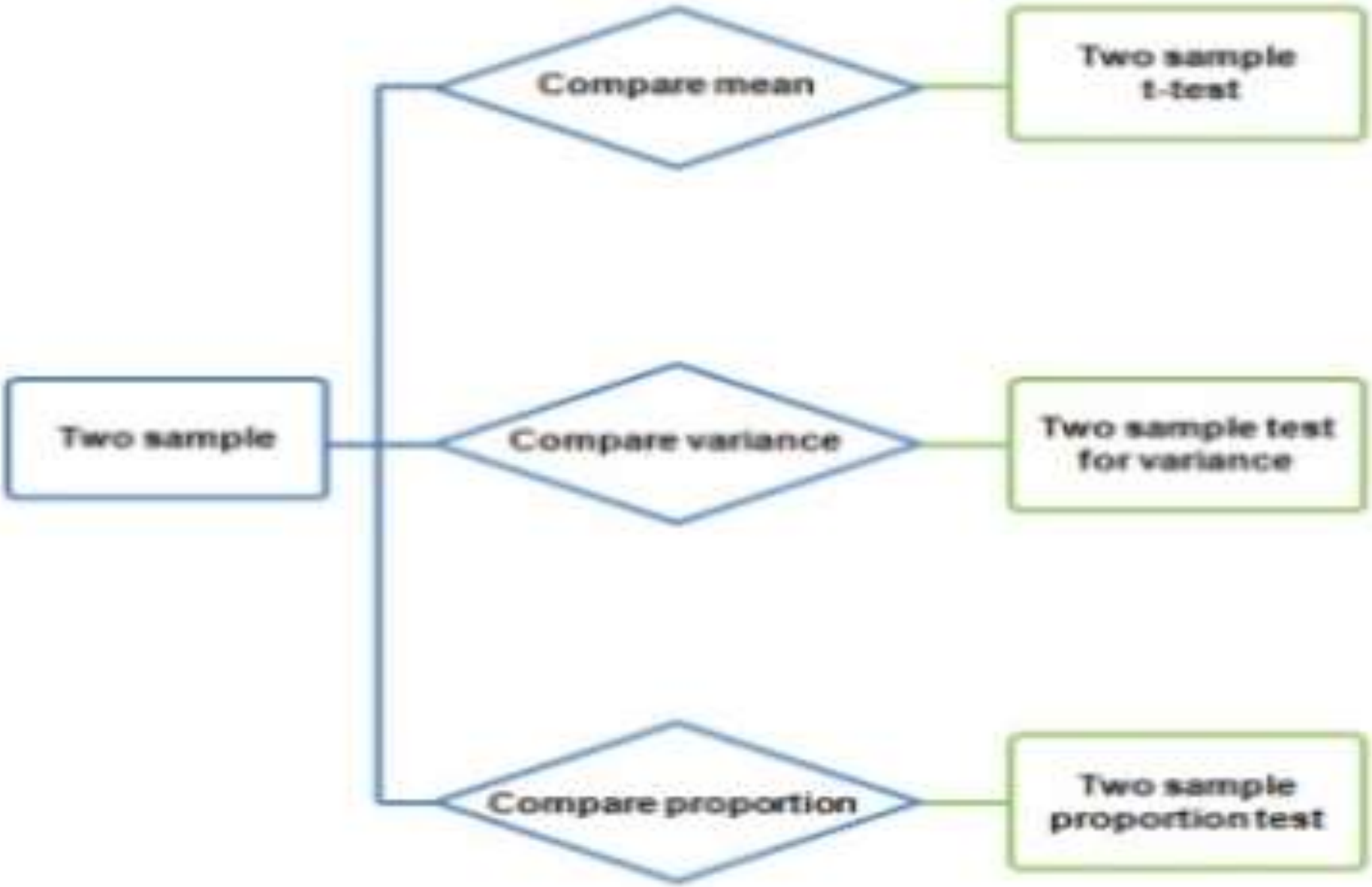


- Pearson's r Correlation
- Z test for large samples (n>30)

One Sample Tests



Two Sample Tests



Student's T-Test

- Developed by Prof W.S. Gossett in 1908, who published statistical papers under the pen name of 'student'. Thus the test is know as student's 't' test.
- **Indications for the test:**
- When samples are small
- Population variance are not known..
- Uses:
- Two means of small independent samples
- Sample mean and population mean
- Two proportions of small independent samples.

Assumption for t-test

- Samples are randomly selected .
- Data utilized is Quantitative .
- Variable follow normal distribution.
- Sample variances are mostly same in both the groups under the study .
- Samples are small, mostly lower than 30.
- A t-test compares the difference between two means of different groups to determine whether that difference is statistically significant.
- Student's 't' test for different purposes .
- 't' test for one sample .
- 't' test for unpaired two samples .
- 't' test for paired two samples.





One sample t-test

- When compare the mean of a single group of observations with a specified value.

Calculation

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

where $\bar{x}$ = sample mean

μ = Population mean

$S/\sqrt{n}$ = standard error

$$S^2 = \frac{\text{Sum } (x - \bar{x})^2}{n-1}$$

where x = element of sample

$\bar{x}$ = sample mean

$n-1$ = degrees of freedom

- In one sample t-test, we know the population mean. We draw a random sample from the population and then compare the sample mean with the population mean and make a statistical decision as to whether or not the sample mean is different from the population.

One sample t-test

- Now we compare calculated value with table value at certain level of significance (generally 5% or 1%).
- If absolute value of ‘t’ obtained is greater than table value then reject the null hypothesis and if it is less than table value, the null hypothesis may be accepted.

EXAMPLE

Research Problem : Comparison of mean dietary intake of a particular group of individuals with the recommended daily intake.

DATA: Average daily energy intake (ADEI) over 10 days of 11 healthy women

sub	1	2	3	4	5	6	7	8	9	10	11
ADEI (KJ)	5260	5470	5640	6180	6390	6515	6805	7515	7515	8230	8770

Mean ADEI value = 6753.6

SD ADEI value = 1142.1

When can we say about the energy intake of these women in relation to a recommended daily intake of 7725 KJ ?

Research Hypothesis

State null hypothesis and alternative hypothesis:

H_0 = there is no difference between population mean and sample mean

OR

$H_0 : \mu = 7725$ KJ

H_1 = there is a difference between population mean and sample mean

OR

$H_1 : \mu \neq 7725$ KJ

- Set the level of significance

  $\alpha = .05, .01$ or $.001$

- Calculate the value of proper statistic
- $t = \dfrac{\text{sample mean} - \text{hypothesized mean}}{\text{standard error of sample mean}}$

- $\dfrac{6753.6 - 7725}{1142.1 / \sqrt{11}} = -0.2564$

Rule of Rejecting Null Hypothesis

State the rule for rejecting the null hypothesis:

Reject H_0 if $t \geq$ +ve Tabulated value

OR

Reject H_0 if $t \leq$ - ve Tabulated value
Or we can say that $p<.05$

In the above example we have seen

t=- .2564 which is less then 2.23

P value suggests that the dietary intake of these women was significantly less than the recommended level (7725 KJ)

Two Sample t-test

- **Unpaired Two sample ‘t’- test**
- Unpaired t- test is used when we wish to compare two means.
- Used when the two independent random samples come from the normal populations having unknown or same variance.
- We test the null hypothesis, that the two population means are same i.e. $\mu 1 = \mu 2$ against an appropriate one sided or two sided alternative hypothesis.
- **Assumptions**
  - The samples are random & independent of each other.
  - The distribution of dependent variable is normal.
  - The variances are equal in both the groups



Two Sample t-test

FORMULA

Test statistic is given by

$$t = \frac{\text{Mean1} - \text{Mean2}}{\text{SE (Mean1 –mean2)}}$$

$$\text{SE(Mean1 –mean2)} = S \sqrt{[1/n_1 + 1/n_2]}$$

$$\mathbf{S = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1+n_2-2)}}$$

Where S_1^2 and S_2^2 are respectively called SD's of first and second group

Research Problem

- Research Problem A study was conducted to compare the birth weights of children born to 15 non-smoking with those of children born to 14 heavy smoking mothers.

Non-smoking mothers (n=15)	Heavy smoking mothers (n=14)
3.99	3.18
3.79	2.84
3.60	2.90
3.73	3.27
3.21	3.85
3.60	3.52
4.08	3.23
3.61	2.76
3.83	3.60
3.31	3.75
4.13	3.59
3.26	3.63
3.54	2.38
3.51	2.34
2.71	

Research Problem

- Research Hypothesis : State null hypothesis and alternative hypothesis
- H_0 = there is no difference between the birth weights of children born to non-smoking and smoking mothers
- H_1 = there is a difference between the birth weights of children born to non-smoking and smoking mothers
- Set the level of significance
  $\alpha = .05, .01$ or $.001$
- Calculate the value of proper statistic
- State the rule for rejecting the null hypothesis
- If $t_{cal} > t_{tab}$ we can say that $P < .05$ then we reject the null hypothesis and accept the Alternative hypothesis.

Decision

- If we reject the null hypothesis so we can say that children born to non-smokers are heavier than children born to heavy smokers.

Paired Two Sample t-test

- Used when we have paired data of observations from one sample only, when each individual gives a pair of observations.
- Same individuals are studied more than once in different circumstances- measurements made on the same people before and after interventions.
- **Assumptions**
- The outcome variable should be continuous.
- The difference between pre-post measurements should be normally distributed.

FORMULA

$$t = \frac{\bar{d}}{SD/\sqrt{n}}$$

Where,

d = difference between x_1 and x_2

$\bar{d}$ = Average of d

SD = Std. deviation for the difference

n = sample size

Research Problem

- A study was carried to evaluate the effect of the new diet on weight loss. The study population consist of 12 people have used the diet for 2 months; their weights before and after the diet are given 33 R.

Patient no.	Weight (Kgs)	
	Before Diet	After Diet
1	75	70
2	60	54
3	68	58
4	98	93
5	83	78
6	89	84
7	65	60
8	78	77
9	95	90
10	80	76
11	100	94
12	108	100

Research Problem

Research Hypothesis

- State null hypothesis and alternative hypothesis
- H_0 = There is no reduction in weight after Diet
- H_1 = There is reduction in weight after Diet
- Further Analysis through Statistical software SPSS as same as previous example

Decision

- If we reject the null hypothesis then there is a statistically significant reduction in weight

How do we compare more than two groups means ??

Example :

Treatments : A, B, C & D

Response : BP level

How does t-test concept work here ?

A versus B	B versus C
A versus C	B versus D
A versus D	C versus D

so the chance of getting the wrong result would be:

$1 - (0.95 \times 0.95 \times 0.95 \times 0.95) = 26\%$

Analysis of Variance (ANOVA) - I

- The principle aim of statistical models is to explain the variation in measurements.
- The statistical model involving a test of significance of the difference in mean values of the variable between two groups is the student's, 't' test.
- If there are more than two groups, the appropriate statistical model is Analysis of Variance (ANOVA).
- **Assumptions for ANOVA**
- Sample population can be easily approximated to normal distribution.
- All populations have same Standard Deviation.
- Individuals in population are selected randomly.
- Independent samples

Analysis of Variance (ANOVA) - II

- ANOVA compares variance by means of a simple ratio, called F-Ratio.

$$F = \frac{\text{Variance between groups}}{\text{Variance within groups}}$$

- The resulting F statistics is then compared with critical value of F (critic), obtained from F tables in much the same way as was done with ‘t’.
- If the calculated value exceeds the critical value for the appropriate level of α, the null hypothesis will be rejected.
- A F test is therefore a test of the Ratio of Variances F Tests can also be used on their own, independently of the ANOVA technique, to test hypothesis about variances.
- In ANOVA, the F test is used to establish whether a statistically significant difference exists in the data being tested.

Analysis of Variance (ANOVA) - III



- One Way ANOVA: If the various experimental groups differ in terms of only one factor at a time- a one way ANOVA is used.
- Example: A study to assess the effectiveness of four different antibiotics on S Sanguis
- Two Way ANOVA: If the various groups differ in terms of two or more factors at a time, then a Two Way ANOVA is performed.
- Example: A study to assess the effectiveness of four different antibiotics on S Sanguis in three different age groups

Pearson's Correlation Coefficient

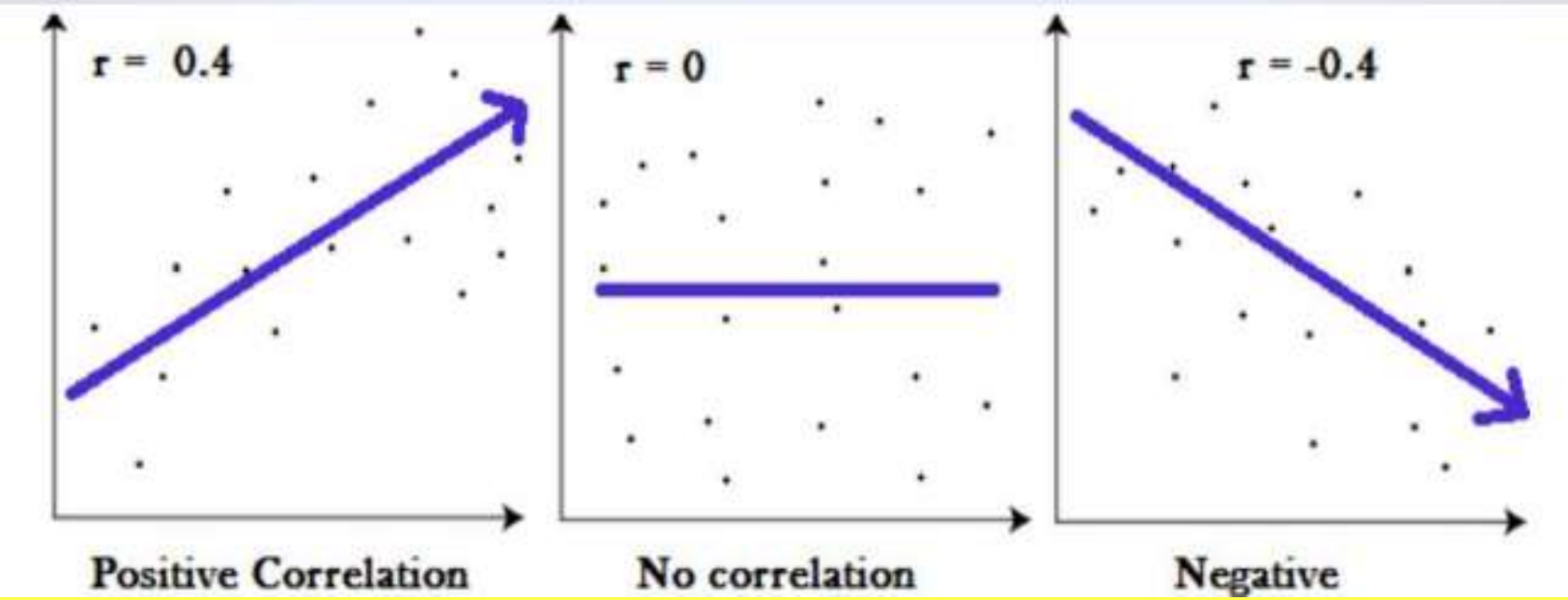
- Correlation is a technique for investigating the relationship between two quantitative, continuous variables.
- Pearson's Correlation Coefficient(r) is a measure of the strength of the association between the two variables.
- Assumptions Made in Calculation of 'r'
- Subjects selected for study with pair of values of X & Y are chosen with random sampling procedure.
- Both X & Y variables are continuous.
- Both variables X & Y are assumed to follow normal distribution

Pearson's Correlation Coefficient Steps

- The first step in studying the relationship between two continuous variables is to draw a scatter plot of the variables to check for linearity.
- The correlation coefficient should not be calculated of the relationship is not linear.
- For correlation only purposes, it does not matter on which axis the variables are plotted.
- However, conventionally, the independent variable is plotted on X axis and dependent variable on Y-axis.
- The nearer the scatter of points is to a straight line, the higher the strength of association between the variables.

Types of Correlation

- Perfect Positive Correlation r = +1
- Partial Positive Correlation 0 < r < + 1
- Perfect negative correlation r = - 1
- Partial negative correlation 0 > r > - 1
- No Correlation



Z - test

- This test is used for testing significance difference between two means (n>30).
- **Assumptions to apply Z test:**
- The sample must be randomly selected.
- Data must be quantitative.
- Samples should be larger than 30.
- Data should follow normal distribution.
- Sample variances should be almost the same in both the groups of study.
- If the SD of the populations is known, a Z test can be applied even if the sample is smaller than 30.
- **Indications for Z Test:**
- To compare sample mean with population mean.
- To compare two sample means.
- To compare sample proportion with population proportion.
- To compare two sample proportions.

Steps in Z - test

- Define the problem.
- State the null hypothesis (H0) & alternate hypothesis (H1).
- Find Z value.

$$Z = \frac{\text{Observed mean} - \text{Mean}}{\text{Standard Error}}$$

- Fix the level of significance.
- Compare calculated Z value with the value in Z table at corresponding degree significance level.
- If the observed Z value is greater than theoretical Z value, Z is significant, reject null hypothesis and accept alternate hypothesis.

Z – Proportionality Test

- Used for testing the significant difference between two proportions,

$$z = \frac{P_1 - P_2}{SE(P_1 - P_2)}$$

- Where, P1 = Prop. rate for 1st population.

  P2 = Prop. rate for 2nd population.
- Where, SE (P1 - P2) is defined SE of difference.
- **One tailed and Two tailed Z tests**
- Z values on each side of mean are calculated as +Z or as -Z.
- A result larger than difference between sample mean will give +Z and result smaller than the difference between mean will give –Z.

Example for one and two tailed test

- **Example for Two Tailed:**

  In a test of significance, when one wants to determine whether the mean IQ of malnourished children is different from that of well nourished and does not specify higher or lower, the P value of an experiment group includes both sides of extreme results at both ends of scale, and the test is called two tailed test.

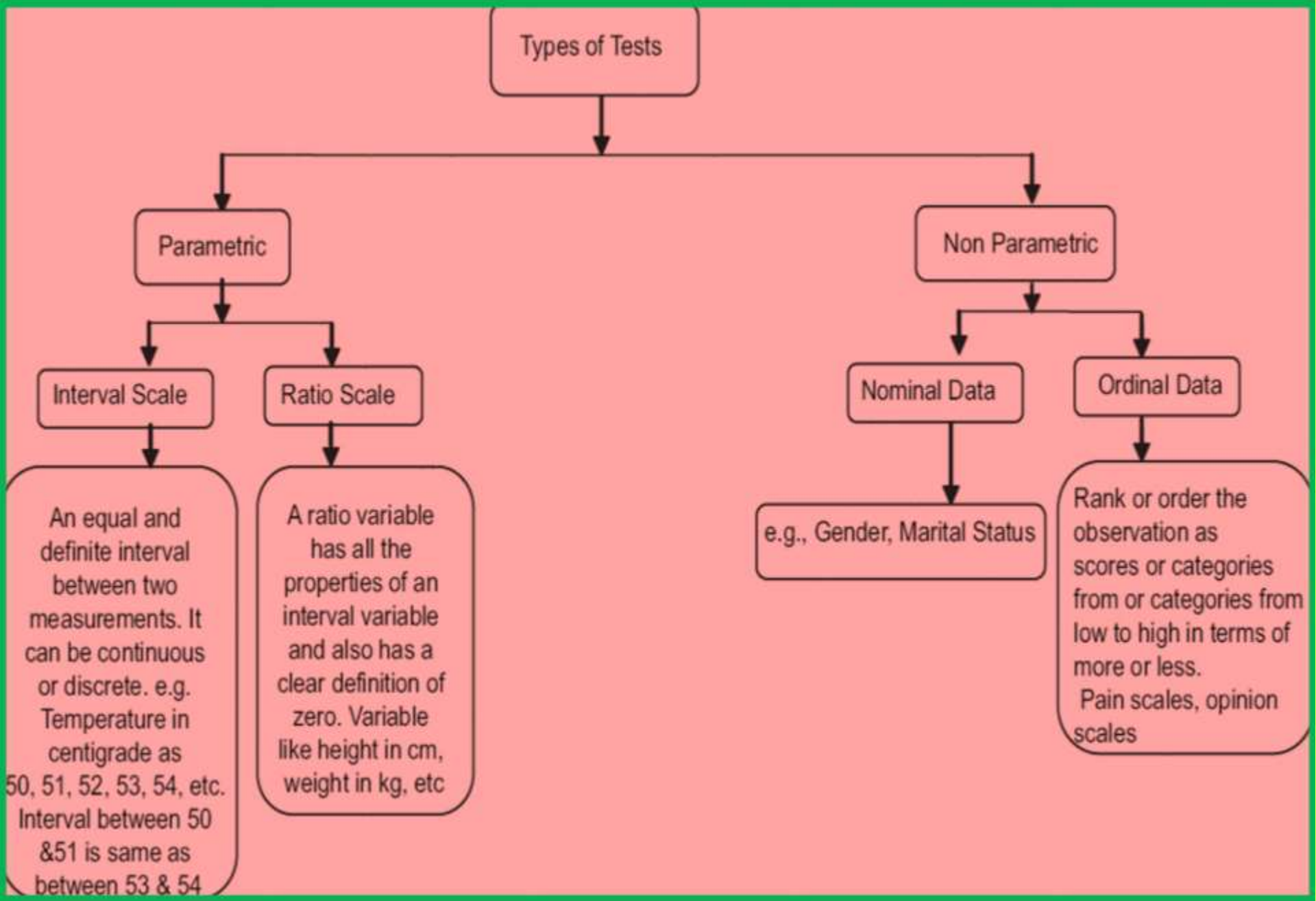
- **Example for single tailed:**

  In a test of significance when one wants to know specifically whether a result is larger or smaller than what occur by chance, the significant level or P value will apply to relative end only e.g. if we want to know if the malnourished have lesser mean IQ than the well nourished, the result will lie at one end (tail)of the distribution, and the test is called single tailed test.

Non Parametric Test

- A non parametric test (sometimes called a distribution free test) does not assume anything about the underlying distribution (for example, that the data comes from a normal distribution). It usually means that you know the population data does not have a normal distribution.
- Nonparametric statistics refers to a statistical method in which the data are not assumed to come from prescribed models that are determined by a small number of parameters; examples of such models include the normal distribution model and the linear regression model.
- Nonparametric tests are also called distribution-free tests because they don't assume that your data follow a specific distribution. You may have heard that you should use nonparametric tests when your data don't meet the assumptions of the parametric test, especially the assumption about normally distributed data.
- Non-parametric tests are experiments which do not require the underlying population for assumptions. It does not rely on any data referring to any particular parametric group of probability distributions. Non-parametric methods are also called distribution-free tests since they do not have any underlying population.

Types of Tests



Difference between Parametric and Non Parametric Test

Properties	Parametric	Non-Parametric
Assumptions	Yes	No
Central tendency value	Mean Value	Median Value
Correlation	Pearson	Spearman
Probabilistic distribution	Normal	Arbitrary
Population Knowledge	Requires	Does not require
Used for	Interval Data	Nominal Data
Applicability	Variables	Attributes & Variables
Examples	z-test, t-test, etc.	Kruskal-Wallis, Mann-Whitney

Difference between Parametric and Non Parametric Test

S. No.	Parametric	Non-Parametric
1.	It can be used on ordinal and nominal scale data (although also on interval and ration scale).	Used mainly on interval and ratio scale data.
2.	It can be used on small samples.	Tend to need larger samples.
3.	It can be used on data that are not normally distributed.	Data should fit a particular distribution; the data can be transformed to that distribution.
4.	It can be used where the samples are not selected randomly.	Samples should be drawn randomly from the population.
5.	Have less power than the equivalent parametric test.	More powerful than non-parametric equivalent.

Difference between Parametric and Non Parametric Test

- The key difference between parametric and nonparametric test is that the parametric test relies on statistical distributions in data whereas nonparametric do not depend on any distribution. Non-parametric does not make any assumptions and measures the central tendency with the median value. Some examples of Non-parametric tests includes Mann-Whitney, Kruskal-Wallis, etc.
- Parametric is a statistical test which assumes parameters and the distributions about the population is known. It uses a mean value to measure the central tendency. These tests are common, and therefore the process of performing research is simple.
- **Note:** **Nonparametric test do not depend on any distribution, hence it is a kind of robust test and have broader range of situations.**
- **Parametric test is completely dependent on statistical data and have more chances of accuracy.**
- **For parametric mean value is taken and for non-parametric test median value is taken into consideration.**
- **T-test and Z-test are the examples of parametric test, in statistics.**
- **Kruskal-Wallis and Mann-Whitney are the examples of Non Parametric Test.**

Definition of Parametric and Non Parametric Test

- **Parametric Test Definition**
- In Statistics, a parametric test is a kind of the hypothesis test which gives generalizations for generating records regarding the mean of the primary/original population. The t-test is carried out based on the students t-statistic, which is often used in that value.
- The t-statistic test holds on the underlying hypothesis which includes the normal distribution of a variable. In this case, the mean is known, or it is considered to be known. For finding the sample from the population, population variance is identified. It is hypothesized that the variables of concern in the population are estimated on an interval scale.
- **Non-Parametric Test Definition**
- The non-parametric test does not require any population distribution, which is meant by distinct parameters. It is also a kind of hypothesis test, which is not based on the underlying hypothesis.
- In the case of the non-parametric test, the test is based on the differences in the median. So, this kind of test is also called a distribution-free test. The test variables are determined on the nominal or ordinal level. If the independent variables are non-metric, the non-parametric test is usually performed.

T-Test Formula

The t-test is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis. It can be used to determine if two sets of data are significantly different from each other, and is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known.

T-test uses means and standard deviations of two samples to make a comparison. The formula for T-test is given below:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{\Delta}}}$$

where

$$s_{\bar{\Delta}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where,

$\overline{x}$ = Mean of first set of values

$\overline{x}_2$ = Mean of second set of values

S_1 = Standard deviation of first set of values

S_2 = Standard deviation of second set of values

n_1 = Total number of values in first set

n_2 = Total number of values in second set.

The formula for standard deviation is given by:

$$S = \sqrt{\frac{\sum (x - \overline{x})^2}{n - 1}}$$

Where,

x = Values given

$\overline{x}$ = Mean

n = Total number of values.

Hypothesis Testing Formula

We run a hypothesis test that helps statisticians determine if the evidence are enough in a sample data to conclude that a research condition is true or false for the entire population. For finding out hypothesis of a given sample, we conduct a Z-test. Usually, in Hypothesis testing, we compare two sets by comparing against a synthetic data set and idealized model.

The Z test formula is given as:

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

Where,

$\overline{x}$ is the sample mean

μ is the population mean

σ is the standard deviation and n is the sample size.

Hypothesis Testing Example

Solved Examples

Question: What will be the z value when the given parameters are sample mean = 600, population mean = 585, the standard deviation is 100 and the sample size is 150?

Solution:

Given parameters are,

Sample mean, $\bar{x}$ = 600

Population mean, μ = 585,

Standard deviation, σ = 100

Sample size, n = 150

The formula for hypothesis testing is given as,

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

$$z = \frac{600 - 585}{\frac{100}{\sqrt{150}}}$$

=1.837

Chi-Square Test

- A **chi-squared test** is basically a data analysis on the basis of observations of a random set of variables.
- Usually, it's a comparison of two statistical data sets. It is also represented as X^2 test.
- This test was introduced by Karl Pearson in 1900 for categorical data analysis and distribution. So it was mentioned as Pearson's chi-squared test.
- With the assumption of the null hypothesis as true, this test is used to estimate how likely the observations that are made would be.
- A hypothesis is a consideration, that a given condition or statement might be true, which we can test afterwards.
- Chi-squared tests are usually created from a sum of squared falsities or errors else via the sample variance.

Chi-Square Distribution

- When we consider, the null speculation as true, the sampling distribution of the test statistic is called as chi-squared distribution.
- The chi-squared test helps to determine whether there is a notable difference between the normal frequencies and the observed frequencies in one or more classes or categories. It gives the probability of independent variables.
- Note: Chi-squared test is applicable only for categorical data, such as men and women falling under the categories of Gender, Age, Height, etc.
- Probability is all about chance or risk or uncertainty. It is the possibility of the outcome of the sample or the occurrence of an event. But when we talk about statistics, it is more about how we handle various data using different techniques.
- It helps to represent complicated data or bulk data in a very easy and understandable way. It describes the collection, analysis, interpretation, presentation, and organization of data. The concept of both probability and statistics is related to the chi-squared test.

Chi-Square Test Formula

The chi-squared test is done to check if there is any difference between the observed value and expected value. The formula for chi-square can be written as;

$$X^2 = \sum \frac{(Observed\ Value - Expected\ Value)^2}{Expected\ Value}$$

- **Chi-Square Test of Independence**
- The chi-square test of independence also known as the chi-square test of association which is used to determine the association between the categorical variables. It is considered as a non-parametric test. It is mostly used to test statistical independence. The chi-square test of independence is not appropriate when the categorical variables represent the pre-test and post-test observations. For this test, the data must meet the following requirements:
- Two categorical variables.
- Relatively large sample size.
- Categories of variables (two or more).
- Independence of observations.

Advantages, Disadvantages and Application of Non-Parametric Test

- **Advantages and Disadvantages of Non-Parametric Test**
- The advantages of the non-parametric test are:
- Easily understandable.
- Short calculations.
- Assumption of distribution is not required.
- Applicable to all types of data.

- **Disadvantages of the non-parametric test are:**
- Less efficient as compared to parametric test.
- The results may or may not provide an accurate answer because they are distribution free.

- **Applications of Non-Parametric Test are:**
- The conditions when non-parametric tests are used are listed below:
- When parametric tests are not satisfied.
- When testing the hypothesis, it does not have any distribution.
- For quick data analysis.
- When unscaled data is available.

Multivariate Data Analysis

- Multivariate data analysis is a set of statistical models that examine patterns in multidimensional data by considering, at once, several data variables.
- As multivariate models consider more variables, they can examine more complex phenomena and find data patterns that more accurately represent the real world.
- **Example:** A researcher has collected data on three psychological variables, four academic variables (standardized test scores), and the type of educational program the student is in for 600 high school students. A doctor has collected data on cholesterol, blood pressure, and weight.
- Multivariate Techniques help determine the relationship between different multiple variables in a quantifiable way.
- Through Multivariate Techniques one can determine how one variable affects the other & finally what effects the variables will have on the final outcome.

Multivariate Data Analysis Techniques

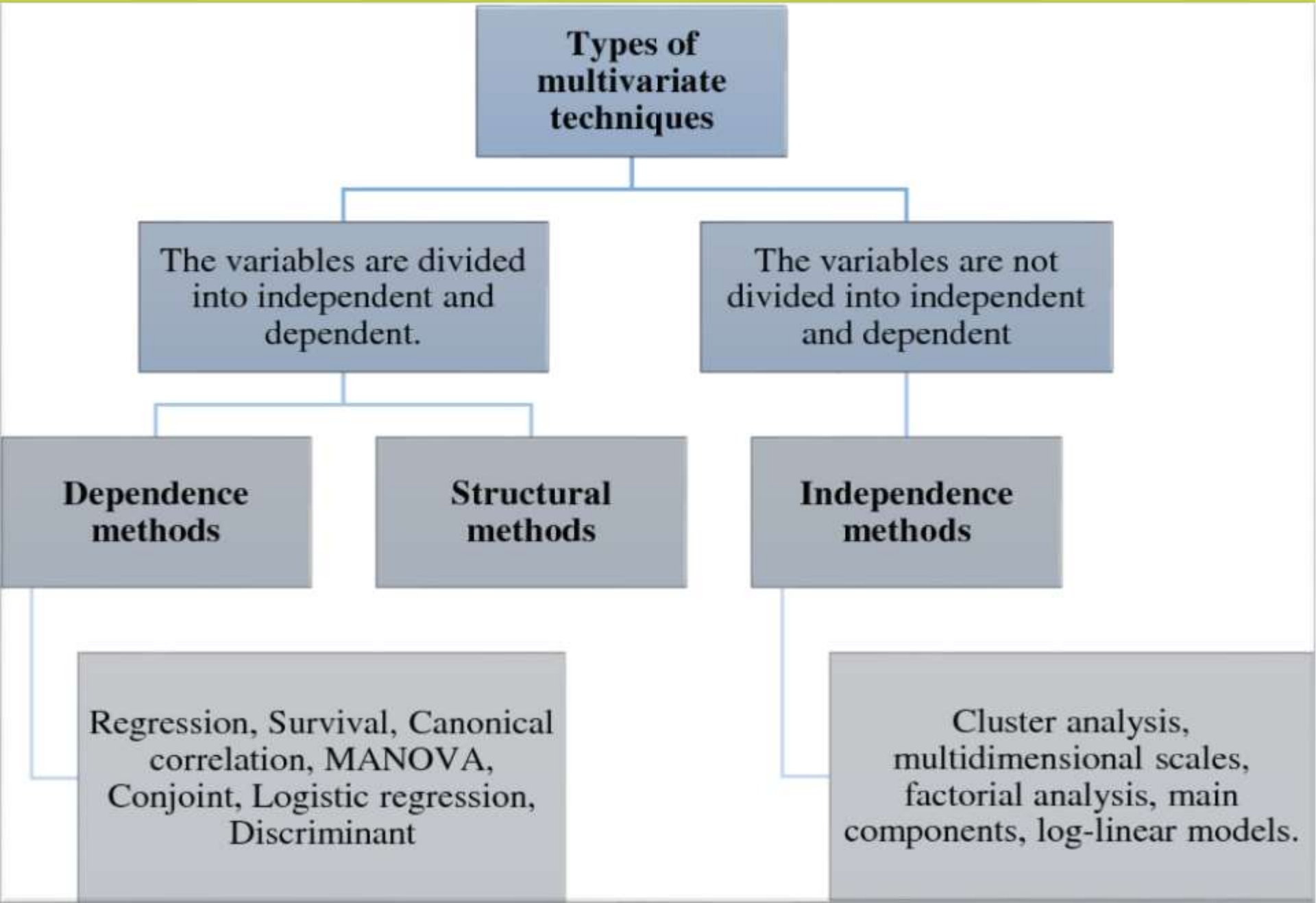
- There are many techniques of Multivariate Analysis starting with quality of the data to structural equation modeling, each one of the techniques has its own purpose, and are used depending on the data and the type of outcome realized by the data analyst. These techniques provide statistical data given a specific data set but requires caution when interpreting and putting them to use remember as I always say people do the most important part than what technology does for us.
- **The techniques are as follows:**
  - ✓ Multiple Regression Analysis
  - ✓ Discriminated Analysis
  - ✓ Multivariate Analysis of Variance (MANOVA)
  - ✓ Factor Analysis
  - ✓ Cluster Analysis
  - ✓ Canonical Correlation
  - ✓ Classification Analysis
  - ✓ Principal Component Analysis

Types of Multivariate Data Analysis Techniques

- Multivariate analysis is an ever-expanding set of techniques for data analysis.
- **More established techniques:**
  - Principal component and common factor analysis.
  - Multiple regression and multiple correlation.
  - Multiple discriminated analysis.
  - Multivariate analysis of variance and covariance.
  - Conjoint analysis.
  - Canonical correlation
  - Cluster analysis.
  - Multidimensional scaling.
- **Emerging Techniques:**
  - Correspondence analysis.
  - Linear probability models.
  - Structural equation modeling.



Types of Multivariate Data Analysis Techniques



Objectives of Multivariate Analysis

- **Data reduction or structural simplification:** The phenomenon being studied is represented as simply as possible without sacrificing valuable information. It is hoped that this will make interpretation easier.
- **Sorting and Grouping:** Groups of " similar " objects or variables are created, based upon measured characteristics. Alternatively, rules for classifying objects into well-defined groups may be required.
- **Investigation of dependence among variables:** The nature of the relationships among variables is of interest. Are all the variables mutually independent or are one or more variables dependent on the others? If so, how?
- **Prediction:** Relationships between variables must be determined for the purpose of predicting the values of one or more variables on the basis of observations on the other variables.
- **Hypothesis construction and testing:** Specific statistical hypotheses, formulated in terms of the parameters of multivariate populations, are tested, this may be done to validate assumptions or to reinforce prior convictions.

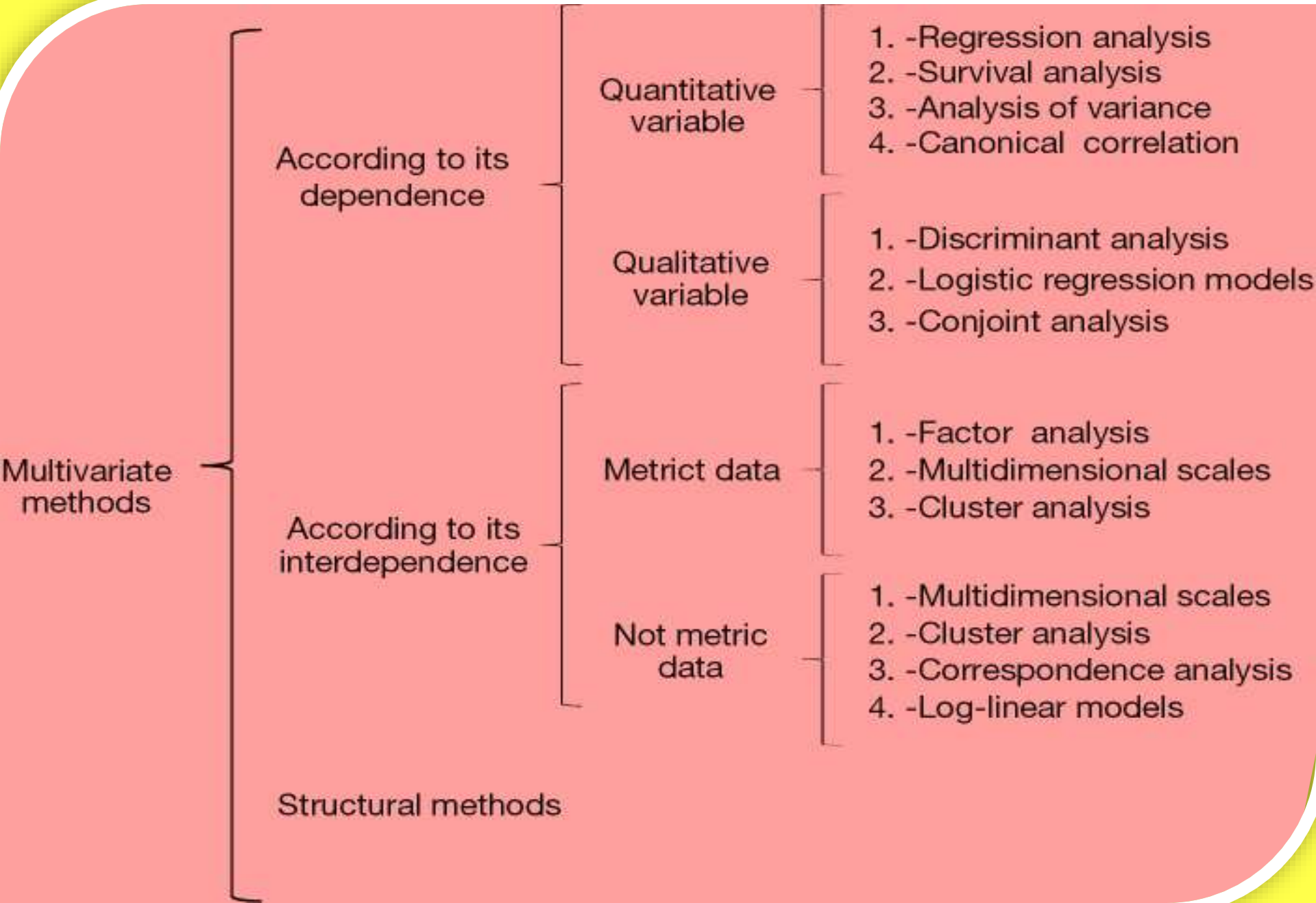
Concepts of Multivariate Data Analysis

- The variate (a linear combination of variables with weights)
- Measurement Scale
  - Non-metric Measurement Scales
    - ✓ Nominal and Ordinal scales
  - Metric Measurement Scales
    - ✓ Interval and ration scales
- Measurement Error and multivariate measurement
    - ✓ Validity and reliability
- Statistical Significance Versus Statistical Power
    - ✓ Type I error (alpha)
    - ✓ Type II error (beta)
    - ✓ Power: Effect size, Alpha, Sample Size.





Multivariate Methods



Data Analysis Matrix

Study Objectives	Types of Variables	Analytic Procedures
To Describe	One Variable (Neither independent or dependent)	Univariate
To Compare	One independent and one dependent variable.	Bi-variate
To Analyze the relative importance	Two or more independent / control variables and one dependent variable	Multivariate

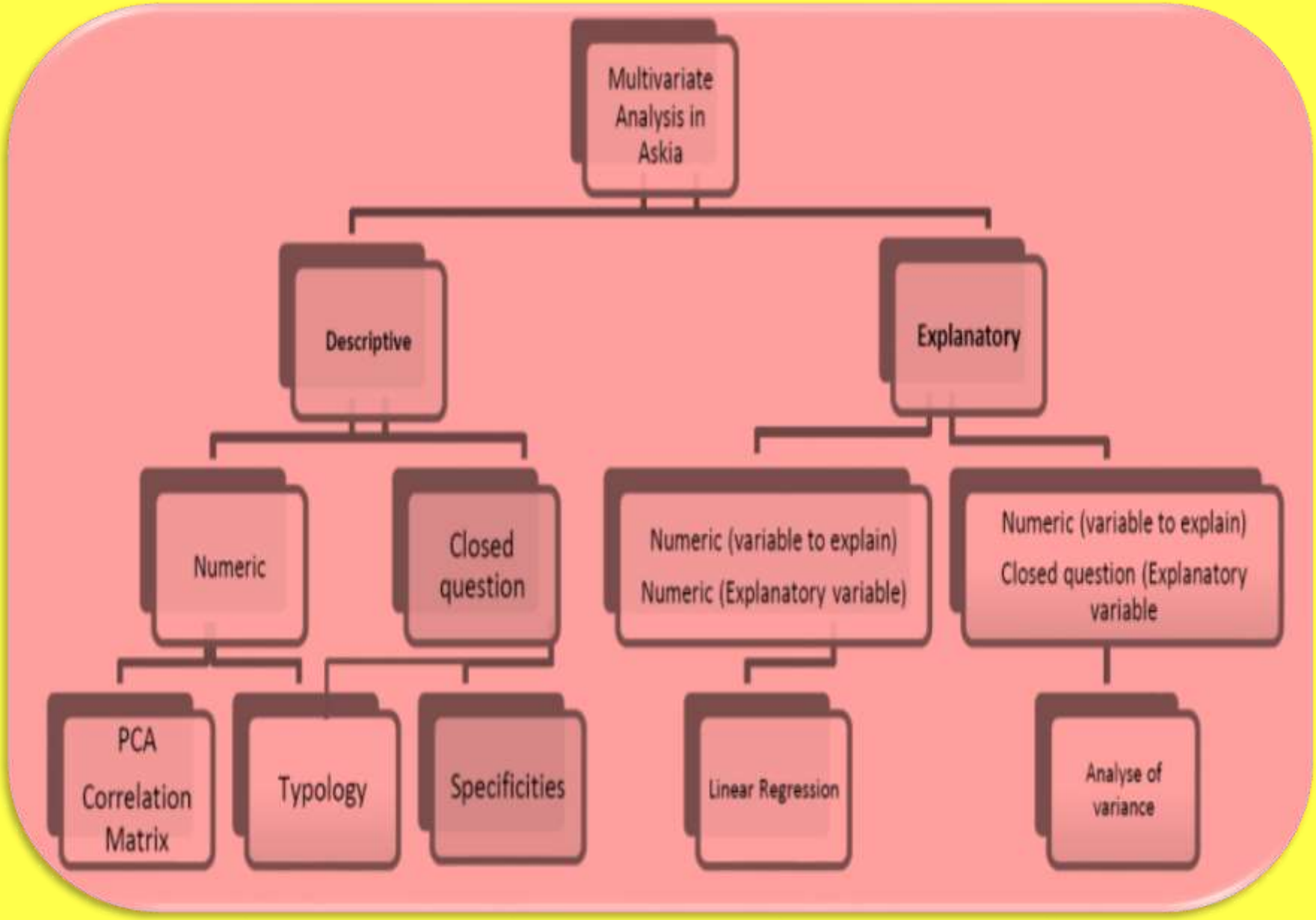
Meaning of Univariate, Bivariate & Multivariate Analysis of Data

- **Univariate Analysis:** In univariate analysis, one variable is analyzed at a time.
- **Bivariate Analysis:** In bivariate analysis two variables are analyzed together and examined for any possible association between them.
- **Multivariate Analysis:** In multivariate analysis, the concern is to analyze more than two variables at a time.
- The type of statistical techniques used for analyzing univariate and bivariate data depends upon the level of measurements of the questions pertaining to those variables.
- Further, the data analysis could be of two types, namely, descriptive and inferential.

Types of Multivariate Models

- **Cross Correlations:** An important exploratory tool for modeling multivariate time series is the cross correlation function (CCF) . The CCF generalizes the ACF to the multivariate case. Thus, its main purpose is t find linear dynamic relationships in time series data that have been generated from stationary processes.
- **Single-equation models:** Two types of so-called single-equation models can be considered for multivariate forecasting: regression models and transfer-function models.
- **Vector auto regressions and VARMA models**
- **Co-integration:** Although VAR modeling traditionally assumes stationary of all series, it is not generally recommended to difference non-stationary components individually, as such a step may destroy important dynamic information.

Types of Multivariate Analysis



Types of Multivariate Tests

- Multivariate Analysis of Variance (MNOVA)
- Discriminant Function Analysis (DFA)
- Principal Components Analysis (PCA)
- Factor Analysis
- Cluster Analysis
- Canonical Correlation Analysis
- Multidimensional Scaling
- Conjoint Analysis
- Multiple Regression Analysis



Types of Multivariate Techniques

First Generation Technique	Cluster Analysis
	Multidimensional Scaling (MDS)
	Logistic Regression
	Multiple Regression
	Multivariate Analysis of Variance (MANOVA)
Second Generation Technique	Structural Equation Model, SEM - Covariance Based - Partial Least Square

Types of Multivariate Techniques - I

S. No.	Types of Analysis	Types of Variable/Test Used	Purpose
1.	Univariate	**Continuous:** Mean, Median, Standard Deviation, Histogram	**Outcome Variable:** To assess normal distribution. **Exposure Variable:** To examine distribution, missing variables, etc.
2.	Univariate	**Categorical:** Frequency Distribution.	**Outcome Variable:** To assess frequency. **Exposure Variables:** To assess frequency (are there enough observations in each category?), missing variables.
3.	Bivariate: For Exposure Groups	**Continuous:** t-test between exposure groups. **Categorical:** Chi-square test.	To assess differences between groups prior to analysis. To look for possible confounding relationships.

