

LECTURE NOTES

ON

BUSINESS RESEARCH METHODS

II MBA III SEMESTER

Ms. SURABHI LAKSHMI

Assistant Professor



INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal, Hyderabad-500043.

Unit-I

INTRODUCTION TO BUSINESS RESEARCH

What is research?

The unique characteristic of human mind is the curiosity to know about the universe. Innumerable questions arise in our mind about our environment, planet and the universe. Most of these questions starting with what, why, how and soon. For example, what are stars? , why day and night alternate? How is rain formed and why the mode of life and activities of human beings vary from place to place? Whenever such questions arise we seek answer to them or we try to find out solutions to them. Seeking answers to questions and finding solutions to the problems have been the basis of human progress. A systematic search for an answer to a question or a solution to a problem is called research.

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.

Therefore, the research is a process of which a person observes the phenomena again and again and collects the data and on the basis of data he draws some conclusions. Research seeks to find out explanations to unexplained phenomena to clarify the doubtful propositions and to correct the misconceived facts. It simply means a search for facts, answer to questions and solutions to problems. The search for facts may be made through either

- (a) arbitrary (unscientific) method or
- (b) scientific method.

Arbitrary method of seeking answer to questions is based on imagination, blind belief or impression. It is vague and inaccurate.

Scientific method is a systematic rational approach to seeking fact. It is objective, precise and arrives at conclusions on the basis of verifiable evidences. Hence research is systematic and logical study of an issue problem or phenomenon through scientific method. Following definitions may reveal the proper meaning of the concept of research.

Definition of Research:

- a) According to **Black and Champion**, "scientific 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".
- b) **Emory** defines research as "any organized inquiry designed and carried out to provide information for solving a problem".
- c) **Kerlinger** defines research as a" systematic, controlled, empirical and critical investigation of hypothetical relations among natural phenomena".
- d) **L.V. Redman and A.V.H. Morry** have defined "systematic effort to gain new knowledge we call research".

Characteristic of Research

The above definitions reveal the following characteristics of research.

- a) Research is a systematic and critical investigation to a phenomenon.
- b) It aims at interpreting and explaining a phenomenon.
- c) It adopts scientific method.
- d) It is based on empirical evidences and observable experience.
- e) It develops generalizations, principles or theories.
- f) It directed towards finding answer to the questions and solutions to the problems.

I. 2: Purpose of Research

The purposes or objectives of research are varied. They are,

- Research extends knowledge of human beings social life and environment.
- Research reveals the mysteries of nature.
- Research establishes generalizations and general laws and contributes to theory building in various fields of knowledge.

Scientific Method

All scientists use common methods for their enquiry. All sciences whether natural or social agree up on methods of studying phenomena. But their materials differ. A biologist studying the structure of some flowers, a chemist studying radio active properties of an element and a sociologist studying crime situation in an urban slum. All follows similar scientific methods of inquiry. But their subjects of study are different. Therefore, they use different techniques of investigation for their study. As their materials are different, their purposes also differ. All of them will observe the phenomenon and analyze them to find out their sequences this is called scientific method. Thus scientific method is a systematic step-by-step procedure (three steps-observation, hypothesis and verification) following logical process of reasoning.

According to **prof. Morgan** “**scientific method being highly elastic, can be applicable to all domain of human activity where the discovery of truth is the objective**”. So the scientific method is means for gaining knowledge of the universe. As Karl Person observed “there is no short-cut to truth, no way to gain a knowledge of the universe expect through the gate way of scientific method”. Two elements of scientific method are,

- a)Procedural components and
- b)Personal Components.

a) Procedural Components.

Observation, hypothesis and verification are the three procedural components. Observation helps to collect data and help to build hypothesis. The second step is formation of one or more hypotheses. A hypothesis is tentative conclusion. It guides collection of data. The third stage is verification of hypothesis. It is done by analytical tools.

b) Personal Components.

The researcher needs imagination, analytical ability resourcefulness, skill, capacity to find out the hearts of the problem. Researcher’s ability and attitude are more important than the method of approach. Ambitions interest and perseverance are very much required to go on successfully with research. Researcher should have an objective scientific and professional qualification and personal quality and interest.

Meaning and essentials of scientific method

Scientific method is a way in which one can test opinion, impressions or guess by examining available evidences fore and against them. So it is controlling lot of things and establishing stable belief.

Essentials of scientific method are,

- Scientific method aims at discovering facts.
- It is itself corrective in nature.
- It is itself based on systematic doubts.
- Scientific theories are abstract in nature.

Characteristics of scientific method

1. Is a very systematic method, offering convenient working.
2. Helps in obtaining very accurate classification of facts.
3. This method is marked by the observation of heavy co relation and sequence.
4. Helps in the discovery of the scientific laws.
5. Depends and aims at achieving actual facts and not the desired ones.
6. Relies on the evidence.
7. Has a definite problem for solving, as every inquiry has a specific sense.

8. Results drawn from the scientific method are capable of being observed and then measured.
9. It links and tries to establish very general propositions.
10. Scientific results can be estimated with sufficient accuracy.
11. Scientific conclusions are very true in nature and working.
12. Observer's own views find no place during the observation as the observation is made in a very true form.

KEY ELEMENTS OF THE SCIENTIFIC METHOD

The goal of the scientific method is to get results that accurately represent the physical processes taking place in the phenomenon. To that end, it emphasizes a number of traits to ensure that the results it gets are valid to the natural world.

- **Objective** – The scientific method intends to remove personal and cultural biases by focusing on objective testing procedures.
- **Consistent** – The laws of reasoning should be used to make hypotheses that are consistent with broader, currently known scientific laws; even in rare cases where the hypothesis is that one of the broader laws is incorrect or incomplete, the hypothesis should be composed to challenge only one such law at a time.
- **Observable** – The hypothesis presented should allow for experiments with observable and measurable results.
- **Pertinent** – All steps of the process should be focused on describing and explaining observed phenomena.
- **Parsimonious** – Only a limited number of assumptions and hypothetical entities should be proposed in a given theory, as stated in Occam's Razor.
- **Falsifiable** – The hypothesis should be something which can be proven incorrect by observable data within the experiment, or else the experiment is not useful in supporting the hypothesis. (This aspect was most prominently illuminated by the philosopher of science Karl Popper.)
- **Reproducible** – The test should be able to be reproduced by other observers with trials that extend indefinitely into the future.

It is useful to keep these traits in mind when developing a hypothesis and testing procedure.

Types of Research Used in the Scientific Method

There are two basic types of research associated with the scientific method.

1) Quantitative Research

Quantitative research is based on collecting facts and figures. This type research is common in biology.

2) Qualitative Research

Qualitative research is based on collecting opinions and attitudes. This type research is common in the social sciences.

Steps in the Scientific Method

1) Identify a Problem or Question

Identify a problem to be solved or a question to be answered. For example, we know that plants require nitrogen fertilizer .If a crop is not growing very well, we might wonder if the lack of growth is due to inadequate nitrogen .Or, the crop may be growing, but we might wonder if we can stimulate the crop to growth even better if we try a new type of nitrogen fertilizer.

2) Review Literature and Gather Information

Determine as much information about the topic as possible. Are there published studies that have investigated the same or similar topic? You do not want to conduct a research project that has already been done. You want to add to the current body of knowledge. The best approach is to conduct an exhaustive review of the scientific literature.

3) Formulate Hypothesis, Null Hypothesis or Research Objective

Develop a hypothesis to be tested. A **hypothesis** is a statement that the experiment will attempt to prove. An example of a hypothesis would be: "Increasing the level of nitrogen fertilizer will increase plant growth." The purpose of the experimentation would be to prove the hypothesis.

Sometimes one develops a null hypothesis. A **null hypothesis** is a statement that the experiment will attempt to disprove. Sometimes one can never "prove a hypothesis", so one attempts to "disprove all possible null hypotheses". An example of a null hypothesis would be: "Nitrogen fertilizer does not effect plant growth." More often than not, researchers in biology develop a **research objective**, such as: "To determine the effect of increasing nitrogen fertilizer on plant growth."

4) Design Experiment

An experiment is designed to test the hypothesis, null hypothesis or satisfy the research objective. This is the critical component of the scientific method. The design of the experiment is what separates the scientific method from testimonials, general observations and assumptions. The scientific method uses the following characteristics to assure creditability.

a) Unbiased

The experiment must be designed and conducted without bias. The experiment is designed such that one treatment is **not favoured** over another. Sometimes the treatments are **blind**, and the researcher does not know which experimental units received which treatments. This is very common in human medical research. In addition, the treatments must be **random**. For example, the experimental units are randomly selected to receive the various treatments, and the treatments are randomly arranged in the area where the experiment is to be conducted. Finally, the experimental units, for example the plants, must be as **uniform** as possible.

b) Control group

There must always be a control. A **control** is a group of experimental units that do not receive the treatment. For example, in a nitrogen fertilizer study, the control group would receive no nitrogen. Sometimes the control must be a **standard** or **normal condition**. For example, if plants are growing in the soil where there is natural nitrogen present, then the control group would be the plants grown with the standard or normal amount of nitrogen present in the soil.

c) Replicates

Each of the treatments is applied to a group of experimental units, for example a group of plants. Single experimental units are never used. Usually, a minimum of 5 uniform experimental units receive each treatment. However, sometimes many more experimental units must be used to collect reliable data. The data collected on the individual experimental units are averaged in step 6) Organize and Analyze data.

d) Repeat experiment

The experiment must always be repeated to make sure the same, or very similar results, are obtained.