Types of Multivariate Techniques - II

S. No.	Types of Analysis	Types of Variable/Test Used	Purpose
4.	Bivariate for outcome groups	**Continuous:** t-test **Categorical:** Odds ratio.	To look for significant differences in the outcome variable by exposure variables 'Crude' analysis.
5.	Multivariate: for continuous variables	Linear regression analysis	To examine the relationship between all the exposure variables and the outcome variable controlling for all the variables in the model High r^2 Desired.
6.	Multivariate for binary (yes/no) outcomes	Logistic regression analysis	To examine the relationship between all the exposure variables and the outcome variable controlling for all the variables in the model 'Adjusted' analysis.

The Multiple Regression Model

Idea: Examine the linear relationship between
1 dependent (y) & 2 or more independent variables (x_i)

Population model:

Y-intercept | Population slopes | Random Err

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$$

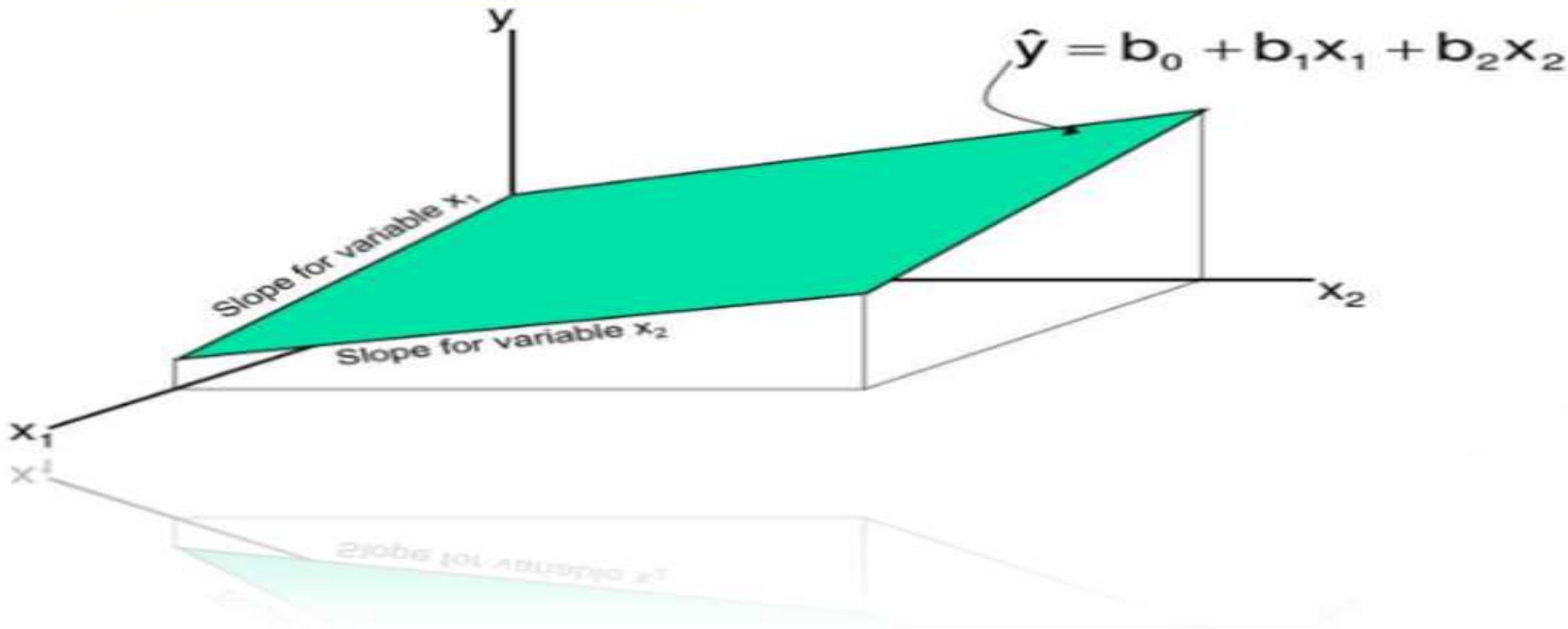
Estimated multiple regression model:

Estimated (or predicted) value of y | Estimated intercept | Estimated slope coefficients

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k$$

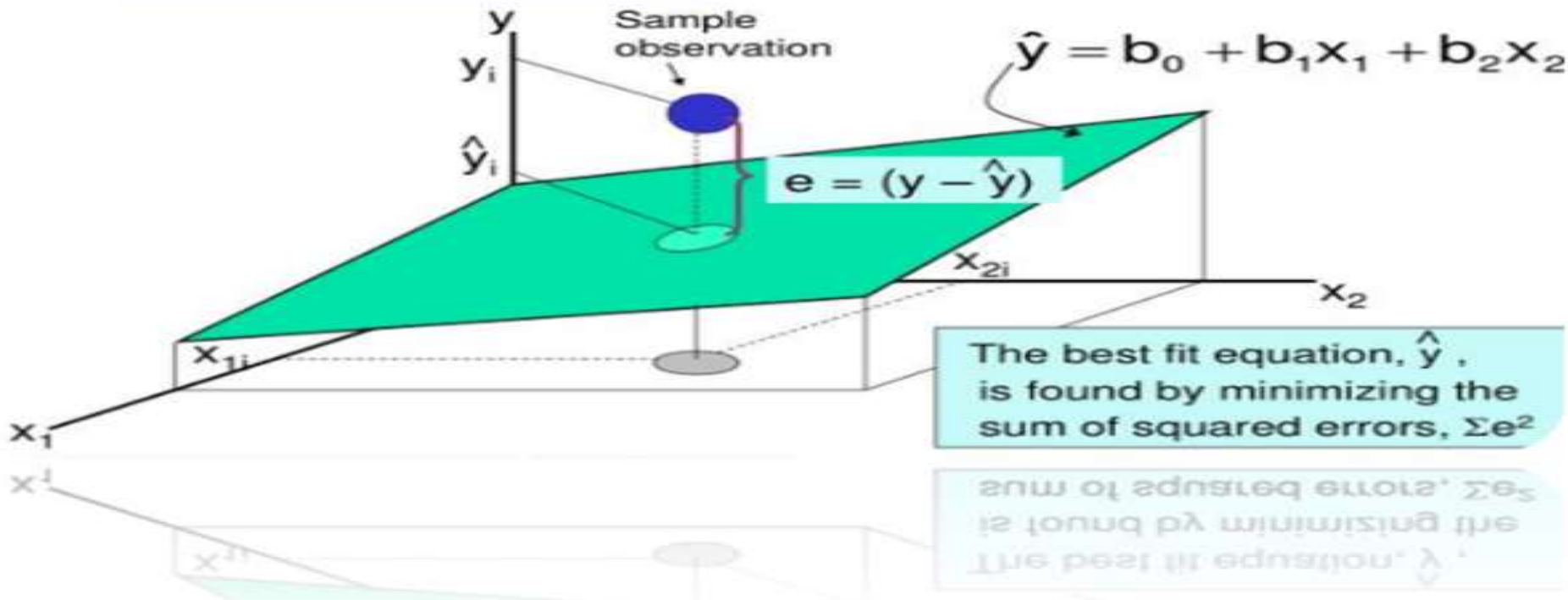
Multiple Regression Model

Two variable model



Multiple Regression Model

Two variable model



Multiple Regression Assumptions

Errors (residuals) from the regression model

$$e = (y - \hat{y})$$

- The errors are normally distributed
- The mean of the errors is zero
- Errors have a constant variance
- The model errors are independent

Model Specification

- Decide what you want to do and select the dependent variable
- Determine the potential independent variables for your model
- Gather sample data (observations) for all variables

The Correlation Matrix

- Correlation between the dependent variable and selected independent variables can be found using Excel:
  - Tools / Data Analysis... / Correlation
- Can check for statistical significance of correlation with a t test

Example

- A distributor of frozen desert pies wants to evaluate factors thought to influence demand
  - Dependent variable: Pie sales (units per week)
  - Independent variables: Price (in $)
    Advertising ($100's)
- Data is collected for 15 weeks

Pie Sales Model

Week	Pie Sales	Price ($)	Advertising ($100s)
1	350	5.50	3.3
2	460	7.50	3.3
3	350	8.00	3.0
4	430	8.00	4.5
5	350	6.80	3.0
6	380	7.50	4.0
7	430	4.50	3.0
8	470	6.40	3.7
9	450	7.00	3.5
10	490	5.00	4.0
11	340	7.20	3.5
12	300	7.90	3.2
13	440	5.90	4.0
14	450	5.00	3.5
15	300	7.00	2.7

Multiple regression model:

$$\widehat{Sales} = b_0 + b_1 (Price) + b_2 (Advertising)$$

Correlation matrix:

	Pie Sales	Price	Advertising
Pie Sales	1		
Price	-0.44327	1	
Advertising	0.55632	0.03044	1

Interpretation of Estimated Coefficients

- Slope (b_i)
  - Estimates that the average value of y changes by b_i units for each 1 unit increase in X_i holding all other variables constant
  - Example: if $b_1 = -20$, then sales (y) is expected to decrease by an estimated 20 pies per week for each \$1 increase in selling price (x_1), net of the effects of changes due to advertising (x_2)
- y-intercept (b_0)
  - The estimated average value of y when all $x_i = 0$ (assuming all $x_i = 0$ is within the range of observed values)

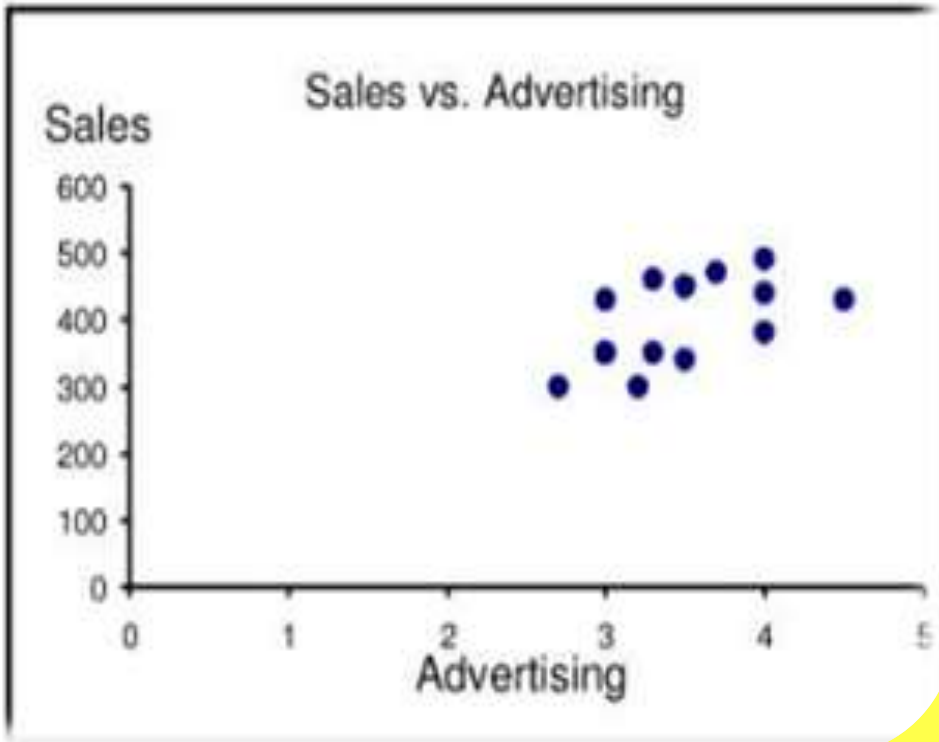
Pie Sales Correlation Matrix

•	•Pie Sales	•Price	•Advertising
•Pie Sales	•1		
•Price	•-0.44327	•1	
•Advertising	•0.55632	•0.03044	•1

- Price vs. Sales : r = -0.44327
  - There is a negative association between price and sales
- Advertising vs. Sales : r = 0.55632
  - There is a positive association between advertising and sales

Scatter Diagrams





Estimating a Multiple Linear Regression Equation

- Computer software is generally used to generate the coefficients and measures of goodness of fit for multiple regression

- Excel:
  - Tools / Data Analysis... / Regression
- PHStat:
  - PHStat / Regression / Multiple Regression…

Multiple Regression Output

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

Sales = 306.526 - 24.975(Price) + 74.131(Advertising)

ANOVA	df	SS	MS	F	Significance F
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888

© 2005 Inc.

The Multiple Regression Equation

$$\widehat{Sales} = 306.526 - 24.975(Price) + 74.131(Advertising)$$

where

Sales is in number of pies per week

Price is in $

Advertising is in $100's.

$b_1 = -24.975$: sales will decrease, on average, by 24.975 pies per week for each $1 increase in selling price, net of the effects of changes due to advertising

$b_2 = 74.131$: sales will increase, on average, by 74.131 pies per week for each $100 increase in advertising, net of the effects of changes due to price

Using The Model to Make Predictions

Predict sales for a week in which the selling price is $5.50 and advertising is $350:

$$\widehat{\text{Sales}} = 306.526 - 24.975(\text{Price}) + 74.131(\text{Advertising})$$

$$= 306.526 - 24.975\,(5.50) + 74.131\,(3.5)$$

$$= 428.62$$

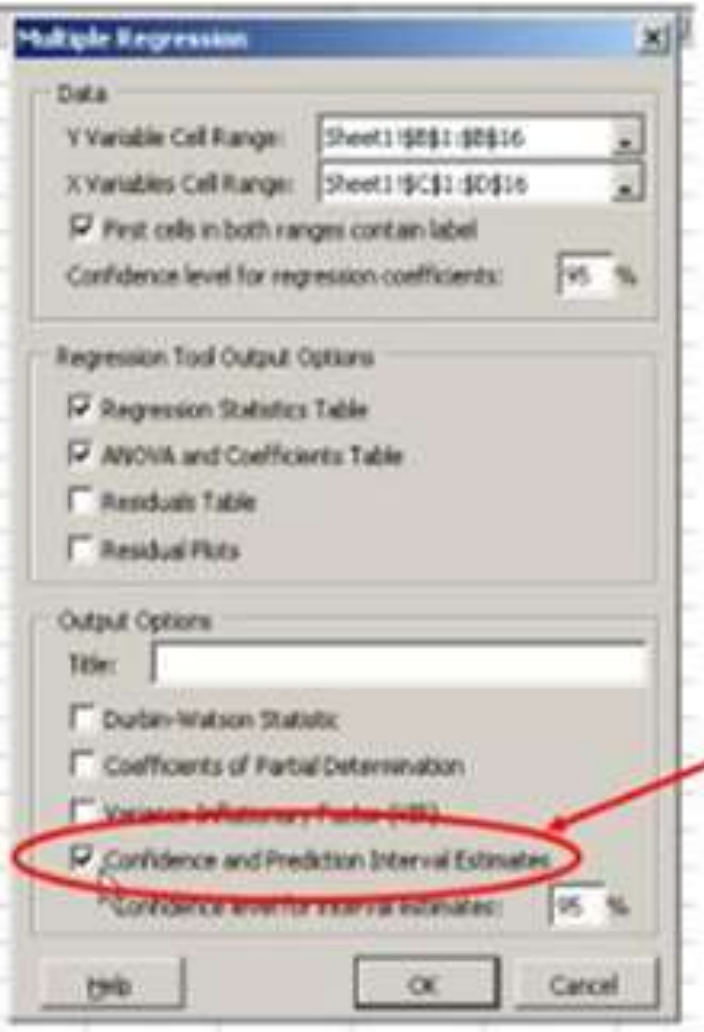
Predicted sales is 428.62 pies

Note that Advertising is in $100's, so $350 means that $x_2 = 3.5$

Predictions in PHStat

- PHStat | regression | multiple regression …

	A	B	C	D
1	Week	Pie Sales	Price	Advertising
2	1	350	5.5	3.3
3	2	460	7.5	3.3
4	3	350	8	3
5	4	430	8	4.5
6	5	350	6.8	3
7	6	380	7.5	4
8	7	430	4.5	3
9	8	470	6.4	3.7
10	9	450	7	3.5
11	10	490	5	4
12	11	340	7.2	3.5
13	12	300	7.9	3.2
14	13	440	5.9	4
15	14	450	5	3.5
16	15	300	7	2.7
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				



Check the "confidence and prediction interval estimates" box

Predictions in PHStat

	A	B
1	Confidence and Prediction Estimate Intervals	
2		
3	Data	
4	Confidence Level	95%
5		
6	Price given value	5.5
7	Advertising given value	3.5
8		
20	t Statistic	2.178813
21	Predicted Y (YHat)	428.6216
22		
23	For Average Predicted Y (Yhat)	
24	Interval Half Width	37.50306
25	Confidence Interval Lower Limit	391.1185
26	Confidence Interval Upper Limit	466.1246
27		
28	For Individual Response Y	
29	Interval Half Width	110.0041
30	Prediction Interval Lower Limit	318.6174
31	Prediction Interval Upper Limit	538.6257

Input values

Predicted $\hat{y}$ value

Confidence interval for the mean $\hat{y}$ value, given these x's

Prediction interval for an individual $\hat{y}$ value, given these x's

Multiple Coefficient of Determination

- Reports the proportion of total variation in y explained by all x variables taken together

$$R^2 = \frac{SSR}{SST} = \frac{\text{Sum of squares regression}}{\text{Total sum of squares}}$$

Multiple Coefficient of Determination

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

$$R^2 = \frac{SSR}{SST} = \frac{29460.0}{56493.3} = .52148$$

52.1% of the variation in pie sales is explained by the variation in price and advertising

ANOVA	df	SS	MS	F	Significance F
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc.

Adjusted R^2

- R^2 never decreases when a new x variable is added to the model
  - This can be a disadvantage when comparing models
- What is the net effect of adding a new variable?
  - We lose a degree of freedom when a new x variable is added
  - Did the new x variable add enough explanatory power to offset the loss of one degree of freedom?

Adjusted R^2

- Shows the proportion of variation in y explained by all x variables adjusted for the number of x variables used

$$R_A^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k-1}\right)$$

(where n = sample size, k = number of independent variables)

- Penalize excessive use of unimportant independent variables
- Smaller than R^2
- Useful in comparing among models

Multiple Coefficient of Determination

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

$R_A^2 = .44172$

44.2% of the variation in pie sales is explained by the variation in price and advertising, taking into account the sample size and number of independent variables

ANOVA	df	SS	MS	F	Significance F
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888

Business Statistics: A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc.

Is the Model Significant?

- F-Test for Overall Significance of the Model
- Shows if there is a linear relationship between all of the x variables considered together and y
- Use F test statistic
- Hypotheses:
  - $H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$ (no linear relationship)
  - H_A: at least one $\beta_i \neq 0$ (at least one independent variable affects y)

F-Test for Overall Significance

- Test statistic:

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} = \frac{MSR}{MSE}$$

where F has (numerator) $D_1 = k$ and

(denominator) $D_2 = (n - k - 1)$

degrees of freedom

F-Test for Overall Significance

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

$$F = \frac{MSR}{MSE} = \frac{14730.0}{2252.8} = 6.5386$$

With 2 and 12 degrees of freedom

P-value for the F-Test

ANOVA	df	SS	MS	F	Significance F
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888

A Decision-Making Approach, 6e © 2005 Prentice-Hall, Inc.

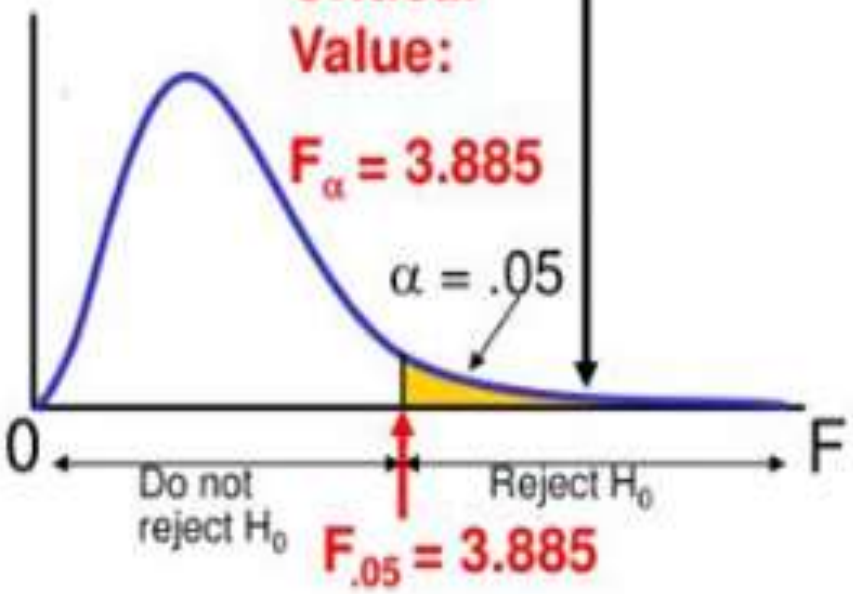
F-Test for Overall Significance

$H_0: \beta_1 = \beta_2 = 0$

$H_A: \beta_1$ and β_2 not both zero

$\alpha = .05$

$df_1 = 2 \quad df_2 = 12$



Test Statistic:

$$F = \frac{MSR}{MSE} = 6.5386$$

Decision:

Reject H_0 at $\alpha = 0.05$

Conclusion:

The regression model does explain a significant portion of the variation in pie sales

(There is evidence that at least one independent variable affects y)

Are Individual Variables Significant?

- Use t-tests of individual variable slopes
- Shows if there is a linear relationship between the variable x_i and y
- Hypotheses:
  - $H_0: \beta_i = 0$ (no linear relationship)
  - $H_A: \beta_i \neq 0$ (linear relationship does exist between x_i and y)

Are Individual Variables Significant?

$H_0: \beta_i = 0$ (no linear relationship)

$H_A: \beta_i \neq 0$ (linear relationship does exist between x_i and y)

Test Statistic:

$$t = \frac{b_i - 0}{s_{b_i}}$$

$(df = n - k - 1)$

Are Individual Variables Significant?

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

t-value for Price is t = -2.306, with p-value .0398

t-value for Advertising is t = 2.855, with p-value .0145

ANOVA	df	SS	MS	F	Significance F
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888

Inferences about the Slope: t Test Example

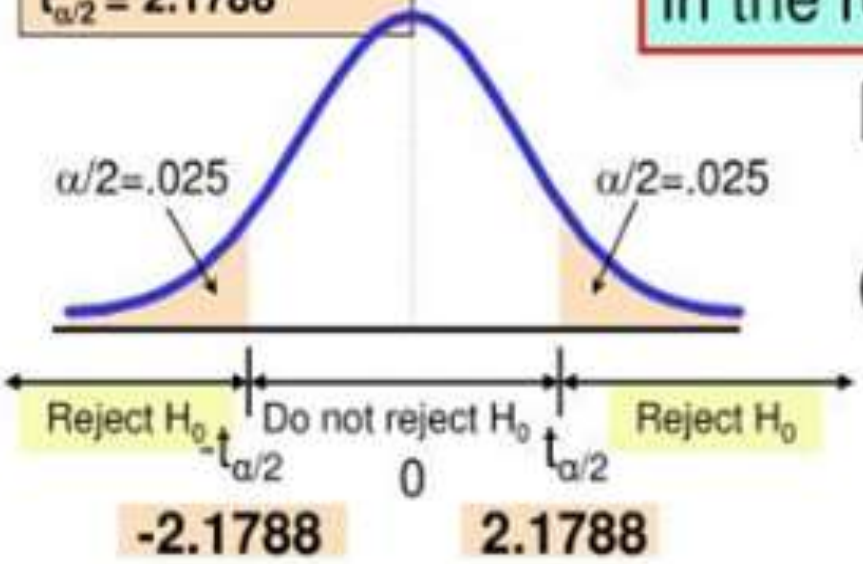
$H_0: \beta_i = 0$

$H_A: \beta_i \neq 0$

d.f. = 15-2-1 = 12

$\alpha = .05$

$t_{\alpha/2} = 2.1788$



From Excel output:

·	·Coefficients	·Standard Error	·t Stat	·P-value
·Price	·-24.97509	·10.83213	·-2.30565	·0.03979
·Advertising	·74.13096	·25.96732	·2.85478	·0.01449

The test statistic for each variable falls in the rejection region (p-values < .05)

Decision:

Reject H_0 for each variable

Conclusion:

There is evidence that both Price and Advertising affect pie sales at $\alpha = .05$

Confidence Interval Estimate for the Slope

Confidence interval for the population slope β_1 (the effect of changes in price on pie sales):

$$b_i \pm t_{\alpha/2} s_{b_i}$$

where t has $(n - k - 1)$ d.f.

·	·Coefficients	·Standard Error	·…	·Lower 95%	·Upper 95%
·Intercept	·306.52619	·114.25389	·…	·57.58835	·555.46404
·Price	·-24.97509	·10.83213	·…	·-48.57626	·-1.37392
·Advertising	·74.13096	·25.96732	·…	·17.55303	·130.70888

Example: Weekly sales are estimated to be reduced by between 1.37 to 48.58 pies for each increase of $1 in the selling price

Standard Deviation of the Regression Model

- The estimate of the standard deviation of the regression model is:

$$s_\varepsilon = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{MSE}$$

- Is this value large or small? Must compare to the mean size of y for comparison

Standard Deviation of the Regression Model

Regression Statistics	
Multiple R	0.72213
R Square	0.52148
Adjusted R Square	0.44172
Standard Error	47.46341
Observations	15

The standard deviation of the regression model is 47.46

ANOVA	df	SS	MS	F	Significance F
Regression	2	29460.027	14730.013	6.53861	0.01201
Residual	12	27033.306	2252.776		
Total	14	56493.333			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	306.52619	114.25389	2.68285	0.01993	57.58835	555.46404
Price	-24.97509	10.83213	-2.30565	0.03979	-48.57626	-1.37392
Advertising	74.13096	25.96732	2.85478	0.01449	17.55303	130.70888

Business A Decision-Making 6e © 2005 Inc.

Standard Deviation of the Regression Model

- The standard deviation of the regression model is 47.46
- A rough prediction range for pie sales in a given week is $\pm 2(47.46) = 94.2$
- Pie sales in the sample were in the 300 to 500 per week range, so this range is probably too large to be acceptable. The analyst may want to look for additional variables that can explain more of the variation in weekly sales

Multicollinearity

- Multicollinearity: High correlation exists between two independent variables
- This means the two variables contribute redundant information to the multiple regression model

Multicollinearity

- Including two highly correlated independent variables can adversely affect the regression results
  - No new information provided
  - Can lead to unstable coefficients (large standard error and low t-values)
  - Coefficient signs may not match prior expectations

Some Indications of Severe Multicollinearity

- Incorrect signs on the coefficients
- Large change in the value of a previous coefficient when a new variable is added to the model
- A previously significant variable becomes insignificant when a new independent variable is added
- The estimate of the standard deviation of the model increases when a variable is added to the model

Detect Collinearity (Variance Inflationary Factor)

VIF_j is used to measure collinearity:

$$VIF_j = \frac{1}{1 - R_j^2}$$

R^2_j is the coefficient of determination when the j^{th} independent variable is regressed against the remaining $k - 1$ independent variables

If $VIF_j > 5$, x_j is highly correlated with the other explanatory variables

Detect Collinearity in PHStat

PHStat / regression / multiple regression …

Check the "variance inflationary factor (VIF)" box

•Regression Analysis •Price and all other X	
•Regression Statistics	
•Multiple R	•0.030437581
•R Square	•0.000926446
•Adjusted R Square	•-0.075925366
•Standard Error	•1.21527235
•Observations	•15
•VIF	**•1.000927305**

Output for the pie sales example:

- Since there are only two explanatory variables, only one VIF is reported
  - VIF is < 5
  - There is no evidence of collinearity between Price and Advertising

Qualitative (Dummy) Variables

- Categorical explanatory variable (dummy variable) with two or more levels:
  - yes or no, on or off, male or female
  - coded as 0 or 1
- Regression intercepts are different if the variable is significant
- Assumes equal slopes for other variables
- The number of dummy variables needed is (number of levels - 1)

Dummy-Variable Model Examp (with 2 Levels)

Let:

y = pie sales

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2$$

x_1 = price

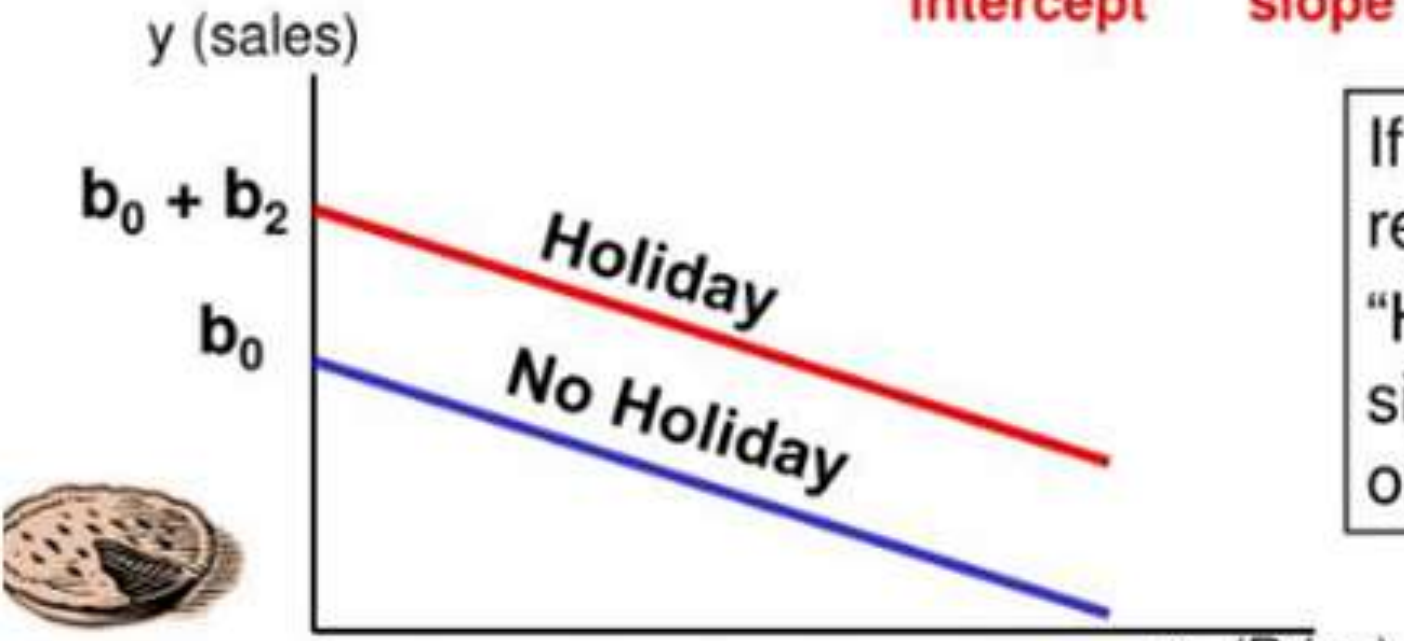
x_2 = holiday (X_2 = 1 if a holiday occurred during the week)
(X_2 = 0 if there was no holiday that week)

Dummy-Variable Model Examp (with 2 Levels)

$$\hat{y} = b_0 + b_1 x_1 + b_2(1) = (b_0 + b_2) + b_1 x_1 \quad \text{Holiday}$$

$$\hat{y} = b_0 + b_1 x_1 + b_2(0) = b_0 + b_1 x_1 \quad \text{No Holiday}$$

Different intercept **Same slope**



If $H_0: \beta_2 = 0$ is rejected, then "Holiday" has a significant effect on pie sales

Interpretation of the Dummy Variable Coefficient (with 2 Levels)

Example: $\text{Sales} = 300 - 30(\text{Price}) + 15(\text{Holiday})$

Sales: number of pies sold per week

Price: pie price in $

Holiday: $\begin{cases} 1 & \text{If a holiday occurred during the week} \\ 0 & \text{If no holiday occurred} \end{cases}$

$b_2 = 15$: on average, sales were 15 pies greater in weeks with a holiday than in weeks without a holiday, given the same price

Dummy-Variable Models (more than 2 Levels)

- The number of dummy variables is **one less than the number of levels**
- Example:

  y = house price ; x_1 = square feet

- The style of the house is also thought to matter

  Style = ranch, split level, condo

Three levels, so two dummy variables are needed

Dummy-Variable Models (more than 2 Levels)

Let the default category be "condo"

$$x_2 = \begin{cases} 1 & \text{if ranch} \\ 0 & \text{if not} \end{cases} \qquad x_3 = \begin{cases} 1 & \text{if split level} \\ 0 & \text{if not} \end{cases}$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3$$

b_2 shows the impact on price if the house is a ranch style, compared to a condo

b_3 shows the impact on price if the house is a split level style, compared to a condo

Interpreting the Dummy Variabl Coefficients (with 3 Levels)

Suppose the estimated equation is

$$\hat{y} = 20.43 + 0.045x_1 + 23.53x_2 + 18.84x_3$$

For a condo: $x_2 = x_3 = 0$

$$\hat{y} = 20.43 + 0.045x_1$$

For a ranch: $x_3 = 0$

$$\hat{y} = 20.43 + 0.045x_1 + 23.53$$

With the same square feet, a split-level will have an estimated average price of 18.84 thousand dollars more than a condo

For a split level: $x_2 = 0$

$$\hat{y} = 20.43 + 0.045x_1 + 18.84$$

With the same square feet, a ranch will have an estimated average price of 23.53 thousand dollars more than a condo.

Nonlinear Relationships

- The relationship between the dependent variable and an independent variable may not be linear
- Useful when scatter diagram indicates non-linear relationship
- Example: Quadratic model
  - $$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon$$
  - The second independent variable is the square of the first variable

Polynomial Regression Model

General form:

$$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \ldots + \beta_p x_j^p + \varepsilon$$

- where:

β_0 = Population regression constant

β_i = Population regression coefficient for variable x_j : j = 1, 2, …k

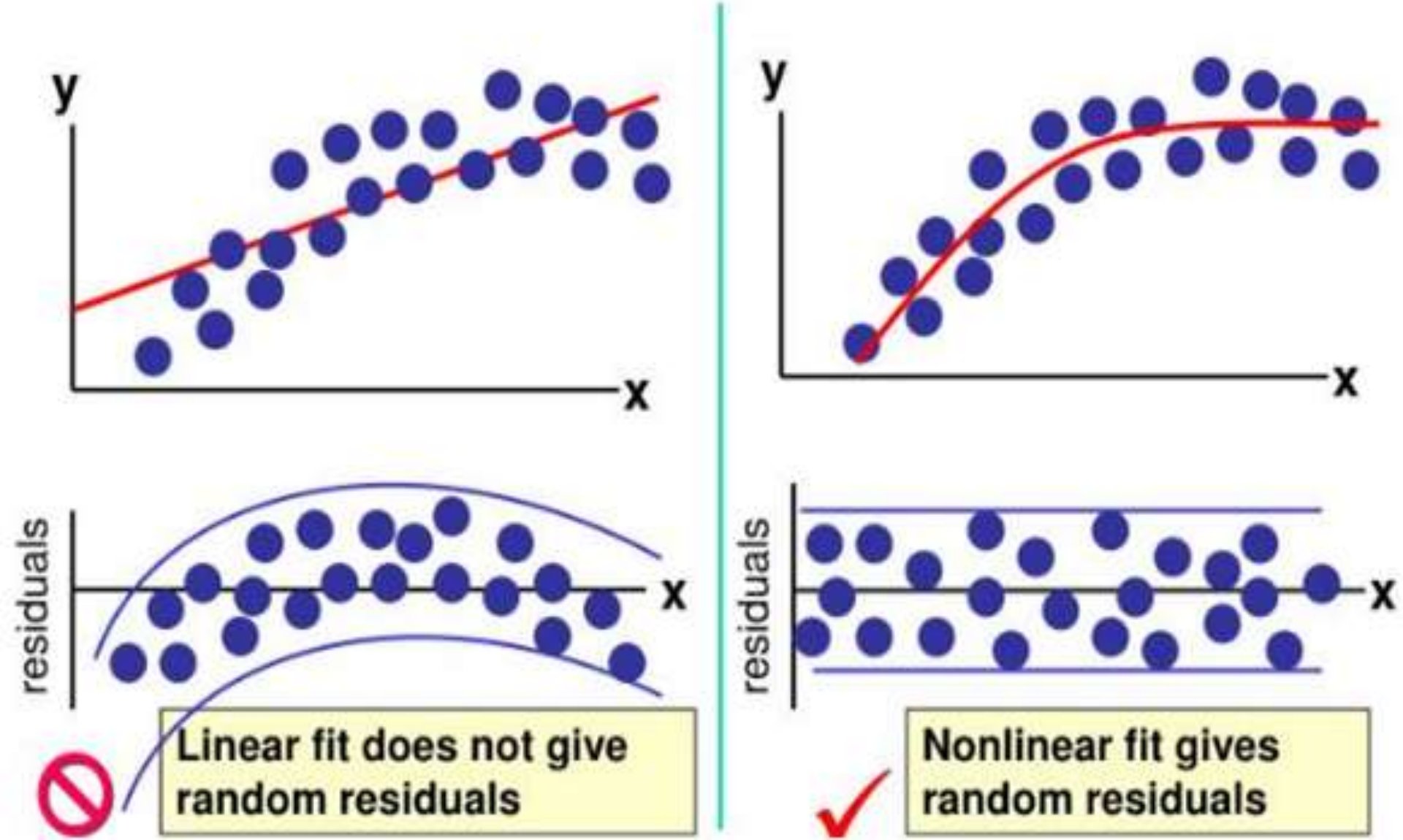
p = Order of the polynomial

ε_i = Model error

If p = 2 the model is a quadratic model:

$$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon$$

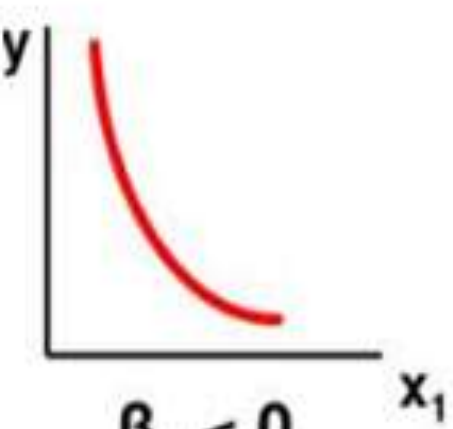
Linear vs. Nonlinear Fit



Quadratic Regression Model

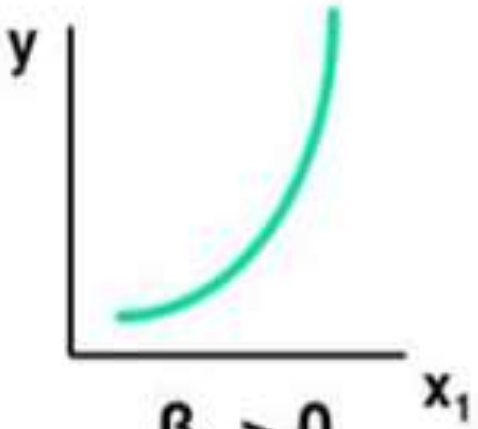
$$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon$$

Quadratic models may be considered when scatter diagram takes on the following shapes:



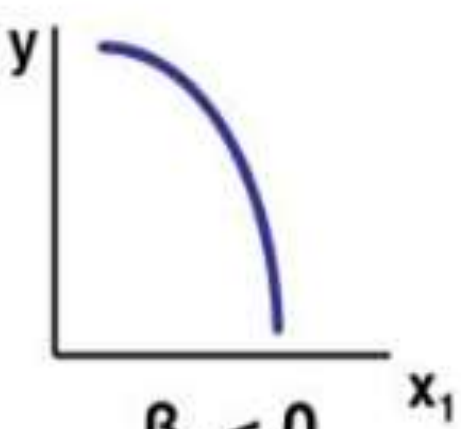
$\beta_1 < 0$

$\beta_2 > 0$



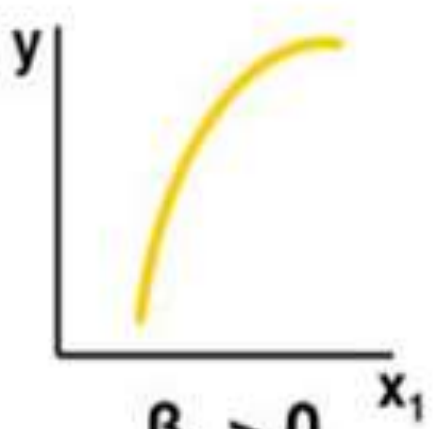
$\beta_1 > 0$

$\beta_2 > 0$



$\beta_1 < 0$

$\beta_2 < 0$



$\beta_1 > 0$

$\beta_2 < 0$

β_1 = the coefficient of the linear term

β_2 = the coefficient of the squared term

Testing for Significance: Quadratic Model

- Test for Overall Relationship
  - F test statistic = $\frac{MSR}{MSE}$
- Testing the Quadratic Effect
  - Compare quadratic model

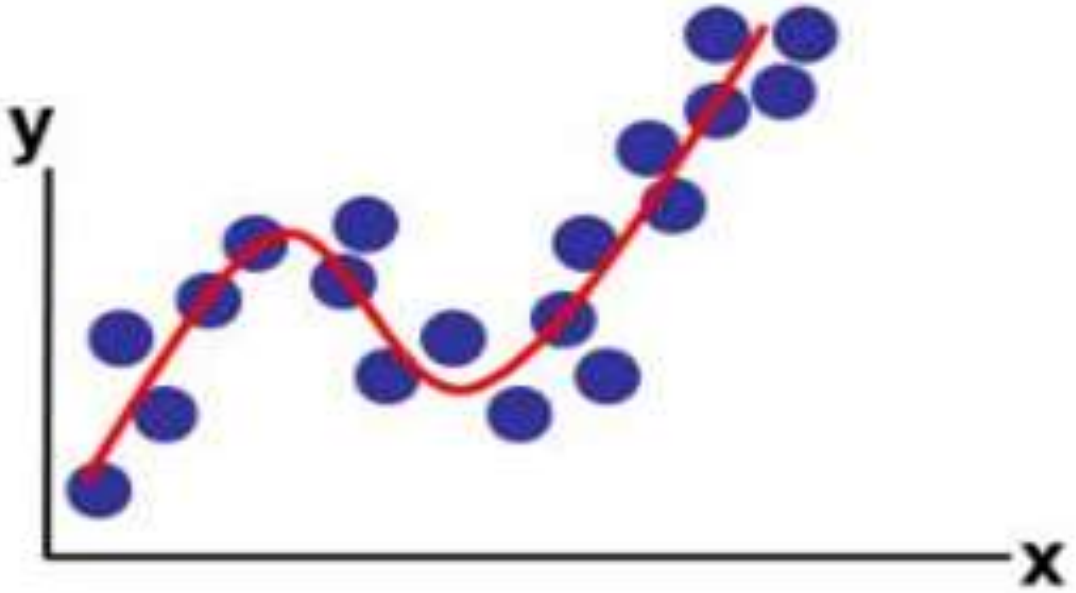
$$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon$$

with the linear model

$$y = \beta_0 + \beta_1 x_j + \varepsilon$$

  - Hypotheses
    - $H_0: \beta_2 = 0$ (No 2nd order polynomial term)
    - $H_A: \beta_2 \neq 0$ (2nd order polynomial term is needed)

Higher Order Models



If p = 3 the model is a cubic form:

$$y = \beta_0 + \beta_1 x_j + \beta_2 x_j^2 + \beta_3 x_j^3 + \varepsilon$$

Interaction Effects

- Hypothesizes interaction between pairs of x variables
  - Response to one x variable varies at different levels of another x variable
- Contains two-way cross product terms

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1^2 x_2$$

Basic Terms

Interactive Terms

Effect of Interaction

- Given:

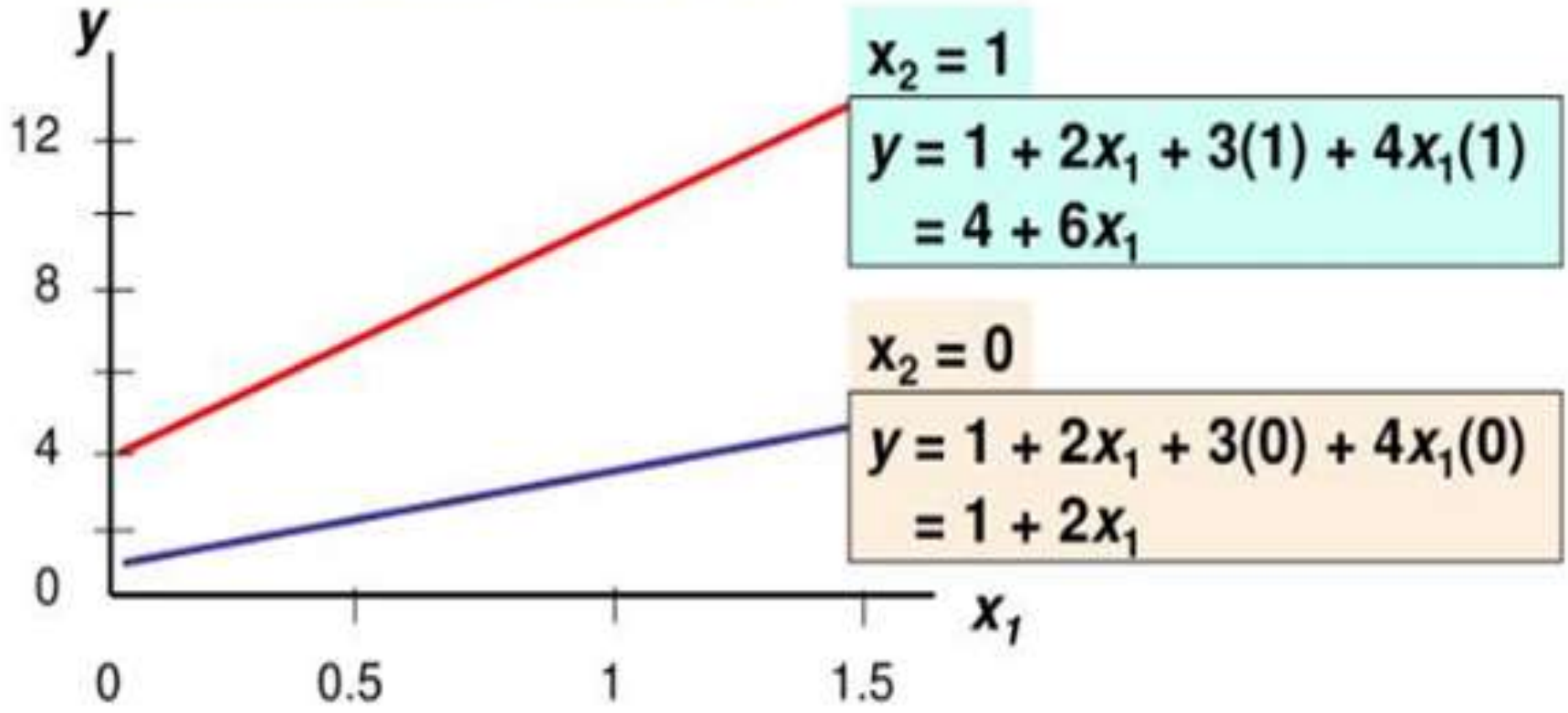
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \varepsilon$$

- Without interaction term, effect of x_1 on y is measured by β_1
- With interaction term, effect of x_1 on y is measured by $\beta_1 + \beta_3 x_2$
- Effect changes as x_2 increases

Interaction Example

$y = 1 + 2x_1 + 3x_2 + 4x_1x_2$

where $x_2 = 0$ or 1 (dummy variable)



Effect (slope) of x_1 on y does depend on x_2 value

Interaction Regression Model Worksheet

Case, i	y_i	x_{1i}	x_{2i}	$x_{1i}\,x_{2i}$
1	1	1	3	3
2	4	8	5	40
3	1	3	2	6
4	3	5	6	30
:	:	:	:	:

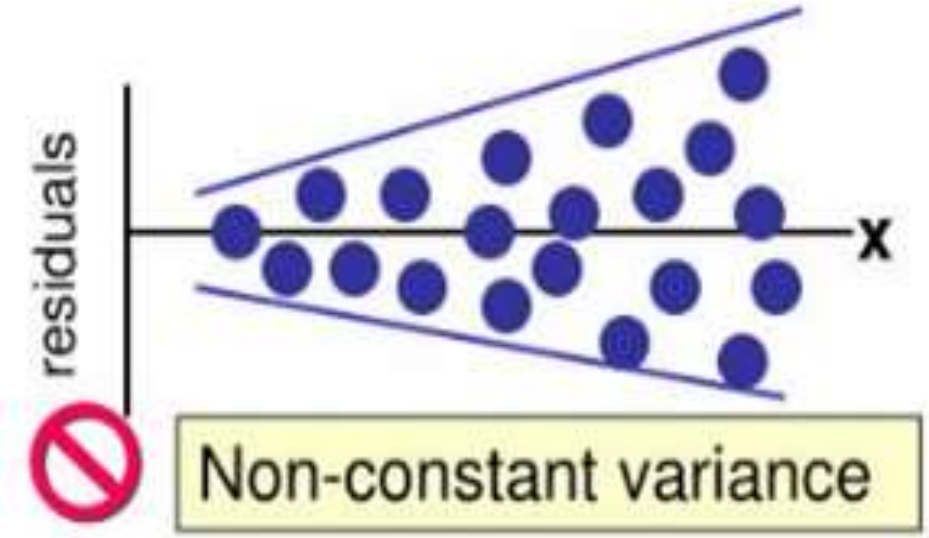
multiply x_1 by x_2 to get x_1x_2, then run regression with y, x_1, x_2 , x_1x_2

Aptness of the Model

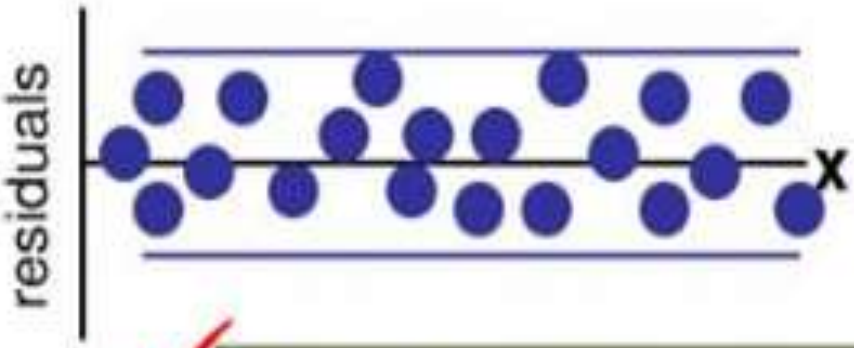
- Diagnostic checks on the model include verifying the assumptions of multiple regression:
  - Each x_i is linearly related to y
  - Errors have constant variance
  - Errors are independent
  - Error are normally distributed

Errors (or Residuals) are given by $e_i = (y - \hat{y})$

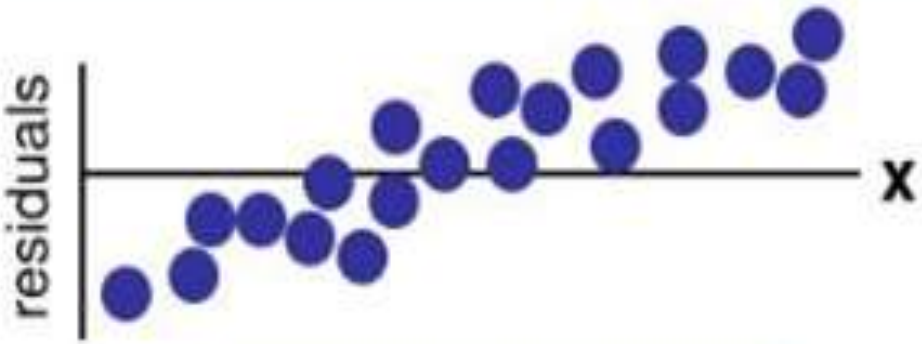
Residual Analysis



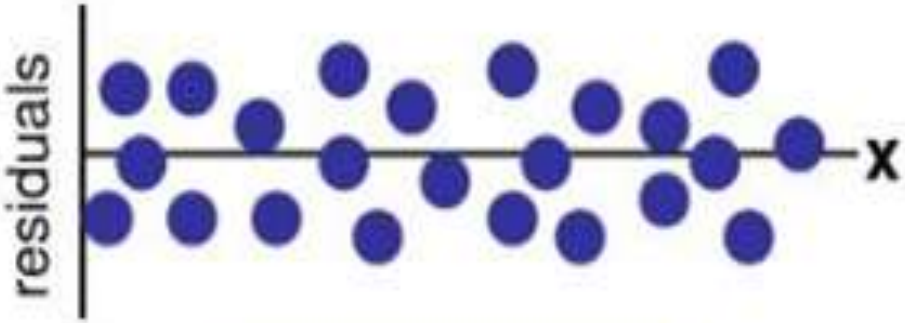
Non-constant variance



Constant variance



Not Independent



Independent

The Normality Assumption

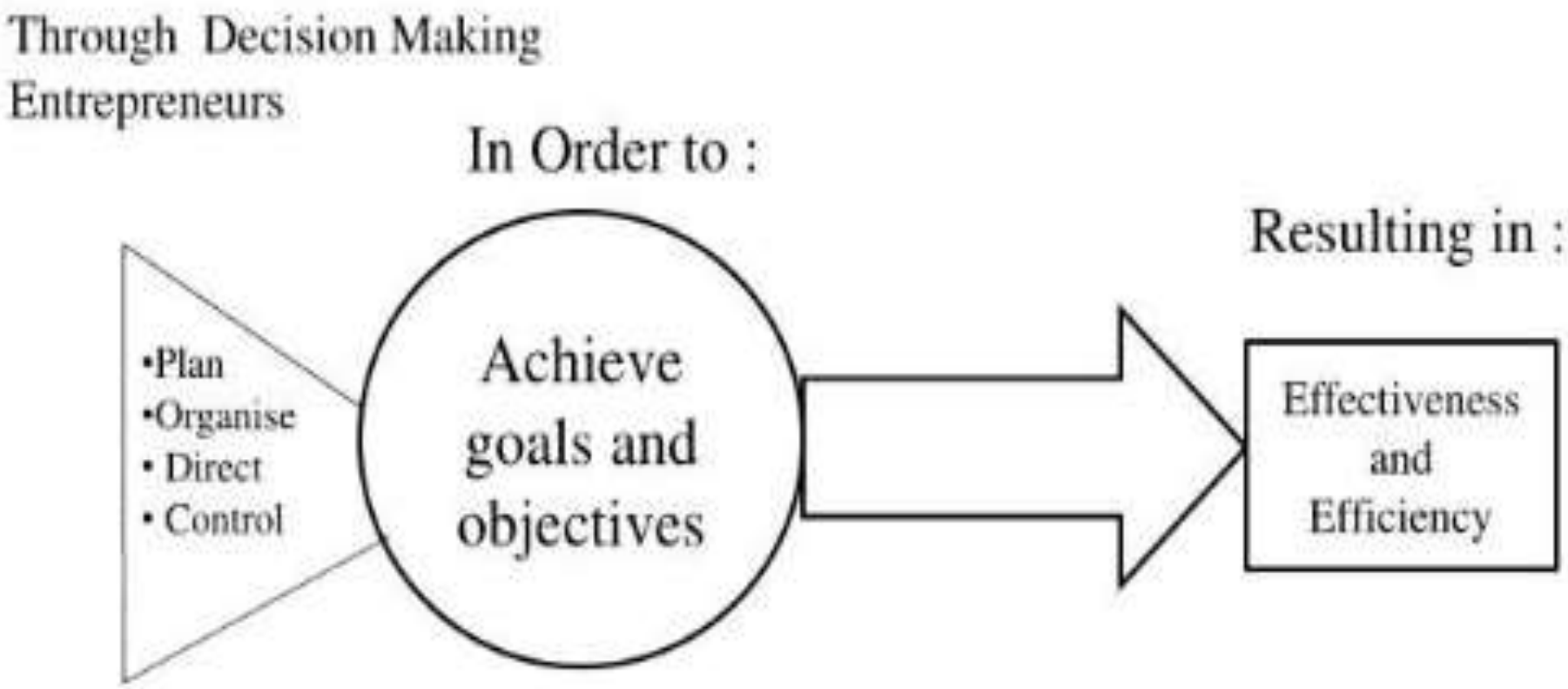
- Errors are assumed to be normally distributed
- Standardized residuals can be calculated by computer
- Examine a histogram or a normal probability plot of the standardized residuals to check for normality

Decision Making

- Decision-making may be described as the process of selecting a course of action from several alternatives so that desired result may be accomplished.
- The purpose of decision-making is to direct human behaviour and commitment towards a future goal.
- Decision making is the process of making choices by identifying a decision, gathering information, and assessing alternative resolutions.
- Using a step-by-step decision-making process can help you make more deliberate, thoughtful decisions by organizing relevant information and defining alternatives.
- Decisions play important roles as they determine both organizational and managerial activities. Decision making process is continuous and indispensable component of managing any organization or business activities.
- Decisions are made to sustain the activities of all business activities and organizational functioning.

Meaning of Decision Making

- **According to George R. Terry:** Decision making is the selection based on some criteria from two or more possible alternatives.
- **According to Earnest Dale:** Entrepreneurial decisions are those decisions, which are always made in the course of one of the true entrepreneurship activities: Planning, Organizing, Staffing, Directing, Controlling, Innovation and Representation.



Characteristics of Decision Making - I

- **Selection of the best alternative:** Decision making is the process of selecting the best alternative among the various alternatives available for the solution of a given problem.
- **Rational Selection:** Rational selection means selection that is balanced, objective and unbiased, based on sound judgement and logical analysis, comparison and evaluation. Only, such a course can lead to the choice of the best solution out of a variety of possible solutions.
- **Responsibilities of Entrepreneur:** All major decisions are responsibility of the entrepreneur for smooth functioning of enterprise.
- **Decision are Means:** The process of entrepreneurship aims at achieving pre-determined objectives and decisions are an important means of doing so but they are not the end. The end is to achieve the objectives through the process of decision making the objectives through the process of decision making and not only to decide and sit

Characteristics of Decision Making - II

- **Risk and Uncertainty:** It is very difficult to forecast the outcome of future activities with complete accuracy. Therefore, decision making involves an element of risk and uncertainty. A wrong decision can lead the whole enterprise unit to losses and even closure.
- **Decision may be positive or negative:** A decision can be both positive and negative. What is right or wrong in doing or not doing, depends upon the situations and conditions of the event.
- **Involves commitment:** A decision reflects the commitment of a person who takes the decision. The person taking the decision is committed to implement it and to prepare the whole network of plans according to the decision making.
- **Time and Psychological Factors:** Time is an important dimension of decision making. The psychological factors like intelligence, experience, habits, temperament, etc. have also impact on the decisions taken by entrepreneur.

Characteristics of Decision Making - III

- **Future Course of Action:** Decisions are made for future course of action based on the basis of past experience and present conditions.
- **Continuous Process:** Decision making is a continuous and regular process because an entrepreneur is required to take decisions continuously for different activities.
- **Communication:** The process of decision making is completed only when the decision are properly communicated in time to all those who are affected by the decision or who are to implement them.
- **Flexible:** Decision making process is flexible, changeable and dynamic in nature. The techniques used for making decisions vary with the types of problems involved. Thus, it changes with the change of situation.

Steps in Decision Making - I



Steps in Decision Making - II

- **Step 1: Identify the decision -** You realize that you need to make a decision. Try to clearly define the nature of the decision you must make. This first step is very important.
- **Step 2: Gather relevant information -** Collect some pertinent information before you make your decision: what information is needed, the best sources of information, and how to get it. This step involves both internal and external "work."
- Some information is internal: you'll seek it through a process of self-assessment. Other information is external: you'll find it online, in books, from other people, and from other sources.
- **Step 3: Identify the alternatives -** As you collect information, you will probably identify several possible paths of action, or alternatives.
- You can also use your imagination and additional information to construct new alternatives. In this step, you will list all possible and desirable alternatives.

Steps in Decision Making - III

- **Step 4: Weigh the evidence -** Draw on your information and emotions to imagine what it would be like if you carried out each of the alternatives to the end.
- Evaluate whether the need identified in Step 1 would be met or resolved through the use of each alternative.
- As you go through this difficult internal process, you'll begin to favor certain alternatives: those that seem to have a higher potential for reaching your goal.
- Finally, place the alternatives in a priority order, based upon your own value system.
- **Step 5: Choose among alternatives -** Once you have weighed all the evidence, you are ready to select the alternative that seems to be best one for you.
- You may even choose a combination of alternatives. Your choice in Step 5 may very likely be the same or similar to the alternative you placed at the top of your list at the end of Step 4.

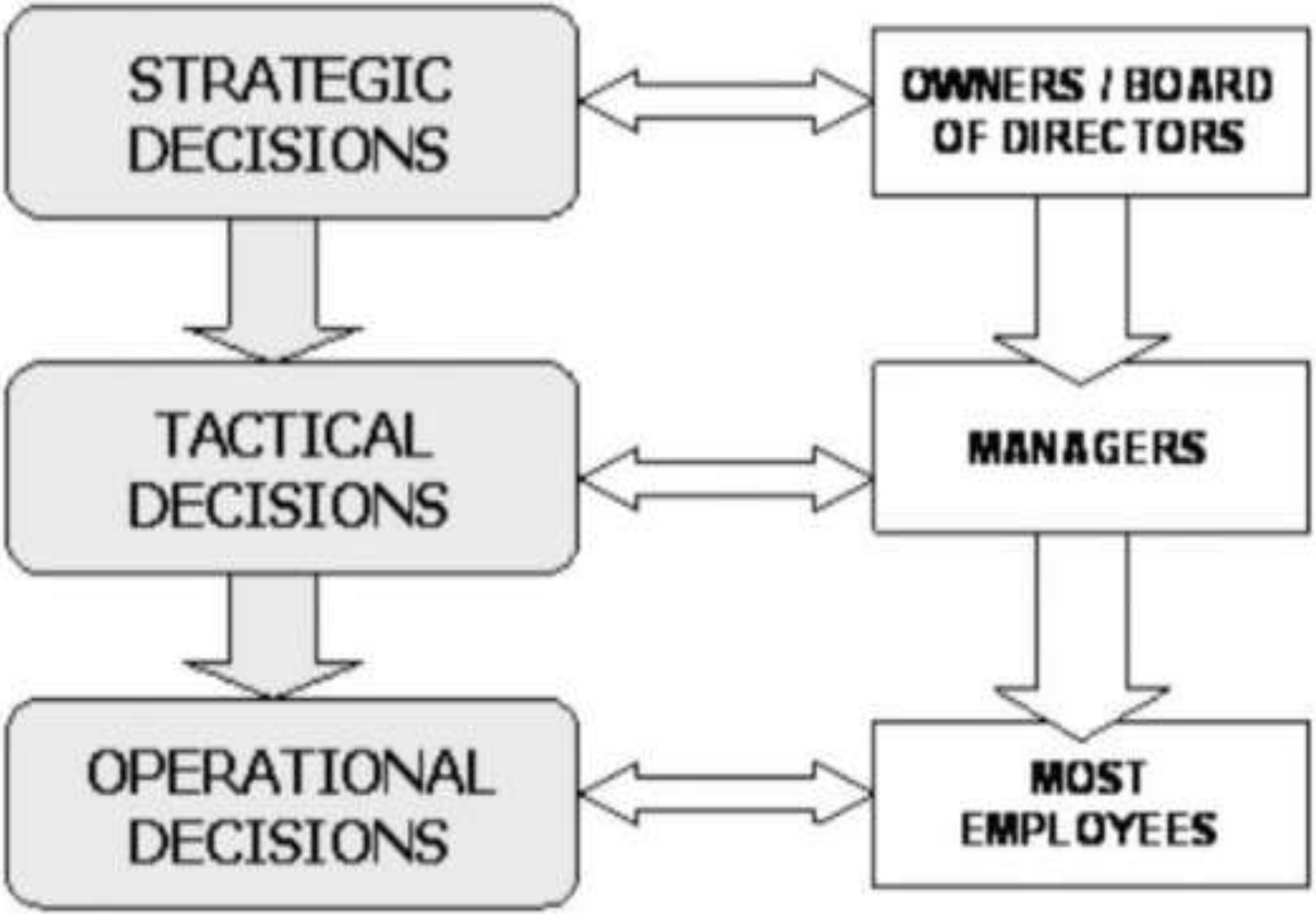
Steps in Decision Making - IV

- **Step 6: Take action -** You're now ready to take some positive action by beginning to implement the alternative you chose in Step 5.
- **Step 7: Review your decision & its consequences -** In this final step, consider the results of your decision and evaluate whether or not it has resolved the need you identified in Step 1.
- If the decision has *not* met the identified need, you may want to repeat certain steps of the process to make a new decision.
- For example, you might want to gather more detailed or somewhat different information or explore additional alternatives.

Importance of Decision Making

- **Heart of Entrepreneurship:** All goals, strategies, policies and programs are determined through decisions. Therefore, the central pole of entrepreneur is to make decisions which determine the future of enterprise.
- **Pervasive Activity:** Decision making is a pervasive activity. It exists in every part and at every level of enterprise.
- **Universal mark of an Entrepreneur:** An entrepreneur is viewed as a specialist in the art of decision making. He is responsible for making decisions rather than doing the work personally. An entrepreneur is a man who decides.
- **Core of Planning:** Decision making is the sole of planning. Planning is deciding in advance what to do.
- **Basis of effective control:** Control is crucial to an enterprise's success. Through decision making, entrepreneurs constantly monitor process, compare actual process with standards and tale corrective action in case of any deviations.
- **To face challenges:** Entrepreneurs can face challenges of modern time through good decisions. In the environment of turbulence, entrepreneurs are forced to make more and more decision at a faster and faster pace.

Types of Decisions - I



Types of Decisions - II

- Strategic Decision : What ?
- Strategic decisions deals with the big picture of the business. The focus of strategic decisions is typically external to the business and usually future oriented. Strategic decision making creates the forward thrust in the business.
- It includes decision about:
- What business are you in ?
- What is your Vision for the business ?
- What is your business identity ?
- What do you stand for ?
- Which direction is the business headed ?
- How will the business compete ?



Types of Decisions - III

- Tactical Decisions – How ?
- Tactical decisions involve the establishment of key initiatives to achieve the overall strategy.
- Example: If you have decided to be the number 1 provider in your market they you will develop tactics (e.g. implement a marketing system, increase number of therapists) to achieve that outcome. In a small business you may have 4 or 5 key tactics that you are going to use to achieve your overall strategy.
- Again this layer of decision-making can sometimes be overlooked yet it is the glue that creates a strong connection between your long-term vision and your day-to-day activities. Tactical decision making is the domain of ‘mission’ statements.



Advantages and Disadvantages of Decision making

- **Advantages of Decision Making:**
- Gives more information.
- Increase people's participation.
- Provide more alternatives.
- Improves the degree of acceptance and commitment.
- Improves the quality of decisions.
- Helps in strengthening the organization.

- **Disadvantages of Decision Making:**
- Costly.
- Time-Consuming.
- Individual Domination.
- Ambiguous Responsibility.





In addition to scale of measurement, we should look at *the population distribution*.

Population is normally distributed

- Parametric
- (may be used)

Not normally distributed population or no assumption can be made about the population distribution

- Nonparametric
- (have to be used)

Normal Distribution

- a very common continuous probability distribution
- All normal distributions are symmetric.
- bell-shaped curve with a single peak.
- **68%** of the observations fall within **1 standard deviation** of the **mean**
- **95%** of the observations fall within **2 standard deviations** of the **mean**
- **99.7%** of the observations fall within **3 standard deviations** of the **mean**
- for a normal distribution, almost all values lie within **3 standard deviations** of the mean

To use parametric tests, stay tuned...

- Interval or ratio data are required.
- Normal distribution is required.

+

Homogeneity of variance

Homogeneity of Variance

- The variance is a measure of the dispersion of the random variable about the mean. In other words, it indicates how far the values spread out.
- It refers to that variance within each of population is equal.
- Homogeneity of Variances is assessed by Levene's test. (T-test and ANOVA use Levene's test.)

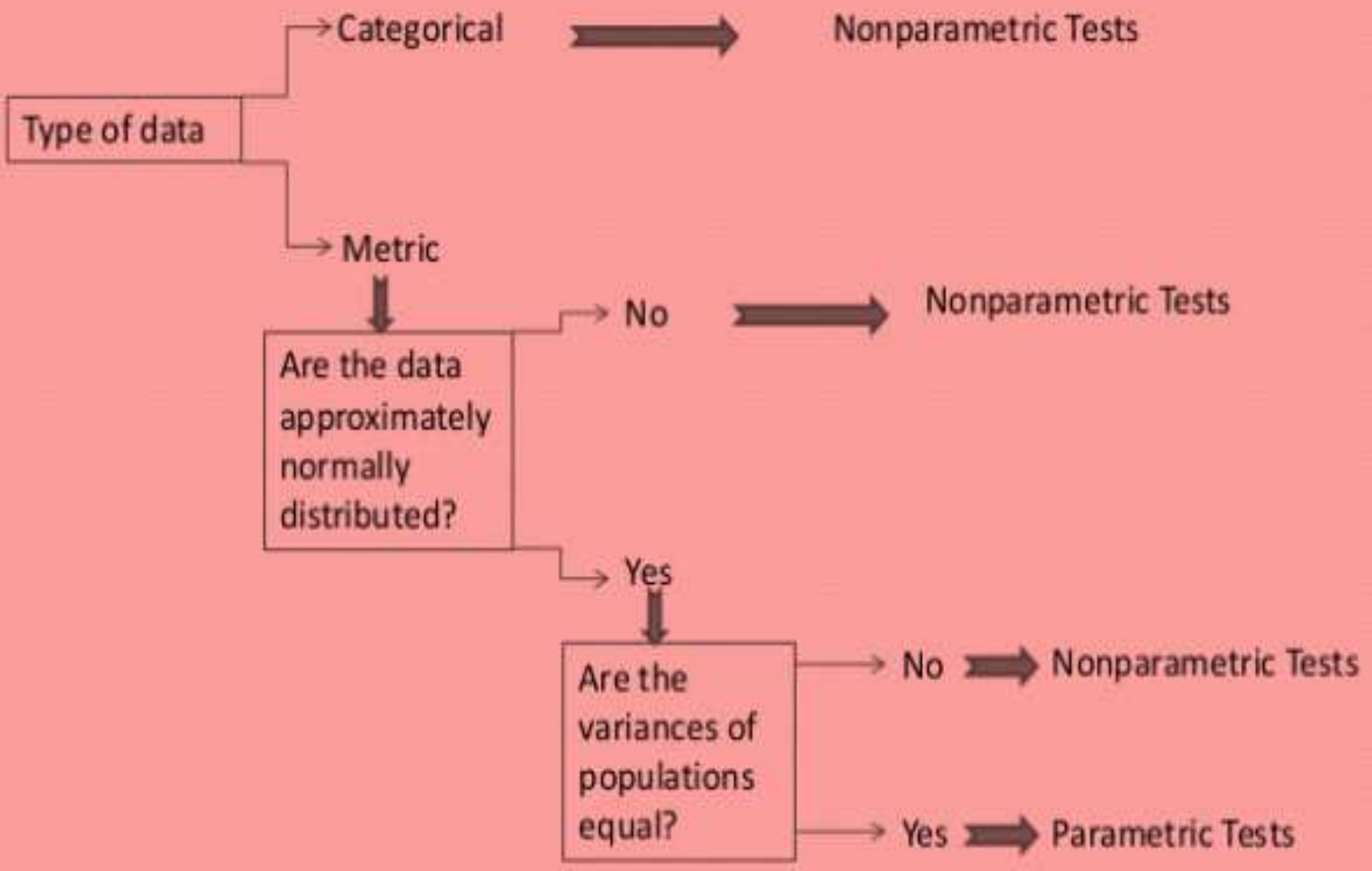
Parametric or nonparametric – *Determination*

- In cases where
  - the data which are measured by interval or ratio scale come from a normal distribution
  - Population variances are equal

  parametric tests are used.
- In cases where
  - the data is nominal or ordinal
  - the assumptions of parametric tests are inappropriate

  nonparametric tests are used.

Parametric or nonparametric – *Determination*



Statistical Test Alternatives: Parametric - Nonparametric

Input variable	Output variable: Nominal	Output variable: Ordinal	Output variable: Interval - Ratio
Nominal	Chi-square	Mann Whitney Kruskal – Wallis	Unpaired t-test or Mann Whitney Paired t-test or Wilcoxon Analysis of variance or Kruskal – Wallis
Ordinal	Chi-square Mann Whitney	Spearman Rank	Linear regression or Spearman
Interval Ratio	Logistic regression	Poisson regression	Pearson's r, Linear regression or Spearman

Wilcoxon signed-rank test

- Nonparametric equivalent of the paired t-test.
- Similar to sign test, but take into consideration the magnitude of difference among the pairs of values. (Sign test only considers the direction of difference but not the magnitude of differences.)



WILCOXON

Wilcoxon signed-rank test

- The 14 difference scores in BP among hypertensive patients after giving drug A were:

  -20, -8, -14, -12, -26, +6, -18, -10, -12, -10, -8, +4, +2, -18

- The statistic T is found by calculating the sum of the positive ranks, and the sum of the negative ranks.
- The smaller of the two values is considered.

Wilcoxon signed-rank test

Score	Rank
+2	1
+4	2
+6	3
-8	4.5
-8	4.5
-10	6.5
-10	6.5
-12	8
-14	9
-16	10
-18	11.5
-18	11.5
-20	13
-26	14

Sum of positive ranks = 6

Sum of negative ranks = 99

T= 6

For N = 14, and $\alpha = .05$, the critical value of T = 21.

If T is equal to or less than T critical, then null hypothesis is rejected i.e., drug A decreases the BP among hypertensive patients.

Mann-Whitney U test

- Mann-Whitney U – similar to Wilcoxon signed-ranks test except that the samples are independent and not paired.

- Null hypothesis: the population means are the same for the two groups.

- Rank the combined data values for the two groups. Then find the average rank in each group.

Mann-Whitney U test

- Then the U value is calculated using formula
- $U = N1*N2 + \frac{Nx(Nx+1)}{2} - Rx$ (where Rx is larger rank total)
- To be statistically significant, obtained U has to be equal to or LESS than this critical value.

Mann-Whitney U test

- 10 dieters following Atkin's diet vs. 10 dieters following Jenny Craig diet
- Hypothetical RESULTS:
- Atkin's group loses an average of 34.5 lbs.
- J. Craig group loses an average of 18.5 lbs.
- Conclusion: Atkin's is better?

Mann-Whitney U test

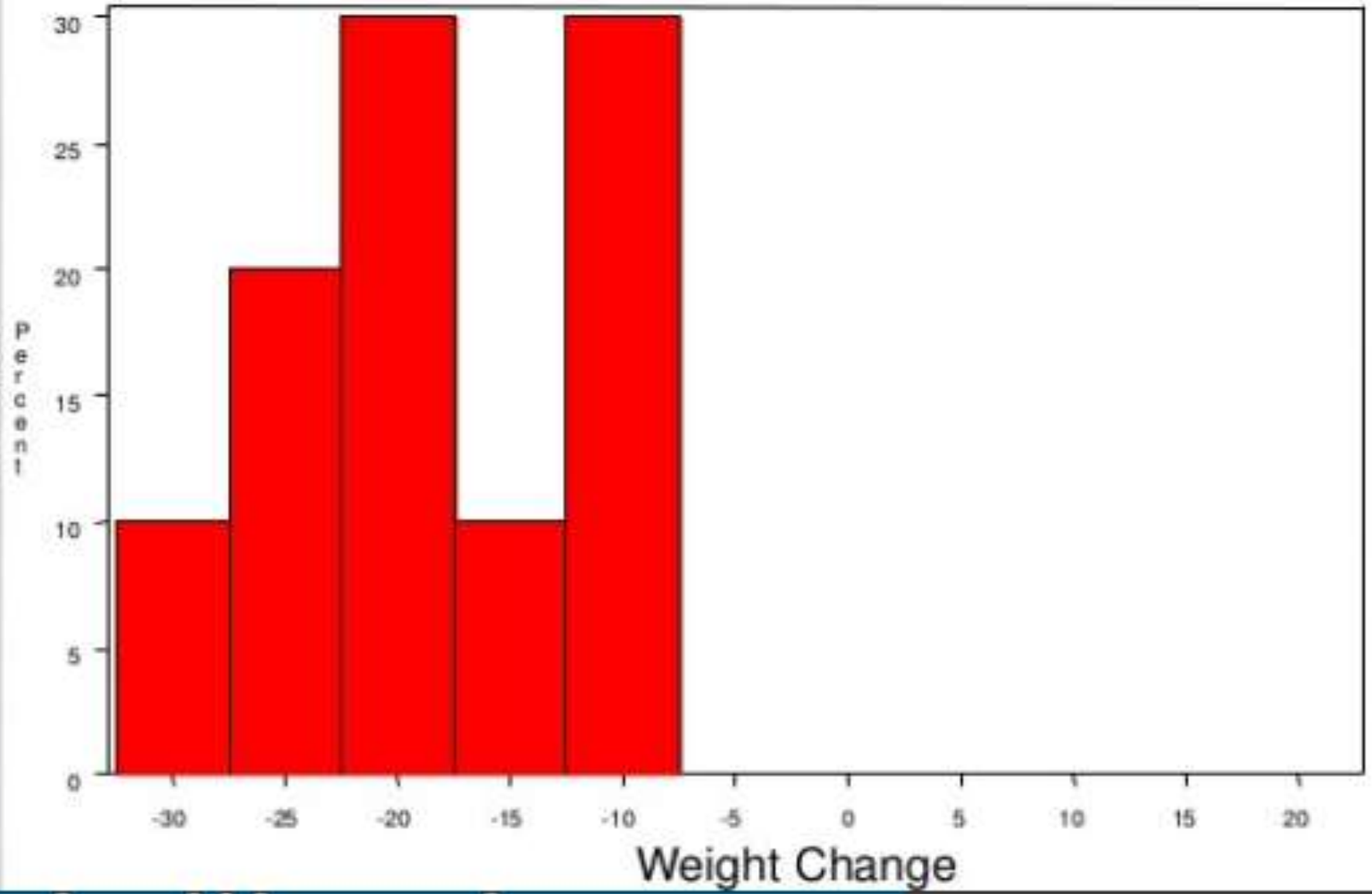
- When individual data is seen
- Atkin's, change in weight (lbs):

  +4, +3, 0, -3, -4, -5, -11, -14, -15, -300

- J. Craig, change in weight (lbs)

  -8, -10, -12, -16, -18, -20, -21, -24, -26, -30

Jenny Craig diet



Mann-Whitney U test

- RANK the values, 1 being the least weight loss and 20 being the most weight loss.
- Atkin's
  - +4, +3, 0, -3, -4, -5, -11, -14, -15, -300
  - 1, 2, 3, 4, 5, 6, 9, 11, 12, 20
- J. Craig
  - -8, -10, -12, -16, -18, -20, -21, -24, -26, -30
  - 7, 8, 10, 13, 14, 15, 16, 17, 18, 19

Mann-Whitney U test

- Sum of Atkin's ranks:

  1+ 2 + 3 + 4 + 5 + 6 + 9 + 11+ 12 + 20=73

- Sum of Jenny Craig's ranks:

  7 + 8 +10+ 13+ 14+ 15+16+ 17+ 18+19=137

- Jenny Craig clearly ranked higher.
- Calculated U value (18) < table value (27), Null hypothesis is rejected.

Kruskal-Wallis One-way ANOVA

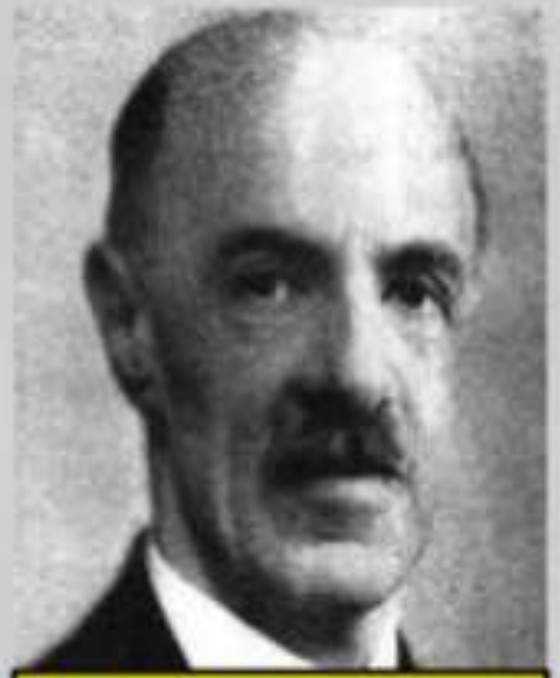
- It's more powerful than Chi-square test.
- It is computed exactly like the Mann-Whitney test, except that there are more groups (>2 groups).
- Applied on independent samples with the same shape (but not necessarily normal).

Friedman ANOVA

- Friedman ANOVA: When either a matched-subjects or repeated-measure design is used and the hypothesis of a difference among three or more (k) treatments is to be tested, the Friedman ANOVA by ranks test can be used.

Spearman rank-order correlation

- Use to assess the relationship between two ordinal variables or two skewed continuous variables.
- Nonparametric equivalent of the Pearson correlation.
- It is a relative measure which varies from -1 (perfect negative relationship) to +1 (perfect positive relationship).



Charles Spearman
(1863–1945)

Cochran's Q test

- Cochran's Q test is a non-parametric statistical test to verify if k treatments have identical effects where the response variable can take only two possible outcomes (coded as 0 and 1)

	Treatment 1	**Treatment 2**	⋯	**Treatment *k***
Block 1	X_{11}	X_{12}	⋯	X_{1k}
Block 2	X_{21}	X_{22}	⋯	X_{2k}
Block 3	X_{31}	X_{32}	⋯	X_{3k}
⋮	⋮	⋮	⋱	⋮
Block *b*	X_{b1}	X_{b2}	⋯	X_{bk}

Advantages of non-parametric tests

- These tests are distribution free.
- Easier to calculate & less time consuming than parametric tests when sample size is small.
- Can be used with any type of data.
- Many non-parametric methods make it possible to work with very small samples, particularly helpful in collecting pilot study data or medical researcher working with a rare disease.

Unit – IV
Research Design and Test of Hypothesis

Dr. G. Yoganandham
Professor and Head,
Department of Economics,
Thiruvalluvar University.

Research Design

- The research design refers to the overall strategy that you choose to integrate the different components of the study in a coherent and logical way, thereby, ensuring you will effectively address the research problem; it constitutes the blueprint for the collection, measurement, and analysis of data.
- The design of a research topic explains the type of research (experimental, survey, correlational, semi-experimental, review) and also its sub-type (experimental design, research problem, descriptive case-study). There are three main types of research design: Data collection, measurement, and analysis.
- The features of **good research design** is often characterized by adjectives like flexible, appropriate, efficient, economical and so on. Generally, the **design** which minimizes bias and maximizes the reliability of the data collected and analyzed is considered a **good design**.

Creation of Research Design

- Identify Intended Learning Outcomes.
- Identify Research Questions.
- Develop Research Design.
- Select Sampling Frame.
- Select Appropriate Data Collection Methods.
- Construct Measurement Instruments.
- Select Appropriate Data Analysis Techniques.
- Consider Communication and Dissemination of Findings.





Research Design - Meaning

- Decisions regarding what, where, when, how much, by what means concerning an inquiry or a research study constitute a research design.
- A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure.
- **Research design have following parts:**
  - Sampling design
  - Observational design
  - Statistical design
  - Operational design.



Parts of Research Design

- **Sampling Designs:**
- Which deals with the methods of selecting items to be observed for the study.
- **Observational design:**
- Which relates to the condition under which the observation are to be made.
- **Statistical Design:**
- Which concern the question of the of How the information and data gathered are to be analyzed ?
- **Operational design:**
- Which deals with techniques by which the procedures satisfied be carried out.



Features of a Good Research Design

- A research design appropriate for a particular research problem, usually involves the consideration of the following factors.
- The means of obtaining information.
- The availability and skills of the researcher and his staff, if any.
- The objective of the problem to be studied.
- The nature of the problem to be studied.
- The availability of time and money for the research work.

Important Concept relating to Research Design

- A concept which can take on different quantitative values is called a variable.
- A phenomena which can take on different qualitatively values even in decimal value are called continues.
- **Dependent and Independent Variables:**
- If one variable is depend on another variable it is termed as a dependent variable.
- The variable that is antecedent to the dependent variable is an independent variable.
- Ex. Height is dependent on age.
- **Extraneous Variables:**
- That are not related to the purpose of the study but may effect on the dependent variables are termed as the extraneous variables.

Example for Extraneous Variable

- Suppose a researcher want to test the hypothesis that there is a relationship between children gains in social studies achievement and their self concept.
- In this case self-concept = independent variable.
- Social studies achievement = dependent variable.
- Intelligence may as well affect on the social achievement.
- But it is not related to the study undertaken by the researcher so it is a Extraneous variable.
- **Control:** One important characteristic of a good research is to minimize the influence or effect of extraneous variables . The technical term ‘control’ is used when we design the study minimizing the effect of extraneous independent variable

Confounded Relationship

- When the dependent variable is not free from the influence of extraneous variables the relationship between the depended and independent variables is said to be confounded by an extraneous variable.
- **Research Hypothesis:**
- When a prediction or a hypothesised relationship is to be tested by scientific methods, it is termed as a research hypothesis.
- The Research Hypothesis is a predicative statement that relates an independent variable to dependent variable.
- **Experimental and Non-experimental Hypothesis Testing:**
- Research in which the independent variable is manipulated is termed as Experimental hypothesis testing research.
- Research in which the independent variable is not manipulated is termed as non- Experimental hypothesis testing research.

Experimental and Control Groups

- When a group is exposed to usual conditions, it is termed as a control group.
- But when the group is exposed to be some special condition, it is termed as Experimental group
- **Treatments:**
- The different conditions under which experimental and control groups are put up usually referred to as treatment.
- **Experiment:**
- The process of examining the truth of a statistical hypothesis, relating to some research problem, is known as an Experiment.
- E.g.:- We can conduct an Experiment to examine the usefulness of a certain newly developed drug.

Different Research Design

- In case of Exploratory Research study
- In case of Descriptive and Diagnostic Research
- In case of Hypothesis-Testing Research studies

Exploratory Research Design

- **Conducted for of gaining better insight into a problem:**
- Evaluate the feasibility of a research project
- Formulation of a statement
- Provide theoretical basis to a hypothesis
- Provide alternative approaches to a problem
- Establish possibilities for further research
- Used when researcher has only a vague idea about the problems during the study.
- Helps in saving time and money.
- Ex:- An advertising company got an account for a new coffee containing chicory, the company started the investigation process with exploratory research in order to identify the situation.
- The researchers found that virtually no one had heard about chicory.
- It wasn't being used, and no-one seemed to know how to put it to use. This resulted in the hypothesis that the advertising could depict the chicory ingredient in whatever way the customer desired.

Descriptive and Diagnostic Research

- Descriptive Research Study's are those studies which are concerned with describing the characteristics of a particular Individual , or a group.
- In diagnostic research study determine the frequency with which some thing occur or is associated with something else.
- Formulating the objective of the study.
- Designing the method of data collection.
- Selecting Sample. (Size)
- Collecting the data.(Where and time)
- Processing and analyzing the data.
- Reporting the findings.
- A descriptive research design should clearly state the subject of the study, object of the study, content of study, time and place of conducting the study and method adopted to conduct the study.

Hypothesis Testing Research Studies

- Hypothesis-testing research studies known as experimental studies are those researcher tests the hypothesis of casual relationship between variables.
- Cause and effect relationship between variable under study.
- Researcher attempts to maintain control and manipulate the variables that affect the study.
- Steps in Experimental Research
- Identification and definition of the problem.
- Formulation of Hypothesis
- Developing experimental design
  - ✓ Selecting the sample subjects
  - ✓ Group or pair subjects
  - ✓ Identify and control non experimental factors
  - ✓ Construct validate an instrument to measure
  - ✓ Determine place, time and duration of experiment
- Conduct the experiment
- Analyse the data and test the hypothesis
- Report the findings.

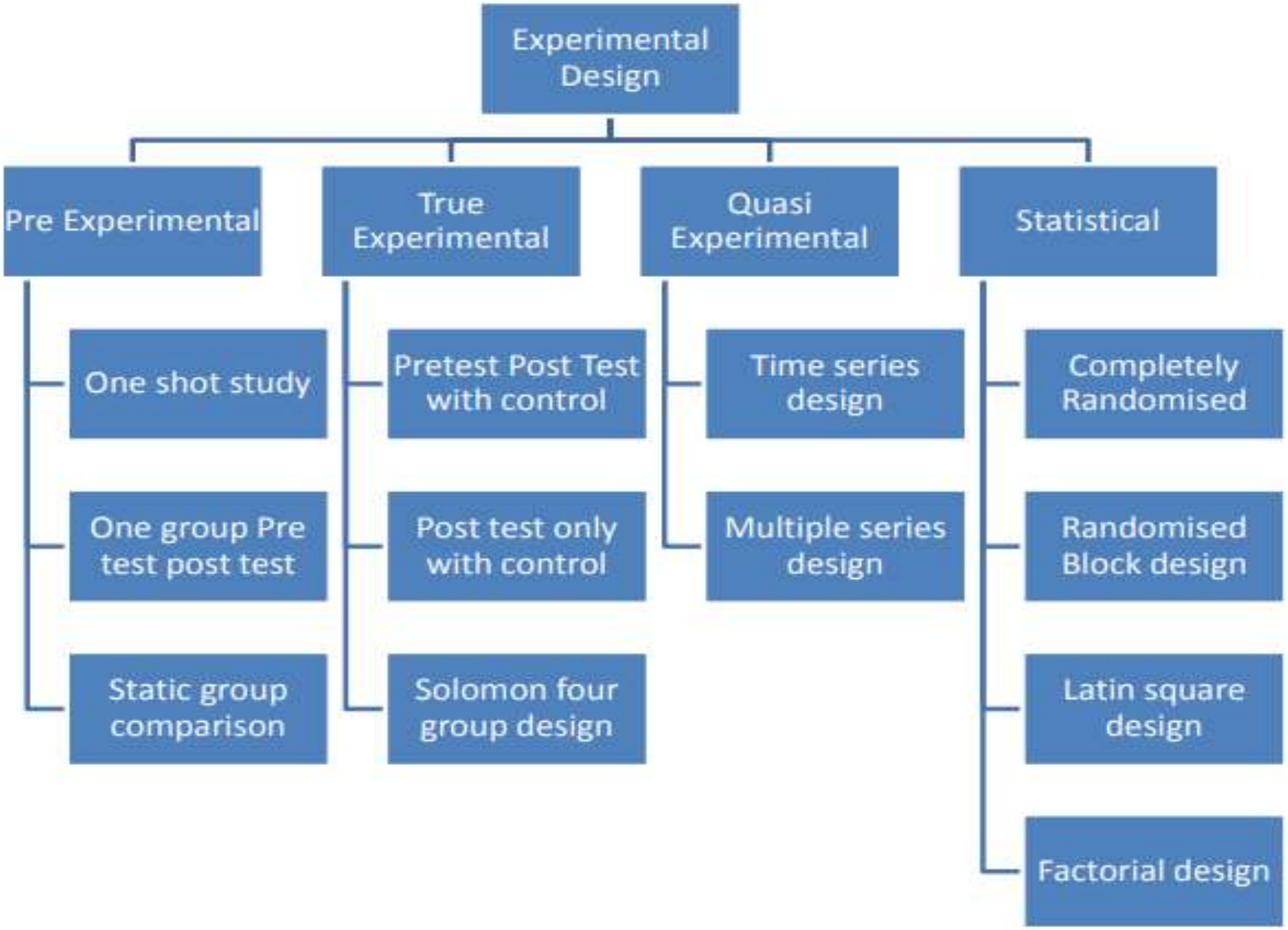


Experimental Research

- Controlling Extraneous Variable
- Randomization
- Matching
- Statistical Control.
- **Principles of Experimental Design**
- Principle of Replication
- Principle of Randomization
- Principle of Local Control.



Types of Experimental Design



Experimental Design

Symbol	Meaning
X	Treatment
O	Observation
R	Randomization
EG	Experimental Group
CG	Control Group
TE	Treatment Effect

Pre Experimental Design

- They follow procedures of Experimental Design but do not include control group.
- One Shot Case Study Researcher attempts to explain a consequence by an antecedent.

$$X \rightarrow O$$

- One Group Pre-Test-Post –Test study.

$$O_1 \xrightarrow{X} O_2$$

$$TE = O_1 - O_2$$

- Static Group Design

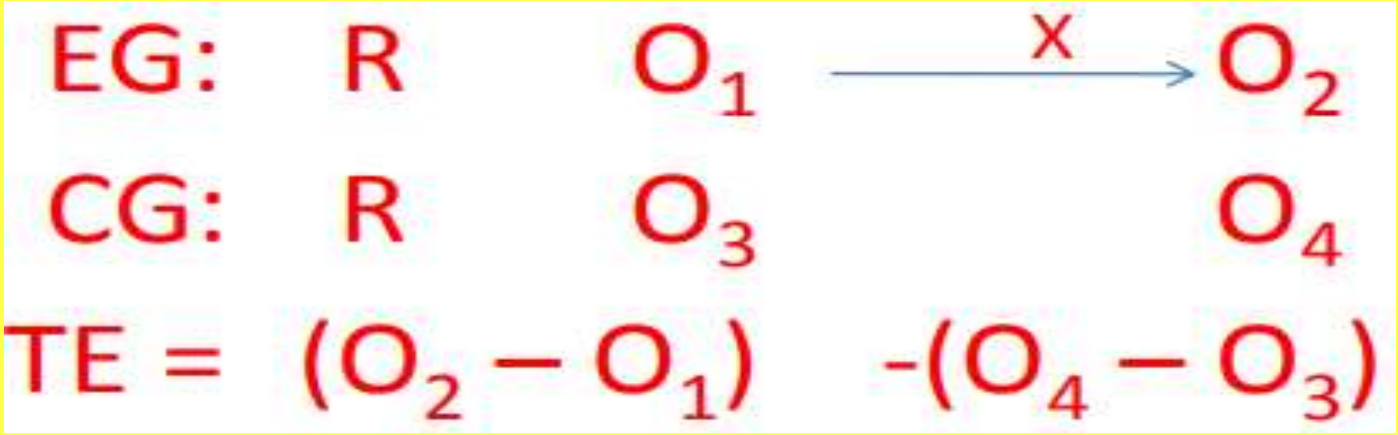
$$EG : \xrightarrow{X} O_1$$

$$CG \qquad O_2$$

$$TE = O_1 - O_2$$

True Experimental Design

- Pre Test Post Test with Control group Design – Study the effects of variable on a carefully controlled sample.