5) Collect data

The experimental units must be measured in order to determine the effect of the treatments. For plants, measurements may be of growth rate, size, color, flowering, yield, internal physiological factors or constituents, or what ever is needed to determine the response to the treatments.

6) Organize and Analyze data

The data must be organized and analysed. The data is averaged and organized into lists, tables, figures and/or graphs.

7) Interpret data

a) Identify trend(s)

The data is studied to identify trends, to determine which treatments caused what types of responses, to determine which treatments are better or worse than others, etc.

b) Determine significant differences

Statistical analysis is used to determine which treatments are different from the others. Oddly enough, in research you can never make a statement that two treatments are "equal", only that they are "not significantly different"!

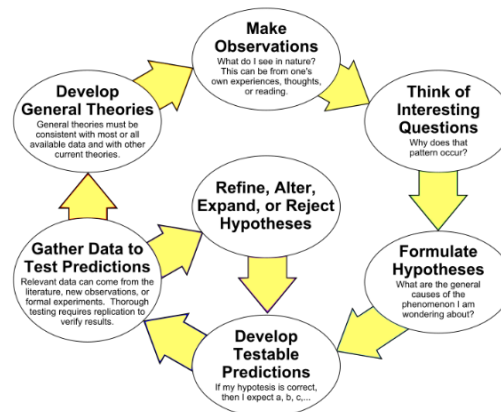
c) Draw conclusions

Finally, conclusions are drawn to support or not support the hypothesis, null hypothesis or research objective.

8) Communicate results

This is the most important part of the research. The research has no value if the results are not communicated with the scientific community, one's colleagues, students and/or the public. The research becomes creditable only if it is reviewed by and accepted by one's scientific peers in the form of a refereed publication. A **refereed publication** is a scientific article that is submitted for publication in a **refereed scientific journal**. The article is sent to several scientists (the referees) for an anonymous review, and they recommend whether or not the article should be published in the refereed journal. If accepted for publication, the results and conclusions from the study have withstood the highest level of scientific scrutiny and are deemed acceptable by the scientific community. After the results are published in the refereed journal, then it is desirable to publish the findings in popular articles, industry journal, etc.

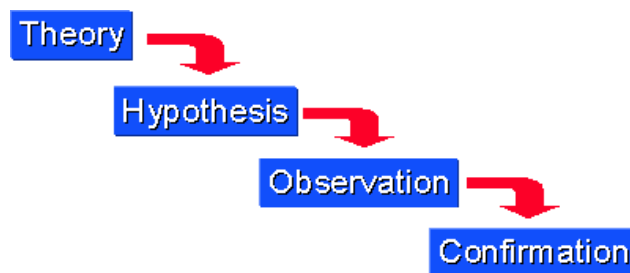
The Scientific Method as an Ongoing Process



Logical reasoning process.

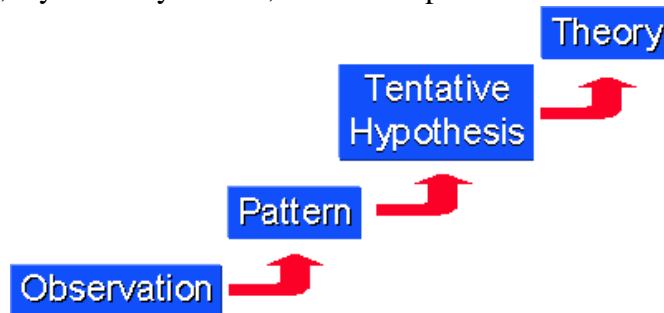
The scientist method involves the logical process of reasoning. This reasoning process is used for drawing inference from the finding of a study or for arriving at conclusion. This logical reasoning process consists of induction and deduction.

Deduction & Induction method In logic, we often refer to the two broad methods of reasoning as the *deductive* and *inductive* approaches. Deductive reasoning works from the more general to the more specific. Sometimes this is informally called a "top-down" approach. We might begin with thinking up a *theory* about our topic of interest. We then narrow that down into more specific *hypotheses* that we can test. We narrow down even further when we collect *observations* to address the hypotheses. This ultimately leads us to be able to test the hypotheses with specific data -- a *confirmation* (or not) of our original theories.



Inductive reasoning works the other way, moving from specific observations to broader generalizations and theories. Informally, we sometimes call this a "bottom up" approach (please note that it's "bottom up" and *not* "bottoms up" which is the kind of thing the bartender says to customers when he's trying to close for the night!). In inductive reasoning, we begin with specific observations and measures, begin to detect patterns and regularities, formulate some tentative hypotheses that we can explore, and finally end up developing some general conclusions or theories.

These two methods of reasoning have a very different "feel" to them when you're conducting research. Inductive reasoning, by its very nature, is more open-ended and exploratory, especially at the beginning.



Deductive reasoning is more narrow in nature and is concerned with testing or confirming hypotheses. Even though a particular study may look like it's purely deductive (e.g., an experiment designed to test the hypothesized effects of some treatment on some outcome), most social research involves both inductive and deductive reasoning processes at some time in the project. In fact, it doesn't take a rocket scientist to see that we could assemble the two graphs above into a single circular one that continually cycles from theories down to observations and back up again to theories. Even in the most constrained experiment, the researchers may observe patterns in the data that lead them to develop new theories.

Social research

Social research is a research conducted by social scientists following a systematic plan. Social research methodologies can be classified as quantitative or qualitative.^[1]

- Quantitative designs approach social phenomena through quantifiable evidence, and often rely on statistical analysis of many cases (or across intentionally designed treatments in an experiment) to create valid and reliable general claims. Related to quantity.
- Qualitative designs emphasize understanding of social phenomena through direct observation, communication with participants, or analysis of texts, and may stress contextual subjective accuracy over generality. Related to quality.

While methods may be classified as quantitative or qualitative, most methods contain elements of both. For example, qualitative data analysis often involves a fairly structured approach to coding the raw data into systematic information, and quantifying inter coder reliability. Thus, there is often a more complex relationship between "qualitative" and "quantitative" approaches than would be suggested by drawing a simple distinction between them.

Social scientists employ a range of methods in order to analyse a vast breadth of social phenomena: from census survey data derived from millions of individuals, to the in-depth analysis of a single agent's social experiences; from monitoring what is happening on contemporary streets, to the investigation of ancient historical documents. Methods rooted in classical sociology and statistics have formed the basis for research in other disciplines, such as political science, media studies, program evaluation and market research.

Objectives of Social Research

The major objectives of social research are listed as follows:-

- a) The aim of social research is to discover new facts and verifying or testing old facts.
- b) It tries to understand the human behavior and its interaction with the environment.
- c) It tries to find out the casual connection between human activities and natural laws governing them.

Methodological assumptions

Social research is based on logic and empirical observations. Charles C. Ragin writes in his *Constructing Social Research* book that "Social research involved the interaction between ideas and evidence. Ideas help social researchers make sense of evidence, and researchers use evidence to extend, revise and test ideas." Social research thus attempts to create or validate theories through data collection and data analysis, and its goal is exploration, description, explanation, and prediction. It should never lead or be mistaken with philosophy or belief. Social research aims to find social patterns of regularity in social life and usually

deals with social groups (aggregates of individuals), not individuals themselves (although science of psychology is an exception here). Research can also be divided into pure research and applied research. Pure research has no application on real life, whereas applied research attempts to influence the real world.

There are no laws in social science that parallel the laws in natural science. A law in social science is a universal generalization about a class of facts. A fact is an observed phenomenon, and observation means it has been seen, heard or otherwise experienced by researcher. A theory is a systematic explanation for the observations that relate to a particular aspect of social life. Concepts are the basic building blocks of theory and are abstract elements representing classes of phenomena. Axioms or postulates are basic assertions assumed to be true. Propositions are conclusions drawn about the relationships among concepts, based on analysis of axioms. Hypotheses are specified expectations about empirical reality derived from propositions. Social research involves testing these hypotheses to see if they are true.

Social research involves creating a theory, operationalization (measurement of variables) and observation (actual collection of data to test hypothesized relationship). Social theories are written in the language of variables, in other words, theories describe logical relationships between variables. Variables are logical sets of attributes, with people being the "carriers" of those variables (for example, gender can be a variable with two attributes: male and female). Variables are also divided into independent variables (data) that influences the dependent variables (which scientists are trying to explain). For example, in a study of how different dosages of a drug are related to the severity of symptoms of a disease, a measure of the severity of the symptoms of the disease is a dependent variable and the administration of the drug in specified doses is the independent variable. Researchers will compare the different values of the dependent variable (severity of the symptoms) and attempt to draw conclusions.

Motivating Factors of Social Research

Factor 1# Stimulation of Respondent:

As the researcher depends upon the respondents for collection of data for the attainment of the research objectives, one of the most important tasks of the researcher is to inspire and stimulate the respondents with zeal to help for the accomplishment of research goals. In other words the motivation of the respondents plays a significant role for the success of any kind of research.

Human motives are based on certain needs which may be primary or secondary and vary in their intensity according to situation and time. The researcher must study these needs, try to understand their intensity and have the responsibility to satisfy them in order to stimulate the respondents for research work.

Motivation means any idea, need, emotion or organic state that prompts a man to an action. Motivation is an internal factor that integrates a man's behaviour. As the motive is within the individual, it is necessary to study the needs, emotions etc. in order to motivate him to co-operate in the research work.

The following are the important inducing factors which influence the respondent's behaviour and induce him for the best performance to meet the need of research:

- (i) The research should be directed towards the solution of respondent's problem.
- (ii) The nature of the problem or topic must have social relevance.
- (iii) The researcher should clearly spell the goals of research.
- (iv) The respondents must be informed about the matters concerning objectives of the research. The more a person knows about its subject matter, the more interest and concern he will develop.
- (v) Respondents can be motivated to involve in research if they get continuous recognition for their efforts. Respondents provide valuable information and suggestion for the success of research work. If the researcher has a praise of words for the respondents' cooperation, it motivates the respondents more and more to be

involved in the research process. Thus recognition tends to motivate the respondents to provide required information for research.

Factor 2# Stimulation of Researcher:

The success of a research work, to a great extent, depends upon the motivation of the researcher as well.

The following are some of the factors which stimulate a researcher to conduct research effectively:

(i) The researcher must have a concrete and complete knowledge of the subject under study. He must be capable of removing the doubts of the respondents regarding the study.

(ii) He must have a personal interest in the study undertaken.

(iii) The researcher must have sufficient knowledge about the respondents.

(iv) The researcher must have the idea of the tools of research.

Factor 3# A Sense or Participation:

Participation in a research activity does not mean simply the involvement of the respondents in giving information on a topic or problem. In real sense, participation is an individual's mental and emotional involvement in research solutions that encourage him to contribute to research and to share the responsibility for it.

From utilitarian point of view the main goal of research is to understand social life and attempt on social welfare. However it cannot be done without the active participation of the people involved in the process of social research.

Any social research, whether it is meant for the development of a specific section of society or for the overall development of the entire society, requires participation of people. Research does not only mean involvement of the researcher but it also requires the conscious participation of the respondents.

The respondents involve themselves in thinking, identifying the needs, fixing priorities of the needs, providing valuable information, implementing and evaluating critically various research programmes. Thus it involves the participation of both the researcher and the respondents.

Factor 4# Growth of Knowledge:

Interest for increasing knowledge motivates people to do research in their own field. Research adds to the existing knowledge in a systematic way. The quest for knowledge is therefore an important motivating factor in social research. Discovering the truth always forces man to undertake research in own society.

Factor 5# Quest for Progress:

Research has proved to be a significant and powerful tool in bringing social progress. Without scientific social research there would be very little progress. The results of social research will provide us with the possible means to bring solution to different social problems. Research opens new avenues and provides a better alternative to us. It enhances the efficiency of all the agencies and organisations engaged in the development of society. So the quest for progress is also another motivating factor of social research.

Factor 6# Curiosity to Understand the Cause and Effect Relationship of Various Social Phenomena:

Research is nothing but a desire to understand the causal explanation of various facts and to explain the natural laws which govern them. Social research tries to discover the cause-effect relationship between different aspects of a social phenomenon. In order to solve a social problem, one must first understand the root cause of that particular problem. Finding the cause of an effect is one of the greatest tasks of research and its quest always motivates people to undertake research.

Social research methods

Social Surveys – are typically structured questionnaires designed to collect information from large numbers of people in standardised form.

Social Surveys are written in advance by the researcher and tend to be pre-coded and have a limited number of closed-questions and they tend to focus on relatively simple topics. A good example is the UK National Census. Social Surveys can be administered (carried out) in a number of different ways – they might be self-completion (completed by the respondents themselves) or they might take the form of a structured interview on the high street, as is the case with some market research.

Experiments – aim to measure as precisely as possible the effect which one variable has on another, aiming to establish cause and effect relationships between variables.

Experiments typically start off with a hypothesis – a theory or explanation made on the basis of limited evidence as a starting point for further investigation, and will typically take the form of a testable statement about the effect which one or more independent variables will have on the dependent variable. A good experiment will be designed in such a way that objective cause and effect relationships can be established, so that the original hypothesis can be verified, or rejected and modified.

There are two types of experiment 1) laboratory and 2) field experiments

– A **laboratory experiment** takes place in a controlled environment, such as a laboratory, whereas a **field experiment** takes place in a real-life setting such as a classroom, the work place or even the high street.

Interviews – A method of gathering information by asking questions orally, either face to face or by telephone.

Structured Interviews are basically social surveys which are read out by the researcher – they use pre-set, standardised, typically closed questions. The aim of structured interviews is to produce quantitative data.

Unstructured Interviews, also known as informal interviews, are more like a guided conversation, and typically involve the researcher asking open-questions which generate qualitative data. The researcher will start with a general research topic and ask questions in response to the various and differentiated responses the respondents give. Unstructured Interviews are thus a flexible, respondent-led research method.

Semi-Structured Interviews consist of an interview schedule which typically consists of a number of open-ended questions which allow the respondent to give in-depth answers. For example, the researcher might have 10 questions (hence structured) they will ask all respondents, but ask further differentiated (unstructured) questions based on the responses given.

Participant Observation – involves the researcher joining a group of people, taking an active part in their day to day lives as a member of that group and making in-depth recordings of what she sees. Participant Observation may be **overt**, in which case the respondents know that the researcher is conducting sociological research, or **covert (undercover)** where the respondents are deceived into thinking the researcher is ‘one of them’ do not know the researcher is conducting research.

Ethnographies and Case Studies

Ethnographies are an in-depth study of the way of life of a group of people in their natural setting. They are typically very in-depth and long-term and aim for a full (or ‘thick’), multi-layered account of the culture of a group of people. Participant Observation is typically the main method used, but researchers will use all other methods available to get even richer data – such as interviews and analysis of any documents associated with that culture.

Case Studies involves researching a single case or example of something using multiple methods – for example researching one school or factory. An ethnography is simply a very in-depth case study.

Longitudinal Studies – studies of a sample of people in which information is collected from the same people at intervals over a long period of time. For example, a researcher might start off in 2015 by getting a sample of

1000 people to fill in a questionnaire, and then go back to the same people in 2020, and again in 2025 to collect further information

Utility of Social Research:

To the question “what use is social research?” one may reply “of what use is a newborn child?” in the manner of Benjamin Franklin who replied thus, when asked the utility of his findings about the relationship between thunderclouds and electricity.

This means that new knowledge like the new-born baby, holds great potential of worth and maturity. Also like the new-born child, it gives us pleasure. It gives us satisfaction of knowing the unknown.

This points to a value that the scientist is committed to, i.e., the self-justifying goodness of new knowledge about anything big or small. “Social research is persistently opening our eyes to the social reality, simplifying the mysterious within the seemingly common place in social life and shattering its garments of make-believe by which pious hands have hidden their uglier features.

The obvious function of research is to add new knowledge to its existing store, but its power of cleansing our minds of clichés and removing the rubbish of inapplicable theory are equally notable. Scientific research is a cumulative process. It is also a rejective process, especially in social sciences ... understanding can be (advanced) not only by gains in knowledge but also by discarding outworn assumptions.”

A social researcher is interested in the discovery and interpretation of social processes, patterns of behaviour, similarities and dissimilarities that apply to typical social phenomena and social systems, generally.

That is the social researcher is concerned with types and classes of social situation, persons or groups of which the unit he is studying at the time, is a specimen or an instance. His facts are selected and related according to their intrinsic nature and the susceptibility to organization into a logical system.

This search for knowledge has a definite relation to people’s basic needs and welfare. The social scientist assumes that all knowledge is potentially useful in the end. It must be remembered, however that science and society have a two way relationship. There is a give and take between science and social conditions. Science helps to create social conditions; social conditions recharge the accumulators of science

Unit -2
The Survey Method

Introduction and Definition

For the classification of the research designs, the approach needed to collect the primary data, is very commonly used. For solving this particular purpose, there are two options we can observe the events, people, behaviour, conditions etc.