- Post Test Only Control Group design.



True Experimental Design

- Solomon Four-Group Design

EG:	R	O_1	X	O_2
CG:	R	O_3		O_4
EG:	R			O_5
CG:	R			O_6

- $TE = O_2 - O_1$
- $TE = O_2 - O_4$
- $TE = O_5 - O_6$
- $TE = O_5 - O_3$
- $TE = [(O_2 - O_1) - (O_4 - O_3)]$

Research Design

- It is the conceptual structure/general plan of study within which research work is conducted,
- it constitute the blue print for the collection, measurement and analysis of data.
- It focuses on the data collection methods, the research instruments utilized and the sampling plan to be used

Research Design

For Example: an Architect prepares a blueprint before he approves a construction.

- An army prepares a strategy before launching an attack.
- An artist makes a design before he executes his ideas.

Research Design

- So, the researcher makes a plan of his study before undertaking the research work
- Thus, It will enable to save time and resources
- such a plan of study or blue print for study is called a research design or research strategy.

Research Design

- it is the outline/the scheme and the paradigm of the operation of the variables related to the study.
- it is the frame of plan on the basis of which a researcher performs his research work.
- Research designs can be taken as strategy for the research work rather than plan.

Research Design

- IT contains the decisions of the following questions related to the area of study of research.

i. Why is the study being made?

ii Where will the study be carried out?

iii. What is the study about?

iv. How the data will be collected?

Research Design

v. What will be the sampling design?

vi. What period of time is required ?

vii What types of observation is taken?

viii. What types of population is being observed?

ix. What techniques of data processing and analysis is taken?

x. In what style the report of research have to prepare?

Research Design

- According to Kerlinger (1986)

- "Research design is the plan, structure and strategy of investigation conceived so as to obtain answers to research questions and to control variance."

Research Design

- According to P V Young

  "A research design is the logical and systematic planning and directing of a piece of research."

- According to R L Aekoff

  "Design is the process of making decisions before the situation arises in which the decision is to be carried out."

Types of Research Design

According to task

i. Sampling design

ii. Observation design

iii. Statistical design

iv. Operational design

Research Design

Sampling design

- Sampling design deals with the different method of selecting items to be included and observed for the given study. It can be broadly divided in to two classes.
  - Non probability sampling
  - Probability sampling

Observational design

- It relates to the conditions under which the observations are to be made. It is concerned with the decision on the following questions related to observations

i. What should be observed?

ii. What type of method of observation should be taken?

iii. How the accuracy of observation can be assured?

Types (cont..)

- IT can be divided into following categories.
- Participant observational design
- Non-participant observational design
- Controlled observational design
- Non-Controlled observational design

Research Design

Statistical design

- It is concerned with
  - How many items are to be observed
  - How the information / data collected
  - which statistical tools are used for data analysis

Operational design

- IT deals with the techniques
  - by which the procedures specified in the sampling, statistical and observational design can be carried out.

Objective of Research Design

- Two basic purposes

1 To provide the solution to the research problem

  - As validly, objectively, accurately and economically

2. To control variance

  - Maximization of experimental variance
  - To control extraneous variables
  - Minimize error variance

Features of good Research Design

- A good research design Minimizes biasness
- Maximizes the reliability of data collected and analysis
- Use appropriate method of data collection and analysis
- Gives smallest experimental error
- Formulate the problem on the basis of the purpose of study
- Indicates the availability of time and money

Advantages of Research Design

- provides more accurate result
- minimizes the expenditure of money and efforts
- provides the best methods and techniques of data collection and analysis
- It minimizes the uncertainty and confusion associated with any research work.
- It minimizes the waste of time
- It gives guideline to the researcher and moves him on a right track

Main steps in preparing Research design

- Title of the study and statement of the problem
- Review of previous literature
- Definition of concepts or theoretical principles
- Coverage and scope of the study
- Objective of the study
- Formulation of hypothesis

Main steps in preparing Research design

- Sampling design
- Construction of schedule or questionnaire
- Data collection
- Analysis of data
- Interpretation of results
- Reporting the findings

Descriptive Research Design

- It is basically interested in detail description of the phenomenon, group or community
- Example: like number of workers in a factory, their age distribution, level of education, state of their physical health, their community wise distribution and so one.

Descriptive Research Design

- It is essentially a fact finding approach relative largely to the present and abstracting generalizations by the cross sectional study of the current situation.
- It is extensively used in the physical and natural sciences
- Example physical measures, zoological dissects and geological studies of rocks, biological classification etc.
- This type of research study is generally conducted to know the views of the people toward any things.

Descriptive Research Design

- objective
  - to acquire knowledge.
  - There are certain areas where knowledge has not yet properly developed. In such a case the investigator has to confine him to descriptive research design.

Types of Research Design

- **Purposes of Descriptive Research Design**
- To collect detailed information that describes existing phenomena.
- To identify problems or justify current conditions and practices.
- To make comparison and evaluation.
- To determine what others are doing with similar problems or situations and benefit from their experience in making future plans and decisions.

Types of Research Design

- **Limitations of descriptive research design**
  - Makes description and end itself.
  - Essentially creative and demands the discovery of facts in order to leads solution of the problems.
  - It dominates statistical tools to present the facts in terms of average, correlation; dispersion .

Exploratory research design

- Also known as Formulative Research design.
- The main purpose of such designs is that of formulating a research problem for more precise investigation or of developing the working hypothesis from an operational point of view.
- Such a study can have other functions as well, namely more structured investigation of phenomena which an investigator wishes, to study in the subsequent etc.
- Helps in clarifying concepts, establishing priorities for further research etc.

Exploratory research design

- *Kuntz,*
  - "Exploratory studies represent the earlier stage of a science".
  - This design is initial step of problem finding
  - In fact much of social research is of exploratory nature.
  - Since most of the theories in social sciences are either too general or too specific therefore these don't provide clear guidance for empirical research

Exploratory research design

Generally following three steps are considered

- **Survey of concerning literature**
- **Experience survey**
- **Case Study:**

Experimental Research Design

- The Experimental Method provides a logical systematic way to answer the question
- "this is done under carefully controlled conditions, what will happen?".
- To provide a precise answer, experimenters manipulate certain influences or variables and observe how the condition or behavior of the subject is affected or changed
- Experimenters control or isolate the variables in such a way that they can be reasonably sure that the effects they observe can be attributed to the variables they have manipulated, rather than to some other uncontrolled influences.
- The classic method of the laboratory is the most powerful method for discovering and developing a body of knowledge about the prediction and control of events.
- The experimental method has been used with some success in the school classroom, where, to some degree, variables cab be controlled.

Experimental Research Design

- Experimental design provides a plan or blueprint for experimentation that enables the researcher to test hypothesis by reaching valid conclusion about relationships between independent and dependent variable.
- Experimentation is a sophisticated technique for problem solving and may not be an appropriate activity for the beginning researcher.

Experimental Research Design

- **Characteristics**
  - **Manipulation:** the researcher does something to one group of subjects on the study.
  - **Control:** The researcher introduces one or more control groups to compare with the experimental group.
  - **Randomization** :The researcher takes care to random assign subjects to the control and experimental groups. Each subject is given an equal chance of being assigned to either group by assigning them number and blindly selecting the number for each group.

Diagnostic Research Design

- Like descriptive research design diagnostic research design basically interested in detailed description of the phenomenon, group or community.
- The diagnostic design is concerned with discovering and testing certain variables with respect to their associations or dissociation.
- A diagnostic study is concerned with an existing social problem and its basic nature and cause. The diagnostic design is concerned with the causes as well as the treatment.

Diagnostic Research Design

- Objectives :
  - to diagnoses the problem, to accurately specify the characteristics, to determine the frequencies of significant variables and to find out whether certain variances are associated.
  - The research deign for this studies must be carefully planned so that subjective bias can be eliminated.
  - The most important step of the research this design would be objectively define questions, which are to be answered.

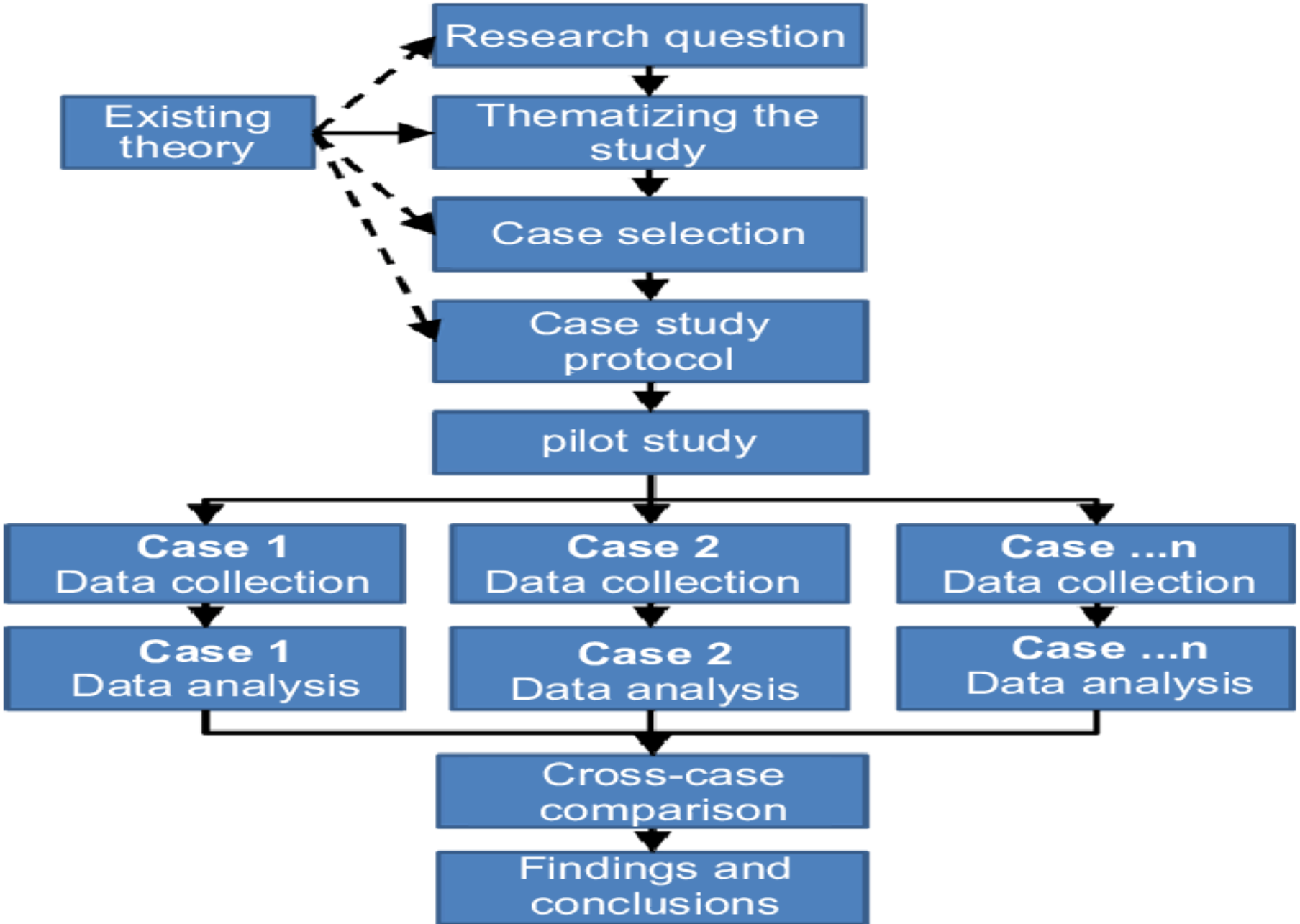
Historical Research Design

- It is connected with past phenomena.
- It is a design of collection, evaluating, verifying, and synthesizing past evidence systematically and objectively to reach a conclusion.
- The present observed phenomenon can be explained if the proper analysis of factors of the past is well done.

Historical Research Design

- It is to be noted that past is the 'starting pointy' of future.
- Hence historical research design is useful for making predictions or for forecasting in future in the field of population, student enrollment, Demand supply, import , export etc

Steps in Research Design



Basic Steps in Research Design - I

- **Defining of a problem:** The problem chosen for research needs to be described clearly in operational terms to make sure that researcher understands positively what information he is searching for and what is strongly related to the study.
- **Data sources:** Once the problem is chosen it is the responsibility of the researcher to mention clearly the different sources of data for example library, personal documents, field work, a specific residential group, etc.
- **Nature of Research:** The research design needs to be expressed with regards to the nature of study to be carried out. The choice of the statistical, experimental or comparative kind of study be made at this time so that the subsequent stages in planning may have relevance to the proposed problem.

Basic Steps in Research Design - II

- **Objective of Study:** Whether the design is aimed at theoretical understanding or presupposes a welfare notion should be explicit at this stage. The objective needs to be very clear in the mind of the investigator as this will lead to the clarity of the design and proper response from the participants.
- **Social – Cultural Context:** The study design needs to be set in the social-cultural context. For instance, in a research of the fertility rate in a people of 'backward' class the context of the so-called backward class of people and the conceptual reference needs to be clarified. Unless the meaning of the term is clearly defined there is often a big deviation in the study since the term backward may have religious, economic and political connotations.
- **Temporal Context:** The physical limit of the design also needs to be known at this step that research linked to be hypothesis does apply to specific social group only.

Basic Steps in Research Design - III

- **Dimension:** It is physically not possible to analyse the data gathered from a big universe. Therefore, the selection of an acceptable and representative sample is a by-word in any study. Based on the dimension of the planned research such a sample of the large people could be chosen to facilitate a practical design.
- **Basis of Selection:** Choosing a right sample is a very important and critical step and this is done with the assistance of some mechanics like drawing a random stratified, deliberate, double cluster or quota sample etc. It will produce a scientifically valid sample in an unbiased manner.
- **Methods of Data Collection:** Related to the research design an appropriate method has to be followed for the collection of necessary data. The relative advantage of observation, interview and questionnaire, when studied collectively will help in the selection of an ideal approach. Once the gathering of data is done, analysis, coding and presentation of the report obviously follow.

Issues in Research Design - I

- **Research Context:**
- One or several organizational contexts.
- Laboratory Vs. Field setting – depends on
  - ✓ Goal of establishing nature of phenomenon (or process) that exists in most contexts.
  - ✓ Generalizability of Results.
  - ✓ Control required of extraneous/nuisance variables.
- **Nature of Sample Size:**
- Random Vs Stratified
- Convenience Sample
- Students Vs Field based sample
- **Size of Sample Determines:**
- Generalizability of results.
- Ability to detect a true effect.

Issues in Research Design - II



- **Data Sources : Self Report**
- ✓ **Common Method Variance:**
  - Similar Response format.
  - Consistency bias of participants.
  - Unreliability of measures.
- ✓ **Social desirability:**
  - Alternative methods to questionnaire (e.g., interviews, behavioral, archival).
  - Multiple times of data collection.
  - Alternative sources (Ex. Supervisor, Peer)
  - Measure of individual's susceptibility to social desirability.

Issues in Research Design - III

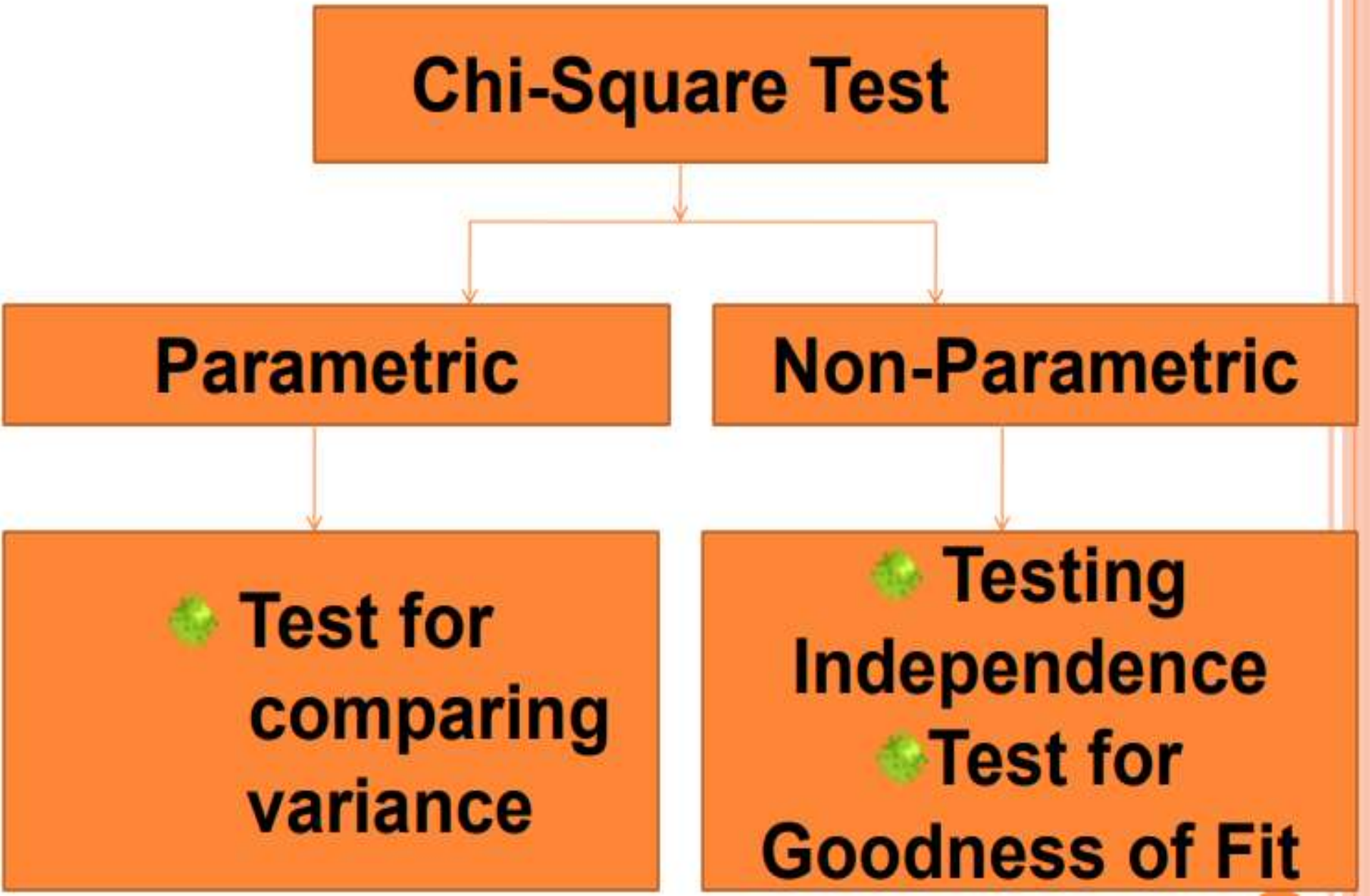
- ✓ **Response shift bias**
  - ➢ Alpha = Real (condition) change.
  - ➢ Beta = Change in frame of reference (change in standards).
  - ➢ Gamma = Change in meaning of construct / variable.
- ❖ **Data Sources: Alternatives**
- ❖ Co-workers (e.g., subordinates, supervisors, peers).
- ❖ Archival
- ❖ Biological
- ❖ Trace
- ❖ Objective term



Chi-Square Test

- The Chi-Square Test of Independence determines whether there is an association between categorical variables (i.e., whether the variables are independent or related).
- It is a nonparametric test. This test is also known as: Chi-Square Test of Association.
- A chi-square goodness of fit test determines if a sample data matches a population.
- A chi-square test for independence compares two variables in a contingency table to see if they are related.
- In a more general sense, it tests to see whether distributions of categorical variables differ from each another.
- If the significance value that is p-value associated with chi-square statistics is 0.002, there is very strong evidence of rejecting the null hypothesis of no fit. It means good fit.

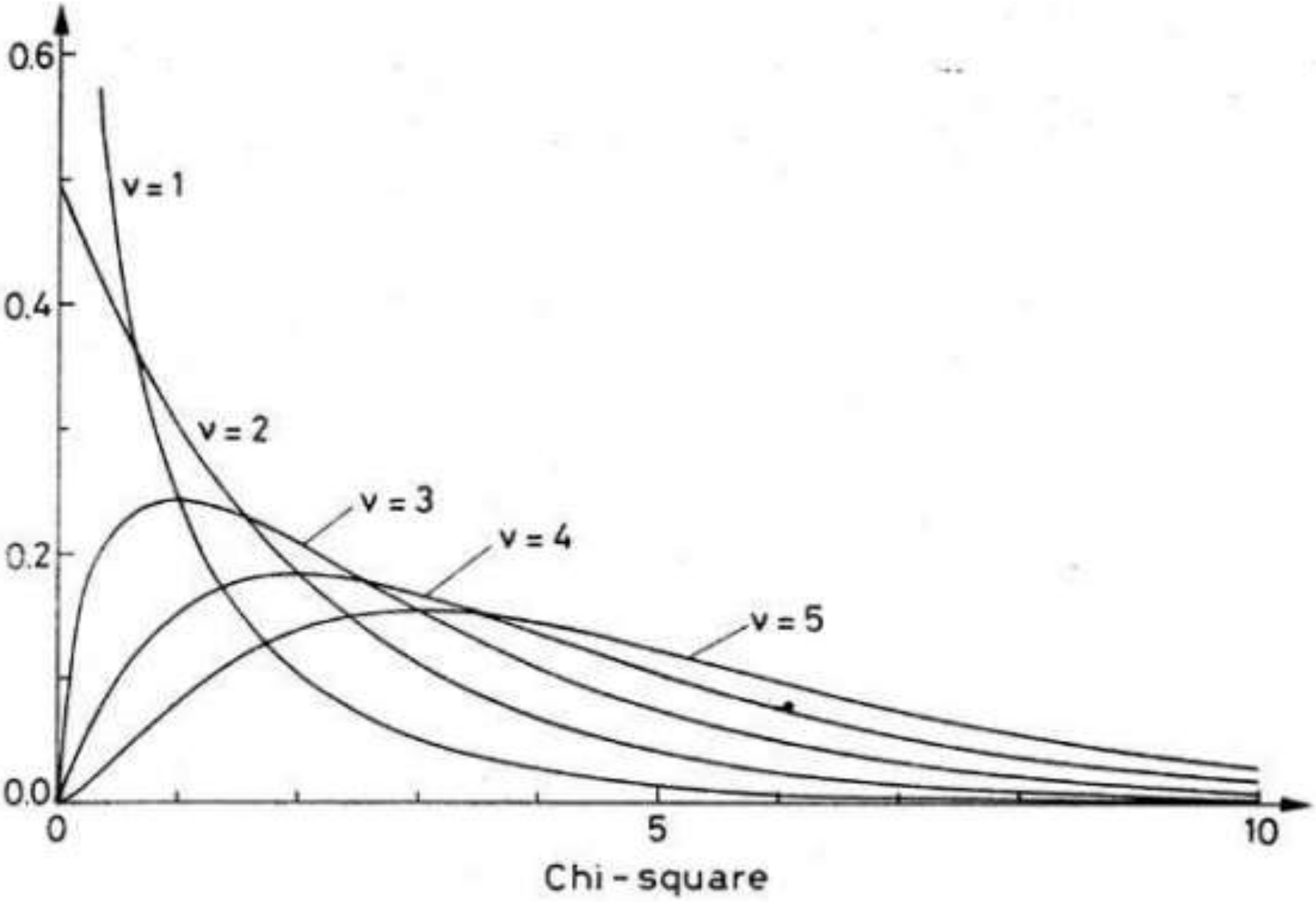
Chi-Square Test



Conditions for the application of $\chi 2$ test

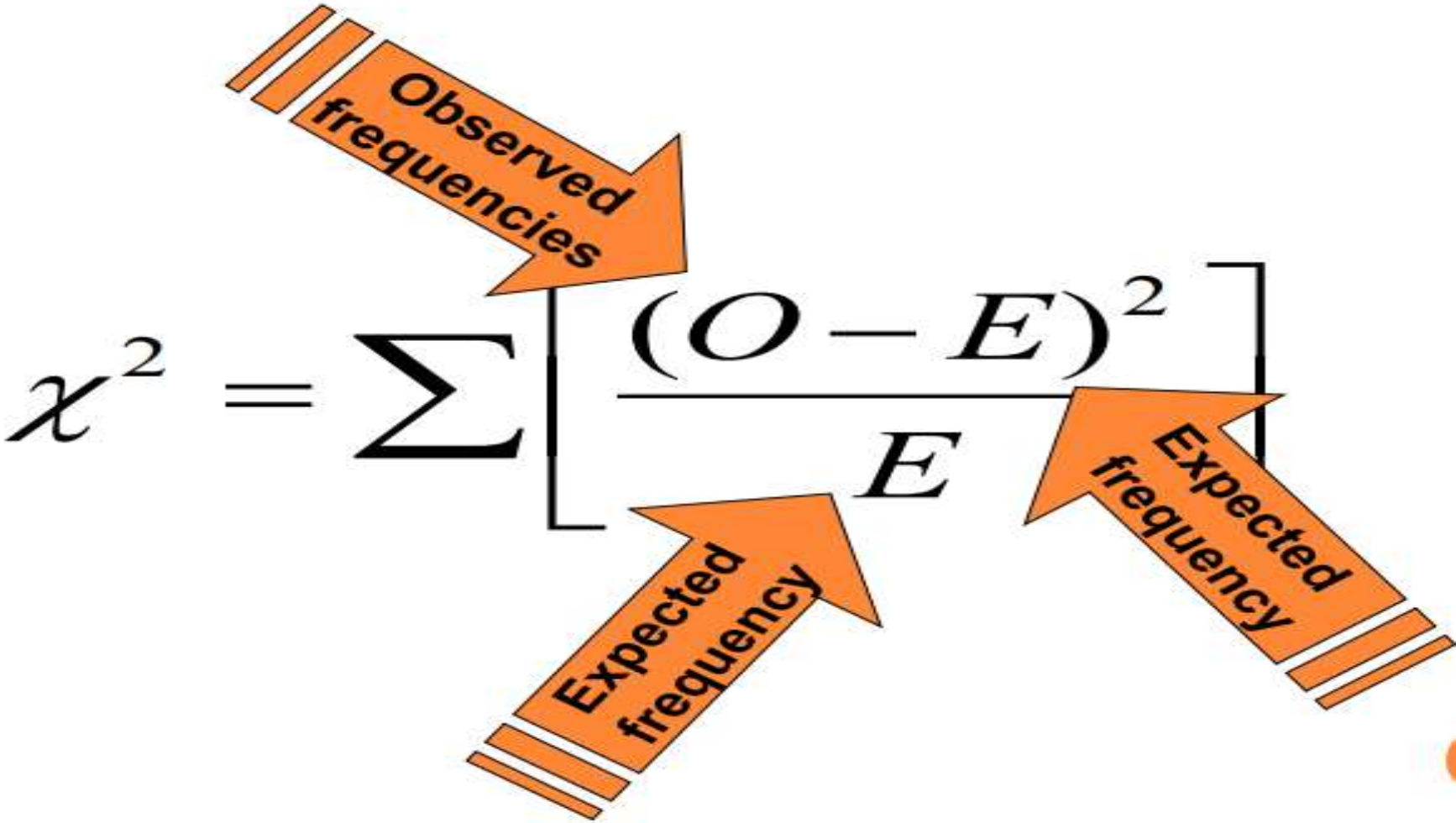
- Observations recorded and collected are collected on random basis.
- All items in the sample must be independent.
- No group should contain very few items, say less than 10.Some statisticians take this number as 5.
- But 10 is regarded as better by most statisticians.
- Total number of items should be large, say at least 50.
- The $\chi 2$ distribution is not symmetrical and all the values are positive. For each degrees of freedom we have asymmetric curves.

Conditions for the application of $\chi 2$ test



Chi – Square Test as a Non- Parametric Test

- Test of Goodness of Fit.
- Test of Independence.



Test of Goodness of Fit

- It enables us to see how well does the assumed theoretical distribution(such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data.
- When the calculated value of $\chi 2$ is less than the table value at certain level of significance, the fit is considered to be good one and if the calculated value is greater than the table value, the fit is not considered to be good.



Test of Independence

- $\chi 2$ test enables us to explain whether or not two attributes are associated.
- Testing independence determines whether two or more observations across two populations are dependent on each other (that is, whether one variable helps to estimate the other.
- If the calculated value is less than the table value at certain level of significance for a given degree of freedom, we conclude that null hypotheses stands which means that two attributes are independent or not associated.
- If calculated value is greater than the table value, we reject the null hypotheses.

Steps involved in Chi - Square Test

- Determine The Hypothesis
  - ✓ Ho : The two variables are independent
  - ✓ Ha : The two variables are associated :
- Select a random sample and record the observed frequencies (O values) in each cell of the contingency table and calculate the row, column, and grand totals.
- Calculate Expected frequencies

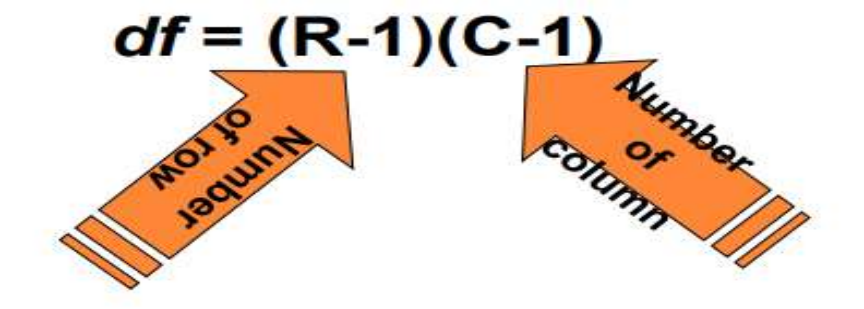
$$E = \frac{(\text{Row total})\,(\text{Column total})}{\text{Grant total}}$$

Steps involved in Chi - Square Test

- Calculate test statistic

$$\chi^2 = \sum \left[\frac{(O-E)^2}{E} \right]$$

- Determine Degrees of Freedom



- Using a level of significance α and df, find the critical (table) value of $x^2\alpha$ (see Appendix). This value of $x^2\alpha$ corresponds to an area in the right tail of the distribution.

Computer computed test statistic against a tabled / critical value

- The computed value of the Pearson chi-square statistic is compared with the critical value to determine if the computed value is improbable.
- The critical tabled values are based on sampling distributions of the Pearson chi-square statistic.
- If calculated $\chi 2$ is greater than $\chi 2$ table value, reject Ho.

F - Test

- An "F Test" is a catch-all term for **any test that uses the F-distribution**. In most cases, when people talk about the F-Test, what they are actually talking about is The *F-Test to Compare Two Variances*. However, the f-statistic is used in a variety of tests including regression analysis, the chow test and the Scheffe test (a post-hoc ANOVA test).

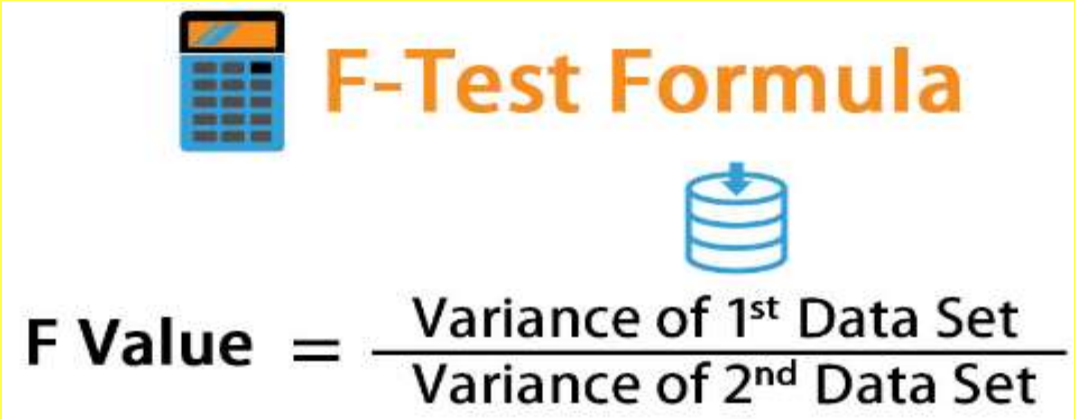


General Steps for F - Test

- If you're running an F Test, you should use Excel, SPSS, Minitab or some other kind of technology to run the test. Why? Calculating the F test by hand, including variances, is tedious and time-consuming. Therefore you'll probably make some errors along the way.
- If you're running an F Test using technology (for example, an F Test two sample for variances in Excel), the only steps you really need to do are Step 1 and 4 (dealing with the null hypothesis). Technology will calculate Steps 2 and 3 for you.
- State the null hypothesis and the alternative hypothesis.
- Calculate the F value. The F Value is calculated using the formula F = (SSE_1 – SSE_2 / m) / SSE_2 / n-k, where SSE = residual sum of squares, m = number of restrictions and k = number of independent variables.
- Find the F Statistic (the critical value for this test). The F statistic formula is: **F Statistic = variance of the group means / mean of the within group variances.** You can find the F Statistic in the F – Table.
- Support or Reject null hypothesis.

F – Test to compare two variances

- A **Statistical F Test** uses an F statistic to compare two variances, s_1 and s_2, by dividing them. The result is always a positive number (because variances are always positive). The equation for comparing two variances with the f-test is: $F = s^2_1 / s^2_2$
- If the variances are equal, the ratio of the variances will equal 1. For example, if you had two data sets with a sample 1 (variance of 10) and a sample 2 (variance of 10), the ratio would be 10/10 = 1.
- You **always** test that the population variances are equal when running an F Test. In other words, you always assume that the variances are equal to 1. Therefore, your null hypothesis will always be that *the variances are equal*.



Assumptions of F – Test

- Several **assumptions** are made for the test. Your population **must be approximately normally distributed** (i.e. fit the shape of a bell curve) in order to use the test. Plus, the samples must be independent events. In addition, you'll want to bear in mind a few important points:
- The larger variance should always go in the numerator (the top number) to force the test into a right-tailed test. Right-tailed tests are easier to calculate.
- For two tailed test, divide alpha by 2 before finding the right critical value.
- If you are given standard deviations, they must be squared to get the variances.
- If your degrees of freedom aren't listed in the F Table, use the larger critical value. This helps to avoid the possibility of Type I Error.

Two Tailed F – Test

- $H_0 : \sigma_1^2 = \sigma_2^2$
- $H_1 : \sigma_1^2 \neq \sigma_2^2$

Sample 1: Variance = 109.63, sample size = 41.
Sample 2: Variance = 65.99, sample size = 21.

① Fcritical value

$\frac{109.63}{65.99} = 1.66$



② DF

41 − 1 = 40 num

21 − 1 = 20 den

③ Alpha Level 0.05 5%

2.5% 0.025

Two Tailed F – Test

- The difference between running a one or two tailed F test is that the alpha level needs to be halved for two tailed F tests. For example, instead of working at $\alpha = 0.05$, you use $\alpha = 0.025$; Instead of working at $\alpha = 0.01$, you use $\alpha = 0.005$.
- With a two tailed F test, you just want to know if the variances are not equal to each other. In notation: $H_a = \sigma^2 1 \neq \sigma^2 2$
- **Example problem:** Conduct a two tailed F Test on the following samples: Sample 1: Variance = 109.63, sample size = 41. Sample 2: Variance = 65.99, sample size = 21.
- Step 1: Write your hypothesis statements: H_o: No difference in variances. H_a: Difference in variances.
- Step 2: Calculate your F critical value. Put the highest variance as the numerator and the lowest variance as the denominator: F Statistic = variance 1/ variance 2 = 109.63 / 65.99 = 1.66
- Step 3: Calculate the degree of freedom: The degrees of freedom in the table will be the sample size -1, so: Sample 1 has 40 df (the numerator). Sample 2 has 20 df (the denominator).

Two Tailed F – Test

- Step 5: Find the critical F Value using the F Table. There are several tables, so make sure you look in the alpha = .025 table. Critical F (40,20) at alpha (0.025) = 2.287.
- Step 6: Compare your calculated value (Step 2) to your table value (Step 5). If your calculated value is higher than the table value, you can reject the null hypothesis:
- F calculated value: 1.66
- F value from table: 2.287.
- 1.66 < .287.
- So we cannot reject the null hypothesis.



Standard Deviation Formula

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

The standard deviation is represented by the symbol δ and variance is square of the standard deviation.

The formula used to calculate F Test is,

F Test Formula

$$F = \frac{\text{estimate of } \sigma^2 \text{ from means}}{\text{estimate of } \sigma^2 \text{ from individuals}}$$

$$F = \frac{\text{Variance between Treatments}}{\text{Variance within Treatments}}$$

$$F = \frac{\text{Variance of Treatments}}{\text{Variance of Error}}$$

Statistics Tip: The F and Chi- Square Test of variance

- Most users of statistics are familiar with the *F*-test for Variances. But there is also a Chi-Square Test for the Variance. What's the difference?
- The *F*-test compares the Variances from 2 different Populations or Processes. It basically divides one Variance by the other and uses the appropriate *F* Distribution to determine whether there is a Statistically Significant difference.
- If you're familiar with *t*-tests, the *F*-test is analogous to the 2-Sample *t*-test. The *F*-test is a Parametric test. It requires that the data from both the 2 Samples each be roughly Normal.
- The following compare-and-contrast table may help clarify these concepts:

	F-test	Chi-Square Test for the Variance
Compares	Variances from 2 Populations or Processed	Variance from 1 Population or Process to a Specified Variance.
Analogous t-test	2-sample	1-Sample
Data restriction	roughly Normal	none

T- Test Vs Chi-Square Test

- **T – Test: (Parametric):**
- Used to compare two groups means ex. Control Vs Experimental Groups.
- Can only be used for two groups (not multiple).
- Assume the data to be have a normal distribution (bell-shaped curve).
- Similar standard deviation.
- Data must be measured values(ex. Height).
- **Chi-Square Test: (Non-parametric):**
- Can be used only on raw counts (no measurements).
- Sample size must be more than 20.
- Only used to compare an experimental result w/a theoretical outcome.
- It aims to test the null hypothesis of No Difference between data sets.



Z-Test	T-Test
Data size is greater than 30	Data size is smaller than 30
Data points isn't related or doesn't affect another data point. For example, in our example, both the control and the experimental group were independent from each other	Data points might be related to each other, where the behaviour or value of one point may affect another
Data is normally distributed (if data size > 30 we can assume it is)	Data is not normally distributed
Data was randomly selected from a broader population	Data wasn't randomly selected
Sample sizes should be equal if possible	Sample sizes are not equally

F Test: Example Solution

- Form the hypothesis test:

  $H_0: \sigma_1^2 = \sigma_2^2$ (there is no difference between variances)

  $H_1: \sigma_1^2 \neq \sigma_2^2$ (there is a difference between variances)

- Find the F critical value for $\alpha = 0.05$:

- Numerator d.f. = $n_1 - 1 = 21 - 1 = 20$

- Denominator d.f. = $n2 - 1 = 25 - 1 = 24$

- $F_{\alpha/2} = F_{.025, 20, 24} = 2.33$

Applications of F – Test

- Identify a research question from your professional life or career specialization that can be addressed by a one-way ANOVA.
- Indicate why a one-way ANOVA would be the appropriate analysis for this research question.
- Describe the variables and their scale of measurement.
- Discuss the expected outcome.
- We can reject the null hypothesis that black men and black women have identical earnings functions.
- Do we really need the interaction terms, or do we get the same explanatory power by simply giving black women a different intercept ?
- Let us test the null hypothesis that the interaction coefficients are both 0.

T – Test

- The t test tells you how significant the differences between groups are; In other words it lets you know if those differences (measured in means) could have happened by chance.
- **A very simple example:** Let's say you have a cold and you try a naturopathic remedy. Your cold lasts a couple of days. The next time you have a cold, you buy an over-the-counter pharmaceutical and the cold lasts a week.
- You survey your friends and they all tell you that their colds were of a shorter duration (an average of 3 days) when they took the homeopathic remedy.
- What you *really* want to know is, are these results repeatable? A t test can tell you by comparing the means of the two groups and letting you know the probability of those results happening by chance.

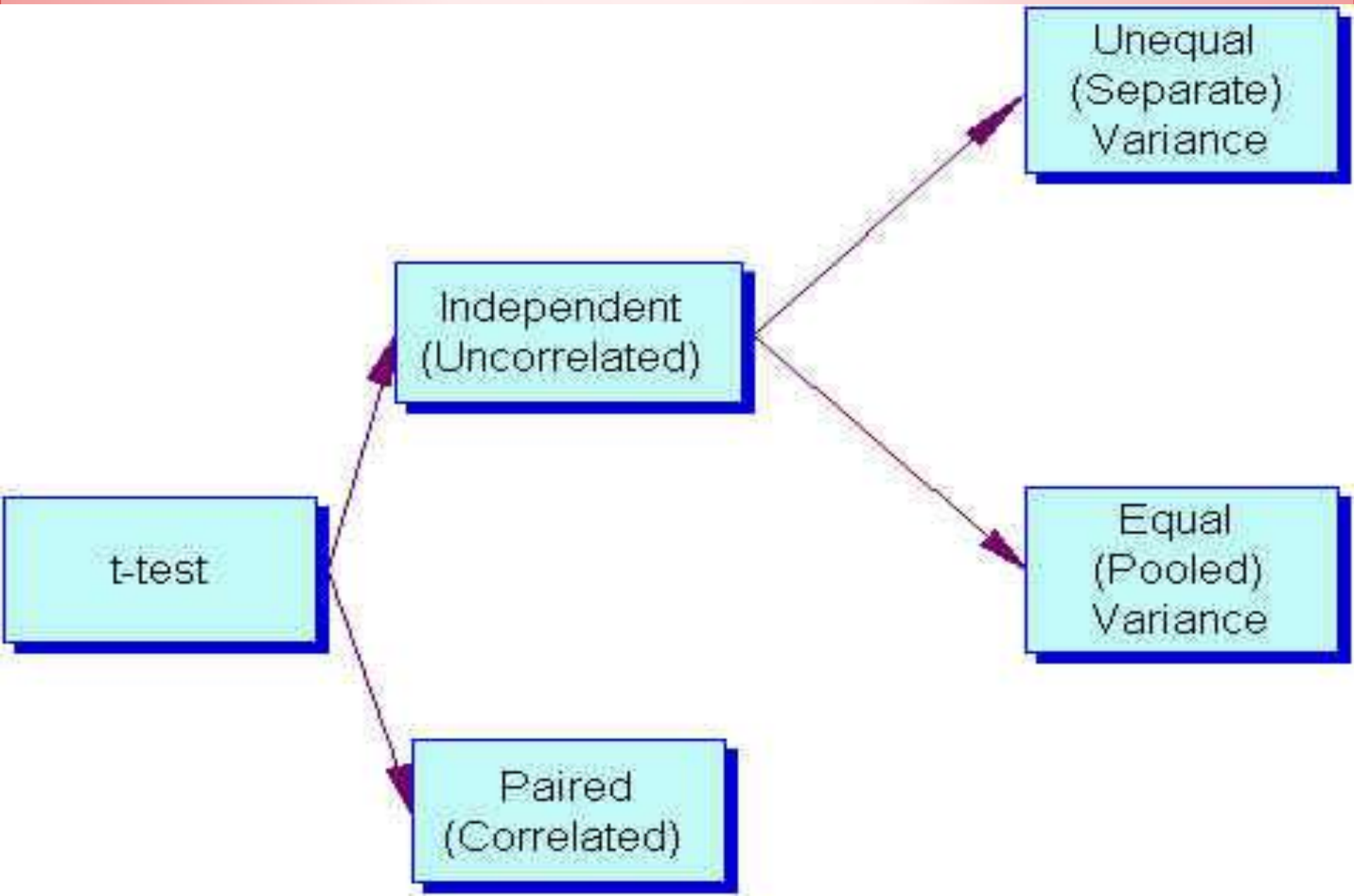
T – Test

- **Another example:** Student's T-tests can be used in real life to compare averages. For example, a drug company may want to test a new cancer drug to find out if it improves life expectancy.
- In an experiment, there's always a control group (a group who are given a placebo, or "sugar pill"). The control group may show an average life expectancy of +5 years, while the group taking the new drug might have a life expectancy of +6 years.
- It would seem that the drug might work. But it could be due to a fluke.
- To test this, researchers would use a Student's t-test to find out if the results are repeatable for an entire population.

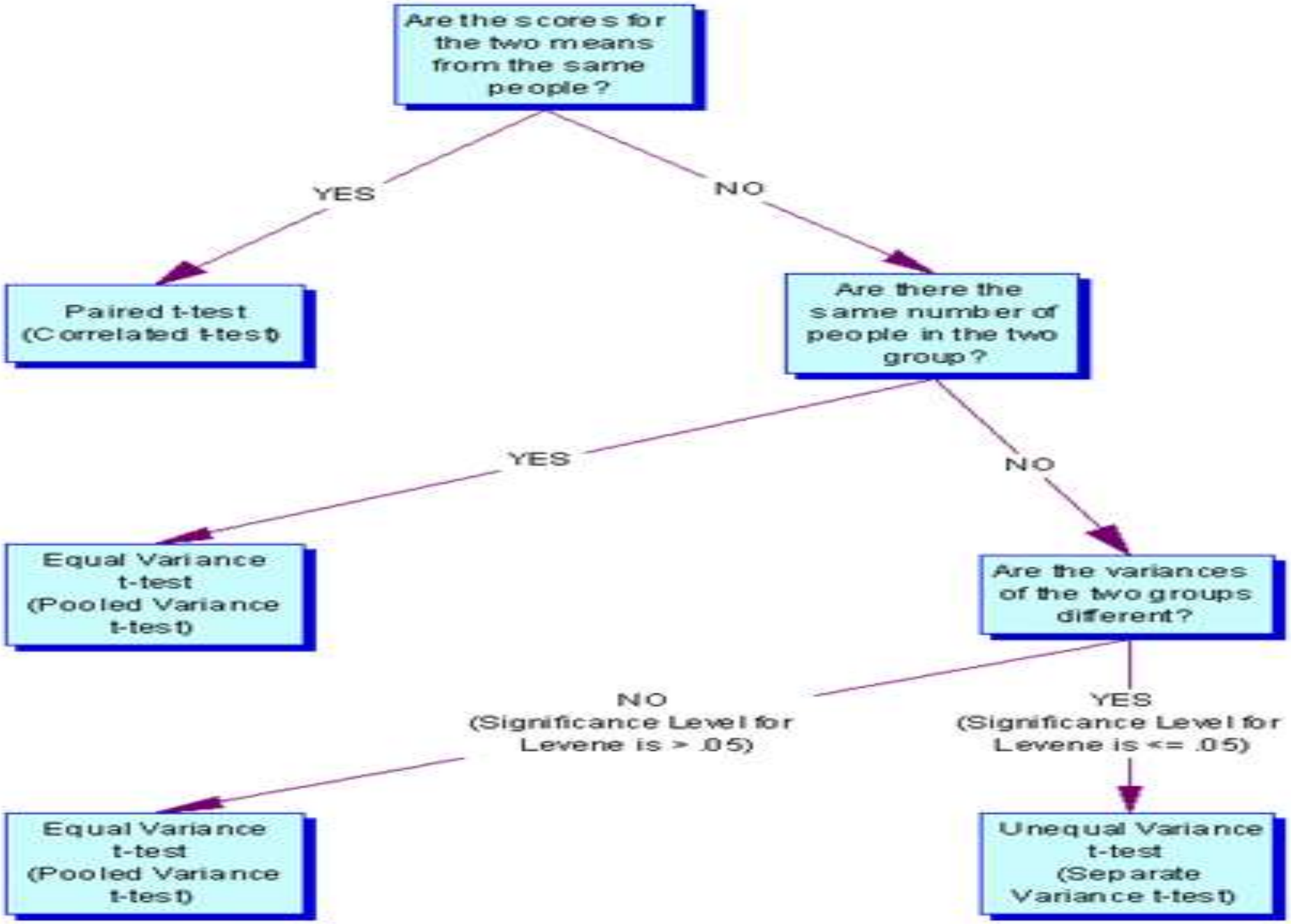
T – Test

- A **t-test** is a type of inferential statistic **used to** determine if there is a significant difference between the means of two groups, which may be related in certain features. The **t-test** is one of many **tests used for** the purpose of hypothesis **testing** in statistics.
- **There are three main types of t-test:**
  - An Independent Samples **t-test** compares the means for two groups.
  - A Paired sample **t-test** compares means from the same group at **different** times (say, one year apart).
  - A One sample **t-test tests** the mean of a single group against a known mean.