Also we can use the other alternative option which involves communicating with the people about the various topics. Now in this article, we will study about the survey method which is used a lot as a social survey and finds maximum use where the published data is available.

So this method is the technique used for carrying out the investigation processes with the help of direct observation or by the collection of the data by conducting interviews etc. The survey method is a very essential and a useful tool to gather the evidence relating to the various social problems.

According to Shelly M. Harrison, social survey is a, “comparative undertaking which applies scientific method to the study and the treatment of the currently related social problems and the conditions having definite geographic limit and bearing, plus such a spreading of facts, conclusions and recommendations as will make them, as far as possible, the common knowledge of the community and a force for the intelligent coordinated action.”

Objectives of the Survey

1. Collecting general information.
2. To provide data that later helps in the formulation of the hypothesis.
3. To enable the researcher to have direct and close contact with the process or the phenomenon, that is being studied.
4. To know the people better by knowing their opinions and their attitudes.
5. To explain the relationship between the various variables.
6. To refine and expand the old theories in order to get the new theories.

Characteristics of the Social Survey

1. Are less intensive in nature and also involve widely dispersed group of people.
2. Data collected in a social survey can act as a proper base for carrying out further the various social research operations.
3. Helps in making the assertion about the distribution of the characteristics in a population.
4. Helps in carrying out the study of specific current problems of the society only.
5. Uses methods like observation, interview etc for collecting the data.

Advantages of Survey

1. Researcher is able to directly get involved with the people from whom he wants to get the information.
2. The results obtained from this method are not based on any type of theory but instead are based on the actual facts of the life.
3. Very helpful in the testing of the validity of the theories.
4. Offers greater objectivity.
5. Very less possibilities of existence of the personal bias.
6. Help in the formulation of the hypothesis.

Disadvantages of survey.

1. Very expensive.
2. Very time consuming.
3. Wasteful method.
4. Less reliable.
5. Sometimes personal bias may vitiate the results.
6. Deals with the problems relating to the immediate importance only.
7. Occurrence of the non response errors.
8. Not an efficient method for conducting the verification of the accuracy of the data, which is collected.

Planning social survey

Well (1935) defined social survey as a fact-finding study, dealing chiefly with working class's poverty, its nature and problem. But the scope of the social survey is much wider than that covered by this definition.

Social survey methods. A social survey may be occasioned simply by a need of administrative facts on some aspects of public life or be designed to, investigate cause-effect relationship or to throw fresh light on some aspects of the sociological theory. The major objective of the surveys is simply to supply information about the social status of the population. Population census, living standard surveys, labour force surveys, market research survey, opinion poses etc all come into this category of study.

Subject matter in Social surveys.

The subject matter of social surveys depends upon the nature of information to be gathered. Some of the subject matters are:

1. Demographic characteristics.
2. Social environment.
3. Social activities.
4. Opinions and attitudes.

1. Demographic characteristics.

Household compositions, marital status, fertility rate, age composition etc.

2. Social environment.

Occupations, income, housing conditions, social amenities.

3. Social activities.

Actions and behaviours of the populations such as a use of leisure time, newspaper reading, travelling habits.

4. Opinions and attitudes.

Opinion and attitudes of the people towards various social factor, motive and expectations of the people etc.

Steps in surveys.

1. Formulation of the problem.
2. Preparation of time schedule.
3. ,Staffing
4. The cost of the survey.
5. Instruments of information collection.
6. Choosing or respondents/Survey units.
7. Pilot survey.
8. Classification and Tabulation.
9. Analysis and interpretations.

A short description of all step is discussed as below.

1. Formulation of the problem.

The first and foremost thing to be decided by the investigation is the identification of subject matter of the intended surveys. The subject matter should be.

1. The area in which the researcher may be deeply interested.
2. Not stray problems.
3. Not necessary new one.
4. Having utility and
5. Practically feasible.

2. Preparation of time scheduling.

In every survey planning, a time schedule for completing the different phases of a survey work should be given. The time scheduling in the form of a chart or a table.

3. Staffing.

A tentative description of manpower to be utilised for the survey work should be given by their man-months utilisation. A competent statistician is always to be included in the survey team.

4. The cost of the survey.

Estimates of the costs for the different components of the survey should be given. The costs may be:

1. Administrative cost\Salary for staff.
2. Consultant's fee.
3. Field cost.
4. Travelling and communicating expenses.
5. Secretarial work /Report preparation.

The overhead cost of about 5%-15% is also included in the total cost for the survey to be conducted for client organisations.

5. Instruments of information collection.

In surveys, information is to be collected from the respondents. The information is collected by using following methods.

1. Interviewing the respondents.
2. Mailed questionnaire administration.
3. Taking notes.

6. Choosing of respondents /Survey units.

The most important technical part of the survey work is the development of sample plan according to which units are selected. Many factors play in this respect Some of them are.

1. Defining survey population.
2. Testing the adequacy of the frame of the survey population.
3. Cost/finance available.
4. Degree of accuracy of the results

7. Pilot survey.

Before conducting the final round of the field survey, a mini-survey called pilot survey/pre-test is to be conducted. The pilot survey gives information as to.

1. The variability of the sample units.
2. Suitability of the methods for data collection.
3. Non-response rates expected.
4. Probable cost and duration of the main survey and its various steps.

8. Classification and tabulation.

Once the field work is over the data collected are classified and tabulated. This may be done either manually or by using computers.

9. Analysis and interpretations.

The various statistical measures are calculated to describe the nature of the survey population. Explorations as regards to interrelationships between various characteristics of the population are also done.

10 Report.

Finally, a report of the whole survey work is prepared as documents. The reports should include all the survey schemes, methods of analysis and the results obtained. The report is prepared in two stages.

1. Final Report and
2. Draft report.

The draft is the preliminary report of the survey work. The draft is prepared to experts for comments and suggestions. And finally, a report is prepared by incorporating their suggestion and comments.

Case Study

Social researches are curious about their social settings. Their interest are virtually unlimited. Any social setting is potential for scientific inquiry. The diversity of social topic and situation made researcher to plan their action. This plan for research is conventionally labelled as research design.

Case study method is considered as one of the popular type of research design used by social scientist. It is an intensive study of a particular case. In sociological investigation a case may be any of the following, taken singly or in combination. (1) A person, (2) A group of person such as family or gang, (3) A class of person such as thieves or professors, (4) An ecological unit such as neighbourhood or community, (5) cultural unit such as fashion or institution.

Definitions:

Kvornvey (1986) defined “case study involves studying individual cases, often in their natural environment and for a long period of time”.

Yin (1991) has defined case study as “an empirical inquiry that investigate contemporary phenomena within its real-life context. This kind of research design usually involves the qualitative method of data collection. It presents holistic account that offers insights in to the case understudy. Thus case study is not a method of data collection; rather it is a research strategy. An empirical inquiry that investigates contemporary phenomenon by using multiple source of evidence.

Characteristics:

Hartfield (1982) has referred to the following characteristics of case study.

- It studies whole unit in their totality.
- It employees several methods in data collection to prevent errors and distortions.
- It often studies a single unit: one unit in one study.
- It perceives the respondent as knowledgeable person, not just as a source of data.
- It studies typically case.

Purpose of case study

Burns (2000) has point out the following purposes of case study.

- It may be a source of hypotheses for future research.
- It helps to establish generalizations about the wider population to which the unit belongs.
- It provides anecdotal evidence that illustrates more general findings.
- To refute a universal generalization, a single case can represent a significant contribution to theory building.
- To test the feasibility of the quantitative study.

Advantages of case study

Black champion (1976) enlisted following advantages of case study.

- It makes in-depth study possible.
- It is flexible in data collection methods.

- It could be used for studying any dimension of the topic; one specific aspect of the problem.
- It could be conducted practically any kind of social setting.
- Case studies are inexpensive.
- It helps to study unique case.

Disadvantages or criticism.

Case study method is generally criticised on the following basis,

1. Subjective bias:

Research subjectivity in collecting data for supporting or refuting a particular explanation, personal view of investigation influences the findings and conclusion of the study.

2. Little evidence for scientific generalization:

The common complaint against case study is; how can generalization be made from a single case? As an answer to this case studies are generalize to theoretical propositions, not to statistical populations. Object of case study is to expand theory and not to undertake statistical generalization.

3. Time consuming:

As it produces a lot of information which is difficult to analyse adequately.

4. Doubtful reliability:

The investigator cannot prove his authenticity for obtaining data or having no bias in analysing them.

5. Missing validity:

For investigator, what seems to be true is more important than what is true. The case study can oversimplify or exaggerate leading to erroneous conclusions.

6. Case study has no representatives, i.e. each case studied does not represent other similar cases.

Criticism

1. Findings of case studies are biased because the research is usually sloppy.

This criticism is probably based on the prejudice that quantitative researchers are against qualitative data. They also believe that qualitative study cannot be replicated.

2. Case studies are not useful for generalization.

- It is not possible to generalize from a single case
- If a number of cases are used for generalizations, it will be extremely difficult to establish their comparability.

3. Case studies take too long time and produce unmanageable amounts of data. The methods of data collection which are time consuming.

Experiment method

An experiment is a procedure carried out to support, refute, or validate a hypothesis. Experiments provide insight into cause-and-effect by demonstrating what outcome occurs when a particular factor is manipulated. Experiments vary greatly in goal and scale, but always rely on repeatable procedure and logical analysis of the results. There also exists natural experimental studies.

A child may carry out basic experiments to understand gravity, while teams of scientists may take years of systematic investigation to advance their understanding of a phenomenon. Experiments and other types of hands-on activities are very important to student learning in the science classroom. Experiments can raise test scores and help a student become more engaged and interested in the material they are learning, especially when used over time.^[1] Experiments can vary from personal and informal natural comparisons (e.g. tasting a range of chocolates to find a favourite), to highly controlled (e.g. tests requiring complex apparatus overseen by many scientists that hope to discover information about subatomic particles). Uses of experiments vary considerably between the natural and human sciences.

Experiments typically include controls, which are designed to minimize the effects of variables other than the single independent variable. This increases the reliability of the results, often through a comparison between control measurements and the other measurements. Scientific controls are a part of the scientific method. Ideally, all variables in an experiment are controlled (accounted for by the control measurements) and none are uncontrolled. In such an experiment, if all controls work as expected, it is possible to conclude that the experiment works as intended, and that results are due to the effect of the tested variable.

Types of experiment method

Experiments might be categorized according to a number of dimensions, depending upon professional norms and standards in different fields of study. In some disciplines (e.g., psychology or political science), a 'true experiment' is a method of social research in which there are two kinds of variables. The independent variable is manipulated by the experimenter, and the dependent variable is measured. The signifying characteristic of a true experiment is that it randomly allocates the subjects to neutralize experimenter bias, and ensures, over a large number of iterations of the experiment, that it controls for all confounding factors.

Controlled experiments

A controlled experiment often compares the results obtained from experimental samples against *control* samples, which are practically identical to the experimental sample except for the one aspect whose effect is being tested (the independent variable). A good example would be a drug trial. The sample or group receiving the drug would be the experimental group (treatment group); and the one receiving the placebo or regular treatment would be the control one. In many laboratory experiments it is good practice to have several replicate samples for the test being performed and have both a positive control and a negative control. The results from replicate samples can often be averaged, or if one of the replicates is obviously inconsistent with the results from the other samples, it can be discarded as being the result of an experimental error (some step of the test procedure may have been mistakenly omitted for that sample). Most often, tests are done in duplicate or triplicate. A positive control is a procedure similar to the actual experimental test but is known from previous experience to give a positive result. A negative control is known to give a negative result. The positive control confirms that the basic conditions of the experiment were able to produce a positive result, even if none of the actual experimental samples produce a positive result. The negative control demonstrates the base-line result obtained when a test does not produce a measurable positive result. Most often the value of the negative control is treated as a "background" value to subtract from the test sample results. Sometimes the positive control takes the quadrant of a standard curve.

An example that is often used in teaching laboratories is a controlled protein assay. Students might be given a fluid sample containing an unknown (to the student) amount of protein. It is their job to correctly perform a controlled experiment in which they determine the concentration of protein in the fluid sample (usually called the "unknown sample"). The teaching lab would be equipped with a protein standard solution with a known protein concentration. Students could make several positive control samples containing various dilutions of the protein standard. Negative control samples would contain all of the reagents for the protein assay but no protein. In this example, all samples are performed in duplicate. The assay is a colorimetric assay in which a spectrophotometer can measure the amount of protein in samples by detecting a colored complex formed by the interaction of protein molecules and molecules of an added dye. In the illustration, the results for the diluted test samples can be compared to the results of the standard curve (the blue line in the illustration) to estimate the amount of protein in the unknown sample.

Controlled experiments can be performed when it is difficult to exactly control all the conditions in an experiment. In this case, the experiment begins by creating two or more sample groups that are *probabilistically equivalent*, which means that measurements of traits should be similar among the groups and that the groups should respond in the same manner if given the same treatment. This equivalency is determined by statistical methods that take into account the amount of variation between individuals and the number of individuals in each group. In fields such as microbiology and chemistry, where there is very little variation between individuals and the group size is easily in the millions, these statistical methods are often bypassed and simply splitting a solution into equal parts is assumed to produce identical sample groups.

Once equivalent groups have been formed, the experimenter tries to treat them identically except for the one *variable* that he or she wishes to isolate. Human experimentation requires special safeguards against outside variables such as the *placebo effect*. Such experiments are generally *double blind*, meaning that neither the volunteer nor the researcher knows which individuals are in the control group or the experimental group until after all of the data have been collected. This ensures that any effects on the volunteer are due to the treatment itself and are not a response to the knowledge that he is being treated.

In human experiments, researchers may give a subject (person) a stimulus that the subject responds to. The goal of the experiment is to measure the response to the stimulus by a test method.

In the design of experiments, two or more "treatments" are applied to estimate the difference between the mean responses for the treatments. For example, an experiment on baking bread could estimate the difference in the responses associated with quantitative variables, such as the ratio of water to flour, and with qualitative variables, such as strains of yeast. Experimentation is the step in the scientific method that helps people decide between two or more competing explanations – or hypotheses. These hypotheses suggest reasons to explain a phenomenon, or predict the results of an action. An example might be the hypothesis that "if I release this ball, it will fall to the floor": this suggestion can then be tested by carrying out the experiment of letting go of the ball, and observing the results. Formally, a hypothesis is compared against its opposite or null hypothesis ("if I release this ball, it will not fall to the floor"). The null hypothesis is that there is no explanation or predictive power of the phenomenon through the reasoning that is being investigated. Once hypotheses are defined, an experiment can be carried out and the results analysed to confirm, refute, or define the accuracy of the hypotheses.

Natural experiments

The term "experiment" usually implies a controlled experiment, but sometimes controlled experiments are prohibitively difficult or impossible. In this case researchers resort to *natural experiments* or *quasi-experiments*.^[12] Natural experiments rely solely on observations of the variables of the system under study, rather than manipulation of just one or a few variables as occurs in controlled experiments. To the degree possible, they attempt to collect data for the system in such a way that contribution from all variables can be determined, and where the effects of variation in certain variables remain approximately constant so that the effects of other variables can be discerned. The degree to which this is possible depends on the observed correlation between explanatory variables in the observed data. When these variables are *not* well correlated, natural experiments can approach the power of controlled experiments. Usually, however, there is some correlation between these variables, which reduces the reliability of natural experiments relative to what could be concluded if a controlled experiment were performed. Also, because natural experiments usually take place in uncontrolled environments, variables from undetected sources are neither measured nor held constant, and these may produce illusory correlations in variables under study.

Much research in several science disciplines, including economics, , ecology, political science, geology, palaeontology, meteorology, and astronomy, relies on quasi-experiments. For example, in astronomy it is clearly impossible, when testing the hypothesis "Stars are collapsed clouds of hydrogen", to start out with a giant cloud of hydrogen, and then perform the experiment of waiting a few billion years for it to form a star. However, by observing various clouds of hydrogen in various states of collapse, and other implications of the hypothesis (for example, the presence of various spectral emissions from the light of stars), we can collect data we require to support the hypothesis. An early example of this type of experiment was the first verification in the 17th century that light does not travel from place to place instantaneously, but instead has a measurable speed. Observation of the appearance of the moons of Jupiter were slightly delayed when Jupiter was farther from Earth, as opposed to when Jupiter was closer to Earth; and this phenomenon was used to demonstrate that the difference in the time of appearance of the moons was consistent with a measurable speed.

Field experiments

Field experiments are so named to distinguish them from laboratory experiments, which enforce scientific control by testing a hypothesis in the artificial and highly controlled setting of a laboratory. Often used in the social sciences, and especially in economic analyses of education and health interventions, field experiments have the advantage that outcomes are observed in a natural setting rather than in a contrived laboratory environment. For this reason, field experiments are sometimes seen as having higher external validity than laboratory experiments. However, like natural experiments, field experiments suffer from the possibility of contamination: experimental conditions can be controlled with more precision and certainty in the lab. Yet some phenomena (e.g., voter turnout in an election) cannot be easily studied in a laboratory.

Trial and error experiment

Trial and error is a fundamental method of problem solving. It is characterised by repeated, varied attempts which are continued until success, or until the agent stops trying.