T – Test


t-test
Independent
(Uncorrelated)
Paired
(Correlated)
Unequal
(Separate)
Variance
Equal
(Pooled)
Variance

Procedure to select which type of T – Test to be used



Difference between Chi-Square & T – Test

- The **t-test** allows you to say either "we can reject the null hypothesis of equal means at the 0.05 level" or "we have insufficient evidence to reject the null of equal means at the 0.05 level."
- A **chi-square test** allows you to say either "we can reject the null hypothesis of no relationship at the 0.05 level" .
- The independent **t-test formula** is used to compare the means of two independent groups.
- The independent samples **t-test** comes in two different forms: the standard Student's **t-test**, which assumes that the variance of the two groups are equal.
- **P-Value in T-Test:** Recall that probability equals the area under the probability curve. The **P-value** is therefore the area under a $\mathbf{t}_{n-1} = \mathbf{t}_{14}$ curve and to the right of the **test** statistic **t*** = 2.5. It can be shown using statistical software that the **P-value** is 0.0127.

Uses of Z – Test, T – Test & ANOVA

- **Z-tests are** statistical calculations that **can** be **used** to compare population means to a sample's. **T-tests are** calculations **used** to **test** a hypothesis, but they **are** most useful when **we** need to determine if there **is** a statistically significant difference between two independent sample groups.
- A **t-test** is a statistical **test** that compares the means of two samples. If you are studying one group, **use** a paired **t-test** to compare the group mean over time or after an intervention, or **use** a one-sample **t-test** to compare the group mean to a standard value. If you are studying two groups, **use** a two-sample **t-test**.
- The **t-test** is a method that determines whether two populations are statistically **different** from each other, whereas **ANOVA** determines whether three or more populations are statistically **different** from each other.

T – Score

- The t score is a ratio between the **difference between two groups and the difference within the groups**. The larger the t score, the more difference there is between groups.
- The smaller the t score, the more similarity there is between groups. A t score of 3 means that the groups are three times as different *from* each other as they are within each other. When you run a t test, the bigger the t-value, the more likely it is that the results are repeatable.
- A large t-score tells you that the groups are different.
- A small t-score tells you that the groups are similar.
- T-Values and P-values
- How big is "big enough"? Every t-value has a p - value to go with it. A p-value is the probability that the results from your sample data occurred by chance. P-values are from 0% to 100%.
- They are usually written as a decimal. For example, a p value of 5% is 0.05. **Low p-values are good**; They indicate your data did not occur by chance. For example, a p-value of .01 means there is only a 1% probability that the results from an experiment happened by chance. In most cases, a p-value of 0.05 (5%) is accepted to mean the data is valid.

Paired T - Test

- A paired t test (also called a **correlated pairs t-test**, a **paired samples t test** or **dependent samples t test**) is where you run a t test on dependent samples. Dependent samples are essentially connected — they are tests on the same person or thing. For example:
  - ✓ Knee MRI costs at two different hospitals,
  - ✓ Two tests on the same person before and after training,
  - ✓ Two blood pressure measurements on the same person using different equipment.
- Choose the paired t-test if you have two measurements on the same item, person or thing. You should also choose this test if you have two items that are being measured with a unique condition.
- For example, you might be measuring car safety performance in vehicle research and testing and subject the cars to a series of crash tests. Although the manufacturers are different, you might be subjecting them to the same conditions.



Paired T - Test

- With a "regular" two sample t test, you're comparing the means for two different samples.
- For example, you might test two different groups of customer service associates on a business-related test or testing students from two universities on their English skills.
- If you take a random sample each group separately and they have different conditions, your samples are independent and you should run an independent sample t-test (also called between-samples and unpaired-samples).
- The null hypothesis for the for the independent samples t-test is $\mu_1 = \mu_2$. In other words, it assumes the means are equal.
- With the paired t test, the null hypothesis is that the *pairwise difference* between the two tests is equal (H_0: $\mu_d = 0$).
- The difference between the two tests is very subtle; which one you choose is based on your data collection method.

Chi-Square Test ($\chi 2$)

- The **Chi-square test** of association evaluates relationships between categorical variables. Like any statistical hypothesis **test**, the **Chi-square test** has both a null hypothesis and an alternative hypothesis. Null hypothesis: There are no relationships between the categorical variables.
- Advantages of the Chi-square include its robustness with respect to distribution of the data, its **ease** of computation, the detailed information that can be derived from the test, its use in studies for which parametric assumptions cannot be met, and its **flexibility** in handling data from both two group and multiple groups.
- A **chi-square** is only a nonparametric criterion. You can make comparisons for each characteristic. You can also use Factorial ANOVA. In Factorial ANOVA, you can investigate the dependence of a quantitative characteristic (dependent variable) on one or more qualitative characteristics (category predictors).

Formula for T- Test

$$s_{X_1X_2} = \sqrt{\frac{(n_1 - 1)s_{X_1}^2 + (n_2 - 1)s_{X_2}^2}{n_1 + n_2 - 2}}.$$

$$s_{x_1x_2} = \sqrt{\frac{(20 - 1) * 12.48^2 + (23 - 1) * 10.51^2}{20 + 23 - 2}} = 11.47$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{9 - 12}{11.47 * \sqrt{\frac{1}{20} + \frac{1}{23}}} = -0.86$$

Calculation T - Statistic

- Below is the formula for the two-sample t-test, where:
- t is the t-statistic.
- x1 is the average NPS for men → 9.
- x2 is the average for women → 12.
- n1 is the number of men who provided a response to the NPS question → say 20 men responded to the survey.
- n2 is the number of women → 23 women responded.
- s1 is the standard deviation of the NPS for men → say the calculated standard deviation is 12.48.
- s2 is the standard deviation of the NPS for women → the calculated standard deviation is 10.51.

Calculation of Degrees of Freedom

- This formula must be used to determine degrees of freedom in two-sample t-tests. The formulas for other types of test are included below.

$$df = n_1 + n_2 - 2 = 20 + 23 - 2 = 41$$

- **Determine critical value:**
- According to this table, for a two-tailed test with an alpha level of 0.05 at 41 degrees of freedom, the critical value is 2.02. Note that most analysts use a two-tailed test instead of a one-tailed test because it's more conservative.
- **Compare absolute value of the t-statistic to critical value:**
- Since the absolute value of the t-statistic is 0.86, which is not larger than the critical value of 2.02, then you can conclude that men do not give a significantly lower NPS ratings than women.

One and Two- tailed test

	one-tailed test		two-tailed test
hypothesis	$H_0 : \mu_1 \geq \mu_0$ $H_1 : \mu_1 < \mu_0$	$H_0 : \mu_1 \leq \mu_0$ $H_1 : \mu_1 > \mu_0$	$H_0 : \mu_1 = \mu_0$ $H_1 : \mu_1 \neq \mu_0$
test statistic (normal distribution)	$z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$		
deg. of freedom	—		
rejection	reject H_0 if $z < -z_\alpha$	reject H_0 if $z > z_\alpha$	reject H_0 if $\lvert z \rvert > z_{\alpha/2}$

Features of T - Test

	Type of *t* Test		
Feature of *t* Test	One-Sample	Independent Measure	Dependent Measure
s^2 is known	No	No	No
μ is known	Yes	No	No
Number of scores for each participant	1	1	2
t test carried out on different scores	No	No	Yes
Formula for *df*	$df = N-1$	$df_{total} = df_1 + df_2$ ($df_1 = N_1-1$; $df_2 = N_2-1$)	$df = N-1$
What is *N*	Number of scores	Number of scores	Number of pairs

Definition of t - Test

- A *t* - test is a hypothesis-testing procedure in which the population variance is unknown; it compares *t* scores from a sample to a comparison distribution called a *t* distribution.
- **Types of t – Test:**
- One sample t-test - used when σ (population standard deviation) is unknown.
- Independent measure t-test - compares the means from 2 totally independent (uncorrelated) groups.
- Dependent t-test - compares the means from 2 related groups.

One sample t-test

- A one-sample t test is used to answer the question: Does the sample mean (*M*) equal the population mean (σ)?
- Additionally, a one-sample *t* test is used when σ is unknown.
- **Formula**
- **Symbols**

$$t = \frac{M - \mu}{SE_M} \qquad SE_M = \frac{s}{\sqrt{n}}$$

- *M* = sample mean
- μ = population mean
- SE_M = estimated standard error of the mean
- *s* = estimated population standard deviation
- *n* = total number of scores
- degrees of freedom (*df*) = *n* - 1

Example for One sample t-test

- In each of the following studies, a single sample's mean is being compared to a population with a known mean but an unknown variance. For each study, decide whether the result is significant.

Study	n	μ	s^2	M	*Tails*	Significance Level
(a)	64	12.40	9.00	11.00	1 (low)	.05

(a)

$H_o: M \leq \mu$

$H_1: M > \mu$

$$t = \frac{11-12.4}{SE_M} = \frac{11-12.4}{0.375} = \frac{-1.4}{0.375} = -3.73$$

$$SE_M = \frac{\sqrt{9}}{\sqrt{64}} = \frac{3}{8} = 0.375$$

$df = 64\text{-}1 = 63$

$t_{critical} = -1.671$

ANSWER: Since $t = -3.73$ falls within the $t_{critical}$ region, the null hypothesis (H_o) is rejected.

Independent measures t - test

- An independent measures t – test compares the means from 2 totally independent groups.

$$t_{IND} = \frac{M_1 - M_2}{SE_D} \quad SE_D = \sqrt{\left(\frac{(df_1)(s_1^2)+(df_2)(s_2^2)}{df_{total}}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

- **Symbols**
- M = sample mean
- SE_D = estimated standard error of the difference
- s^2 = estimated population variance
- n = total number of scores
- degrees of freedom ($df = df_{total}$) = ($df_1 + df_2$) = [(N_1-1)+(N_2-1)]

Example for Independent measures t - test

- A communication researcher randomly assigned 82 volunteers to 1 of 2 experimental groups. Sixty-one were instructed to get their news for a month only from television, and 21 were instructed to get their news for a month only from the Internet. (Why the researcher didn't assign equal numbers to the two conditions is a mystery).

- After the month was up, all participants were tested on their knowledge of several political issues. The researcher did not have a prediction as to which news source would make people more knowledgeable. That is, the researcher simply predicted that there is some kind of difference. These were the results of the study:

- TV Group Internet Group
- $M = 24$ $M = 26$
- $S^2 = 4$ $S^2 = 6$
- *Using the .01 level of confidence, what should the researcher conclude?*

Example for Independent measures t - test

H_o: M=μ

H_1: M≠μ

	TV Group (1)	Internet Group (2)
M	24	26
S^2	4	6
n	61	21
df	60	20
df_{total}	80	

$$t = \frac{24-26}{SE_D} = \frac{24-26}{0.5367} = \frac{-2}{0.5367} = -3.7265$$

$$SE_D = \sqrt{\left(\frac{(60)(4)+(20)(6)}{80}\right)\left(\frac{1}{61}+\frac{1}{21}\right)} = \sqrt{(4.5)(0.0640)} = \sqrt{0.288} = 0.5367$$

.01 level of confidence (2 tail)

$t_{critical}$ = 2.639

ANSWER: Since t = -3.7265 falls within the $t_{critical}$ region, the null hypothesis (H_o) is rejected.

Dependent measures t - test

- The dependent measures *t* test is used to compare means from 2 related groups.
- In order to be considered related, both groups must meet one of the following criteria:
- the same people are in each group (i.e. pre/post, before/after).
- different people are in each group, but there is some sort of relationship between them (i.e. couples in a relationship with each other) the two groups are matched.

Formula

$$t_{DEP} = \frac{M_D}{SE_{MD}} \quad \Rightarrow \quad SE_{MD} = \frac{s_D}{\sqrt{n}}$$

Symbols

M_D = mean difference

SE_{MD} = estimated standard error of the mean difference

s_D = estimated (population) standard deviation of the difference

n = total number of ***pairs*** of scores

Example for Dependent measures t - test

Example:

A researcher assesses the level of a particular hormone in the blood in 5 patients before and after they begin taking a hormone treatment program. Results for the five are as follows:

Patient	Before	After
A	.20	.18
B	.16	.16
C	.24	.20
D	.22	.19
E	.17	.16

Using the .05 level of significance, was there a significant change in the level of this hormone?

$H_0: M_1 = M_2$

$H_1: M_1 \neq M_2$

Example for Dependent measures t - test

Patient	Before	After	D	D^2
A	.20	.18	(.20-.18)= 0.02	(0.02^2)= 0.0004
B	.16	.16	(.16-.16)= 0.00	(0.00^2)= 0.0000
C	.24	.20	(.24-.20)= 0.04	(0.04^2)= 0.0016
D	.22	.19	(.22-.19)= 0.03	(0.03^2)= 0.0009
E	.17	.16	(.17-.16)= 0.01	(0.01^2)= 0.0001
Σ			0.10	0.003
M			0.02	0.0006
s			0.0255	

$$t = \frac{0.02}{SE_{MD}} = \frac{0.02}{0.114} = 0.175$$

$$SE_{MD} = \frac{0.0255}{\sqrt{5}} = 0.114$$

You can use either the computational or the conceptual formula to calculate the estimated standard deviation of the difference. For a review of these formulas, click the tab on the left sidebar.

level of confidence = .05 (2 tail)
df = 5-1 = 4
$t_{critical}$ = 2.776

ANSWER: Since t = 0.175 falls outside of the $t_{critical}$ region, the null hypothesis (H_o) is accepted.

Analysis of Variance (ANOVA)

- An ANOVA test is a way to find out if survey or experiment results are significant, in other words, they help you to figure out if you need to reject the null hypothesis or accept the alternative hypothesis.
- **Examples of testing different groups:**
- A group of psychiatric patients are trying three different therapies: counseling, medication and biofeedback. You want to see if one therapy is better than the others.
- A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.
- Students from different colleges take the same exam. You want to see if one college outperforms the other.

$$F = \frac{\text{Estimate of population variance based on between samples variance}}{\text{Estimate of population variance based on within samples variance}}$$

One-way and Two-way ANOVA

- **One-way** or **two-way** refers to the number of independent variables (IVs) in your Analysis of Variance test.
- One-way has one independent variable (with 2 levels). For example: *brand of cereal*,
- Two-way has two independent variables (it can have multiple levels). For example: *brand of cereal, calories*.
- **Groups or Levels:**
- Groups or levels are different groups within the same independent variable. In the above example, your levels for "brand of cereal" might be Lucky Charms, Raisin Bran, Cornflakes — a total of three levels. Your levels for "Calories" might be: sweetened, unsweetened — a total of two levels.
- Let's say you are studying if an alcoholic support group and individual counseling combined is the most effective treatment for lowering alcohol consumption. You might split the study participants into three groups or levels:
  - ✓ Medication only,
  - ✓ Medication and counseling,
  - ✓ Counseling only.
- Your dependent variable would be the number of alcoholic beverages consumed per day.
- If your groups or levels have a hierarchical structure (each level has unique subgroups), then use a nested ANOVA for the analysis.

One way ANOVA

- A one way ANOVA is used to compare two means from two independent (unrelated) groups using the F-distribution. The null hypothesis for the test is that the two means are equal. Therefore, a significant result means that the two means are unequal.
- **Examples of when to use a one way ANOVA**
- **Situation 1:** You have a group of individuals randomly split into smaller groups and completing different tasks. For example, you might be studying the effects of tea on weight loss and form three groups: green tea, black tea, and no tea.
- **Situation 2:** Similar to situation 1, but in this case the individuals are split into groups based on an attribute they possess. For example, you might be studying leg strength of people according to weight. You could split participants into weight categories (obese, overweight and normal) and measure their leg strength on a weight machine.
- **Limitations of the One Way ANOVA**
- A one way ANOVA will tell you that at least two groups were different from each other. But **it won't tell you which groups were different.** If your test returns a significant f-statistic, you may need to run an ad hoc test (like the least significance different test) to tell you exactly which groups had a difference in means.

ANOVA Table for One-Way ANOVA

Source of Variation	Sum of Squares (SS)	Degrees of Freedom (d.f.)	Mean Square (MS)	F-ratio
Between samples	$= n_1(\overline{X}_1 - \overline{\overline{X}})^2 + n_2(\overline{X}_2 - \overline{\overline{X}})^2 + \ldots\ldots n_k(\overline{X}_k - \overline{\overline{X}})^2$	k-1	$\frac{\text{SS Between}}{k-1}$	$\frac{\text{MS Between}}{\text{MS Within}}$
Within samples	$= \Sigma(X_{1i} - \overline{X}_1)^2 + \Sigma(X_{2i} - \overline{X}_2)^2 + \ldots\ldots \Sigma(X_{ki} - \overline{X}_k)^2$	n-k	$\frac{\text{SS Within}}{n-k}$	
Total	$\Sigma(X_{ij} - \overline{\overline{X}})^2$	n-1		

Two way ANOVA

- A Two Way ANOVA is an extension of the One Way ANOVA. With a One Way, you have one independent variable affecting a dependent variable. With a Two Way ANOVA, there are two independents. Use a two way ANOVA when you have one measurement variable (i.e. a qualitative variable) and two nominal variables.
- In other words, if your experiment has a quantitative outcome and you have two categorical explanatory variables, a two way ANOVA is appropriate.
- For example, you might want to find out if there is an interaction between income and gender for anxiety level at job interviews. The anxiety level is the outcome, or the variable that can be measured. Gender and Income are the two categorical variable. These categorical variables are also the independent variables, which are called **factors** in a Two Way ANOVA.
- The factors can be split into **levels**. In the above example, income level could be split into three levels: low, middle and high income. Gender could be split into three levels: male, female, and transgender. Treatment groups are all possible combinations of the factors. In this example there would be 3 x 3 = 9 treatment groups.

Multiple Correlation Coefficient

- The multiple correlation coefficient generalizes the standard coefficient of correlation. It is used in multiple regression analysis to assess the quality of the prediction of the dependent variable. It corresponds to the squared correlation between the predicted and the actual values of the dependent variable.
- The multiple correlation coefficient for the kth variable with respect to the other variables in R1 can be calculated by the formula =SQRT(Rsquare (R1, k)).
- Correlation is primarily used to quickly and concisely summarize the direction and strength of the relationships between a set of 2 or more numeric variables.
- It measures the strength of association between the independent (explanatory) variables and the dependent variable (the variable we wish to forecast). Its value varies between 0 and 1; the higher value, the stronger the association.

Correlation

- The range of the correlation coefficient is from -1 to $+1$.
- If there is a **strong positive linear relationship** between the variables, the value of r will be close to $+1$.
- If there is a **strong negative linear relationship** between the variables, the value of r will be close to -1.

Correlation Coefficient

The formula for the correlation coefficient is

$$r = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[n\left(\sum x^2\right) - \left(\sum x\right)^2\right]\left[n\left(\sum y^2\right) - \left(\sum y\right)^2\right]}}$$

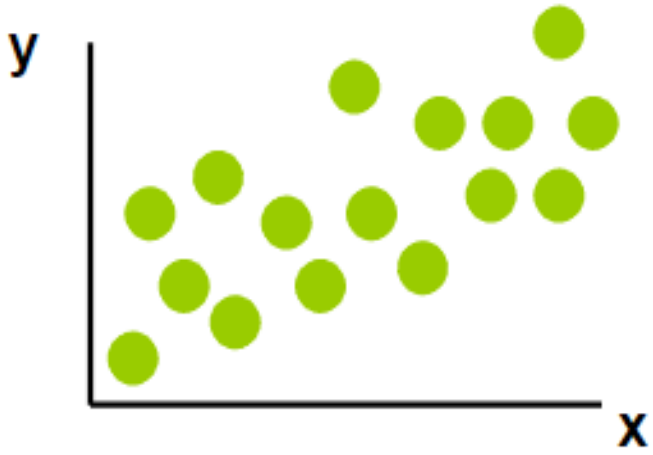
where n is the number of data pairs.

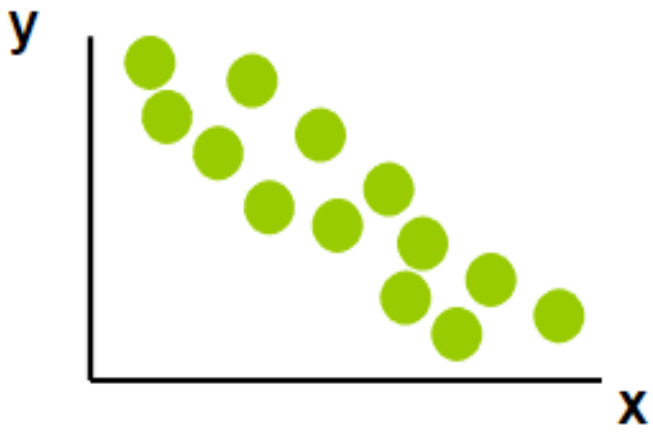
Scatter Plots and Correlation

- A scatter plot (or scatter diagram) is used to show the relationship between two variables
- Correlation analysis is used to measure strength of the association (linear relationship) between two variables
  - Only concerned with strength of the relationship
  - No causal effect is implied

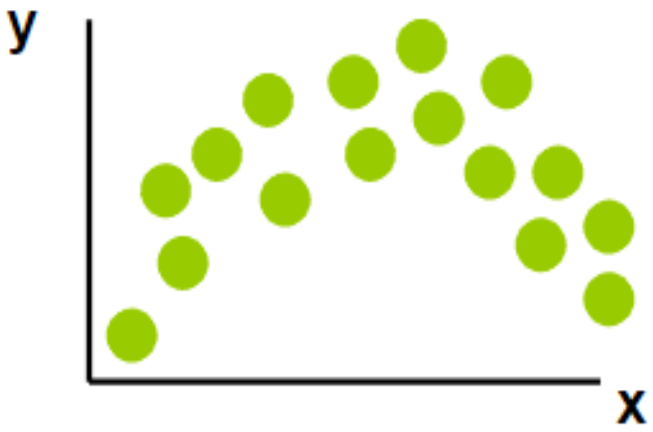
Scatter Plot Examples

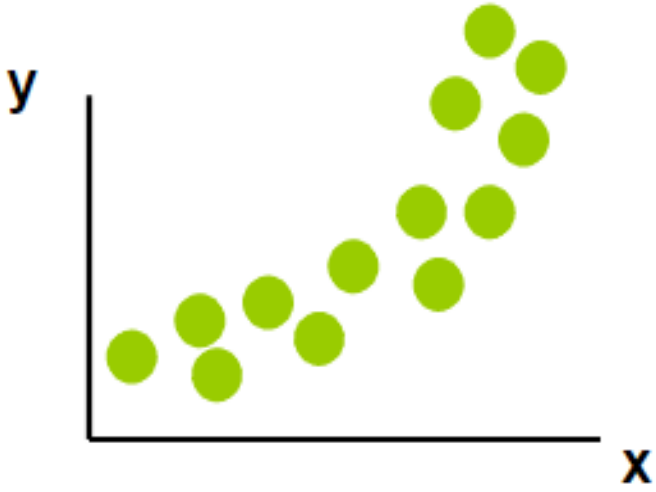
Linear relationships



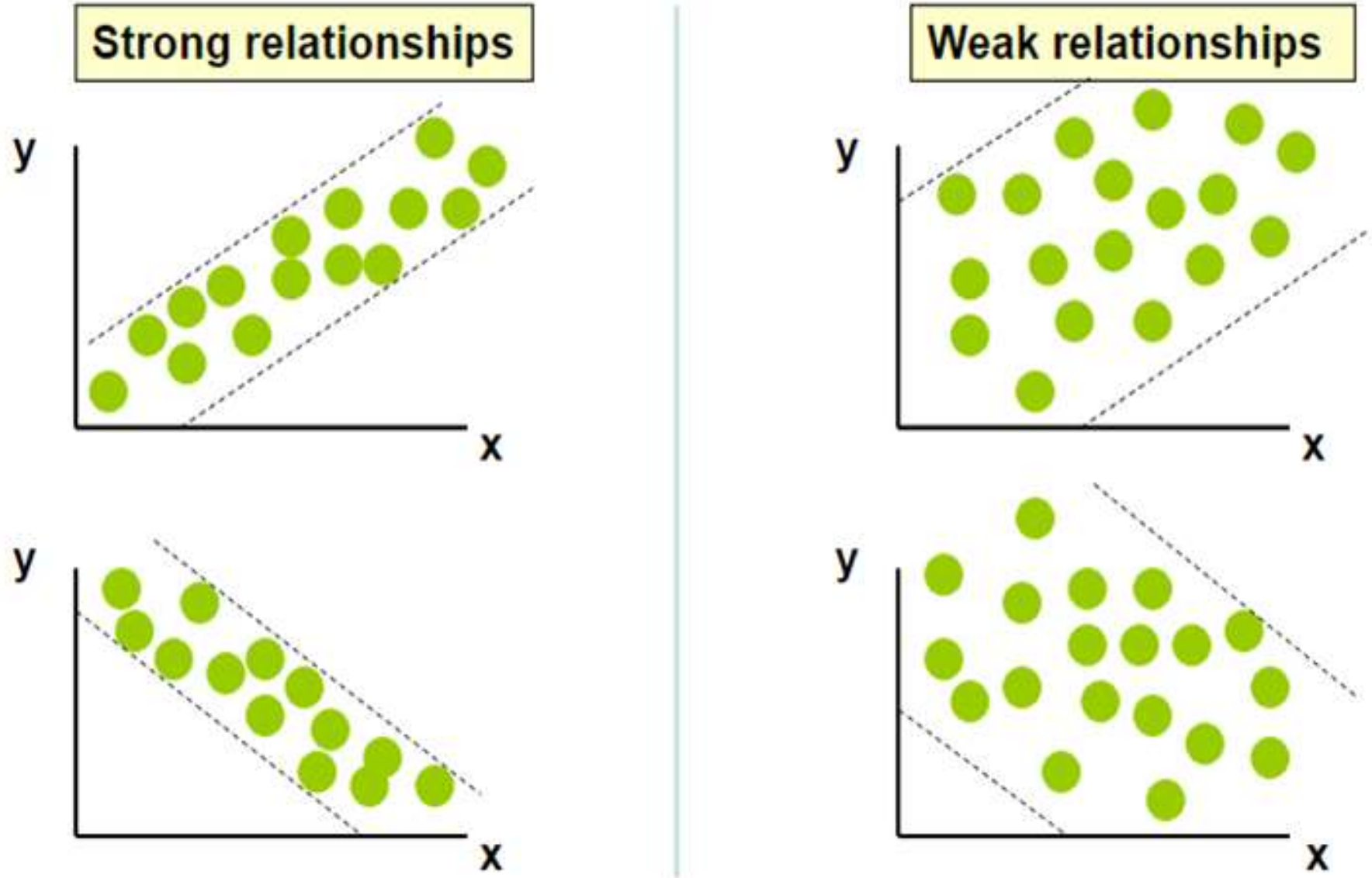


Curvilinear relationships

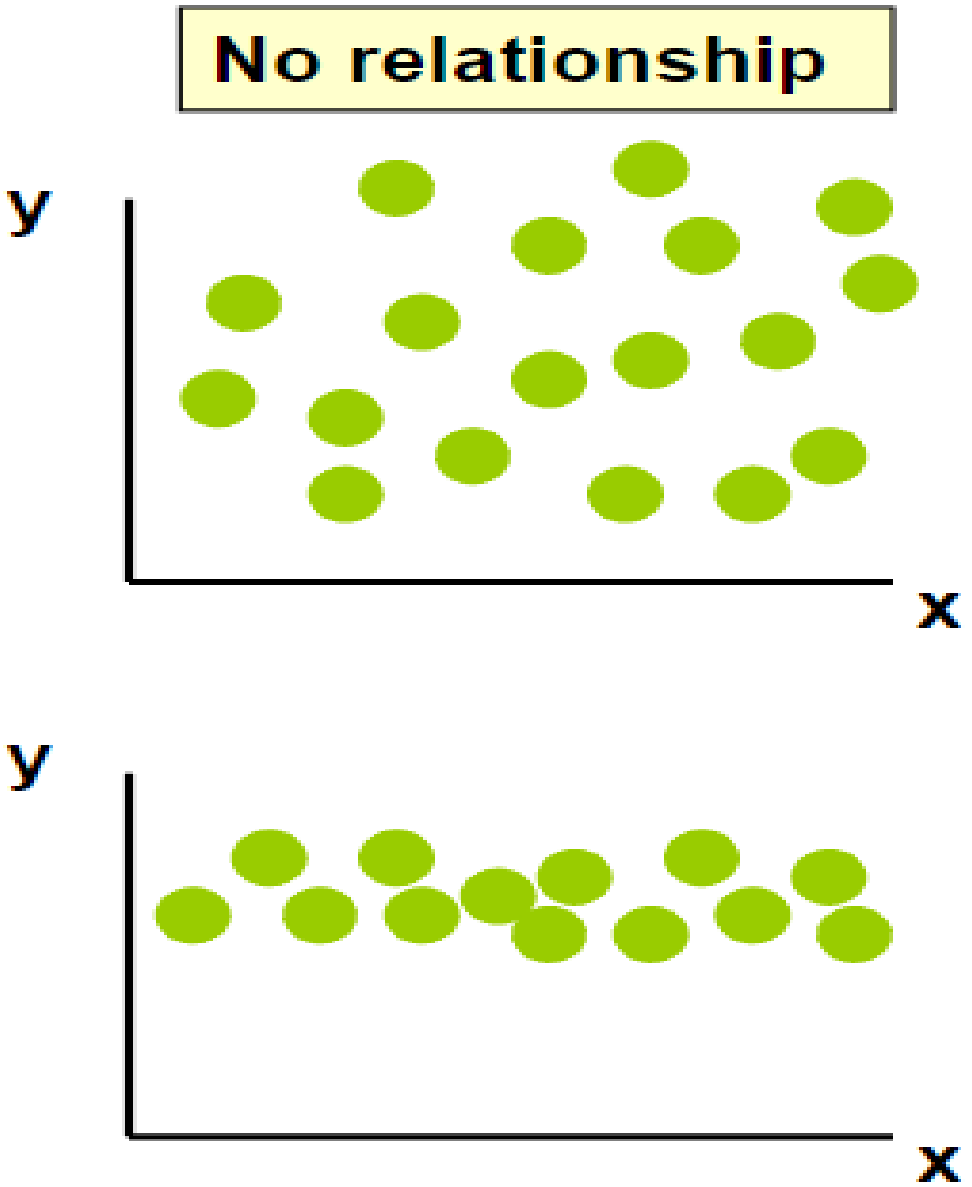




Scatter Plot Examples



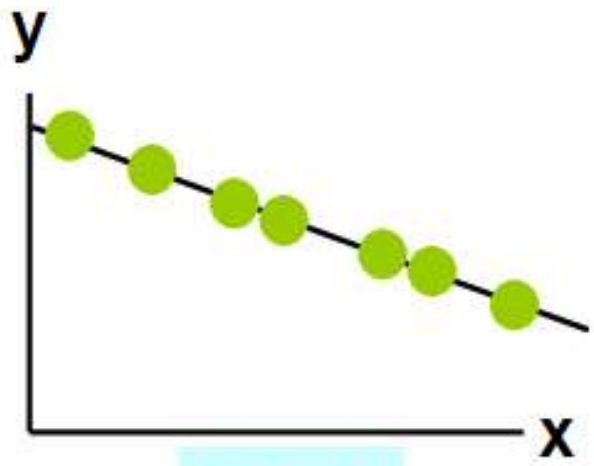
Scatter Plot Examples



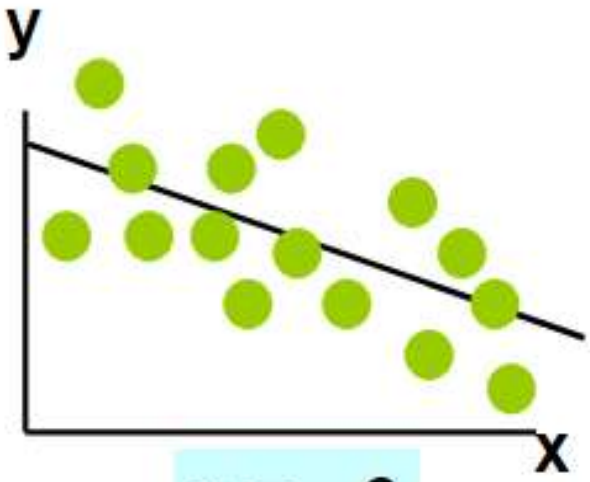
Correlation Coefficient

- The correlation coefficient r is used to measure the strength of the linear relationship in the observations
- Unit free
- Range between -1 and 1
- The closer to -1, the **stronger the negative linear relationship**
- The closer to 1, the **stronger the positive linear relationship**
- The closer to 0, the **weaker the linear relationship**

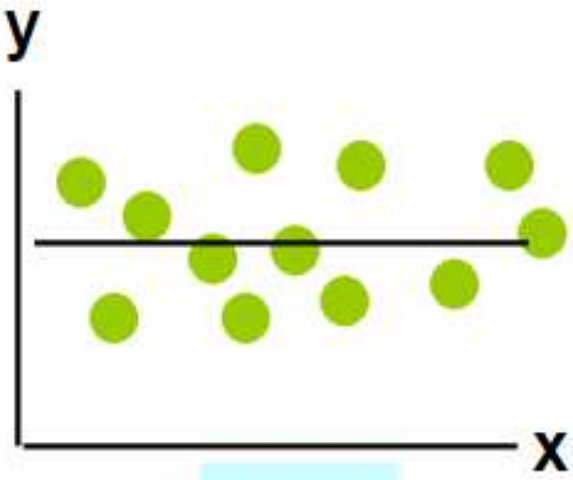
Examples of Approximate r Values



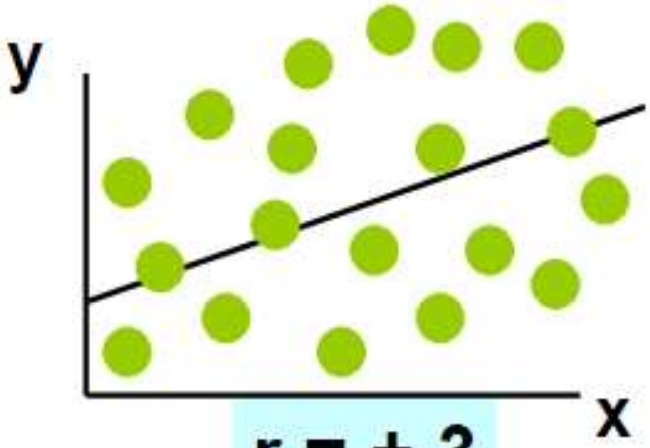
r = -1



r = -.6



r = 0



r = +.3



r = +1

Calculating the Correlation Coefficient

Actual Mean Method

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

or the algebraic equivalent:

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

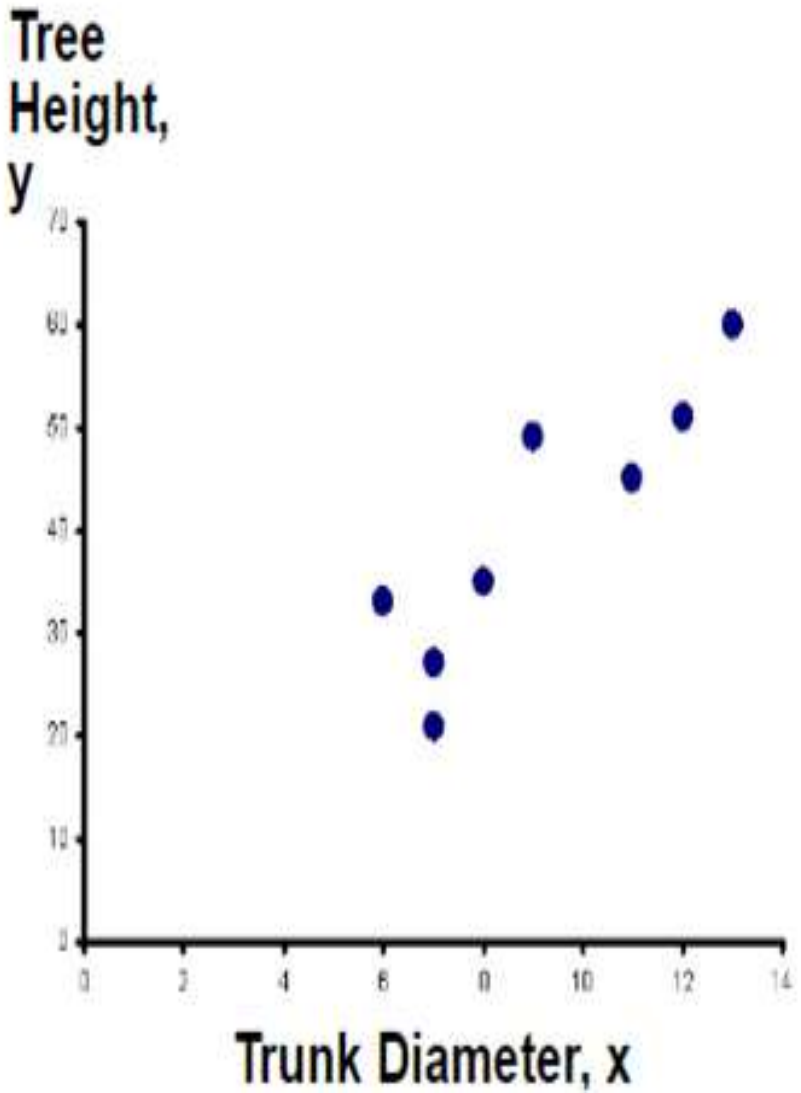
where:

r = Sample correlation coefficient
n = Sample size
x = Value of the independent variable
y = Value of the dependent variable

Calculation Example

Tree Height	Trunk Diameter			
y	x	xy	y^2	x^2
35	8	280	1225	64
49	9	441	2401	81
27	7	189	729	49
33	6	198	1089	36
60	13	780	3600	169
21	7	147	441	49
45	11	495	2025	121
51	12	612	2601	144
Σ=321	**Σ=73**	**Σ=3142**	**Σ=14111**	**Σ=713**

Calculation Example



$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

$$= \frac{8(3142) - (73)(321)}{\sqrt{[8(713) - (73)^2][8(14111) - (321)^2]}}$$

$$= 0.886$$

r = 0.886 → relatively strong positive linear association between x and y

Actual Mean Method

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Direct Method

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$

Assumed Mean Method

$$r = \frac{N\sum d_x d_y - \sum d_x \sum d_y}{\sqrt{N\sum d_x^2 - (\sum d_x)^2}\sqrt{N\sum d_y^2 - (\sum d_y)^2}}$$

Bivariate frequency table

$$r = \frac{N\sum fd_x d_y - \sum fd_x \sum fd_y}{\sqrt{N\sum fd_x^2 - (\sum fd_x)^2}\sqrt{N\sum fd_y^2 - (\sum fd_y)^2}}$$

$$R = 1 - \frac{6\sum D^2}{N^3 - N}$$

D= difference of rank between the paired items in two series
N= No. of paired items in the two series

$$R = 1 - \frac{6\left(\sum D^2 + \frac{m^3 - m}{12} + \frac{m^3 - m}{12} + \ldots\ldots\right)}{N^3 - N}$$

m= No. of items whose ranks are common

Properties of the Correlation Coefficient

- The correlation coefficient formula is based on a ratio of "*X*-units × *Y*-units" (i.e., their respective deviations from the mean) divided by "*X*-units × *Y*-units" (i.e., their respective SDs).
  - This means that all units of measurement cancel out and the correlation coefficient a *pure number* that is *not expressed in any units*.
  - For example, suppose the correlation between the height (measured in inches) and weight (measured in pounds) of students in the class is $r = +.45$.
    - This is *not* $r = +.45$ *inches* or $r = +.45$ *pounds*, or $r = +.45$ *pounds per inch*, etc.,
    - rather it is just $r = +.45$.
  - Moreover, *the magnitude of the correlation coefficient is independent of the units used to measure the two variables*.
    - If we measured students' heights in feet, meters, etc. (instead of inches) and/or measured their weights in ounces, kilograms, etc. (instead of pounds), and then calculated the correlation coefficient based on the new numbers, it would be just the same as before, i.e., $r = +.45$.
- In addition, if you check the formula above, you can see that *r* is symmetric, i.e., it is unchanged if we interchange *X* and *Y*.
  - Thus, *the correlation between two variables is the same regardless of which variable is considered to be independent and which dependent.*

Calculating the Correlation Coefficient

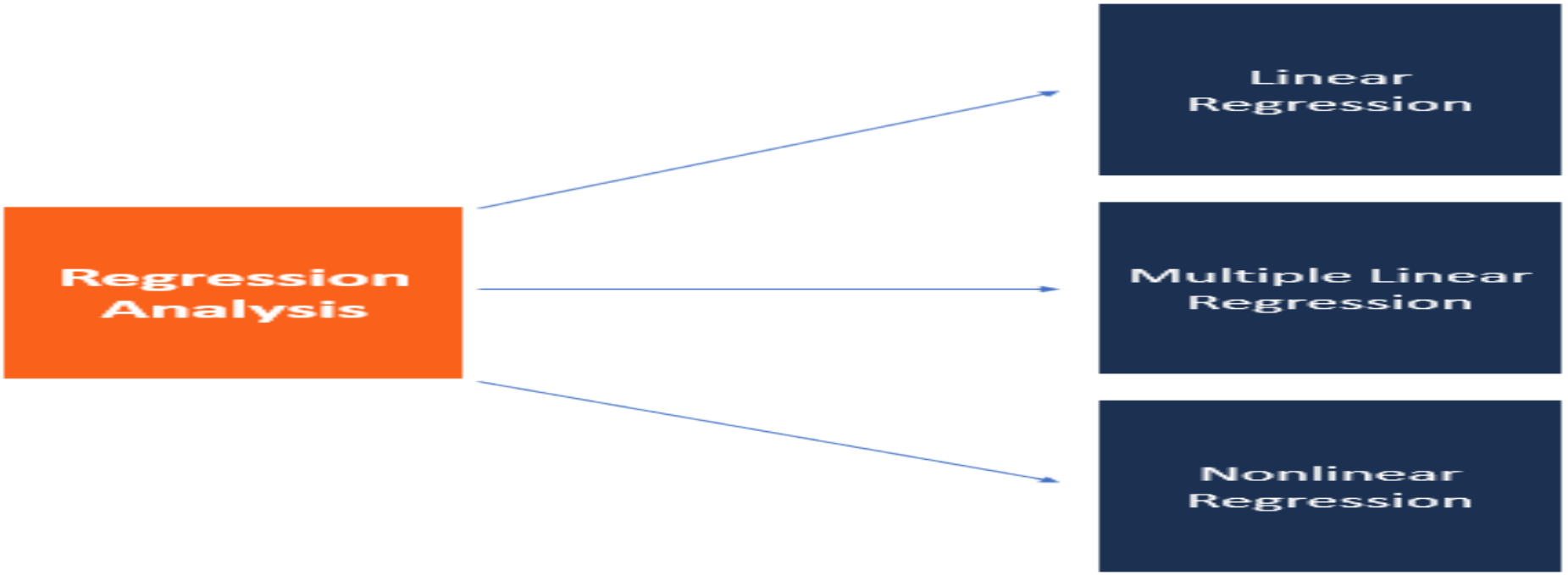
Case #	X	Y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	1	0	-3	-5	9	25	+15
2	2	2	-2	-3	4	9	+6
3	4	3	0	-2	0	4	0
4	5	2	+1	-3	1	9	-3
5	5	11	+1	+6	1	36	+6
6	7	12	+3	+7	9	49	+21
Sum	24	30	0	0	24	132	+45
Average	4 [= $\bar{x}$]	5 [= $\bar{y}$]			4 [=Var(X)]	22 [=Var(Y)]	+7.5 [= Cov(X,Y)]
SQRT(Average)					2 [=SD(X)]	$\sqrt{22}$ = 4.7 [= SD(Y)]	

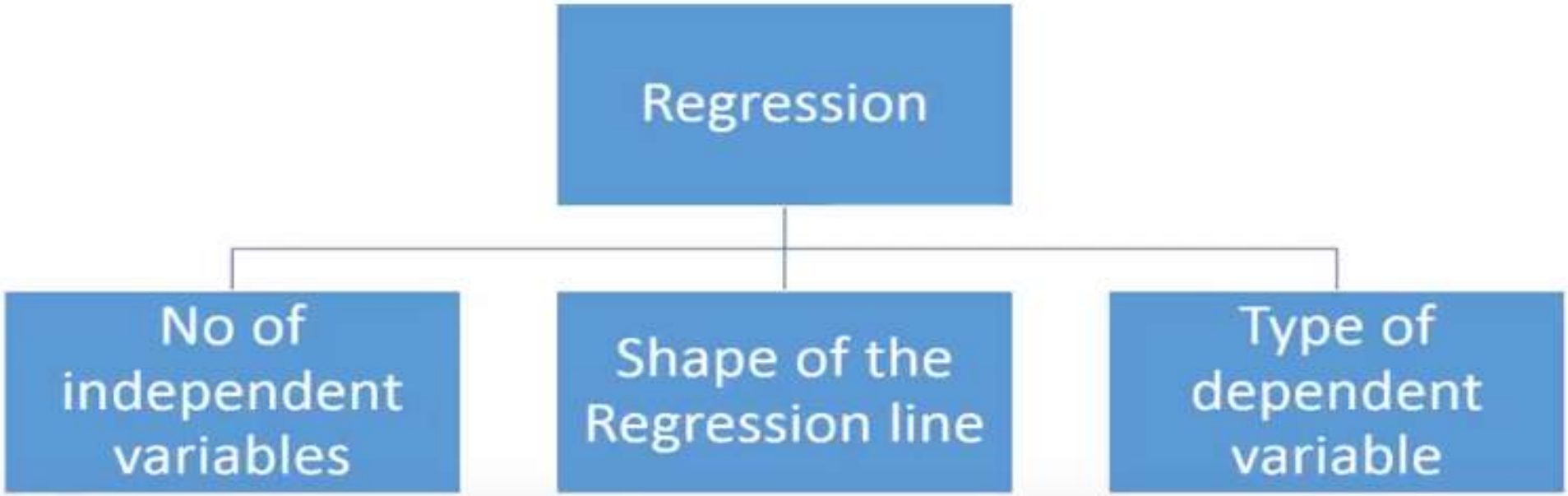
$$\textit{Correlation coefficient} = r = \frac{\text{Cov}(X,Y)}{\text{SD}(X)\text{SD}(Y)} = \frac{+7.5}{2\sqrt{22}} = +0.7995 = \frac{+750}{20\times 47}$$

Regression

- Regression takes a group of random variables, thought to be predicting Y, and tries to find a mathematical relationship between them. This relationship is typically in the form of a straight line (linear regression) that best approximates all the individual data points.
- Simple regression analysis uses a single x variable for each dependent "y" variable. For example: (x_1, Y_1). Multiple regression uses multiple "x" variables for each independent variable: $(x1)_1$, $(x2)_1$, $(x3)_1$, Y_1).
- Regression analysis is a reliable method of identifying which variables have impact on a topic of interest. The process of performing a regression allows you to confidently determine which factors matter most, which factors can be ignored, and how these factors influence each other.
- These are often quantities, such as amounts and sizes. For example, a house may be predicted to sell for a specific dollar value, perhaps in the range of $100,000 to $200,000. A regression problem requires the prediction of a quantity.