According to W.H. Thorpe, the term was devised by C. Lloyd Morgan (1852–1936) after trying out similar phrases "trial and failure" and "trial and practice". Under Morgan's Canon, animal behaviour should be explained in the simplest possible way. Where behaviour seems to imply higher mental processes, it might be explained by trial-and-error learning. An example is the skilful way in which his terrier Tony opened the garden gate, easily misunderstood as an insightful act by someone seeing the final behaviour. Lloyd Morgan, however, had watched and recorded the series of approximations by which the dog had gradually learned the response, and could demonstrate that no insight was required to explain it.

Edward Thorndike showed how to manage a trial-and-error experiment in the laboratory. In his famous experiment, a cat was placed in a series of puzzle boxes in order to study the law of effect in learning. He plotted learning curves which recorded the timing for each trial. Thorndike's key observation was that learning was promoted by positive results, which was later refined and extended by B.F. Skinner's operant conditioning.

Trial and error is also a heuristic method of problem solving, repair, tuning, or obtaining knowledge. In the field of computer science, the method is called generate and test. In elementary algebra, when solving equations, it is "guess and check".

This approach can be seen as one of the two basic approaches to problem solving, contrasted with an approach using insight and theory. However, there are intermediate methods which for example, use theory to guide the method, an approach known as *guided empiricism*.

Examples

Trial and error has traditionally been the main method of finding new drugs, such as antibiotics. Chemists simply try chemicals at random until they find one with the desired effect. In a more sophisticated version, chemists select a narrow range of chemicals it is thought may have some effect using a technique called structure-activity relationship. (The latter case can be alternatively considered as a changing of the problem rather than of the solution strategy: instead of "What chemical will work well as an antibiotic?" the problem in the sophisticated approach is "Which, if any, of the chemicals in this narrow range will work well as an antibiotic?") The method is used widely in many disciplines, such as polymer technology to find new polymer types or families.

The scientific method can be regarded as containing an element of trial and error in its formulation and testing of hypotheses. Also compare genetic algorithms, annealing and reinforcement learning – all varieties for search which apply the basic idea of trial and error.

Biological evolution can be considered as a form of trial and error. Random mutations and sexual genetic variations can be viewed as trials and poor reproductive fitness, or lack of improved fitness, as the error. Thus after a long time 'knowledge' of well-adapted genomes accumulates simply by virtue of them being *able* to reproduce.

Bogosort, a conceptual sorting algorithm (that is extremely inefficient and impractical), can be viewed as a trial and error approach to sorting a list. However, typical simple examples of bogosort do not track which orders of the list have been tried and may try the same order any number of times, which violates one of the basic principles of trial and error. Trial and error is actually more efficient and practical than bogosort; unlike bogosort, it is guaranteed to halt in finite time on a finite list, and might even be a reasonable way to sort extremely short lists under some conditions. Jumping spiders of the genus *Portia* use trial and error to find new tactics against unfamiliar prey or in unusual situations, and remember the new tactics. Tests show that *Portia fimbriata* and *Portia labiata* can use trial and error in an artificial environment, where the spider's objective is to cross a miniature lagoon that is too wide for a simple jump, and must either jump then swim or only swim

Controlled observation study In fields such as epidemiology, social sciences, psychology and statistics, an observational study draws inferences from a sample to a population where the independent variable is not under the control of the researcher because of ethical concerns or logistical constraints. One common observational study is about the possible effect of a treatment on subjects, where the assignment of subjects into a treated group versus a control group is outside the control of the investigator. This is in contrast with experiments, such as randomized controlled trials, where each subject is randomly assigned to a treated group or a control group.

Types of observation study

- Case-control study: study originally developed in epidemiology, in which two existing groups differing in outcome are identified and compared on the basis of some supposed causal attribute.
- Cross-sectional study: involves data collection from a population, or a representative subset, at one specific point in time.
- Longitudinal study: correlational research study that involves repeated observations of the same variables over long periods of time.
- Cohort study or Panel study: a particular form of longitudinal study where a group of patients is closely monitored over a span of time.
- Ecological study: an observational study in which at least one variable is measured at the group level.

Ex post facto

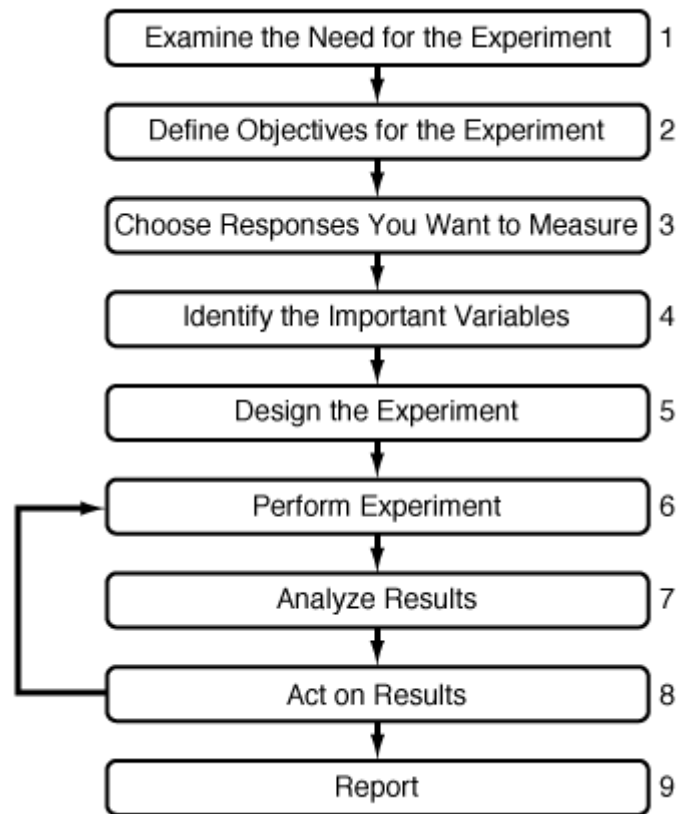
An *ex post facto* law (corrupted from Latin: *ex postfacto*, lit. 'out of the aftermath') is a law that retroactively changes the legal consequences (or status) of actions that were committed, or relationships that existed, before the enactment of the law. In criminal law, it may criminalize actions that were legal when committed; it may aggravate a crime by bringing it into a more severe category than it was in when it was committed; it may change the punishment prescribed for a crime, as by adding new penalties or extending sentences; or it may alter the rules of evidence in order to make conviction for a crime likelier than it would have been when the deed was committed. Conversely, a form of *ex post facto* law commonly called an *amnesty law* may decriminalize certain acts. A pardon has a similar effect, in a specific case instead of a class of cases. Other legal changes may alleviate possible punishments (for example by replacing the death sentence with lifelong imprisonment) retroactively. Such legal changes are also known by the Latin term *in mitius*.

A law may have an *ex post facto* effect without being technically *ex post facto*. For example, when a previous law is repealed or otherwise nullified, it is no longer applicable to situations to which it had been, even if such situations arose before the law was voided. The principle of prohibiting the continued application of such laws is called *nullum crimen, nulla poena sine praevia lege poenali*, especially in European Continental systems. This is related to the principle of legality.

Some common-law jurisdictions do not permit retroactive criminal legislation, though new precedent generally applies to events that occurred before the judicial decision. *Ex post facto* laws are expressly forbidden by the United States Constitution in Article 1, Section 9, Clause 3 (with respect to federal laws) and Article 1, Section 10 (with respect to state laws). In some nations that follow the Westminster system of government, such as the United Kingdom, *ex post facto* laws are technically possible, because the doctrine of parliamentary supremacy allows Parliament to pass any law it wishes. In a nation with an entrenched bill of rights or a written constitution, *ex post facto* legislation may be prohibited.

Experimental Planning

Four to six weeks in the lab can save you an hour in the library--G.J. Quartered, Dow Chemical Co. So far, this chapter has presented various methods of analysing rate data. It is just as important to know in which circumstances to use each method as it is to know the mechanics of these methods. In this section we discuss a heuristic to plan experiments to generate the data necessary for reactor design.



Flowchart for experimental projects

Do You Really Need the Experiments?

When you are preparing to initiate an experimental program, be sure to question yourself and others to help guide your progress. The following questions will help you dig deeper into your project.

- Why perform the experiments?
- Can the information you are seeking be found elsewhere (such as literature journals, books, company reports, etc.)?
- Can you do some calculations instead?
- Have sufficient time and money been budgeted for the program?
- Are you restricted to specific materials or equipment?
- Will the safety of the investigators be endangered to such a degree that the program should not be carried out?

These and other appropriate questions must be answered prior to beginning the experimental program so that the need for the experiments is clearly established. Text Figure 5-12 shows a flow sheet for experimental planning.

Define the Objectives of the Experiment

Prepare a list of all the things you want to accomplish. Next try to prioritize your list, keeping in mind the following:

- What questions regarding your problem would you most like to answer?
- Are you sure you are not losing sight of the overall objectives and other possible alternative solutions ("can't see the forest for the trees" syndrome)?
- How comprehensive does the program need to be? Are you looking at an exhaustive study or a cursory examination of a narrow set of conditions?