Regression Analysis





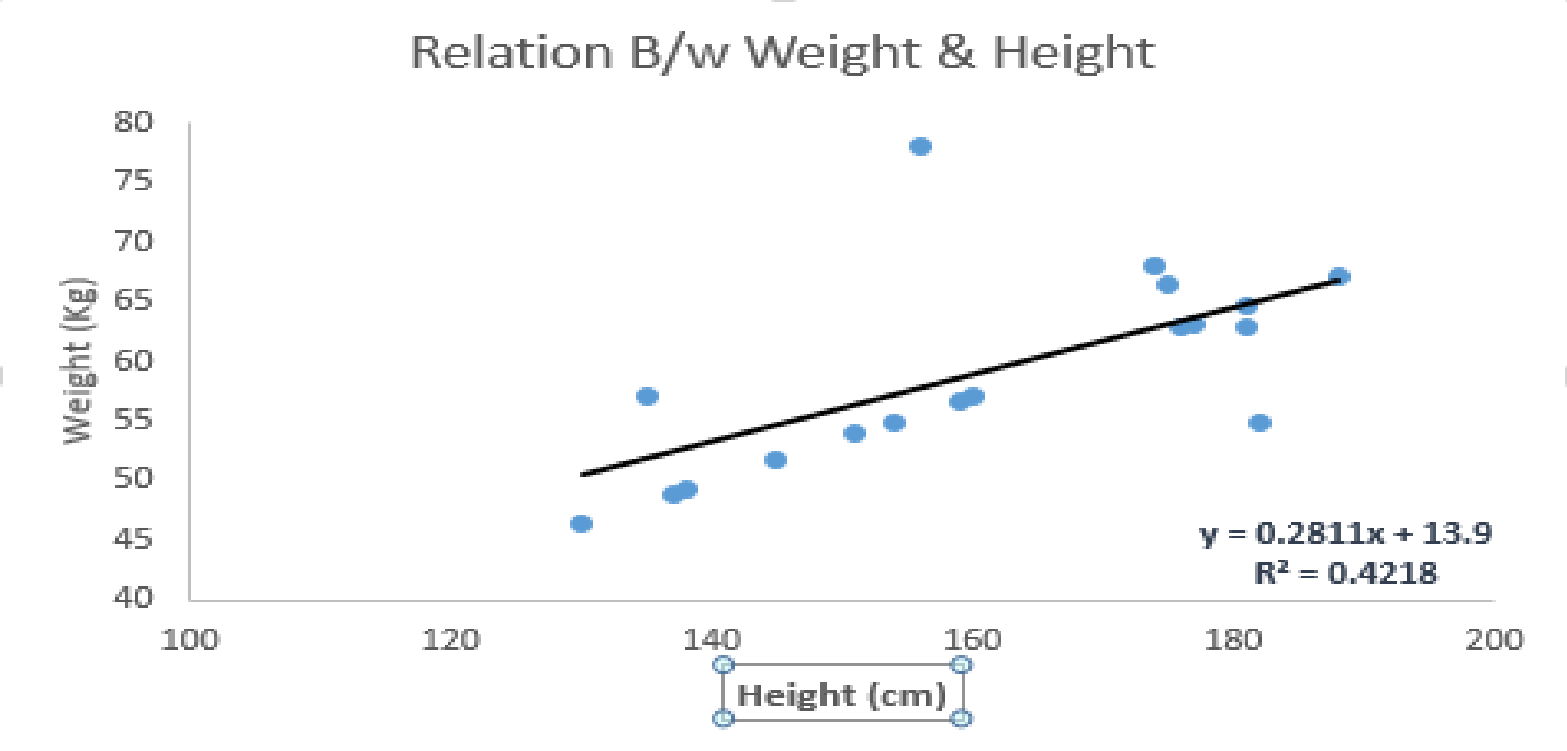
Linear Regression

- It is one of the most widely known modeling technique. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear.
- Linear Regression establishes a relationship between **dependent variable (Y)** and one or more **independent variables (X)** using a **best fit straight line** (also known as regression line).
- It is represented by an equation **Y=a+b*X + e**, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).



Linear Regression

- The difference between simple linear regression and multiple linear regression is that, multiple linear regression has (>1) independent variables, whereas simple linear regression has only 1 independent variable. Now, the question is "How do we obtain best fit line?".



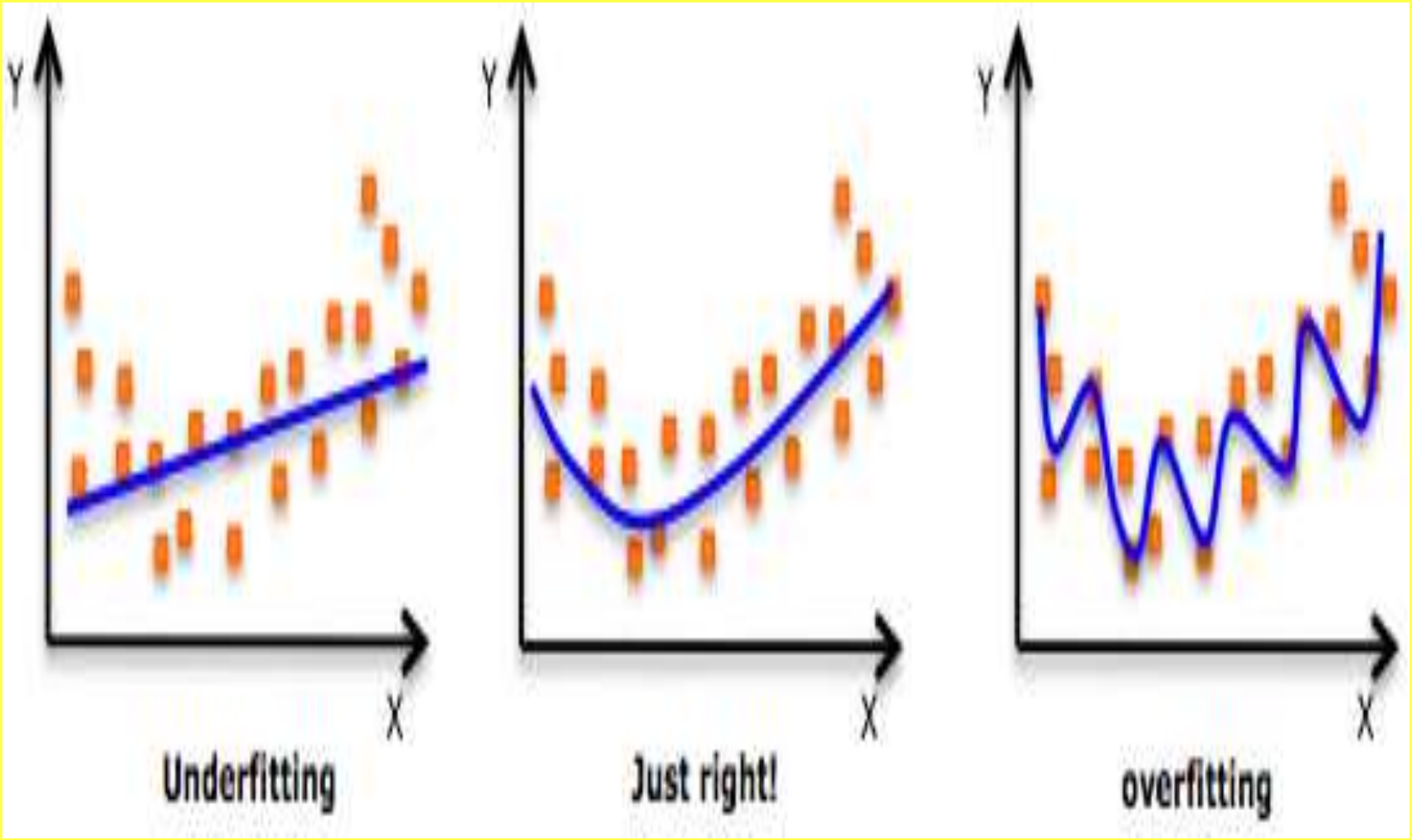
Logistic Regression

- Logistic regression is used to find the probability of event=Success and event=Failure. We should use logistic regression when the dependent variable is binary (0/ 1, True/ False, Yes/ No) in nature. Here the value of Y ranges from 0 to 1 and it can represented by following equation.
- Logistic regression is widely used for classification problems.
- Logistic regression doesn't require linear relationship between dependent and independent variables. It can handle various types of relationships because it applies a non-linear log transformation to the predicted odds ratio.
- To avoid over fitting and under fitting, we should include all significant variables. A good approach to ensure this practice is to use a step wise method to estimate the logistic regression.
- It requires large sample sizes because maximum likelihood estimates are less powerful at low sample sizes than ordinary least square.

Polynomial Regression

- A regression equation is a polynomial regression equation if the power of independent variable is more than 1. The equation below represents a polynomial equation:
- y=a+b*x^2
- While there might be a temptation to fit a higher degree polynomial to get lower error, this can result in over-fitting. Always plot the relationships to see the fit and focus on making sure that the curve fits the nature of the problem.
- Especially look out for curve towards the ends and see whether those shapes and trends make sense. Higher polynomials can end up producing wierd results on extrapolation.

Polynomial Regression



Properties of the Regression Coefficient

- So, from one point of view, the regression coefficient answers this question:
  - How big is the covariance of the IND and DEP variables compared with the variance of the IND variable?
- But (as we have already seen) from a more practical point of view, the regression coefficient (i.e., the slope of the regression line [or line of averages]) answers this question:

How many units, on average, does the dependent variable increase when the independent variable increases by one unit?

  - If the dependent variable decreases (has a "negative increase") when the independent variable increases, the regression coefficient is negative.
- Thus the *magnitude of the regression coefficient* [unlike the correlation coefficient] *depends on*
  - *which variable is considered to be independent and which dependent*, and also
  - *the units in which each variable is measured.*

Properties of the Regression Coefficient (cont.)

- The regression coefficient is a ratio of "*X*-units × *Y*-units" (i.e., their respective deviations from the mean) divided by "*X*-units × *X*-units" (i.e., the variance of *X*).

- Thus *the regression coefficient is expressed in units of the dependent variable per unit of the independent variable*; that is,
  - it is a *rate*,
    - like "miles per hour" [rate of speed] or
    - "miles per gallon" [rate of fuel efficiency], and so
  - its numerical value changes as these units change.

Properties of the Regression Coefficient (cont.)

- Remember that we informally calculated regression coefficients (by eyeball methods) in the INCOME by EDUCATION example.
  - In society A, the regression coefficient was about +$1000 (of INCOME [DEP]) per YEAR (of EDUCATION [IND]).
  - In society B, the regression coefficient was about +$4000 (of INCOME [DEP]) per YEAR (of EDUCATION [IND]).
- Likewise we informally calculated the regression coefficient in the Pearson (SON'S HEIGHT by FATHER'S HEIGHT)
  - It was about +0.5 INCHES (of SH [DEP]) per INCH (of FH [IND]).
  - In this (rather special) case, both variables are "in the same currency" and measured in the *same* units (INCHES), so
  - if we change the unit of measurement for *both* variables to FEET, CENTIMETERS, etc.), the magnitude of the regression coefficient does not change.

Properties of the Regression Coefficient (cont.)

- But suppose we measure the *height in inches* and *weight in pounds* of all students in the class, suppose we treat height as the independent variable, and suppose the regression coefficient is $b = +6$.
  - This means +6 *pounds* (dependent variable units) of weight *per inch* (independent variable units) of height.
    - That is, if we lined students up in order from shortest to tallest, we would observe that their weights increase, on average, by about 6 pounds for each additional inch of height.
  - This also means that students' weights increase, on average, by about 2.7 kilograms (since there are about .45 kilograms to a pound) for each additional inch of height.
    - So if weight were measured in kilograms instead of pounds, the regression coefficient would be $b = +6 \times .45 = +2.7$ *kilograms per inch*.
    - Likewise if we measured weight in pounds but height in feet, the regression coefficient would be about $b = +6 \times 12 = 72$ *pounds per foot*.
- Remember that the magnitude of the correlation coefficient is unchanged by such changes in units.

Relationship between Calculations for Correlation and Regression Coefficients

- Note from the formulas that the sign (i.e., "+" or "-") of *b* and *r* are both determined by the sign of *cov*(*X*,*Y*), from which it follows that
  - *b* and *r* themselves always have the same sign, and
  - if one is zero, the other is also zero.
- Notice also that the regression and correlation coefficients are equal in the event the independent and dependent variables have the same dispersion, i.e., same SD.
  - For example, in the SON'S HEIGHT by FATHER'S HEIGHT scattergram, by eyeball methods we can determine the regression coefficient *b* is about +0.5,
  - It is also apparent, both from common expectations and examination of the scattergram, that the dispersions (SDs) of SON's HEIGHT and FATHER's HEIGHT are just about the same, so we also know that the correlation coefficient is also about +0.5.
- In general, $b = r \times \frac{SD(Y)}{SD(X)}$ and $r = b \times \frac{SD(X)}{SD(Y)}$

Difference between Correlation and Regression - I

Particulars	Correlation	Regression
Meaning	Correlation is a statistical measure which determines a co-relationship or association of two variables.	Regression descries how an independent variable (y).
Main Purpose	Correlation analysis lets experimenters know the association or the absence of the relationship between two variables under study; if the variables are correlated, it allows measuring the strength of their association.	Regression analysis helps determine a functional relationship between two variables so as to estimate the unknown variable with the help of known variable and make future projection on events.
Objective	To find a numerical value that expresses the relationship between the variables.	To estimate the values of a random variable on the basis of the value of a fixed variable.
Usage	Regression the linear relationship between two variables.	Fits the best line and estimates one variable on the basis of another variable.

Continue….

Difference between Correlation and Regression - II

Particulars	Correlation	Regression
Nature of Variables	The variables are not designated as dependent or independent.	One variable is dependent and another variable is independent.
Indicates	Correlation coefficient indicates the extent to which two variables move together.	Regression indicates the impact of a unit change in the known variable on the estimated variable.
Range	Correlation coefficients can range from -1.00 to +1.00.	In regression analysis, if byx > 1, then bxy < 1
Exceptional Cases	Non-sense correlation may exist in the correlation analysis.	Non-sense regression does not exist in regression analysis.
Association	The correlation coefficient measures the extent and direction of a linear association between two variables.	Linear regression allows experimenters to describe one variable as a linear function of another variable.
Relationship	Correlation is confined to the linear relationship between variables only.	Regression studies linear and non-linear relationships.
Scope	Correlation analysis has limited applications.	Regression analysis has wider applications.

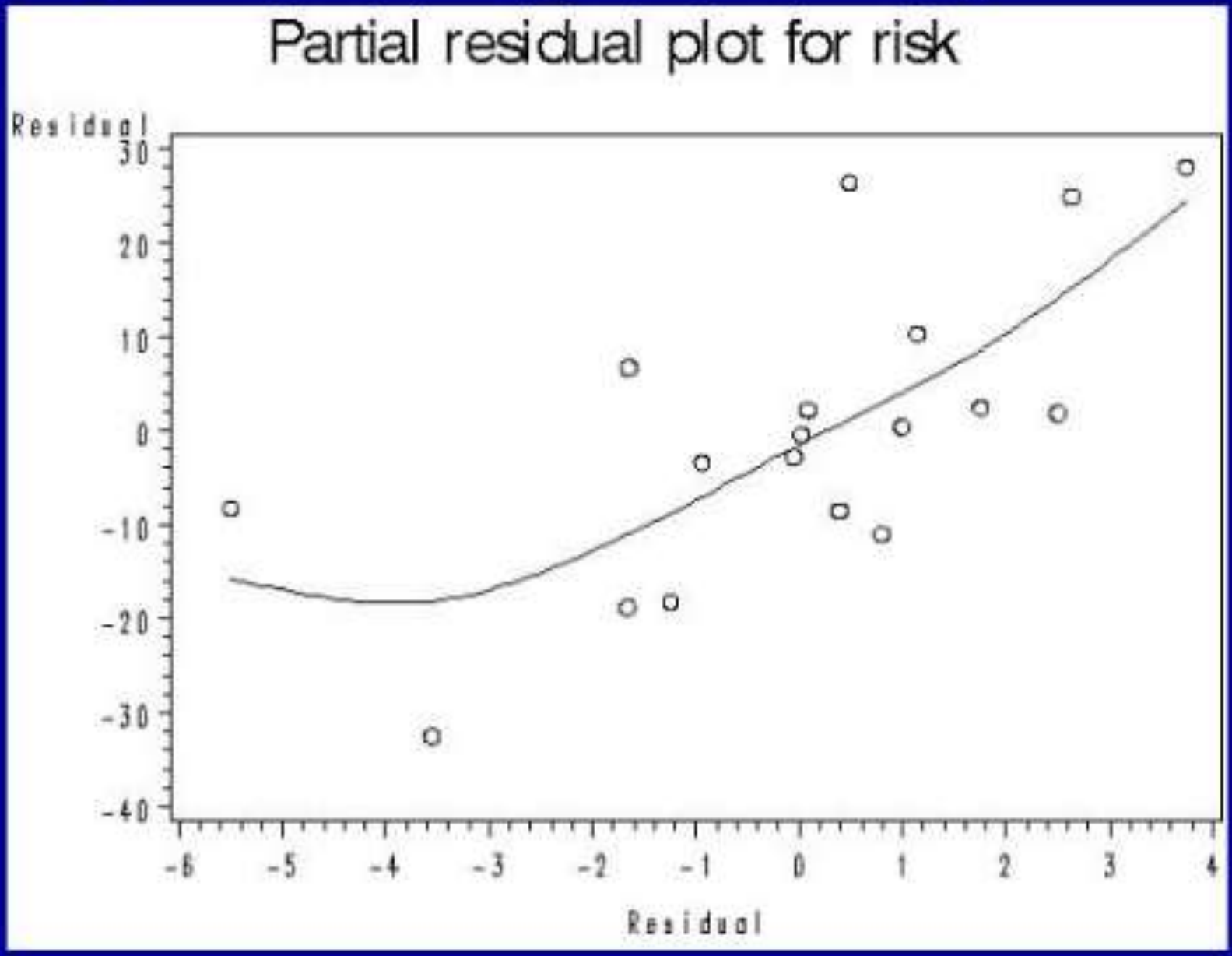
Partial Regression

- In applied statistics, a partial regression plot attempts to show the effect of adding another variable to a model that already has one or more independent variables.
- Partial regression plots are also referred to as added variable plots, adjusted variable plots, and individual coefficient plots.'
- Partial regression coefficients are the most important parameters of the multiple regression model.
- They measure the expected change in the dependent variable associated with a one unit change in an independent variable holding the other independent variables constant.
- Partial effects distinguish between dummy variables and continuous variables.
- For a dummy variable, the effect is computed as the difference in the estimated probabilities with the dummy variable equal to one and zero and other variables at their means. For continuous variables, the effect is the derivative.

The partial regression plots

- To generate the partial regression plots
  - Regress Y and X_1 each on X_2.
  - Get the residual from each regression namely $e(Y|X_2)$ and $e(X_1|X_2)$
  - Plot $e(Y|X_2)$ against $e(X_1|X_2)$
- Do the same for Y and X_2 each on X_1.

The partial regression plots (2)



Multiple Regression

- Multiple regression is an extension of simple linear regression. It is used when we want to predict the value of a variable based on the value of two or more other variables.
- The variable we want to predict is called the dependent variable (or sometimes, the outcome, target or criterion variable).
- For example, if you're doing a multiple regression to try to predict blood pressure (the dependent variable) from independent variables such as height, weight, age, and hours of exercise per week, you'd also want to include sex as one of your independent variables.
- There must be a linear relationship between the outcome variable and the independent variables.
- Scatter plots can show whether there is a linear or curvilinear relationship. Multivariate Normality–Multiple regression assumes that the residuals are normally distributed.

Multiple Regression

The "least squares" regression equation when there are multiple intercorrelated predictor (x) variables is found by calculating "partial regression coefficients" for each x

A partial regression coefficient for x_1 shows the relationship between y and x_1 while statistically controlling for the other x variables (or holding the other x variables constant)

Multiple Regression

The formula for the partial regression coefficient is :

$b_1 = (r_{Y1} - r_{Y2} r_{12}) / (1 - r_{12}^2) * (s_Y / s_1)$

Where

r_{Y1}=correlation of x_1 and y

r_{Y2}=correlation of x_2 and y

r_{12}=correlation of x_1 and x_2

s_Y=standard deviation of y, s_1=standard deviation of x_1

Multiple Regression

- Multiple correlation coefficient (R) is an estimate of the relationship between the dependent variable (y) and the best linear combination of predictor variables (correlation of y and y-pred.)
- R^2 tells you the amount of variance in y explained by the particular multiple regression model being tested.

Cautions in Multiple Regression

- We can use as many predictors as we wish but we should be careful not to use more predictors than is warranted.
  - Simpler models are more likely to generalize to other samples.
  - If you use as many predictors as you have participants in your study, you can predict 100% of the variance. Although this may seem like a good thing, it is unlikely that your results would generalize to any other sample and thus they are not valid.
  - You probably should have at least 10 participants per predictor variable (and probably should aim for about 30).

Multiple Regression Model with 2 Independent Variables (First-Order)

- The simplest multiple regression model is one constructed with two independent variables, where the highest power of either variable is 1 (first-order regression model).
- In multiple regression analysis, the resulting model produces a response surface.

Multiple Regression Model with 2 Independent Variables (First-Order)

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Population Model

where: β_0 = the regression constant

β_1 = the partial regression coefficient for independent variable 1

β_2 = the partial regression coefficient for independent variable 2

ε = the error of prediction

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2$$

Estimated Model

where: $\hat{Y}$ = predicted value of Y

b_0 = estimate of regression constant

b_1 = estimate of regression coefficient 1

b_2 = estimate of regression coefficient 2

Response Plane for First-Order Two-Predictor Multiple Regression Model

- In multiple regression analysis, the resulting model produces a response surface.
- In the multiple regression model shown on the next slide with two independent first-order variables, the response surface is a response plane.
  - The response plane for such a model is fit in a three-dimensional space (x_1, x_2, y).

Factor Analysis

- Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors.
- Factor analysis is a way to condense the data in many variables into a just a few variables. For this reason, it is also sometimes called “dimension reduction.”
- You can reduce the “dimensions” of your data into one or more “super-variables.” The most common technique is known as Principal Component Analysis (PCA).
- The relationship of each variable to the underlying factor is expressed by the so-called factor loading. Here is an example of the output of a simple factor analysis looking at indicators of wealth, with just six variables and two resulting factors.

Factor Analysis

- There are two types of factor analyses, exploratory and confirmatory. Exploratory factor analysis (EFA) is method to explore the underlying structure of a set of observed variables, and is a crucial step in the scale development process. The first step in EFA is factor extraction.
- Minimum Sample Size Recommendations for Conducting Factor Analyses. There is no shortage of recommendations regarding the appropriate sample size to use when conducting a factor analysis. Suggested minimums for sample size include from 3 to 20 times the number of variables and absolute ranges from 100 to over 1,000.
- As a data analyst, the goal of a factor analysis is to reduce the number of variables to explain and to interpret the results. This can be accomplished in two steps: factor extraction.

Purpose of Factor Analysis

- To identify underlying dimensions called factors, that explain the correlations among a set of variables.
- Lifestyle statements may be used to measure the psychographic profile of consumers.
- To identify a new, smaller set of uncorrelated variables for subsequent analysis such as Regression of Discriminant analysis.
- Psychographic factors may be used as independent variables explain the difference between loyal and non loyal customers.

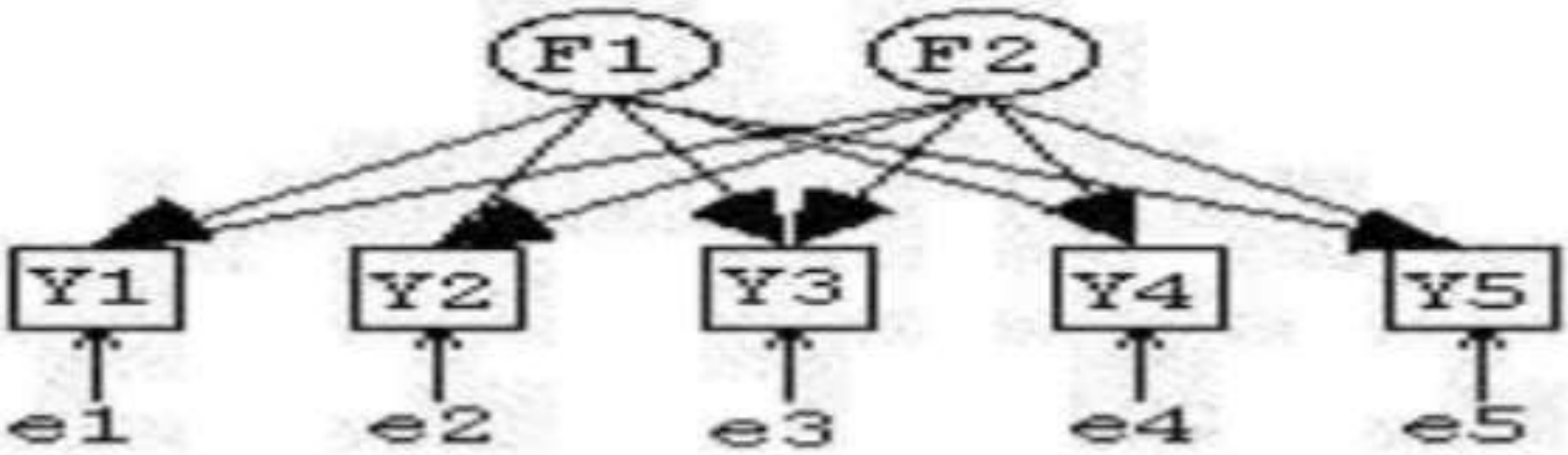
Types of Factor Analysis

- **Exploratory Factor Analysis**
  - ✓ Summarizing data by grouping correlated variables.
  - ✓ Investigating sets of measured variables related to theoretical constructs.
  - ✓ Usually done near the onset of research.
- **Confirmatory Factor Analysis**
  - ✓ More advanced technique.
  - ✓ When factor structure is known or at least theorized.
  - ✓ Testing generalization of factor structure to new data, etc.

Assumptions of Factor Analysis

- Models are usually based on linear relationships.
- Models assume that the data collected are interval scaled.
- Multi-collinearity in the data is desirable because the objective is to identify interrelated set of variables.
- The data should be amenable for factor analysis. It should not be such that a variable is only correlated with itself and no correlation exists with any other variables. This is like an identity matrix. Factor analysis cannot be done on such data.

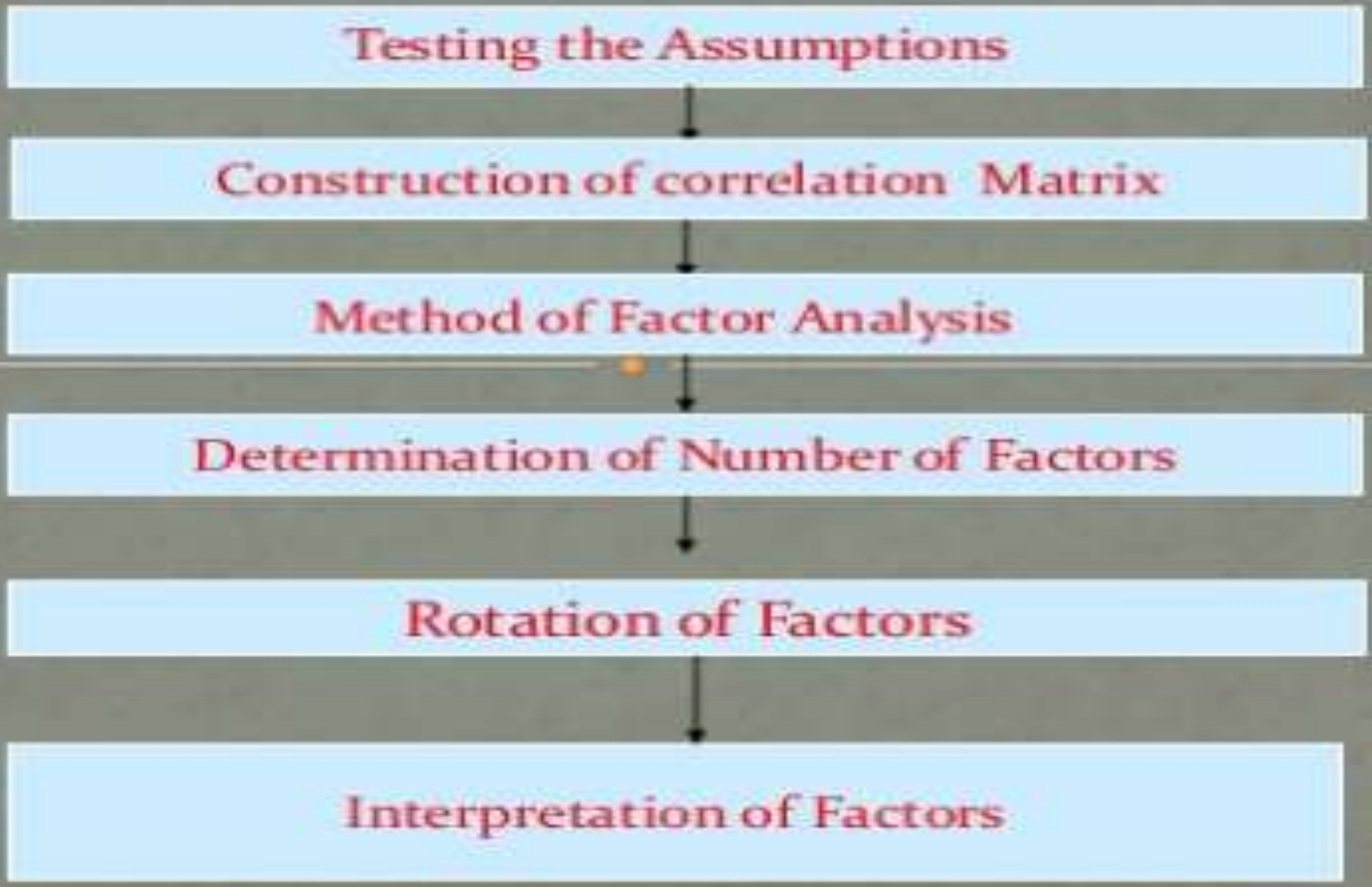
SIMPLE PATH DIAGRAM FOR A FACTOR ANALYSIS MODEL



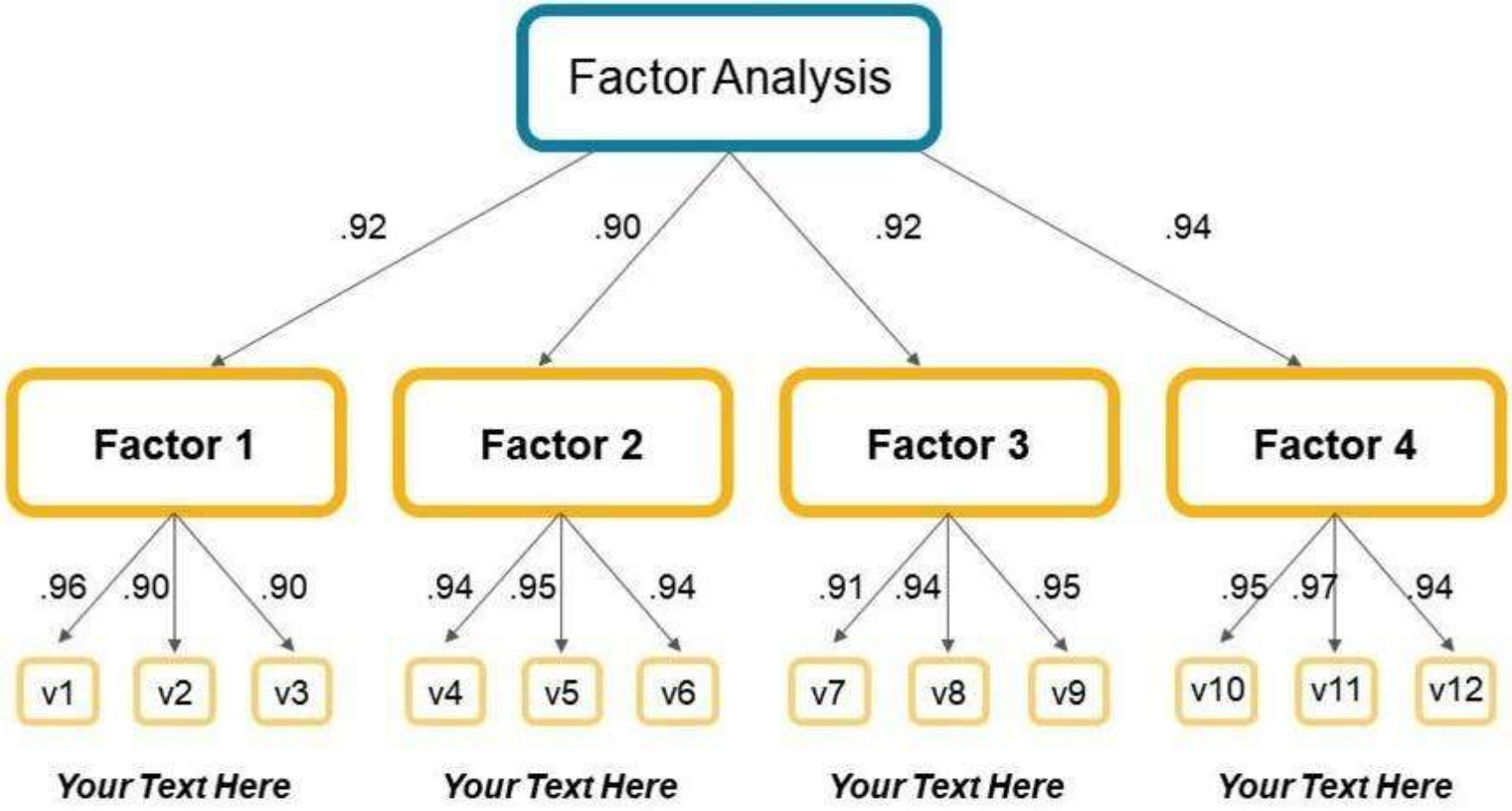
•F1 and F2 are two common factors. Y_1, Y_2, Y_3, Y_4, and Y_5 are observed variables, possibly 5 subtests or measures of other observations such as responses to items on a survey.

• e1,e2,e3,e4, and e5 represent residuals or unique factors, which are assumed to be uncorrelated with each other.

Conducting Factor Analysis

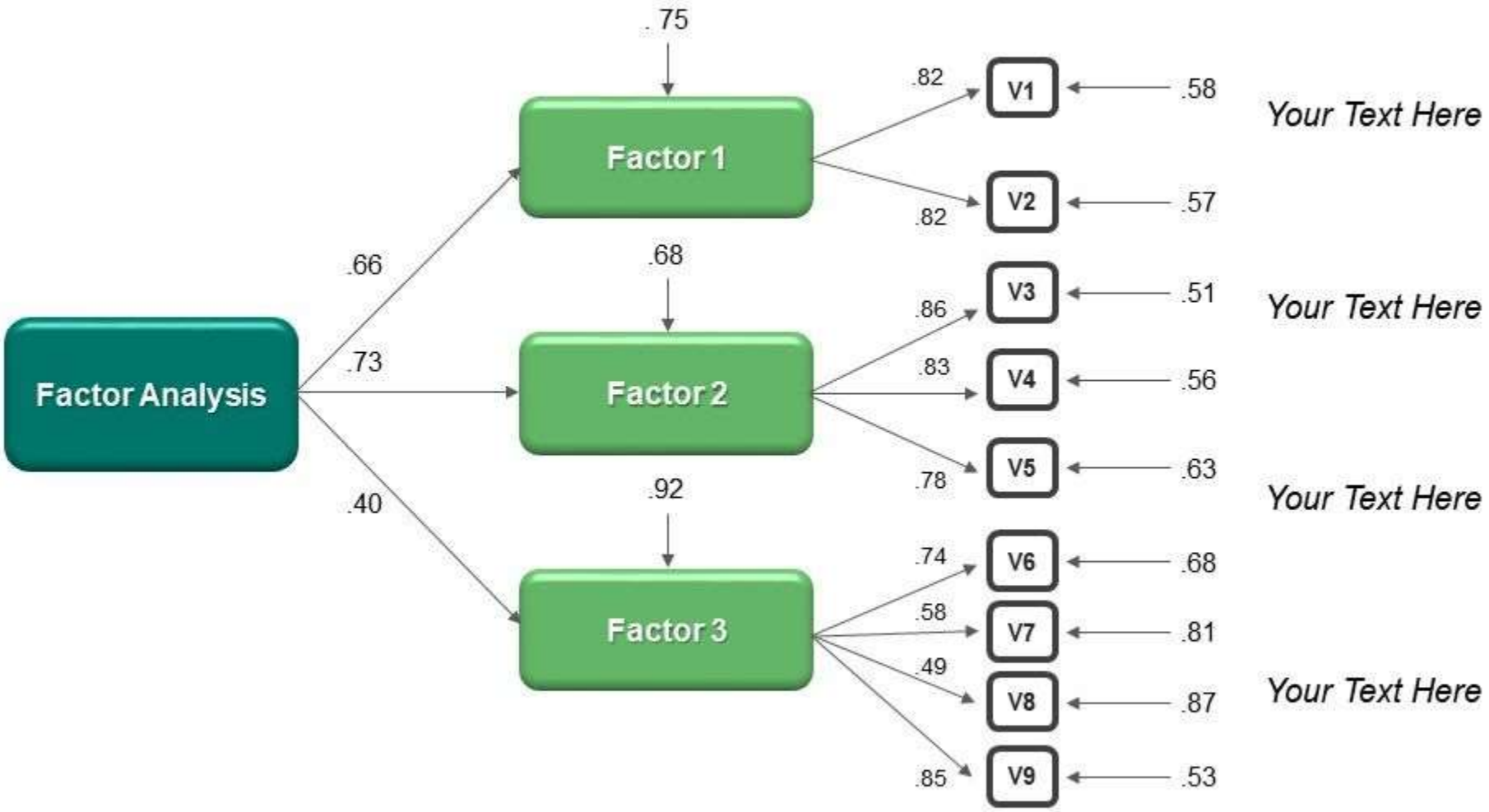


Factor Analysis Connecting Chart



This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

Factor Analysis Horizontal Flow



This slide is 100% editable. Adapt it to your needs and capture your audience's attention.

Computer Aided Research

- Computer is an essential tool for research, whether for academic purpose or for commercial purpose. Computers play a major role today in every field of scientific research from genetic engineering to astrophysics research. Various programs and applications have eased our way into computing our research process.
- Social science research provides authentic information to the people that are scientifically proven so that end users can rely upon it. It also helps us in finding the solutions to different social problems.
- Computer can be used for, deciding population sample, questionnaire designing and data collection. These are different internet sites which help to design questionnaire. Several softwares can be used to calculate the sample size. It makes pilot study of the research possible.

Computer Aided Research

- Scientists, engineers and researches alike are able to compile vast amounts of data and leave it to the computer to work through the data while focusing on another area of the research project. This creates research results with fewer errors and better-engineered products.
- Computers have revolutionized the teaching profession in multiple ways. Teachers use computers to record grades, calculate averages, manage attendance and access data on student performance in online programs and assessments. Computers have also made it easier for teachers to vary their instructional delivery.
- Computer have always assisted to solve the problems faced by the mankind since the time of invention, the size of the computers have drastically reduced from that of a room to that can be accommodated in a human palm. The word computer means something which computes or a machine for performing calculations automatically, but, today computer means not merely a calculator. It does vast variety of jobs with tremendous speed and efficiency.

Computer Aided Research

- The importance of computers in scientific research is exceptionally high and the use of a computer can help scientific research immensely, and is an almost invaluable tool. There are many reasons why computers are so important in scientific research and here are some of the main reasons:
- **SPEED:** computer can process numbers and information in a very short time. So researcher can process and analyze data quickly. By saving time researcher can conduct further research. A calculation that may take a person several hours to process will take computer mere minutes, if not seconds.
- **ACCURACY:** Computer is incredibly accurate. Accuracy is very much important in scientific research. Wrong calculation could result an entire research or project being filled with incorrect information.

Computer Aided Research

- **ORGANIZATION:** We can store millions of pages of information by using simple folders, word processors & computer programs. Computer is more productive & safer than using a paper filing system in which anything can be easily misplaced.
- **CONSISTENCY:** computer cannot make mistakes through "tiredness" or lack of concentration like human being. This characteristic makes it exceptionally important in scientific research.
- Research process consists of series of actions or steps necessary to effectively carry out research and the desired sequencing of these steps. The following order concerning various steps provides a useful procedural guideline regarding the research process: (1) formulating the research problem; (2) extensive literature survey; (3) developing the hypothesis; (4) preparing the research design; (5) determining sample design; (6) collecting the data; (7) execution of the project; (8) analysis of data; (9) hypothesis testing; (10) generalisations and interpretation, and (11) preparation of the report or presentation of the results, i.e., formal write-up of conclusions reached.

Five major Phases of the research process

- Conceptual phase
- Design and planning phase
- Data collection phase
- Data Analysis phase and
- Research Publication phase
- **Role of Computer in Conceptual Phase:** The conceptual phase consists of formulation of research problem, extensive literature survey, theoretical frame work and developing the hypothesis.
- **Use of computers in extensive literature review:** computers help for searching the literatures (for review of literature) and bibliographic reference stored in the electronic database of the world wide webs. It can thus be used for storing relevant published articles to the retrieved whenever needed. This has the advantage over searching the literatures in the form of books, journals and other newsletters at the libraries which consume considerable amount of time and effort.

Role of Computers in Design and Planning Phase

- This phase consists of research design preparation and determining sample design. Design and planning phase also consists of population, research variables, sampling plan, reviewing research plan and pilot study. Role of Computers for Sample Size Calculation: Several software's are available to calculate the sample size required for a proposed study. The standard deviation of the data from the pilot study is required for the sample size calculation.
- **Role of Computers in Data collection phase:** This Empirical phase consists of collecting and preparing the data for analysis: In research studies, the preparation and inputting data is the most labor-intensive and time consuming aspect of the work. Typically the data will be initially recorded on a questionnaire or record for suitable for its acceptance by the computer. To do this the researcher in conjunction with the statistician and the programmer, will convert the data into Microsoft word file or excel spreadsheet or any statistical software data file. These data can be directly opened with statistical software's for analysis.

Role of Computers in Data Analysis

- This phase consist of the analysis of data, hypothesis testing and generalizations and interpretation. Data analysis phase mainly consist of statistical analysis of the data and interpretation of results.
- **Data analysis:** many software's are now available to perform the mathematical part of the research process i.e. the calculations using various statistical methods.
- Software like SPSS and spreadsheets are the widely used. They can be like calculating the sample size for a proposed study, hypothesis testing and calculating the power of the study.
- Familiarity with any one package will suffice to carry out the most intricate statistical analysis. Computers are useful not only for statistical analysis, but also to monitor the accuracy and completeness of the data as they are collected. These software's also display the results in graphical char or graph form.

Role of Computers in Research Publication

- This phase consists of preparation of the report or presentation of the results, i.e., formal write-up of conclusions reached. This is the research publication phase.
- The research article, research paper, research thesis or research dissertation is typed in word processing software and converted to portable data format (PDF) and stored and/or published in the world wide web.
- Online sites are available through we can convert our word file into any format like html, pdf etc. Various online applications are also available for this purpose. Even we can prepare our document using online word processing software and can store/edit/access it from anywhere using internet.

Role of Computers in Scientific Research - I

- There are various computer applications used in scientific research. Some of the most important applications used in scientific research are data storage, data analysis, scientific simulations, instrumentation control and knowledge sharing.
- **Data Storage:** Experimentation is the basis of scientific research. Every experiment in any of the natural sciences generates a lot of data that needs to be stored and analyzed to derive important conclusions, to validate or disprove hypotheses. Computers attached with experimental apparatuses, directly record data as it's generated and subject it to analysis through specially designed software. Data storage is possible in SPSS data file, lotus spreadsheet, excel spreadsheet, ASCII/DOS text file etc.
- **Data Analysis:** Analyzing tons of statistical data is made possible using specially designed algorithms that are implemented by computers. This makes the extremely time-consuming job of data analysis to be a matter of a few minutes. In genetic engineering, computers have made the sequencing of the entire human genome possible. Data from different sources can be stored and accessed via computer networks set up in research labs, which makes collaboration simpler.

Role of Computers in Scientific Research - II

- **Scientific Simulations:** One of the prime uses of computers in pure science and engineering projects is the running of simulations. A simulation is a mathematical modeling of a problem and a virtual study of its possible solutions. Problems which do not yield themselves to experimentation can be studied through simulations carried out on computers. For example, astrophysicists carry out structure formation simulations, which are aimed at studying how large-scale structures like galaxies are formed. Space missions to the Moon, satellite launches and interplanetary missions are first simulated on computers to determine the best path that can be taken by the launch vehicle and spacecraft to reach its destination safely.
- **Instrumentation Control:** Most advanced scientific instruments come with their own on-board computer, which can be programmed to execute various functions. For example, the Hubble Space Craft has its own onboard computer system which is remotely programmed to probe the deep space. Instrumentation control is one of the most important applications of computers.

Tools and Application used in Research Process

- **Statistical Analysis Tool:** SPSS is the most popular tool for statisticians. SPSS stands for Statistical Package for Social Sciences. The latest version of SPSS is IBM SPSS STATISTICS 20 (purchased by IBM after version 19). It provides all analysis facilities like following and many more.
- Provides Data view & variable view
- Measures of central tendency & dispersion
- Statistical inference
- Correlation & Regression analysis
- Analysis of variance
- Non parametric test
- Hypothesis tests: T-test, chi-square, z-test, ANOVA, Bipartite variable
- Multivariate data analysis
- Frequency distribution
- Data exposition by using various graphs like line, scatter, bar, ogive, histogram, pie chart

Data Base Management Packages

- Database is an organized collection of information. A DBMS is a software designed to manage a database.
- Various Desktop Databases are Microsoft Access, Paradox, Dbase or DbaseIII+, FoxBase, Foxpro/ Visual Foxpro, FileMaker Pro.
- Commercial Database Servers that supports multiuser are Oracle, Ms-SQL Server, Sybase, Ingres, Informix, DB2 UDB (IBM), Unify, Integral, etc.
- Open source Database packages are MySQL, PostgreSQL, Firebird etc.
- **Browsers:** A web browser is a software application which enables a user to display and interact with text, images, videos, music, games and other information typically located on a Web page at a website on the World Wide Web or a local area network.
- **Examples:** Microsoft Internet explorer, Mozilla firefox, Opera, Netscape navigator, Chrome (google browser), Safari.

Unit – V
Report Writing and Presentation

Dr. G. Yoganandham
Professor and Head,
Department of Economics,
Thiruvalluvar University.

Writing a Report

- During your studies you may be required to research a particular area and particular area and produce a report.
- For instance, depending on your area of study, you might be asked to write a report on the performance characteristics of an aircraft engine, the classroom management practices of a teacher, the cost benefits of using a particular type of software in a company, etc.
- **Some of the reasons we write reports are to:**
  - Inform
  - Make proposals or recommendation for change.
  - Analyze and solve problems.
  - Present the findings of an investigation or project.
  - Record progress.
- **Your lecturer will usually provide you with the following information:**
  - The topic or subject of the report.
  - The required length and due date.
  - A clear idea of its purpose and who will read it.
  - The format headings to be used and their order.

Format of a Report - I

- Although, there is no set report writing format, however, there are general sections that should be included.
- Unlike essays, reports are written in sections with headings and sub-headings, which are usually numbered.
- There are numerous possible formats available for writing a report, and it mostly depends on the context of topic.
- **Below Given are the components of a report:**
  - **Title Page:** It should include the title, your name and the name of the tutor to whom it is being submitted, date of submission, your course/department. The logo of the organization should also be printed.
  - **Acknowledgements:** A list of people and organizations who have helped you in the compilation of report and other related work.
  - **Content Page:** A clear, well-formatted list of all the sections and sub-sections of the report. Page numbers should be marked correctly.

Format of a Report - II

- **Abstract:** A summary of the major points, conclusions, and recommendations should be written to give a general overview of report.
- **Introduction:** The first page of the report needs to have an introduction. You will explain the problem and show the reader why the report is being made.
- **Body:** This is the main section of the report. There needs to be several sections, with each having a subtitle. The various sections include Review of Literature, Materials and Methods and Results. A discussion section can also be included at the end of the body to go over by findings and their significance.
- **Conclusion:** A conclusion should draw out the implications of your findings, with deductions based on the facts described in your main body. The significance and relevance of study is discussed in this section.
- **References:** This is a list giving the full details of all the sources to which you have made references within your text.

Standard Format for Report Writing

Format guidelines:

- font: 12-point Times New Roman
- double-spaced lines
- 1 in (2.5 cm) margins all around
- title page
  - title of the report
  - your name
  - the date
  - instructor name
- optional
  - table of contents
  - abstract or summary

I. Title or cover page: this includes the followings:

- 1-report title : states the purpose of the report . consist of a challenging statement or question,
- 2-student's name,
- 3-the module name,
- 4-the supervisor’s name,
- 5-the institution,
- 6-the academic year, etc.

2-Table of contents:

- (if the report is longer than10 pages).
- It is an essential List of headings given to each section of the report, together with its page number.
- Begin numbering the pages at the introduction.
- Anything before the introduction is given a Roman numeral.
- It provides the reader a quick overview of the major sections of report, with page references

Table of contents (example):

- Section..... Page Number
- introduction 1
- Literature review.....3
- _Methodology section.....5
- _Results.....7
- Discussion 11
- conclusion.....15
- _Recommendations.....16
- appendices(list tables, figures and appendices separately).....17
- bibliography.....20
- List of Figures: Bar Chart, Pie Chart .....
- List of Tables, figures

3-List of abbreviations and glossary:

- glossary: is an alphabetical list of technical terms, with brief explanations of their meanings,
- Abbreviations: technical terms should be stated in full when used for the first time in the report. If there are many, they should be listed in alphabetical order.

4-Abstract:

- It is a summary of the report, in which few sentence for the main section are included.
- It follows the title page;
- it is formatted as one paragraph.
- Use complete sentences; be concise, but not telegraphic.
- It should not be just a list of the headings.
- It helps the reader to make a decision, but not to substitute reading the whole article.

Abstract contains: -

- 1- Very brief description of the problem (WHAT has been studied)
- 2- brief description of the quality of the literature
- 3- the main objectives (WHY this study was needed)
- 4- the place of study (WHERE)
- 5- the type of study and methods used; subjects, apparatus and procedure used (HOW)
- 6- major findings and conclusions
- 7- the major recommendations.
- It is about 250 words.

Introduction

- is extremely important brief description of the research question, topic, focus and the method of its investigation.
- It should contain some relevant background data related to the problem, and then the statement of the problem should follow.
- It should contain a paragraph on what researcher hoped to achieve with the results of the study.

Introduction

- It serves to justify the study;
- it is first opportunity to highlight importance of the work.
- Introduction establishes the need for research within the current knowledge of the discipline, clarifying how this work will contribute to knowledge in the field.

B- Statement of the Research Problem/justification

- Statement of the problem clearly identified in nature and scope in the first paragraph or the gaps in the previous research, the unresolved conflicts in the field that still require investigation, new developments that are required by the current state of knowledge in your field.

B- Statement of the Research Problem/justification

- Statement of the problem clearly identified in nature and scope in the first paragraph or the gaps in the previous research, the unresolved conflicts in the field that still require investigation, new developments that are required by the current state of knowledge in your field.

C- Objectives & Hypothesis

- The general and specific objectives should be included as stated in the proposal.
- If necessary, adjust them slightly, but not to change their basic nature.
- They determine the methodology and how to structure the reporting of the findings.
- If some of the objectives are not met, this should be stated in the methodology section and in the discussion of the findings.

Language and Presentation

Verbs

- Use past tense for research actions (occasionally shifting to present tense for other purposes)
- Use active when possible, switching to passive voice to downplay researcher role as required
- Aim for precision

Visual design

- Use informative headings
- Include useful visuals (e.g., photos or illustrations of samples, equipment and procedures)

Example (modified from Belcher, 2005, p.94)

5.1 Multi-Method Charcoal Analysis

In order to maximise the amount of information gained from the charcoal record, three separate methods of preparation have been used: 1) polished blocks were collected, 2) demineralised sediment samples were captured, and 3) these sediment samples were sieved into two separate fraction sizes, macroscopic and microscopic.

general to specific description

Studying charcoal in polished blocks is essential in order to establish the exact stratigraphic location of any charcoal. They will, however, only represent a "snapshot" of the section's charcoal content. Residues from demineralised sediment have a lower spatial resolution but will reveal the sizes and shapes of particles and the overall charcoal content of each horizon (e.g. particles of charcoal per gram of sediment).

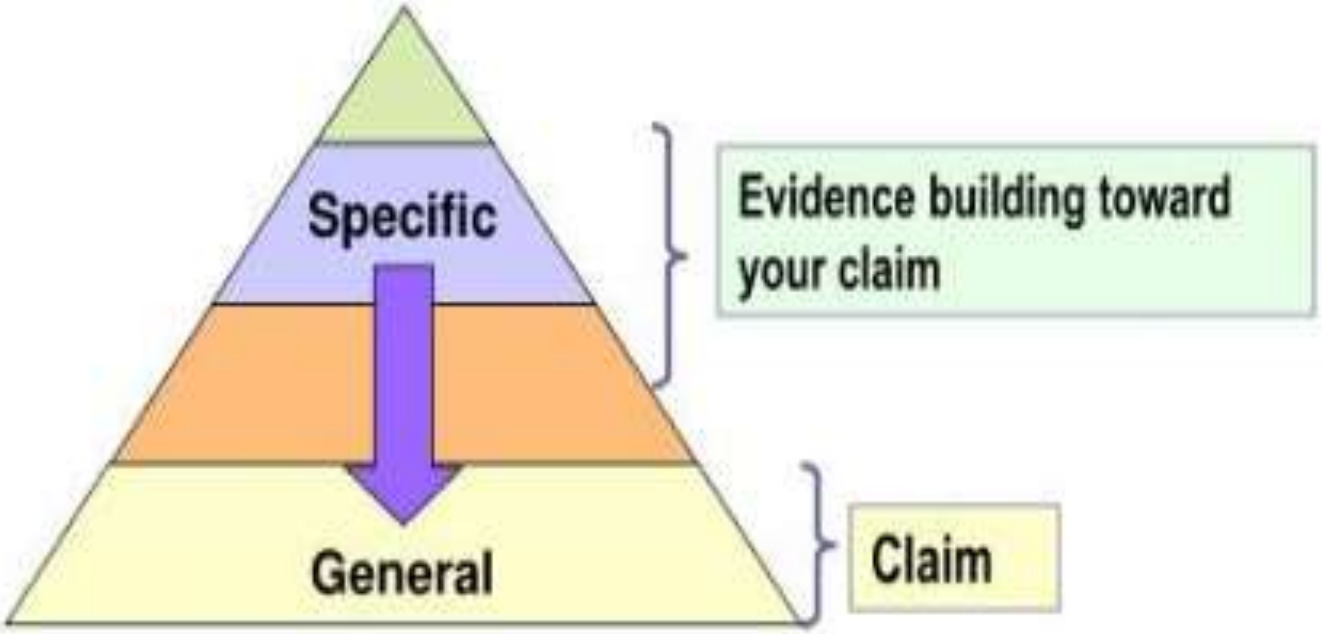
Justification of method

Presenting the Results

- Separate results from methods and discussion
- Summarize the research, reminding readers of key aspects; describe the sample
- Present the results first and analyze them later in the chapter (or section)
- Provide detailed data in labelled tables / figures; use appendices if necessary for raw data or detailed results
- Refer to visuals and appendices explicitly (e.g., Figure 2; Appendix A) where appropriate in the discussion.
- Provide conceptual links between results and the research design: connect findings to your research

Discussing Results: Inductive Logic

- Developing a new claim (Evidence→ Claim)
- Presenting contested claims
- Justifying decisions; discussing weaknesses



Discussing Results & Concluding

- Summarize major findings
- Connect findings to your research questions, objectives, and hypotheses
- Highlight the meaning of the findings and relate the findings to the literature
  - *Do they support previous research or theory?*
- Discuss limitations of the research
- Discuss contributions to knowledge in your field, applications to theory development, and implications for practice
- Indicate directions for future research

References: Citation Resources

- UofC Effective Writing Program writing resources: http://efwr.ucalgary.ca/citationstyles

Resources for

- **APA**
- **Chicago**
- **CSE**
- **Harvard**
- **IEEE**
- **MLA**
- **Vancouver**

Appendices

- Use appendices for detailed information related to research methods or results
- Label as Appendix A, B, C,... & title each
- List by name in your Table of Contents

References (Samples)

Belcher, C. (2005). Assessing the evidence for extensive wildfires at the Cretaceous-Tertiary boundary. PhD dissertation: Department of Geology, Royal Holloway University of London.

Chiang, K.K., (2002). Geochemistry of the Cenozoic igneous rocks of Borneo and tectonic implications. PhD dissertation: Department of Geology, Royal Holloway University of London.

Hunton, J. &, Rose, J. (2005) Cellular telephones and driving performance: The effects of attentional demands on motor vehicle crash risk. *Risk Analysis, 25, No. 4, 855-866.*

Klapperich, C. M. & Bertozzi. (2004) Global gene expression of cells attached to a tissue engineering scaffold, C. *Biomaterials*, 25, No. 25, p5631.

Nam, S. *et al.* (2009). An inkjet-printed passivization layer based on a photocrosslinkable polymer for long-term stable pentacene field-effect transistors. *Organic Chemistry*. 10, 67–72.

Peat, I. U. (2003). *Volcanostratigraphy, geochronology and geochemistry of silicic volcanism in the Afro-Arabian flood volcanic province (Yemen and Ethiopia).* PhD Thesis. University of London.

Romer, U. (2005). "This seems somewhat counterintuitive, though..." Negative evaluation in book reviews by male and female authors. In Tognini-Bonelli, E. & Camiciotti, G. (eds.) *Strategies in academic discourse.* Amsterdam: John Benjamins Publishing.

Ullmer, B. A. (2002). *Tangible interfaces for manipulating aggregates of digital information.* Retrieved September 14, 2007, from http://alumni.media.mit.edu/~ullmer/dissertation/

Purpose of Research Report

- The purpose of research report is to convey the interested persons the whole result of study in sufficient detail and to determine himself the validity of the conclusions.
- The research purpose is a statement of "why" the study is being conducted, or the goal of the study. The goal of a study might be to identify or describe a concept or to explain or predict a situation or solution to a situation that indicates the type of study to be conducted.
- Strengthen Your Credibility and Reputation. By using objective analysis to make decisions, you will not only develop better business strategies, you will also improve your own professional reputation in the field and help others to have more confidence in your conclusions.
- Complete and Compact Document : **Report** is a complete and compact written document giving updated information about a specific problem. Systematic Presentation of Facts : **Report** is a systematic presentation of facts, figures, conclusions and recommendations.

Classification of Research Report - I

- **Based on Purpose**
  - **Basic Research / Pure or Fundamental Research:** To develop new ways of thinking, theories, principles and ideas.
  - **Applied Research:** To apply research or past theories, knowledge and methods for solving an existing problems.
  - **Evaluation Research:** To facilitate decision making regarding the relative worth of 2 or more actions.
  - **Research & Development:** To develop effective products demanded by the existing clients.
  - **Action Research:** To solve problems through application of scientific method.

Classification of Research Report - I

- **Based on Method**
  - **Historical Research:** Involves studying, understanding & explaining past events.
  - **Descriptive Research:** Collecting data in order to test or answer questions concerning current subject of the study (Survey, Development Studies, Fellow up).
  - **Co relational Research:** Determines whether and to what degree, a relationship exists between 2 or more quantifiable variables.
  - **Casual-Comparative & Experimental Research:** Establishes causes & effect relationships, cause already occurred.
- **Based on Approach**
  - **Qualitative Research:** Exploration, description, generation & construction of theories using qualitative data.
  - **Quantitative Research:** Testing theories & hypothesis using qualitative data to see if they are confirmed or not.

Research Report Purposes

- To explore a new topic/issue in order to learn about it.To satisfy the researcher’s curiosity and desire for better understanding, To test the feasibility of undertaking a more extensive study. Addresses the "what" question: "what is this phenomena really about?" Experience surveys, Secondary data analysis, Case studies, Pilot studies, Focus groups or small group discussions Study on what stem cell therapy is all about.
- **Descriptive:** To describe situations and events. To present a picture of the specific details (or gather info) of a situation, social setting, or relationship Focus on `who,' `what,' `when,' `where,' and `how' but not why? Data, gathering techniques, surveys, field research, content analysis Labor Force Surveys, Population Census, (including race/ethnicity, age, sex, household size, income, etc.) and Educational surveys.

Research Report Purposes

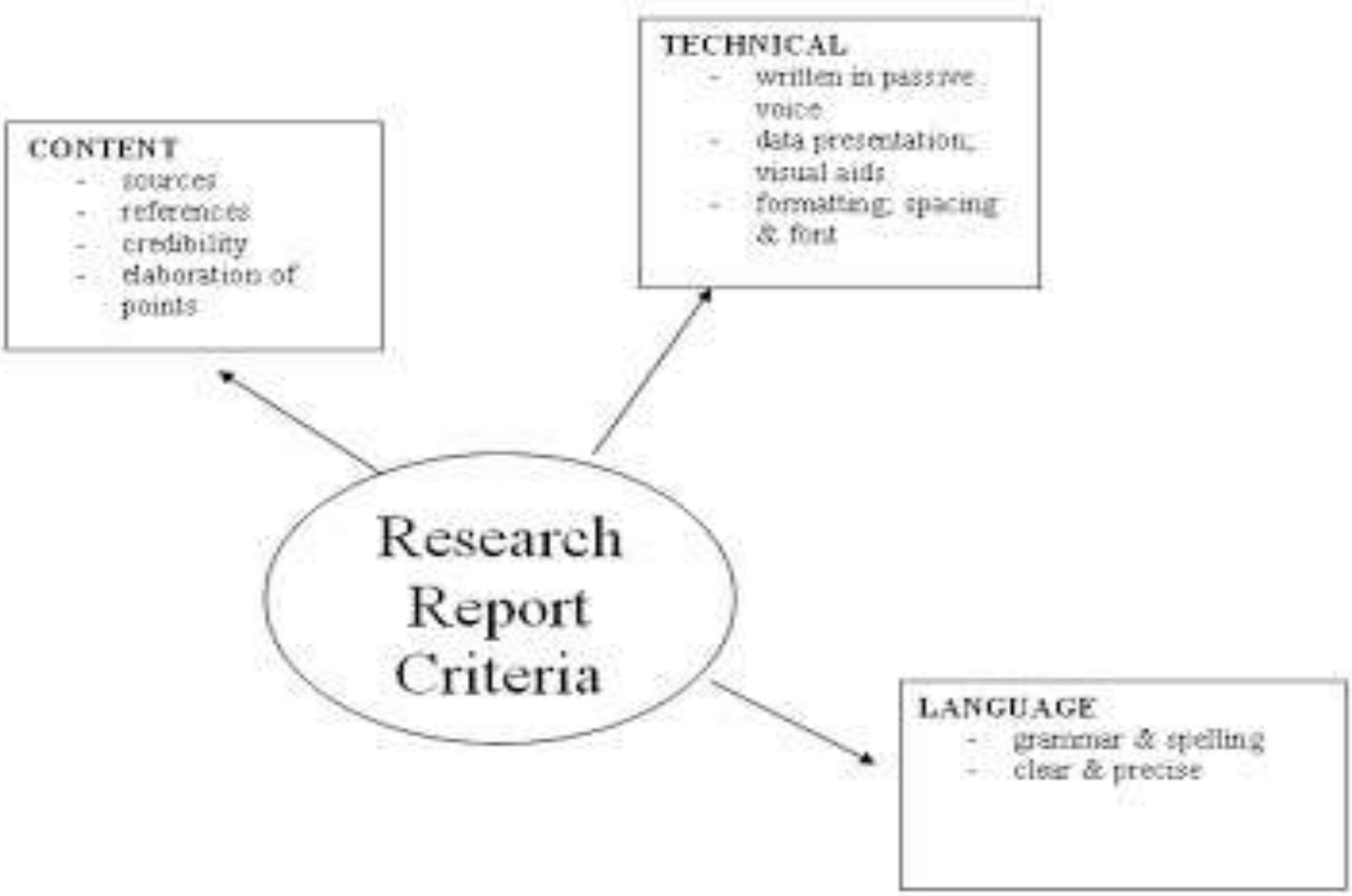
- **Explanatory:** To know "why," to explain things, looks for causes and reasons -Builds on exploratory and descriptive research and goes on to identify the reasons for something that occurs.
- To extend a theory or principle to new areas, new issues, new topics.
- To provide evidence to support or refute an explanation or prediction.
- To test a theory's predictions or principle All of the above For example, reporting the autism rates of different cities is descriptive.
- Identifying the variables that explain why some cities have higher autism rates than others involves explanation.
- Likewise, reporting the frequency of school attendance is descriptive, but reporting why some people attend school while others don’t is explanatory.

Criteria to write a Good Research Report



- **Write in a positive manner.**
- **Keep it short.**
- **Try to be original.**
- **Use simple language.**
- **Data should be accurate.**
- **Wherever possible use graphs, pictures.**
- **Have consistency of logic.**
- **Don’t repeat the same ideas.**
- **Detail the design of the study.**
- **Organize the material in major themes.**
- **Acknowledge the work of others through references or footnotes.**
- **The cover page should catch the attention.**
- **Heading, sub-headings should be clear.**

Criteria for Research Report



Criteria for Evaluating Research Reports

- Fundamental criterion: Communication with the reader or readers.
- Writing Criteria
- Completeness: Whether the report provides all the information readers need in a language they understand.
- Accuracy: Whether the reasoning in the report is logical and the information correct.
- Clarity: Whether the phrasing in the report is precise.
- Conciseness: Whether the writing in the report is crisp and direct.
- The review of previous research.
- The problem and the purpose.
- The hypothesis.
- The method.
- The sample.
- Results and discussion.
- References.
- General comments about the report.



Criteria of a Good Research

- Purpose should be clearly defined.
- Research process/procedure should be detailed.
- Research design should be thoroughly planned objectively.
- High ethical standards should be applied/report frankly.
- Limitations and Assumptions should be frankly revealed.
- Data, Analysis and Reliability.
- Findings should be presented unambiguously.
- Conclusions should be justified.
- Researcher’s experience and integrity should be reflected.
- Systematic.
- Logical.
- Empirical.
- Replicable.



Structure of the Research Report - I

- A key feature of reports is that they are formally structured in sections. The use of sections makes it easy for the reader to jump straight to the information they need.
- Unlike an essay which is written in a single narrative style from start to finish, each section of a report has its own purpose and will need to be written in an appropriate style to suit – for example, the methods and results sections are mainly descriptive, whereas the discussion section needs to be analytical.
- Understanding the function of each section will help you to structure your information and use the correct writing style. Reports for different briefs require different sections, so always check carefully any instructions you've been given.
  - Title
  - Abstract
  - Introduction
  - Literature Survey
  - Methods
  - Results
  - Discussion
  - Summary, Findings, Conclusion, Suggestions and Future Research.
  - Bibliography.
  - Appendices.



Structure of the Research Report - II

- **Title:**
- The title needs to concisely state the topic of the report. It needs to be informative and descriptive so that someone just reading the title will understand the main issue of your report. You don't need to include excessive detail in your title but avoid being vague and too general.
- **Abstract (Also called Summary or Executive Summary):**
- This is the 'shop window' for your report. It is the first (and sometimes the only) section to be read and should be the last to be written. It should enable the reader to make an informed decision about whether they want to read the whole report. The length will depend on the extent of the work reported but it is usually a paragraph or two and always less than a page.

Structure of the Research Report - III

- A good way to write an abstract is to think of it as a series of brief answers to questions. These would probably include:
  - What is the purpose of the work?
  - What methods did you use for your research?
  - What were the main findings and conclusions reached as a result of your research?
  - Did your work lead you to make any recommendations for future actions?
  - What is the purpose of the work?
  - What methods did you use for your research?
  - What were the main findings and conclusions reached as a result of your research?
  - Did your work lead you to make any recommendations for future actions?

Structure of the Research Report - IV

- **Introduction (Also called Background or Context):**
- In this section you explain the rationale for undertaking the work reported on, including what you have been asked (or chosen) to do, the reasons for doing it and the background to the study.
- It should be written in an explanatory style. State what the report is about - what is the question you are trying to answer? If it is a brief for a specific reader (e.g. a feasibility report on a construction project for a client), say who they are.
- Describe your starting point and the background to the subject, for instance: what research has already been done (if you have been asked to include a Literature Survey later in the report, you only need a brief outline of previous research in the Introduction); what are the relevant themes and issues; why are you being asked to investigate it now? Explain how you are going to go about responding to the brief.
- If you are going to test a hypothesis in your research, include this at the end of your introduction. Include a brief outline of your method of enquiry. State the limits of your research and reasons for them, for example; “Research will focus on native English speakers only, as a proper consideration of the issues arising from speaking English as a second language is beyond the scope of this project”.

Structure of the Research Report - V

- **Literacy Survey (Also called Literature Review or Survey / Review of Research):**
- This is a survey of publications (books, journals, authoritative websites, sometimes conference papers) reporting work that has already been done on the topic of your report.
- It should only include studies that have direct relevance to your research. A literature survey should be written like an essay in a discursive style, with an introduction, main discussion grouped in themes and a conclusion.
- Introduce your review by explaining how you went about finding your materials, and any clear trends in research that have emerged. Group your texts in themes.
- Write about each theme as a separate section, giving a critical summary of each piece of work and showing its relevance to your research. Conclude with how the review has informed your research (things you'll be building on, gaps you'll be filling etc).

Structure of the Research Report - VI

- **Methods (Also called Methodology):**
- You need to write your Methods section in such a way that a reader could replicate the research you have done. There should be no ambiguity here, so you need to write in a very factual informative style.
- You need to state clearly how you carried out your investigation. Explain why you chose this particular method (questionnaires, focus group, experimental procedure etc), include techniques and any equipment you used.
- If there were participants in your research, who were they? How many? How were they selected? Write this section concisely but thoroughly – go through what you did step by step, including everything that is relevant. You know what you did, but could a reader follow your description?

Structure of the Research Report - VII

- **Results (Also called Data or Findings):**
- This section has only one job which is to present the findings of your research as simply and clearly as possible.
- Use the format that will achieve this most effectively e.g. text, graphs, tables or diagrams.
- When deciding on a graphical format to use, think about how the data will look to the reader.
- Choose just one format - don't repeat the same information in, for instance, a graph and a table. Label your graphs and tables clearly. Give each figure a title and describe in words what the figure demonstrates.
- Writing in this section should be clear, factual and informative. Save your interpretation of the results for the Discussion section.

Structure of the Research Report - VIII

- **Discussion:**
- This is probably the longest section and worth spending time on.
- It brings everything together, showing how your findings respond to the brief you explained in your introduction and the previous research you surveyed in your literature survey.
- It should be written in a discursive style, meaning you need to discuss not only what your findings show but why they show this, using evidence from previous research to back up your explanations.
- This is also the place to mention if there were any problems (for instance, if your results were different from expectations, you couldn't find important data, or you had to change your method or participants) and how they were or could have been solved.

Structure of the Research Report - IX

- **Conclusion:**
- Your conclusions should be a short section with no new arguments or evidence. Sum up the main points of your research - how do they answer the original brief for the work reported on?
- This section may also include:
- Recommendations for action
- Suggestions for further research





Structure of the Research Report - X

- **References (Also called Reference List or Bibliography):**
- List here full details for any works you have referred to in the report, including books, journals, websites and other materials.
- You may also need to list works you have used in preparing your report but have not explicitly referred to - check your instructions for this and for the correct style of referencing to use.
- You can find information about how to reference more unusual materials (television programmes, blogs etc) from various websites including the Learn Higher website on referencing.
- If you're not sure, the rule is to be consistent and to give enough details that a reader can find the same piece of information that you used.
- **Appendices:**
- The appendices hold any additional information that may help the reader but is not essential to the report's main findings: anything that 'adds value'.
- That might include (for instance) interview questions, raw data or a glossary of terms used. Label all appendices and refer to them where appropriate in the main text (e.g. 'See Appendix A for an example questionnaire').

Suggested order to write the main sections of the Report

- **Methods and Data/Results:** As a rough guide, the more factual the section, the earlier you should write it. So sections describing ‘what you did and what you found’ are likely to be written first.
- **Introduction and Literature Survey:** Sections that explain or expand on the purpose of the research should be next. What questions are you seeking to answer, how did they arise, why are they worth investigating? These will help you to see how to interpret and analyse your findings.
- **Discussion:** Once you’ve established the questions your research is seeking to answer, you will be able to see how your results contribute to the answers and what kind of answers they point to. Write this early enough that you still have time to fill any gaps you find.
- **Conclusions and Recommendations:** These should follow logically from your Discussion. They should state your conclusions and recommendations clearly and simply.
- **Abstract/Executive Summary:** Once the main body is finished you can write a succinct and accurate summary of the main features.

Bibliography

- A bibliography is a list of all of the sources you have used (whether referenced or not) in the process of researching your work. In general, a bibliography should include:
  - The authors' names.
  - The titles of the works.
  - The names and locations of the companies that published your copies of the sources.
  - The dates your copies were published.
  - The page numbers of your sources (if they are part of multi-source volumes).
- A bibliography is a list of the books and other sources that are referred to in a scholarly work-such as an essay, term paper, dissertation, or a book. The bibliography comes at the end of the work.
- The bibliography usually does list the references in alphabetical order, regardless of the style or format.

Types of Bibliography

S. No.	Modern Language Association (MLA)	Likenesses	American Psychological Association (APA)
1.	Most commonly used in the Arts & Humanities fields	Both papers are double spaced, this includes the work cited or reference pages. Hanging indents are used for citation.	Most commonly used in the social sciences fields.
2.	There is a 'Work Cited' page at the end of the paper to cite all works used in the research process.	Every piece of information used in the text of the paper must be included in the reference or works cited pages.	There is a 'reference page' at the end of the paper to cite all works used in the research process.
3.	Whenever information is cited, and the name of the author is listed in the same sentence, the page number must be placed at the end.	Body necessitate the use of parenthetical citations in the body of the paper.	Whenever information is cited, and the name of the author is listed in the same sentence, the year of publication should also be included.
4.	Whenever information is listed and the name of the author is not listed, enter the surname of the author and the page number at the end.	List citations in alphabetical order on both the reference and works cited pages.	Whenever information is cited and the author's name is not listed, place the author's surname and the year of publication in the sentence.

References

- A reference list is a list of the publication information for the sources you've cited in your paper and is intended to give your readers all the information they need to find those sources. In other publication styles, this list may be called a bibliography or a works cited page, but APA uses the term reference list.
- **Reference Contains:**
  - Author/Editor (if it is an editor always put (ed.)
  - Title (this should be in italics).
  - Series title and number (if part of a series).
  - Edition (if not the first edition).
  - Place of publication (if there is more than one place listed, use the first named).
  - Publisher.
  - Year of publication.

Formatting the Reference List

- Start the reference list on a new page.
- **Bold** the section label "**References**" and center it at the top of the page.
- **Double-space** the entire reference list, both within and between entries.
- **Alphabetize** entries by the **first word** of the reference (usually the first author's surname).
- Place multiple works by the **same authors in chronological order**, with the earliest first.



Footnotes

- Footnotes are notes placed at the bottom of a page.
- They cite references or comment on a designated part of the text above it.
- For example, say you want to add an interesting comment to a sentence you have written, but the comment is not directly related to the argument of your paragraph.
- **How to insert footnotes:**
  - ✓ Place the cursor where you would like the superscript number to appear.
  - ✓ Click on “Insert Footnote” in the “References” tab. ...
  - ✓ The corresponding number will be automatically inserted in the footer ready for you to add the footnote citation.
  - ✓ Type in your footnote citation.

Footnotes	Endnotes
Both are additional notes to what is mentioned in the text.	
Include a superscript (example: [1]) next to reference or quote.	
Notes are located at the bottom or "foot" of the page.	Notes are located at the end of the section, chapter, or book.
PRO: Let readers see the sources and notes as they read.	PRO: Reduce distraction on the page.
CON: Could interrupt document flow since they take up page space.	CON: References are not as accessible, and the reader must search for another page/section to find the reference.

BIBLIOGRAPHY

VERSUS

WORKS CITED

BIBLIOGRAPHY	WORKS CITED
List of the external sources or materials referred to in a scholarly work, typically printed as an appendix	A list of only the sources or literature referenced in the body of an academic article
Necessarily includes all the consulted or referred literature sources in the book or academic article	Only includes the sources the student have cited in writing
Usually used in works with a higher academic purpose such as a published book or an academic paper	Used in works with a comparatively lower academic purpose such as a student essays, or academic research papers of undergraduates
At the end of the academic paper, or published book	At a separate page at the end of the writing
Written either in the MLA, APA, Chicago or Turabian style depending on the style of writing followed	Written in the MLA style

Visit www.PEDIAA.com

Difference between Reference and Bibliography - I

Parameters of Differences	References	Bibliography
Content	Comprises the sources of the materials such as texts, books, websites, magazines, periodicals and scientific papers which have actually been quoted in the book or essay.	Comprises all the research materials such as text, books, websites, magazines, periodicals and scientific papers which are not necessarily quoted in the book or essay.
Order	Appears immediately beneath the essay and above the bibliography.	Appears below the reference list.
Scope	Only limited to the materials and contents that have actually been quoted in the essay work. It is hence quite limited in scope.	Pretty broad in scope in that it comprises those contents which have been quoted in the essay work and some others which are relevant but not quoted.

Difference between Reference and Bibliography - II

Parameters of Differences	References	Bibliography
Arrangement	A reference may be arranged numerically or alphabetically. This means the reference list is more flexible and broader than that of a bibliography.	The bibliography may only be arranged alphabetically. As such, it is not as flexible as the reference list.
Contents	A reference showcases the source of the information such as a website or book followed by the year of publication or the time when you visited the web page.	A bibliography, on the other hand, includes much more information than a reference. It contains the last and first names of the author, title of the book, the year of publication, place and name of the publisher.

Reference Vs. Bibliography

*The terms References and Bibliography are often used same, but there is a difference in meaning between them.

*References are the items you have read and specifically referred to (or cited) in your work, and your list of sources at the end of the assignment will be headed References.

*Bibliography is a list of everything you read -whether or not you referred specifically to it .

* A bibliography can give a tutor an overview of which authors have influenced your ideas and arguments even if you do not specifically refer to them.

Reference vs. Bibliography

- Bibliography is a list of everything you read - whether or not you referred specifically to it .
- Normally contain sources that have been cited and also those found to be influential, but decided not to cite.
- A bibliography can give a tutor an overview of which authors have influenced your ideas and arguments even if you do not specifically refer to them.

CITATION VERSUS FOOTNOTE

CITATION	FOOTNOTE
Citation refers to a reference to a book, paper, or author, given in the body of the text itself	Footnote refers to a piece of information printed at the bottom of a page
Contains bibliographical information	Contain bibliographical information, author's comment or additional information
Allowed in all style guides	Not allowed in APA style
Located in the text itself	Found at the bottom of the page

Pediaa.com

Differences between footnotes and bibliography entries:

- *Footnotes* format the author's name **first name last name**. *Bibliography entries* format the author's name **last name, first name**.
- *Footnotes* **indent the first line**, and any subsequent lines are flush left. *Bibliography entries* have a **hanging indent**.
- *Footnotes* use **commas** and **parentheses** to separate elements in the entry. *Bibliography entries* use **periods**.

Differences between footnotes and bibliography entries:

- *Footnotes* include **specific page numbers** where the information was found. *Bibliography entries* only include page numbers if they indicate the pages of a complete **article** or **chapter**.
- *Footnotes* are **numbered** and ordered according to where the information is introduced in the text. *Bibliography entries* are always **alphabetized** and **unnumbered**.

	Notes and Bibliography Style	Author-Date Style
Area of scholarship	Humanities (literature, history, the arts)	Social sciences and sciences
Format of references	Sources are cited in footnotes or endnotes with either a full or shortened citation, depending on whether a full bibliography is provided.	Sources are cited briefly in-text, noting the author's last name and year of publication.
Bibliography	Usually included, but not always necessary if sources are cited in full in the footnotes or endnotes. If included, it is called the "bibliography."	Always required so that reader can identify sources briefly identified in-text. This is called the "reference" list in author-date style.
Flexibility	Highly flexible and can accommodate many types of sources.	Less flexible and best for typical academic sources.

Example in-text citation of a book using the Chicago Referencing Style

This is an example of in-text citation for the purposes of demonstrating the correct format to be used for Chicago referencing.[1] In-text citation is "important to acknowledge the work of others".[2]

A superscript number indicates the in-text citation, which corresponds to the same number in the footnotes at the bottom of the page.

The superscript number for a footnote appears after punctuation (including full stops, bracket and commas).

The first citation of a source is written in full.

---

[1] Margherita Letizia, *Example Footnote in the Chicago Style* (Melbourne: Monash University, 2016), 13.

[2] Letizia, *Example Footnote*, 14.

Subsequent citations of the same source are abbreviated.

Example for footnotes and Bibliography - I

Citation Structure for footnotes	Author First Name Author Last Name (user name), “Title,” Platform, Date Posted, URL.
Footnote example	1. Bill Gates, “As Melinda and I reflected on our foundation’s work over the last 20 years in our annual letter,” Facebook, February 23, 2020, https://www.facebook.com/BillGates/photos/a.10150331291841961/10156880911936961/?type=3&theater.
Citation structure for bibliography	Author Last Name, Author First Name (user name). “Title.” Platform, Date Posted, URL.
Bibliography example	Gates, Bill. “As Melinda and I reflected on our foundation’s work over the last 20 years in our annual letter.” Facebook, February 23, 2020. https://www.facebook.com/BillGates/photos/a.10150331291841961/10156880911936961/?type=3&theater.

Footnote formatting:

- The *first* time a source is used, list all the publication information within the footnote (this looks a lot like a bibliography entry).

[3]Daniel Miller, *Standard Footnotes* (Dallas: DBU Press, 2011), 45.

- In each *subsequent* entry that uses the same source, include the author's last name, a comma, and the page number.

[4]Miller, 88.

Information to include (if available):

- Author's name
- Article or chapter title
- Book title, journal title, website title, etc.
- Volume, issue, edition, etc.
- Editor's name
- Place of publication
- Publishing company
- Copyright date
- Page number

Some examples:

- Book with one author:

[1]James Smith, *My Book* (Ennis, TX: Some Press, 2004), 38.

- Book with two authors:

[1]Bert Tall and Ernie Short, *Streets of Home (Boston: Big Bird Publications, 1995), 24.*

- A multivolume work with editor/compiler as author:

[1]Bilbo Baggins, ed., *The Red Book of Westmarch (London: Tolkien Press, 19), 2:1439.*

Some examples:

- Article from a scholarly journal:

[1]Max Deluch "Mind from Matter," *American Scholar* 47, no. 7 (Spring 1978): 343.

- Electronic journal article:

[1]Novak, Ivana, "Keeping up with Bicarbonate," *The Journal of Physiology 528, no.* 2 (October 20, 2000), under "Medications," http://www.jphysiol.org/cig/content/full/528/235 (accessed October 24, 2000).

Basic bibliography formatting:

- Place a page titled "**Bibliography**" at the end of the paper.
- This includes an alphabetized list of each source consulted for the paper.
- Single-space each entry.
- Place a blank line between each entry.
- Each entry should have a hanging indent.

Difference between Ibid., & Op. Cit., - I

- Whilst ibid refers to the reference that has just been mentioned, op. cit. usually refers to a reference which is earlier in the text or in the bibliography (i.e. not the one that has just been mentioned). The author's name and date will help you to identify the correct source.
- Op.Cit., stands for opere citato - cit. is an abbreviation of the Latin phrase opere citato, meaning "in the work cited." It is used in an endnote or footnote to refer the reader to a previously cited work, standing in for repetition of the full title of the work.
- **The use ibid, Op. Cit., & Loc. Cit.,**
- Loc. cit. is used in place of ibid. when the reference is not only to the work immediately preceding, but also refers to the same page. Loc.

Difference between Ibid., & Op. Cit., - II

- An abbreviation for the Latin 'ibidem' which **means** 'in the same place'. When you see the word '**ibid**. ' in a reading list it is referring you to material in a source just mentioned.
- If you consecutively cite the same source two or more times in a note (complete or shortened), you may **use** the word "**Ibid**" instead. **Ibid** is short for the Latin ibidem, which means "in the same place". If you're referencing the same source but different page, follow '**Ibid**' with a comma and the new page number(s).

Difference between Footnotes, References, Endnotes and Citations

- The one difference between footnotes and endnotes is that footnotes appear at the bottom of the same page, while endnotes appear at the end of the paper. Footnotes and endnotes are used in printed documents to explain, comment on, or provide references for text in a document.
- The terms reference and citation are also often used to refer to the same thing although a citation tends to mean the part of the text within your assignment where you acknowledge the source; whilst a reference usually refers to the full bibliographic information at the end.
- The style of Chicago/Turabian we use requires footnotes rather than in-text or parenthetical citations. Footnotes or endnotes acknowledge which parts of their paper reference particular sources.

Examples for

- **Footnotes** are notes placed at the bottom of a page. They cite references or comment on a designated part of the text above it. For example, say you want to add an interesting comment to a sentence you have written, but the comment is not directly related to the argument of your paragraph.
- A **bibliography** is a list of the books and other sources that are referred to in a scholarly work-such as an essay, term paper, dissertation, or a book. The **bibliography** comes at the end of the work. The **bibliography** usually does list the references in alphabetical order, regardless of the style or format.
- When using **endnotes**, your quoted or paraphrased sentence or summarized material is followed by a superscript number. **Example**: Let's say that you have quoted a sentence from Lloyd Eastman's history of Chinese social life.
- An **example of reference** is a professor who will write a letter recommending a student for an internship. **Reference** means someone or something which is a source of information about a subject. An **example of reference** is an encyclopedia.

Notes in Research

- **Definition: Note** citing a particular source or making a brief explanatory comment placed at the end of a **research paper** and arranged sequentially in relation to where the reference appears in the **paper**.
- **Taking notes:**
  - ✓ Use abbreviations, acronyms, or incomplete sentences to record information to speed up the note taking process.
  - ✓ **Write** down only the information that answers your **research** questions.
  - ✓ Use symbols, diagrams, charts or drawings to simplify and visualize ideas.
- **Note making** is not just about writing down everything you hear or read. It is a **process** of reviewing, connecting and synthesizing ideas from your lectures or reading.
- **Making notes** helps you to: stay active and engaged during your lectures, reading and revision. be selective and identify key ideas.

Strategies for Organizing your Research

- Assemble printed sources and interact with them.
- Consider other methods of gathering data.
- Choose a system for keeping notes.
- Use your sources to generate ideas.
- Organize your ideas.
- Write your paper.
- Evaluate your argument.



Tips for Writing notes in Research

- Date **your notes** and make the main topic visible.
- Don't **write** everything down – **write** down the important points.
- Make short **notes** of the examples given.
- Use color.
- Use illustrations and drawing.
- Use headings and sub-headings.
- Keep **your** sentences short.
- If you need to see how things are connected, consider using mind maps.

Source Notes

- **Source note** information allows you to keep a formal reminder of the origin of your information. For the reader, it stands as a verification of the factual basis for the statement or assertion. Write about your reference in the main text only briefly, citing the author's name, and the page number, in parentheses.
- **BMC Research Notes** is an open access journal publishing peer-reviewed contributions from across all scientific and clinical disciplines, including intriguing initial observations, updates to previous work and established methods, valid negative results, and scientific data sets and descriptions.
- There is no sharp **distinction between** the two categories: **notes** can consist of or contain **references**, and **references** may be annotated. ... **References**, on the other hand, tend to be listed in appendices whose titles and locations are usually given on the contents page of the work in question.

Computer Assisted Research

- Computer-Assisted Research in the Humanities describes various computer-assisted research in the humanities and related social sciences.
- It is a compendium of data collected between November 1966 and May 1972 and published in Computer and the Humanities.
- The book begins with an analysis of language teaching texts including the DOVACK system, a program used for remedial reading instruction.
- It then discusses the objectives, types of computer used, and status of the Bibliographic On-line Display (BOLD), semiotic systems, augmented human intellect program, automatic indexing, and similar research.
- The remaining chapters present computer-assisted research on language and literature, philosophy, social sciences, and visual arts. Students who seek a single reference work for computer-assisted research in the humanities will find this book useful.

COMPUTER ASSISTED INSTRUCTION





OLD THEORETICAL MODEL OF LEARNING



NEW THEORETICAL MODEL OF LEARNING

INTRODUCTION

- Computers have become omnipresent and ubiquitous today and found in every walk of life.
- Computer influence every sphere of human activity and bring many changes in education, health care, scientific research, social sciences, law, music and painting.



DEFINITIONS:

- Computer-assisted instruction (CAI) is an interactive instructional technique whereby a computer is used to present the instructional material and monitor the learning that takes place.

OR

- A self-learning technique, usually offline/online, involving the interaction of students with programmed instructional materials.

TERMINOLOGY

Terminology	Meaning
CBT	Computer Based Training/Test
CAI	Computer Assisted Instruction
CAL	Computer Assisted Learning
WBT	Web Based Training/Tutorial
CALL	Computer Assisted Language Learning
WBI	Web Based Instruction

COMPUTER ASSISTED LEARNING



INTRODUCTION

- E-learning, the way a person learn and communicate electronically, is emerging as a key application of the internet. E-learning covers a wide set of application and processes, such as web –based learning, computer based learning, virtual class rooms, and digital collaboration



DEFINITION

- "Computer assisted learning is to convey a vast amount of information in a very short period if time.it is powerful of reinforcing concepts and topics first introduced to students through textbook and discussion in the classroom. Computer assessed learning enables in a powerful way to comprehend complex concepts"

  -Prof .Hussein Ashram

- It refers to anything delivered, enabled, or mediated by electronic technology for the explicit purpose of learning

TYPES OF COMPUTER ASSISTED LEARNING

- **Computer Assisted Instruction**-When computer is used as an aid to teaching method of instruction, it is called computer assisted instruction.
- **Computer Managed Instruction**-It is defined as the use of computer programs for the online management of the instructional process.



MODELS OF COMPUTER ASSISTED LEARNING

- DRILL AND PRACTICE MODE
- TUTORIAL METHOD
- LABORATORY MODE
- CASE-SIMULATION MODE
- CONSULTANT ROLE

PURPOSES OF CAL

- It supports educational and organisational goals
  - Improved training cost
  - Decreased material cost
  - Increased productivity
  - Standardization



ADVANTAGES OF COMPUTER ASSISTED LEARNING

- Each student receives instruction at his own pace.
- Each student responds continuously as he receives instruction.
- All units learning are broken down into sub units and small elements of learning.
- Students can access the computer at any place.
- Students can learn through their own ways and style. i.e., through examples through case studies or through problems.
- Students can test their learning at any time of progress.
- The computer is consistent, relevant and approving of every student.
- The program author can update a computer lesson as anew knowledge is identified.

DISADVANTAGES OF COMPUTER ASSISTED LEARNING

- Cost Factors
- Content control
- Lack of formal communication among user
- With excessive use of CAL, artistic competence of students takes a back seat.
- It produces more mental and physical stress as compared to other methods of teaching.
- It does not help in developing language proficiency which depends more on direct experience with the reality.
- As students cannot interact with the computers in a human language, the mechanical response by punching keys can become dull and frustrating.