Specific answers to these questions will guide the rest of the project.

Choose the Responses You Want to Measure

There are generally two different types of variables that are considered in an experimental program. The *independent* variables make things happen. Changes in the independent variables cause the system to respond. The *responses* are the *dependent* variables. Changing any one of the independent variables will change the system response (the dependent variable). As the experimental program is designed, the important dependent variables to be measured must be identified.

- What are the controlled or independent variables?
- What are the dependent variables?
- Are instruments or techniques available to make the measurements?
- Do they need to be calibrated? If so, have they been?
- Will the accuracy and precision of the expected results be sufficient to distinguish between different theories or possible outcomes?

Identify the Important Variables

In any experimental program there will always be many, many quantities you can measure. However, you must decide which independent variables have the greatest influence on the dependent variable.

- What are the *really* important measurements to make?
- What are the ranges or levels of these variables to be examined?
- Instead of changing each independent variable separately, can dimensionless ratios or groups be formed (i.e., Schmidt or Sherwood numbers) and varied so as to produce the same end results with fewer measurements?

Design the Experiment

To obtain the maximum benefit from a series of experiments, they must be properly designed. How can the experimental program be designed to achieve the experimental objectives in the simplest manner with the minimum number of measurements and the least expense? A *successfully designed experiment* is a series of organized trials which enables one to obtain the most experimental information with the least amount of effort. Three important questions to consider when designing experiments are:

- What are the types of errors to avoid?
- What is the minimum number of experiments that must be performed?
- When should we consider repeating experiments?

Types of Errors

There are two types of errors that should be avoided in experimental design. A type I error is one in which you declare that a variable has an effect on the experimental outcome when in fact it really doesn't. A type II error occurs when we fail to discover a *real* effect. A type II error results in lost information; a variable gets incorrectly classified as insignificant to the process or ignored and as a result, no further examination of it takes place. Type II errors can be avoided by researching fundamental principles related to the experiments, gathering sufficient information, and planning thoughtfully.

The Minimum Number of Experiments (or . . . "Getting the Most Bang for Your Buck")

The minimum number of experiments that must be performed is related to the number of important independent variables that can affect the experiment and to how precisely we can measure the results of the experiment. One of the most important strategies to remember is to carry out first experiments at the extremes (maximum and minimum setting) of the range of the controlled variables. For example, if the range of pressures that can be used to determine the rate law of a gas-phase reaction is from 1 to 100 atm, it is somewhat best to determine the

rate at 1 atm and then at 100 atm. If the independent variables have no effect on the dependent variables at the extremes, it is somewhat doubtful that there will be an effect in the intermediate range. Consequently, a lot of time, money, and energy would be lost if we progressed from a setting of 1 atm to 2 atm and found no effect, then to 10 atm and found no change, then to 50 atm and 80 atm with similar results.

In designing the experiments, we will first choose two levels (i.e., settings) for each independent variable. Because these levels are usually at the extremes of the variable range, we refer to these settings as high and low (e.g., on/off, red/green, 100 psi/14.7 psi, 100°C/0°C, etc.). For example, consider an experimental program where the dependent variable is a function of three independent variables (A, B, and C), each of which can take on two possible values or levels.

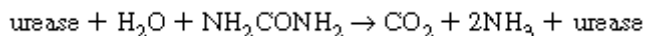
<i>Independent Variable Names</i>	<i>Possible Levels</i>	
A	High	Low
B	High	Low
C	High	Low

If all possible variable combinations were to be tested, the number of experiments is equal to the number of levels, N , raised to the power of the number of independent variables, n . For the example for variables A, B, and C, the number of experiments necessary to test all combinations of independent variables is equal to $N^n = 2^3 = 8$ experiments. These are detailed in Table A and Text Figure 5-13 [(+) indicates a high level, while (-) indicates a low level of a particular variable].

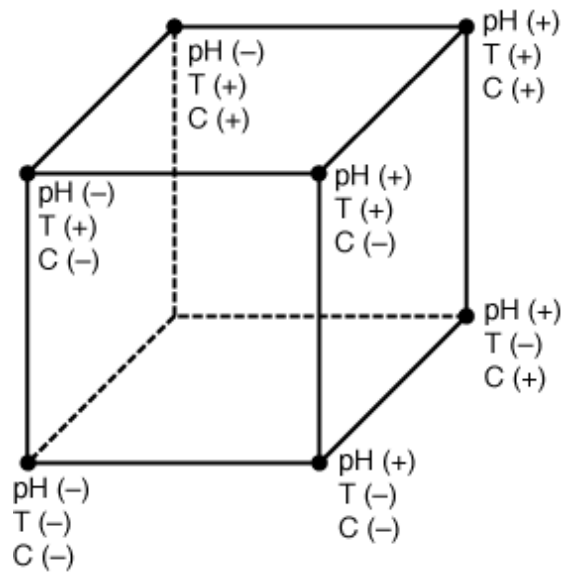
TABLE A. CONTROLLED VARIABLE SETTINGS

<i>Experiment No.</i>	<i>pH</i>	<i>Temperature</i>	<i>Concentration</i>
1	-	-	-
2	+	-	-
3	-	+	-
4	-	-	+
5	+	+	+
6	-	+	+
7	+	-	+
8	+	+	-

For example, let's design a series of experiments to determine the effect of pH and temperature on the rate of the enzyme-catalyzed decomposition of urea.



Enzyme degradation is believed to occur at temperatures above 50°C and pH values above 9.5 and below 3.0. The rate of reaction is negligible at temperatures below 6°C. For a urea concentration below 0.001 M, the reaction will not proceed at a measurable rate and the rate appears to be



Figure

5-13

Placement (high/low) of controlled variables. independent of concentration above 0.1 M. Consequently, the following high/low values of the parameters were chosen:

A (-) pH 4	(+) pH 8
B (-) 10°C	(+) 40°C
C (-) 0.005 M	(+) 0.1 M

If there is no interaction among the variables (which may not be known beforehand), experiments 1-4 will yield all the necessary information (Table B). By no interaction it is meant that each of the variables affects the outcome of the experiment independently and there is no synergistic effect of a combined interaction. When there is no interaction, the effect of the pH variable changing from a high value to a low value is always the same, regardless of whether or not the values of temperature and urease concentration are high or low. Experiments 1-4 explore the effect of raising each variable, in turn, from its low level to its high level. In this type of situation, the minimum number of experiments that must be run is the number of independent variables plus one ($3 + 1 = 4$), and we can predict the results of the other experiments (5 through 8) by combinations of the appropriate responses. However, if there is interaction among the variables, it will not be detected from the tables. For example, if there is *interaction* between temperature and pH, the higher temperature may cause the pH to have a stronger effect on the reaction than at low temperatures. Then all eight experiments would be required to detect an interaction such as this one.

TABLE B: FIRST FOUR EXPERIMENTS

<i>Experiment No.</i>	<i>pH</i>	<i>Temperature</i>	<i>Concentration</i>	<i>Comments</i>
1	-	-	-	Base case
2	+	-	-	Reveals effect of high pH
3	-	+	-	Reveals effect of high temperature
4	-	-	+	Reveals effect of high

				concentration
--	--	--	--	---------------

A full factorial design (all eight experiments in this case) is also useful for developing a model to predict the outcome of experiments whose independent variables can change continuously (i.e., they can assume a continuous range of values and not just two discrete values). Two levels (at least) of each of the variables are examined and the results can be interpreted in the form of a model to predict the outcome of future experiments. Deming⁵ discusses this method of statistically designing experiments.

R5.3.8 Performing the Experiment: How Many Times?

If there is some error associated with measuring the outcome of an experiment, we must consider repeating some of the trials to be sure we have accurate information. But how much data is enough? The answer to this question depends on how precise (reproducible) the experiments are and on how small a change in the outcome or result of an experiment we wish to detect. Obviously, the less precise the measurements (i.e., the more error that is present) and the smaller the change we are interested in, the more data we must collect and average to be confident in our result. Averaging several runs under the same conditions is the best way to deal with such a situation to ensure reliable results. The required number of times that each run should be repeated prior to averaging can easily be calculated using a statistical procedure discussed by Hendrix.

How good are the measurements? What modifications, if any, of the existing equipment are necessary to improve the accuracy or precision of the measurements or to better achieve the overall experimental objectives? In the experiments, we might find that the urease reaction results are inconclusive with respect to pH in the high-temperature range, and additional runs will be necessary. Is there software available to perform least-squares analysis (see Appendix A.4), set confidence limits, or other statistical analyses? Is there any mathematical model or theory available that suggests how the data might be plotted or correlated? What generalizations can be made from the data? Should other experiments be run to extend the data into different regions? Has an error analysis been performed, and sources of error listed and discussed in relation to how they affect the final result (i.e., by what magnitude and in what direction?). Finally, *have all experimental objectives been satisfied?*

Benefits and Limitations Of Experimental Research

There are many benefits and limitations to experimental research and many of them have been alluded to in previous modules in this series. Following is more detailed discussion regarding both the advantages and the limitations or disadvantages.