Computer application in research

- Model processing
- Sample size calculation
- Managing missing data
- Performing computations
- In estimating accurate p-value
- Data analysis and interpretation

Computer application in research

Software Packages

SPSS – Statistical Package for Social Sciences

SAS – Statistical Analysis Software

Medcalc – Medical calculation

Computer and Researcher

- Performing calculation at the speed of light
- Data analysis concerning large research projects
- Huge storage of data
- Faster retrieval
- Added quality to research activity
- Trial and error process frequently employed

Computer and Researcher

- Can easily calculate computation of means, SD, correlation coefficients, 't' tests, analysis of variance, multiple regression and various nonparametric analyses
- Multivariate analysis can be computed
- Only work of researcher is to feed in the data

Computer and Researcher

- Facilitate research work
- Innumerable data can be processed and analyzed with greater ease and speed
- Results obtained are accurate and reliable
- Pictorial graphing and reporting

Limitation of computer-based analysis

- Requires expensive computerised analysis setting, collection and feeding of data (GAITRE)
- Not economical in case of small projects
- Various small details which are not fed into computers may get lost
- Does not think only execute
- "Garbage in, garbage out"

Importance of

- The **bibliography** appears at the end. The main purpose of a **bibliography** entry is to give credit to authors whose work you've consulted in your research. It also makes it easy for a reader to find out more about your topic by delving into the research that you used to write your paper.
- Referencing allows you to acknowledge the contribution of other writers and researchers in your work. Any university assignments that draw on the ideas, words or research of other writers must contain citations. Referencing is a way to provide evidence to support the assertions and claims in your own assignments.
- **Footnotes** are the **important** points present in the form of notes at the end of a page or bottom of a page. It elaborates as well as simplifies a particular point mentioned in the body of text and also shows its accountability by acknowledging the source of information taken out.

Importance of

- An **endnote** is source citation that refers the readers to a specific place at the end of the paper where they can find out the source of the information or words quoted or mentioned in the paper. When using **endnotes**, your quoted or paraphrased sentence or summarized material is followed by a superscript number.
- **Endnotes** and **footnotes** are the **important** elements that are required to write very carefully in order to compose an effective assignment. These notes will help your readers know the various resources that you've used so that they can verify your research work.

Format for Seminar Report

Seminar Report

Each student is required to write a comprehensive report about the seminar. The report should consist of 15 to 20 pages describing the topic selected. The report should be in the format as described below. It is important that you adhere to these guidelines.

1. ARRANGEMENT OF CONTENTS:

The sequence in which the project report material should be arranged and bound should be as follows:

1. Cover Page & Title Page
2. Abstract
4. Table of Contents
5. Acknowledgment
6. Chapters
7. Appendices
8. References

2. PAGE DIMENSION AND BINDING SPECIFICATIONS:

The dimension of the seminar report should be in A4 size. The project report should be **hard bound using flexible cover of the thick art paper**. Two copies of the report (hard copy only) should be submitted.

Format for Seminar Report

3 TYPING INSTRUCTIONS:

One and a half spacing should be used for typing the general text. The general text shall be **justified and typed in the Font style 'Times New Roman' and Font size 12.**

Subheading shall be typed in the Font style 'Times New Roman' and Font size 12 and bold.

Heading shall be typed in the Font style 'Times New Roman' and Font size 14 and bold.

Format for Seminar Report

3. PREPARATION FORMAT:

3.1 **Cover Page & Title Page** – A specimen copy of the Cover page & Title page of the project report are given in **Appendix 1.**

3.2 **Abstract** – Abstract should be one page synopsis, it should summarize the aims, conclusions and implications of the topic of your seminar, typed with double line spacing, Font Style Times New Roman and Font Size 14.

3.3 **Table of Contents** – The table of contents should list all material following it as well as any material, which precedes it. The page numbers of which are in lower case Roman letters. One and a half spacing should be adopted for typing the matter under this head. A specimen copy of the Table of Contents of the project report is given in **Appendix 3.**

3.4 **Chapters** – The chapters may be broadly divided into 3 parts (i) Introductory chapter i.e. developing the main theme of the topic (ii) Literature review of the topic selected (iii) and Summary or Conclusions

The main text will be divided into several chapters and each chapter may be further divided into several divisions and sub-divisions.

- Each chapter should be given an appropriate title.
- Tables and figures in a chapter should be placed in the immediate vicinity of the reference where they are cited.

Format for Seminar Report

Appendix 3.

Chapters – The chapters may be broadly divided into 3 parts (i) Introductory chapter i.e. developing the main theme of the topic (ii) Literature review of the topic selected (iii) and Summary or Conclusions

The main text will be divided into several chapters and each chapter may be further divided into several divisions and sub-divisions.

- Each chapter should be given an appropriate title.
- Tables and figures in a chapter should be placed in the immediate vicinity of the reference where they are cited.

List of References – The listing of references should be typed 2 spaces below the heading "REFERENCES" in alphabetical order in single spacing left – justified. Should be numbered consecutively (in square [] brackets, throughout the text and should be collected together in the reference list at the end of your report. The references should be numbered in the order they are used in the text. The name of the author/authors should be immediately followed by the year and other details.

A typical illustrative list given below relates to the citation example quoted above.

APPENDIX 1
(A typical Specimen of Cover Page & Title Page)

TITLE OF SEMINAR

<Font Size 22><BOLD><Centralized>

A Seminar Report
<Font Size 14><BOLD><Centralized>

Submitted by
<Font Size 14><Italic><BOLD><Centralized>

NAME OF THE CANDIDATE
<Font Size 16><BOLD><Centralized>

in partial fulfillment for the award of the degree

of
<Font Size 14><1.5 line spacing><Italic><BOLD><Centralized>

NAME OF THE DEGREE
<Font Size 16><BOLD><Centralized>

IN
BRANCH OF STUDY
<Font Size 14><BOLD><Centralized>
At

DEPARTMENT OF ECONOMICS

THIRUVALLUVAR UNIVERSITY (A STATE UNIVERSITY.)

SERKKADU, VELLORE DISTRICT, TAMIL NADU- 632 115.
<Font Size 16><BOLD><Centralized>
MONTH & YEAR
<Font Size 14><BOLD><Centralized>

SPECIMEN

STATUS, SURVIVAL AND CURRENT DILLEMA OF PANIYAN TRIBES IN NILGIRIS DISTRICT OF TAMIL NADU

A SEMINAR REPORT

Submitted by

Mr. XYZ

in partial fulfillment for the award of the degree

of

MASTER OF ARTS

IN

ECONOMICS

At

DEPARTMENT OF ECONOMICS

THIRUVALLUVAR UNIVERSITY (A STATE UNIVERSITY.)

SERKKADU, VELLORE DISTRICT, TAMIL NADU- 632 115.

SEPTEMBER - 2020

PARTS OF A REPORT

- Title Page
- Table of Contents
- Executive Summary
- **Introduction**
- **Findings**
- **Conclusions**
- **Recommendations**
- Bibliography/References
- Appendices

Body of report


Title Page
Table of Contents
Executive Summary
Introduction
Findings
Conclusions
Recommendations
References
Appendices

PARTS OF A REPORT

- Title Page
- Table of Contents
- Executive Summary
- **Introduction**
- **Findings**
- **Conclusions**
- **Recommendations**
- Bibliography
- Appendices



PARTS OF A REPORT

Part	Description
Title Page	Title, Author's name, Date of submission
Table of Contents	List of sections, sub-sections, appendices, etc.
Executive Summary	Summary of entire report
Introduction	Purpose, Background, Methods of Investigation, Scope
Findings	Analysis of facts obtained
Conclusions	Deductions based on findings
Recommendations	Suggestions and advice based on conclusions
Bibliography/References	List of references
Appendices	Supplementary materials e.g. other supporting data

BODY OF REPORT

- Introduction
- Findings
- Conclusions
- Recommendations

INTRODUCTION

- Four sections: **Purpose, Background, Method of Investigation, Scope**
- Take note of language conventions (i.e. typical words and phrases that are used)
- Uses present and past tense
- Format – numbered headings

INTRODUCTION

- **1.1 Purpose**
  - States the purpose of the report
  - Includes what the report will recommend

Example 1

The purpose of this report is to investigate the reasons behind the rise in computer gaming addiction among teenagers. The report will also recommend preventive measures for computer gaming addiction.

INTRODUCTION

Example 2

This report looks into the factors influencing students' behavior in lecture theatres and suggests ways to promote positive learning behavior among students.

Example 3

This report studies the causes of unhealthy eating among teenagers and recommends ways to promote more healthy eating habits.

INTRODUCTION

- **1.2 Background**
  - Provides information on the problem or situation that gave rise to the investigation
  - May refer to **secondary data**, e.g. newspaper report (i.e. data that was discovered by someone else)

INTRODUCTION

Example

A recent study on game technology (Tan, 2010) shows that nearly 75% of teenagers in Singapore play computer games regularly and the majority of them show signs of addiction. The rise in computer gaming addiction among teenagers is a growing concern and the Ministry of Media Development (MMD) would like to find out the reasons behind this trend.

The report was commissioned by the Director of Corporate Communications, Ms Julie Sim, on 31 May 2010.

INTRODUCTION

- **1.3 Method of Investigation**
  - States how investigation was carried out, e.g. questionnaires issued to [how many people] for [how long]
  - This is called **primary data** (i.e. data that is discovered by you, the writer of the report)
  - May include references to secondary data

INTRODUCTION

Example

Questionnaires were issued to 100 teenagers from 25 educational institutions on 25 June 2010. Three teenagers who were former addicts were also interviewed to get more in-depth views.

References were also made to books and newspaper articles.

QUIZ TIME

There are four sources of data.

Which is primary data? Which is secondary data?

Questionnaires were issued to 100 teenagers from 25 educational institutions on 25 June 2010. Three teenagers who were former addicts were also interviewed to get more in-depth views.

References were also made to books and newspaper articles.

ANSWER

PRIMARY DATA

Questionnaires were issued to 100 teenagers from 25 educational institutions on 25 June 2010. Three teenagers who were former addicts were also interviewed to get more in-depth views. References were also made to books and newspaper articles.

SECONDARY DATA

INTRODUCTION

- **1.4 Scope**
  - States the areas of investigation, e.g. reasons, consequences, etc., i.e. what information you need in order to meet your purpose
  - First scope item is **respondents' profile**
  - There should be at least 3 other scope items

INTRODUCTION

Example

Besides respondents' profile, this report looks into four possible reasons for the rise in gaming addiction among teenagers: boredom, stress, societal influence and addictive game features.

A Report on the Rise in Computer Gaming Addiction Among Teenagers

1. Introduction

1.1 Purpose

The purpose of this report is to investigate the reasons behind the rise in computer gaming addiction among teenagers. The report will also recommend preventive measures for computer gaming addiction.

1.2 Background

A recent study on game technology (Tan, 2010) shows that nearly 75% of teenagers in Singapore play computer games regularly and the majority of them show signs of addiction. The rise in computer gaming addiction among teenagers is a growing concern and the Ministry of Media Development (MDD) would like to find out the reasons behind this trend.

The report was commissioned by the Director of Corporate Communications, Ms Julie Sim, on 31 May 2010.

1.3 Method of Investigation

Questionnaires were issued to 100 teenagers from 25 educational institutions on 25 June 2010. Three teenagers who were former addicts were also interviewed to get more in-depth views. References were also made to books and newspaper articles.

1.4 Scope

Besides respondents' profile, this report looks into four possible reasons for the rise in gaming addiction among teenagers: boredom, stress, societal influence and addictive game features.

A Report on the Rise in Computer Gaming Addiction Among Teenagers



1. Introduction

Numbered heading – whole number

1.1 Purpose

Numbered sub-heading – numerical-decimal

The purpose of this report is to investigate the reasons behind the rise in computer gaming addiction among teenagers. The report will also recommend preventive measures for computer gaming addiction.

1.2 Background

A recent study on game technology (Tan, 2010) shows that nearly 75% of teenagers in Singapore play computer games regularly and the majority of them show signs of addiction. The rise in computer gaming addiction among teenagers is a growing concern and the Ministry of Media Development (MDD) would like to find out the reasons behind this trend.

The report was commissioned by the Director of Corporate Communications, Ms Julie Sim, on 31 May 2010.

1.3 Method of Investigation

Questionnaires were issued to 100 teenagers from 25 educational institutions on 25 June 2010. Three teenagers who were former addicts were also interviewed to get more in-depth views. References were also made to books and newspaper articles.

1.4 Scope

Besides respondents' profile, this report looks into four possible reasons for the rise in gaming addiction among teenagers: boredom, stress, societal influence and addictive game features.

Informal Report



- The informal report functions to inform, analyze, and recommend.
- It usually takes the form of a memo, letter or a very short international document like a monthly financial report, monthly activities report, research and development report, etc.
- This report differs from the formal report in length and formality.
- It is written according to organization style and rules, but usually does not include the preliminary (front) and supplemental (back) material.
- The informal report is usually more controversial in tone and typically deals with everyday problems and issues addressed to a narrow readership inside the organization.

TYPES OF INFORMAL REPORT

There are many embodiments of the informal report:

i. Progress report

ii. Sales activity report

iii. Personnel evaluation

iv. Financial report

v. Feasibility report

vi. Literature review

vii. Credit report

Formal Report

- The formal report is the collection and interpretation of data and information.
- The formal report is complex and used at an official level.
- It is often a written account of a major project.
- Examples of subject matter include new technologies, the advisability of launching a new project line, results of a study or experiment, an annual report, or a year old review of developments in the field.

TYPES OF FORMAL REPORT

They can be categorized as:

- Informational reports
- Analytical reports
- Recommendation reports



INFORMATIONAL REPORTS

- Informational reports present results so readers can understand a particular problem or situation.

- Example: Manager of a city's website might prepare an informational report for the city council; the report would provide statistics on the number of people who pay their city water and sewage bills online etc.

- Informational reports might:

A. Present information on the status of current research or of a project.
B. Present an update of the operation in your division.
C. Explain how your organization or division does something.
D. Present the results of a questionnaire or research.

ANALYTICAL REPORTS

- This type goes a step beyond presenting results. Analytical reports present results, analyze those results, and draw conclusions based on those results.
- These reports attempt to describe why or how something happened and then to explain what it means.
- Like informational reports, analytical reports can be formal or informal.
- Explain what cause a problem or situation – Present the results of a traffic study showing accidents at an intersection – the report explains what it means.
- Explain the potential results of a particular course of action.
- Suggest which option, action, or procedure is best.

RECOMMENDATION REPORTS

- This type advocate a particular course of action. This usually present the results and conclusions that support the recommendations.
- This type is identical to analytical report.
- For example, your analytical report suggests using treatment X is more efficient than treatments Y and Z. However, that does not mean that you will use treatment X as cost and other considerations might recommend treatment Y.
- What should we do about a problem?
- Should we or can we do something?
- Should we change the method or technology we use to do something?

THE PLAN FOR PREPARING A FORMAL REPORT

- Identify the readers
- Determine your purpose
- Formulate specific questions
- Conduct research to answer the questions
- Draw valid conclusions (for analytical or recommendation reports)
- Decide on recommendations (for recommendations)
- Write the report

FORMAT OF A REPORT

- Although, there is no set report writing format, however, there are general sections that should be included.

- Unlike essays, reports are written in sections with headings and sub-headings, which are usually numbered.

- There are numerous possible formats available for writing a report, and it mostly depends on the context of topic

Below given are the components of a report in which they would occur:

1. **Title page:** It should include the title, your name and the name of the tutor to whom it is being submitted, date of submission, your course/department. The logo of the organisation should also be printed.

2. **Acknowledgements:** A list of people and organisations who have helped you in the compilation of report and other related work.

3. **Contents page:** A clear, well-formatted list of all the sections and sub-sections of the report. Page numbers should be marked correctly.

4. **Abstract:** A summary of the major points, conclusions, and recommendations should be written to give a general overview of report.

5. **Introduction:** The first page of the report needs to have an introduction. You will explain the problem and show the reader why the report is being made.

6. **Body:** This is the main section of the report. There needs to be several sections, with each having a subtitle. The various sections include Review of Literature, Materials and Methods and Results. A discussion section can also be included at the end of the body to go over by findings and their significance.

7. **Conclusion**: A conclusion should draw out the implications of your findings, with deductions based on the facts described in your main body. The significance and relevance of study is discussed in this section.

8. **References**: This is a list giving the full details of all the sources to which you have made reference within your text.

1. Report as a means of internal communication

- A report acts as an effective means of communication within the organization.
- It provides feedback to employees.
- It is prepared for the information and guidance of others connected with the matter / problem.

2. Report facilitates decision making and planning



- Report provide reliable data which can be used in the planning and decision making process.
- It acts as a treasure house of reliable information for long term planning and decision making.

3. Report discloses unknown information

- Reports provide information, which may not be known previously.
- The committee members collect data, draw conclusions and provide information which will be new to all concerned parties.
- Even new business opportunities are visible through unknown information available in the reports.

4. Report gives Information to Employees

- Reports are available to managers and departments for internal use.
- They are widely used by the departments for guidance.
- Report provide a feedback to employees and are useful for their self-improvement.

WHAT IS A REPORT WRITING

- A means to convey some information to others.
- An organized, factual, objective presentation of information.
- 'Organized' since it follows a systematic pattern.
- 'Objective' means it should not be influenced by our personal feelings.
- Report is a comprehensive document and covers all aspects of the subject matter of study.

REPORTS AND ESSAYS.!!!

- Similarities:

  - formal style
  - careful proof-reading and neat presentation
  - introduction, body and conclusion
  - analytical thinking

KEY DIFFERENCES:

A report:

- presents information, not an argument
- is meant to be scanned quickly by the reader
- uses numbered headings and sub-headings
- uses short, concise paragraphs and dot-points where applicable
- uses graphics wherever possible (tables, graphs, illustrations)
- may need an abstract (sometimes called an executive summary)
- does not always need references and bibliography
- is often followed by recommendations and/or appendices

TYPES OF REPORTS :



1. Information only Reports

- Provides basic information
- Very very short
- Report writing on monthly budgets,staff absentees and so on..

FORMAT OF THE REPORT:



SECTIONS OF THE REPORT



Parts of the report	Contents	Functions	Types of writing	Language characteristics
Title page	Name of the report, date	To inform	Discursive	Factual
Table of contents	What you find in your report	To inform	Discursive	Factual
Summary	A summary of the report	To inform	Discursive	Factual, impersonal use of 3rd person pronouns, passive verbs
Introduction	Background, problem, approach, definition of special words used	To inform	Discursive	Same as the above
Methods	Methods or procedures which lead to the findings	To inform	Discursive	Same as the above
Findings	Results of investigation, research, and calculation	To inform	Discursive	Sa as the above
Conclusion	Conclusions drawn from the findings	To convince	Argument	It seems that … / The result indicates tthat … / It is probable that…
Recommen-dation	Things that should be done as a result	To persuade reason	Argument, persuasive	Personal use of "I" and "We"
Bibliography	Books, articles, magazines, reports , any documents that has used for reference	To inform		

EXAMPLE OF REPORT WRITING



Computer system- SCANNERS

Abstract

This report investigates the current state of scanner tech. and examines the predicted future advancements of scanners.

The performance of these scanners is examined in relation to four main criteria; resolution, bit-depth, dynamic range and software.



WHAT IS A THESIS?

noun, plural the·ses

1. a proposition stated or put forward for consideration, especially one to be discussed and proved or to be maintained against objections: He vigorously defended his thesis on the causes of war.
2. a subject for a composition or essay.
3. a dissertation on a particular subject in which one has done original research, as one presented by a candidate for a diploma or degree.



INITIAL PAGES

- Title Page
- Approval Sheet
- Abstract
- Acknowledgment
- Dedication
- Table of Contents
- List of Tables
- List of Figures



TITLE PAGE

The following information needs to be on the title page:

- The title (and possibly the subtitle) of your thesis
- First name and surname of the author(s)
- Whether it is a 'Bachelor's thesis' or a 'Master's thesis'
- Faculty and department
- Place and date of completion



APPROVAL SHEET

- This is to prove that the authors have passed the requirements needed for the thesis.
- This is signed by the thesis/FS adviser, panel and the Dean.
- This also states the grade obtained by the author/s.



ABSTRACT

An abstract presents a brief summary of your thesis.

The aim of the abstract is to briefly provide the reader with the most important information from the entire text.

An abstract never contains new information.

This summary is no longer than 2 pages of A4.



ACKNOWLEDGMENT

- This is a page focused on expressing gratitude to organizations, agencies or individuals who, in one way or another, have aided the researchers in finishing the thesis.



DEDICATION

- This is the page for dedicating the thesis to certain people or groups who have inspired the researchers while doing the thesis.



TITLE OF CHAPTERS

1. Problem and Its Background
2. Review of Related Literature and Studies
3. Methodology of the Study
4. Presentation, Analysis and Interpretation of Data
5. Summary, Conclusions and Recommendations



INTRODUCTION

- The first chapter of your thesis is your introduction.
- This is where you provide an introduction to the topic of your thesis: you give the context in terms of content of the research project.

CONCEPTUAL FRAMEWORK



- A conceptual framework elaborates the research problem in relation to relevant literature. This section may summarize the major (dependent and independent) variables in your research. The framework may be summarized in a schematic diagram that presents the major variables and their hypothesized relationships. It should also cover the following:
- •Existing research and its relevance for your topic
- •Key ideas or constructs in your approach
- •Identify and discuss the variables related to the problem.
- •Conceptualized relationships between variables
- Independent variables (presumed cause)
- Dependent variables (presumed effect)
- Intervening variables (other variables that influence the effect of the independent variable)

Instructional Materials and Equipment commonly used in teaching Science

1Traditional Materials

2Technological Equipment



Effects of the use of instructional Materials and equipment

Profile of the respondents

1Age

2Gender

3Civil Status

4Educational Attainment

5Years of Service

RELATED LITERATURE



- In this part you must get your data and information from any books, magazines, and news papers. You must label your published material with local or foreign.

- 1. Must be also organized to cover specific problems.
- 2. Must take all the evidences about the problem with the author's experiences.
- 3. As much as possible, get the latest published materials. Avoid old published materials.
- 4. It must be related to your topic. If not, do not get it.
- 5. On the last part of this part you must have a statement how this old published material helps the researcher in their current study and relate it to your study.



RELATED STUDIES

- In this part you must get your data and information from unpublished material such as previous or old study, research or thesis. In some format, you must label your unpublished material with local or foreign.

- 1. This should be organized to cover the specific problems.
- 2. You must take note all of the evidences that the previous researcher came up.
- 3. The unpublished material should not be older than 5 years if possible.
- 4. It must be related to your topic. If not, do not get it.
- 5. On the last part of this part you must have a statement how this old unpublished material helps the researcher in their current study and relate it to your study.



ANALYSIS OF DATA

- Data may be analyzed quantitatively or qualitatively depending on the level of measurement and the number of dimensions and variables of the study.

- Analyze in depth to give meaning to the data presented in the data presented in the table. Avoid table reading.

- State statistical descriptions in declarative sentences, e.g. in the studies involving:



INTERPRETATION OF DATA

- Establish interconnection between and among data
- Check for indicators whether hypothesis/es is/are supported or not by findings.
- Link the present findings with the previous literature.
- Use parallel observations with contemporary events to give credence presented in the introduction.



SUMMARY OF FINDINGS

- This describes the problem, research design, and the findings (answer to the questions raised). The recommended format is the paragraph form instead of the enumeration form.
- For each of the problems, present:
  - The salient findings,
  - The results of the hypothesis tested

CONCLUSIONS



- These are brief, generalized statements in answer to the general and each of the specific sub-problems.
- These contain generalized in relation to the population. These are general inferences applicable to a wider and similar population.
- Flexibility is considered in making of conclusions. It is not a must to state conclusions on a one-to-one correspondence with the problems and the findings as all variables can be subsume in one paragraph.
- Conclusions may be used as generalizations from a micro to a macro-level or vice versa (ZOOM LENS approach).

Steps in Data Analysis

- Before Data Collection, the researcher should accomplish the following:
  - Determine the method of data analysis
  - Determine how to process the data
  - Consult a statistician
  - Prepare dummy tables
- After Data Collection:
  - Process the data
  - Prepare tables and graphs
  - Analyze and interpret findings
  - Consult again the statistician
  - Prepare for editing
  - Prepare for presentation

1. Descriptive Analysis

- refers to the description of the data from a particular sample;
- hence the conclusion must refer only to the sample.
- In other words, these summarize the data and describe sample characteristics.
- **Descriptive Statistics**
  - are numerical values obtained from the sample that gives meaning to the data collected.

Classification of Descriptive Analysis



a. Frequency Distribution

- A systematic arrangement of numeric values from the lowest to the highest or highest to lowest.

- Formula: Ef = N

  - Where:

    E = sum of

    f = frequency

    N= sample size

Classification of Descriptive Analysis



b. Measure of Central Tendency

- A statistical index that describes the average of the set values.

- **Kinds of Averages**

1. **Mode** - a numeric value in a distribution that occurs most frequently.

- 2. **Median** - an index of average position in a distribution of numbers.

- 3. **Mean** - the point on the score scale that is equal to the sum of the scores divided by the total number of scores.

- Formula:

$$X = \frac{\Sigma}{n}$$

Where:
X = the mean
Σ = the sum of
X = each individual raw score
n = the number of cases

Classification of Descriptive Analysis



c. Measure of Variability

- Statistics that concern the degree to which the scores in a distribution are different from or similar to each other.

Two Commonly Used Measures of Variability

1. Range

- the distance between the highest score and the lowest score in a distribution.

Example:

The range for learning center A 500 (750 - 250) and the range for learning center is about 300 (650 - 350)

2. Standard Deviation

- the most commonly used measure of variability that indicates the average to which the scores deviate from the mean.

Classification of Descriptive Analysis

d. Bivariate Descriptive Statistics

- Derived from the simultaneous analysis of two variables to examine the relationships between the variables.

- **Two Commonly Used Bivariate Descriptive Analysis**

5. **Contingency tables**

   - is essentially a two-dimensional frequency distribution in which the frequencies of two variables are cross-tabulated.

- **Correlation**

  - the most common method of describing the relationship between two measures

2. Inferential Analysis

- The use of statistical tests, either to test for significant relationships among variables or to find statistical support for the hypotheses.

- **Inferential Statistics**
  - are numerical values that enable the researcher to draw conclusion about a population based on the characteristics of a population sample.
  - This is based on the laws of probability.

Level of Significance

- An important factor in determining the representativeness of the sample population and the degree to which the chance affects the findings.
- The level of significance is a numerical value selected by the researcher before data collection to indicate the probability of erroneous findings being accepted as true.
- This value is represented typically as 0.01 or 0.05. (Massey,1991)

- **Uses of Inferential Analysis**
  - Cited some statistical test for inferential analysis.

1. t-test

is used to examine the difference between the means of two independent groups.

2. Analysis of Variance (ANOVA) - is used to test the significance of differences between means of two or more groups.

3. Chi-square - this is used to test hypotheses about the proportion of elements that fall into various cells of a contingency table.

- Interpretation of Data
  - After analysis of data and the appropriate statistical procedure, the next chapter of the research paper is to present the interpretation of the data, which is the final step of research process.

- The three areas:
  - Summary of Findings
  - Conclusions
  - Recommendations



3. Table of Contents

- from the word itself, it contains all the parts of the research paper including the pages.

4. List of Tables

- this follows the table of content and indicates the title of the tables in the research paper.

5. List of Figures

- is composed of paradigms, diagrams, graphs and charts or flowcharts.

Main Body



- Chapter I. Introduction
- This section refers to:
  - "What this study is all about" or "What makes the researcher interested in doing the study".
  - Purpose: to introduce the reader to the subject matter.
  - The introduction serves as a springboard for the statement of the problem as stated by Dr. Barrientos-Tan.

- Chapter II. Review of Related Literature
- Chapter III. Research Design and Methodology
- Chapter IV. Analysis and Interpretation of Data
- Chapter V. Summary, Conclusions, and Recommendations Supplementary Pages
- Bibliography
- Appendix



Sample of Table of Contents

- Table of Contents
  - Acknowledgment
  - List of Tables
  - List of Figures
  - Chapter I The Problem and its Background
    - Introduction
    - Background of the Study
    - Conceptual Framework/Theoretical
    - Statement of the Problem
    - Hypotheses
    - Significance of the Study
    - Scope and Delimitation of the Study
    - Definition of Terms

- Chapter II Review of Related Literature and Studies
  - Literature (Foreign/Local)
  - Studies (Foreign/Local)
  - Justification of the Present Study
- Chapter III Research Design and Methodology
  - Research Design
  - Research Subject
  - Instrumentation
  - Data Gathering Procedure
  - Statistical Treatment of Data
- Chapter IV Analysis and Interpretation of Data
- Chapter V Summary, Conclusion and Recommendations
- Bibliography
- Appendix
- Curriculum Vitae

Application of Innovative Research in Social Science

- The Social Science subject deals with diverse concerns of society and include a wide a range of content, drawn from the disciplines of history, geography, political science, economics social work and sociology.
- In order to make the process of learning participative there is a need to shifts from mere imparting of information in research to debate and discussion.
- Application of new research trends in this regards is very much helpful.
- The research student become active learners rather than being passive recipient of information.

- The student get the opportunity to explore by their own so that the learning becomes joyful and long lasting.
- It is generally assumed that an innovation should not just be gimmickry, attracting the favorable opinion of the book or grant reviewers in the same way as novelties are marketed in the retail industry, nor be in response to latest wave of enthusiasm.
- It is the major resource that the library provides.
- The library is therefore; set aside to meet the appetite for learning and to feed the hunger for understanding.
- The student assigned to read some items in the library to broaden their understanding of research, what the teacher has taught in the classroom.



Broad classification of Innovative Methods in Research

The Innovative Methods are classified into three broad categories:

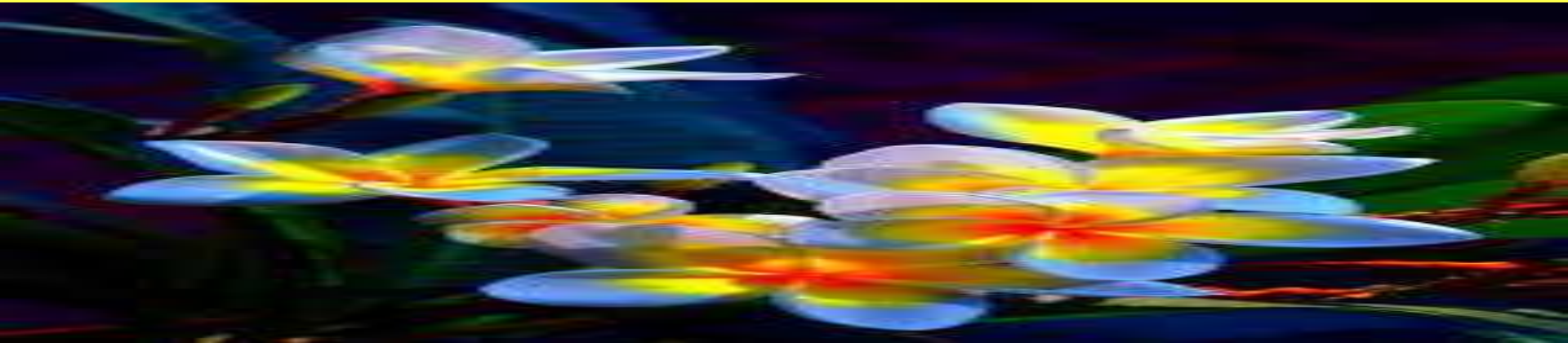
- Technological Innovative
- Crossing Boundaries
- Intra- disciplinary/ Intra- methodological



❑Technological Innovation



Technological invocation refers to advances in technology (usually in the form of software and/ or internet- related) that either constitutes a method in itself or pushes the boundaries of a methodology in contributing to or amending previous research practice.



❑Crossing Boundaries

Crossing boundaries refers to methodological innovation that develop because of addressing research questions beyond or across disciplinary boundaries. This usually involve a combination of theoretical and methodological approaches.



❑Intra- disciplinary/ Intra-methodological

Intra disciplinary/ Intra – methodological refers to developments in methodology that observed within discipline or that extend a method or applied it in different it in different ways.

- Innovator's Details: Name, Affiliation, Country
- Title of Innovative Research Practice
- Introduction/ Background
- Description and Explanation (Published if available Development



- Example of Application (Published if available or short descriptions)
- Inspiration and Rationale
- Application and Applicability
- Colleagues and /or Relevant/ Related Research
- Material (Usually in the form of ‘References)

Contribution of Research to Theory

The relationships between theory and research is contributory. Research contributes to the development of theory.

- Research initiates theory
- Research tests an existing theory
- Reformulation of an existing theory
- Research refocuses theory
- Research clarifies theory

Qualitative and Quantitative Research

- Qualitative research is a broad term that encompasses a variety of approaches to inter preparative research.
- It can be historical, sociological, education and much more.
- Qualitative research style in social research in not much related to the scientific logic of research , but more close to the work around.
- Quantitative research focus upon human behavior for better understandings about the world around them.

- Therefore, the emphasis given by quantitative researches in their studies involve an examination of the perspectives of the people or group; e.g. their ideas, attitudes, motive, and intensions.
- In qualitative research, the researcher's primary goals are an understanding of social processes rather than obtaining a representative sample.
- The study of one or small number of cases, often over a lengthy period.
- Qualitative research methods such as in- depth interviews and participant observation have some of the following characteristics.

- Research is carried out in reliable settings.
- In qualitative research, the objectives are to take detailed descriptions of people behavior and thought. The focus of the research may change during the course of research.
- The qualitative approach involves theory constriction rather than theory testing.
- In qualitative research, when compared to quantitative based research, research designs are often relatively small, loosely structured.

Difficulties in the Use of Scientific Methods in Social Science Research

Some theorists argues that scientific methods is more applicable to physical or natural sciences and it cannot applicable to social sciences.

- Human behavior is different. It is very difficult to categories.
- When human behavior is studied and analyzed by another human, there may be personal problems.
- Psychological nature of human behavior and predictable.





- Human behavior is not uniform and predictable. Uncertainty is present in human behavior.
- Differences in choice and decision.
- Personal prejudices and biases.

A researcher's attitudes towards socio- economic issues are influenced by his values. All these problems cause biases and invalidate the research finding and conclusions.



Ethical Issues in Social Sciences

- An ethic is more than presence of a basic value or values.
- It is base of action in any sciences.
- The conflict between the ethics of science and personal respects of researcher is the major problems in social science research.
- When we talk about ` ethics' in social research, we are addressing those issues that concern the behavior of social researcher and the consequences that their research bring to the people they study.

Three Main Challenges for Social Scientists in the Era of Big Data

How we can develop the very best tools to help us interrogate, analyze and understand the vast quantities and varieties of data that exist.

The challenge of developing insightful research questions and the need to focus energy on the substantive evidence that can be gleaned from the empirical material.

The ethical questions raised by new forms of data and new approaches to analysis.

New Trends in Research Method Applications

The innovations focused on a range of different type of research methods or approaches, which describe the focus of innovation claimed with an `other' category for those papers, which were not easily classifiable.



Creative Methods

- Different paper that employed art, drama, dance, poetry, photography or a combination of these with the aim of engaging participation or audience in a more holistic way, giving scope for emotional and moral a well as intellectual responses.



Narrative Methods

Paper describing techniques for collecting, analyzing or presenting narratives, including auto-ethnography. Several of these paper explored participatory approaches to research afforded by narrative methods.



Mixed Methods

Paper, describing techniques for combining and analyzing different types of qualitative data or qualitative and quantitative data.

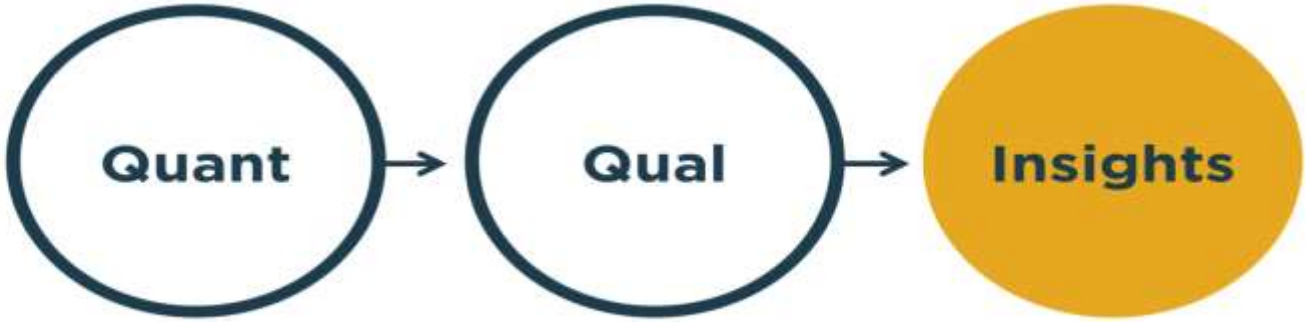
Mixed Methods Research is defined as a type of user research that combines qualitative and quantitative methods into a single study. Companies like Spotify, Airbnb and Lyft are using Mixed Methods Research to combine rich user insights with actionable statistics for deeper user insights.

Mixed Methods

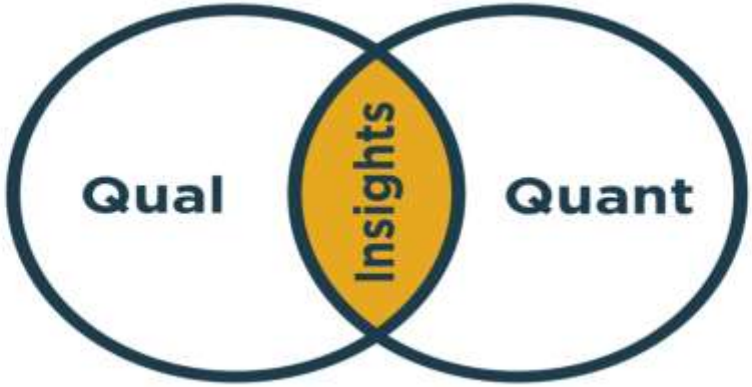
Exploratory Research



Explanatory Research



Dynamic Research





Online and e- Research Methods - I

Paper widening the scope of online research, using synchronous and asynchronous text facilities, blogs and a graphical online environment.

E-research encompasses the use of information and communication technology to facilitate research. Such type of research makes use of internet-based tools and techniques. Internet-based research has made it possible for researchers to study topics that could not be researched using traditional methods.

Online and e- Research Methods - II

Online research methods (**ORMs**) are ways in which researchers can collect data via the internet. They are also referred to as Internet research , Internet Science or Science, or Web-based methods. Many of these online research methods are related to existing research methodologies but re-invent and re-imagine them in the light of new technologies and conditions associated with the internet. The field is relatively new and evolving. With the growth of social media, a new level of complexity and opportunity has been created. The inclusion of social media research can provide unique insights into consumer and societal segments and gaining an "emotional" measure of a population on issues of interest.

Some specific types of method include

- Cyber-ethnography.
- Online content analysis.
- Online focus groups.
- Online interviews.
- Online qualitative research.
- Online questionnaires.
- Social network analysis.
- Web-based experiments.



Cyber-ethnography.

cyber-ethnography attempts **to adapt participant-observation procedures** such as making cultural entrée, collecting and analyzing data, and conducting ethical research to these computer-mediated contingencies. Cyber-ethnography has significantly developed with the emergence of new technologies.



Online content analysis

Online content analysis or online textual analysis refers to a collection of research techniques used to describe and make inferences about online material through systematic coding and interpretation. Online content analysis is a form of content analysis for analysis of Internet-based communication.



Online focus groups

An online focus group is **a method of data collection that enables the researcher to host a discussion between a number of respondents through an online platform.** B2B International has the ability to host and moderate focus groups through our online qualitative package.



Online interviews

Online interviews are used for primary Internet-mediated research (IMR), that is, they are used to gather original data via the Internet with the intention of subjecting them to analysis to pro- vide new evidence in relation to a specific research question (Hewson, 2010).



Online qualitative research

Heck out these online approaches to traditional qualitative research. Online qualitative research methods include **online group chats**, online and phone in-depth interviews, online focus groups, online bulletin boards, mobile diaries, and social media tracking/listening.



Online questionnaires.

An online survey is a structured questionnaire that your target audience completes over the internet generally through a filling out a form. The data is stored in a database and the survey tool generally provides some level of analysis of the data in addition to review by a trained expert.



Social network analysis

Social network analysis (SNA) is **the process of investigating social structures through the use of networks and graph theory.** It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them.



Web-based experiments

A web-based experiment or Internet-based experiment is **an experiment that is conducted over the Internet**. This form of experimental setup has become increasingly popular because researchers can cheaply collect large amounts of data from a wider range of locations and people.



Software Tools

Paper, describing software to assist in or enhance, the analysis or sharing of a range of qualitative data.

Research tools are specific mechanisms or strategies that the researcher uses to collect, manipulate, or interpret data. Six general tools of research: 1) the library and its resources, 2) the computer and its software, 3) techniques of measurement, 4) statistics, 5) the human mind, and 6) language.

Software Tools for social science research

There is a whole range of software packages and tools for data analyses and visualisation - from Access or Excel to dedicated packages, such as SPSS, Stata and R for statistical analysis of quantitative data, Nvivo for qualitative (textual and audio-visual) data analysis (QDA), or ArcGIS for analysing geospatial data.

In planning studies' designs, sociologists generally choose from four widely used methods of social investigation: **survey, field research, experiment, and secondary data analysis**, or use of existing sources.

software is used for data analysis for social science

SPSS is a widely used program for statistical analysis in social science. It is also used by market researchers, health researchers, survey companies, government, education researchers, marketing organizations, data miners, and others.

Graphpad is an amazing statistical software which can guides your for statiscal tests and graphics analysis. SPSS. IBM SPSS software. XLSTAT. XLSTAT is the leading data analysis and statistical solution for Microsoft Excel.

Methodology in Social Science Research

Methodology in the social sciences is usually characterized by the following: (1) **it defines the information to be analyzed**; (2) it provides the conceptual tools and procedures necessary to perform an analysis; and (3) it sets forth the limits of the analysis.

The term methodology may be defined in at least three ways: (1) a body of rules and postulates that are employed by researchers in a discipline of study; (2) a particular procedure or set of procedures; and (3) the analysis of the principles of procedures of inquiry that are followed by researchers in a discipline of study. This entry will first discuss each of these definitions. It will then cover the debate among philosophers of science about general methodological assumptions. Finally, the entry will review some of the issues pertaining to the quantitative versus qualitative debate about methods.

Most useful software to learn for research in social science

- Let's assume we're talking about quantitative research and you're using a text editor to convey your findings (Ms Office, LYX, Libre Office, Scientific Work Place, etc.). Let's start with the quantitative analysis.

- SPSS (Statistical Processing for Social Sciences) is a popular programme for doing conventional studies on huge datasets. However, it is not inexpensive and may be overkill for a single user. Stata, Statistica, and JMP all have similar features but are different in cost and quality.

- R is open-source, extremely flexible, and rapidly spreading, but it has a steep learning curve. Simpler tools may be available for standard analyses. If you need flexibility, MatLab and Octave are also good options.

- Gephi is also a useful tool for network data (social graphs).A serious researcher will have a variety of tools at their disposal. Perhaps something for a quick and dirty examination, as well as something for non-standard data handling and tactics.

Focus Group Methodology

- Papers that addressed different aspects of the method including managing the process of data collection and enhancing the trustworthiness of analysis.
- The `other' category comprised paper, one describing the use of a radio phone – in programme in research (Weller, 2006); and few presenting methods to improve an aspect of the research process, including validity (Cho and Trent, 112006 -2008

Computer Aided Learning (Cal)

- In this method, specially designed soft wear application are used to learn a research.
- This creates an environment, where learning and assessment is fun and the opportunities to learns is equitable among the rural and urban setting.
- The main objective of this programme is to attract the students, retain them in the institute and to improve the quality of research in the education through animated multimedia – based educational content.

Social Science Lab

- Well – equipped Social Science Lab only helps in creating and maintain a much – needed congenial atmosphere for effective learning, it provides a workroom for the students.
- Because activities and the practical solution of the problem characterize every topic in detail.
- The permanent display of important maps, and globes, picture and charts and constant reference to them by teaching research of research is sure to make the teaching of research in social science is effective, lively and interesting.

Expanding Methods

- Expanding the scope of a method (its concepts and practices) within its context and discipline of origin, though low on a continuum of methodological innovation, can respond to the changing concerns of a discipline or area, raise new research questions, enhance a method's contribution, and be a step towards further methodological innovation.
- However, or will not unsettle the ontological or epistemological basis a methods leading to debate as to the extent to which expending a methods counts as innovation.

- For example, while the methods of sensory ethnography (e.g. sensory walks and interviews) place new emphasis on the sensory and visual aspects of embodied meaning and are considered innovative by some(Pink, 2009), for others they maintain the fundamental theoretical assumptions and methodological practices of anthropology/ethnography (Howe's & Classen, 2014),
- These points to the difficulty of defining innovation, the struggle between tradition and innovation, and how disciplinary communities monitor and regulate methodological boundaries..

Summary and conclusion

- There are many research methods available to teach research in social science subject.
- There is a need to select appropriate methods to achieve the desired objectives.
- The resources can be determined by what is available and relevant to the topic and students` ability and interest.
- A few of the methods available for use in the teaching of research in social science were discussed.
- It is also obvious that no single methods of learning can adequately fit all learning situations.
- However, it should be noted that there is no best method of teaching research in social sciences but combination of the other methods would definitely help in achieving the desired objectives.

Ethics With Respect to Social Science Research

- Research ethics is a **codification of scientific morality in** practice. Guidelines for research ethics specify the basic norms and values of the research community. They are based on general ethics of science, just as general ethics is based on the morality of society at large.
- Research ethics are important for a number of reasons. They **promote the aims of research**, such as expanding knowledge. They support the values required for collaborative work, such as mutual respect and fairness. They support important social and moral values, such as the principle of doing no harm to others.
- Ethics in science include: a) **standards of methods and process that address research design, procedures, data analysis, interpretation, and reporting**; and b) standards of topics and findings that address the use of human and animal subjects in research.

Publication Ethic in Social Science Research

- Researchers, authors, sponsors, editors and publishers all have ethical obligations with regard to the publication and dissemination of the results of research. Sources of funding, institutional affiliations and conflicts of interest must be declared in the publication.
- The term research ethics refers to a wide variety of values, norms, and institutional arrangements that help constitute and regulate scientific activities. Research ethics is **a codification of scientific morality in practice**. Guidelines for research ethics specify the basic norms and values of the research community.
- Ethical integrity is essential to writing and publication. Important ethical concerns to consider while writing a manuscript include **etiquette, fraudulent publication, plagiarism, duplicate publication, authorship, and potential for conflict of interest**.

References

- Alasuutari, P. (2006) The globalization of qualitative research. In C . Seale, G. Goba, J. Gubrium and D. Silverman (Eds.), *Qualitative Research Practice (pp. 507-520). London: Sage.*
- Bessell, A. G., Burke, M.C., Plaza, M. P,. Lee, O., and Schumm, J. S. (2008) The Educational Reform Rating Rubric: Example of a New Tool are Evaluating Complex Scholl Reform Initiatives', *Field Methods 20(3) : 283 – 295.*
- Borum, V. (2012). Reading and Witting Poetic Prose – African American Mothers with Deaf Daughters', *Qualitative Inquiry 12(2) : 340- 352*
- Brannen, J., and Edwards, R. (2013) 10th Aniversary Editorial', *International*
- Journal of Social Research Methodology 10(1): 1-4.


Thank you!