Benefits and Advantages:

- Experimental research is the most appropriate way for drawing causal conclusions, regarding interventions or treatments and establishing whether or not one or more factors causes a change in an outcome. This is largely due to the emphasis in controlling extraneous variables. If other variables are controlled, the researcher can say with confidence that manipulation independent variable caused a changed in the dependent variable.
- It is a basic, straightforward, efficient type of research that can be applied across a variety of disciplines.
- Experimental research designs are repeatable and therefore, results can be checked and verified.
- Due to the controlled environment of experimental research, better results are often achieved.
- In the case of laboratory research, conditions not found in a natural setting can be created in an experimental setting that allows for greater control of extraneous variables. Conditions that may take longer to occur in a natural environment may occur more quickly in an experimental setting.
- There are many variations of experimental research and the researcher can tailor the experiment while still maintaining the validity of the design.

Limitations and Disadvantages:

- Experimental research can create artificial situations that do not always represent real-life situations. This is largely due to fact that all other variables are tightly controlled which may not create a fully realistic situation.
- Because the situations are very controlled and do not often represent real life, the reactions of the test subjects may not be true indicators of their behaviours in a non-experimental environment.
- Human error also plays a key role in the validity of the project as discussed in previous modules.
- It may not be really possible to control all extraneous variables. The health, mood, and life experiences of the test subjects may influence their reactions and those variables may not even be known to the researcher.
- The research must adhere to ethical standards in order to be valid. These will be discussed in the next module of this series.
- Experimental research designs help to ensure internal validity but sometimes at the expense of external validity. When this happens, the results may not be generalizable to the larger population.
- If an experimental study is conducted in its natural environment, such as a hospital or community, it may not be possible to control the extraneous variables.
- Experimental research is a powerful tool for determining or verifying causation, but it typically cannot specify “why” the outcome occurred.

Unit -3
THE SOURCES AND COLLECTION OF DATA

Census and Sample Survey

All items in any field of inquiry constitute a 'Universe' or 'Population.' A complete enumeration of all items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry, when all items are covered, no element of chance is left and highest accuracy is obtained. But in practice this may not be true. Even the slightest element of bias in such an inquiry will get larger and larger as the number of observation increases. Moreover, there is no way of checking the element of bias or its extent except through a resurvey or use of sample checks. Besides, this type of inquiry involves a great deal of time, money and energy. Therefore, when the field of inquiry is large, this method becomes difficult to adopt because of the resources involved. At times, this method is practically beyond the reach of ordinary researchers. Perhaps, government is the only institution which can get the complete enumeration carried out. Even the government adopts this in very rare cases such as population census conducted once in a decade. Further, many a time it is not possible to examine every item in the population, and sometimes it is possible to obtain sufficiently accurate results by studying only a part of total population. In such cases there is no utility of census surveys.

However, it needs to be emphasised that when the universe is a small one, it is no use resorting to a sample survey. When field studies are undertaken in practical life, considerations of time and cost almost invariably lead to a selection of respondents i.e., selection of only a few items. The respondents selected should be as representative of the total population as possible in order to produce a miniature cross-section. The selected respondents constitute what is technically called a 'sample' and the selection process is called 'sampling technique.' The survey so conducted is known as 'sample survey'. Algebraically, let the population size be N and if a part of size n (which is $< N$) of this population is selected according to some rule for studying some characteristic of the population, the group consisting of these n units is known as 'sample'. Researcher must prepare a sample design for his study i.e., he must plan how a sample should be selected and of what size such a sample would be.

Implications Of a Sample Design

A sample design is a definite plan for obtaining a sample from a given population. It refers to the technique or the procedure the researcher would adopt in selecting items for the sample. Sample design is determined before data are collected. There are many sample designs from which a researcher can choose. Some designs are relatively more precise and easier to apply than others. Researcher must select/prepare a sample design which should be reliable and appropriate for his research study.

Steps In Sample Design

While developing a sampling design, the researcher must pay attention to the following points:

(i) **Type of universe:** The first step in developing any sample design is to clearly define the set of objects, technically called the Universe, to be studied. The universe can be finite or infinite. In finite universe the number of items is certain, but in case of an infinite universe the number of items is infinite, i.e., we cannot have any idea about the total number of items. The population of a city, the number of workers in a factory and the like are examples of finite universes, whereas the number of stars in the sky, listeners of a specific radio programme, throwing of a dice etc. are examples of infinite universes.

(ii) **Sampling unit:** A decision has to be taken concerning a sampling unit before selecting sample. Sampling unit may be a geographical one such as state, district, village, etc., or a construction unit such as house, flat, etc., or it may be a social unit such as family, club, school, etc., or it may be an individual. The researcher will have to decide one or more of such units that he has to select for his study.

(iii) **Source list:** It is also known as 'sampling frame' from which sample is to be drawn. It contains the names of all items of a universe (in case of finite universe only). If source list is not available, researcher has to prepare it. Such a list should be comprehensive, correct, reliable and appropriate. It is extremely important for the source list to be as representative of the population as possible.

(iv) **Size of sample:** This refers to the number of items to be selected from the universe to constitute a sample. This is a major problem before a researcher. The size of sample should neither be excessively large, nor too small. It should be optimum. An optimum sample is one which fulfills the requirements of efficiency, representativeness, reliability and flexibility. While deciding the size of sample, researcher must determine the desired precision as also an acceptable confidence level for the estimate. The size of population variance needs

to be considered as in case of larger variance usually a bigger sample is needed. The size of population must be kept in view for this also limits the sample size. The parameters of interest in a research study must be kept in view, while deciding the size of the sample. Costs too dictate the size of sample that we can draw. As such, budgetary constraint must invariably be taken into consideration when we decide the sample size.

(v) **Parameters of interest:** In determining the sample design, one must consider the question of the specific population parameters which are of interest. For instance, we may be interested in estimating the proportion of persons with some characteristic in the population, or we may be interested in knowing some average or the other measure concerning the population. There may also be important sub-groups in the population about whom we would like to make estimates. All this has a strong impact upon the sample design we would accept.

(vi) **Budgetary constraint:** Cost considerations, from practical point of view, have a major impact upon decisions relating to not only the size of the sample but also to the type of sample. This fact can even lead to the use of a non-probability sample.

(vii) **Sampling procedure:** Finally, the researcher must decide the type of sample he will use i.e., he must decide about the technique to be used in selecting the items for the sample. In fact, this technique or procedure stands for the sample design itself. There are several sample designs (explained in the pages that follow) out of which the researcher must choose one for his study. Obviously, he must select that design which, for a given sample size and for a given cost, has a smaller sampling error.

Criteria Of Selecting A Sampling Procedure

In this context one must remember that two costs are involved in a sampling analysis viz., the cost of collecting the data and the cost of an incorrect inference resulting from the data. Researcher must keep in view the two causes of incorrect inferences viz., systematic bias and sampling error. A *systematic bias* results from errors in the sampling procedures, and it cannot be reduced or eliminated by increasing the sample size. At best the causes responsible for these errors can be detected and corrected. Usually a systematic bias is the result of one or more of the following factors:

1. Inappropriate sampling frame: If the sampling frame is inappropriate i.e., a biased representation of the universe, it will result in a systematic bias.

2. Defective measuring device: If the measuring device is constantly in error, it will result in systematic bias. In survey work, systematic bias can result if the questionnaire or the interviewer is biased. Similarly, if the physical measuring device is defective there will be systematic bias in the data collected through such a measuring device.

3. Non-respondents: If we are unable to sample all the individuals initially included in the sample, there may arise a systematic bias. The reason is that in such a situation the likelihood of establishing contact or receiving a response from an individual is often correlated with the measure of what is to be estimated.

4. Indeterminacy principle: Sometimes we find that individuals act differently when kept under observation than what they do when kept in non-observed situations. For instance, if workers are aware that somebody is observing them in course of a work study on the basis of which the average length of time to complete a task will be determined and accordingly the quota will be set for piece work, they generally tend to work slowly in comparison to the speed with which they work if kept unobserved. Thus, the indeterminacy principle may also be a cause of a systematic bias.

5. Natural bias in the reporting of data: Natural bias of respondents in the reporting of data is often the cause of a systematic bias in many inquiries. There is usually a downward bias in the income data collected by government taxation department, whereas we find an upward bias in the income data collected by some social organisation. People in general understate their incomes if asked about it for tax purposes, but they overstate the same if asked for social status or their affluence. Generally in psychological surveys, people tend to give what they think is the 'correct' answer rather than revealing their true feelings.

Sampling errors

Sampling errors are the random variations in the sample estimates around the true population parameters. Since they occur randomly and are equally likely to be in either direction, their nature happens to be of compensatory type and the expected value of such errors happens to be equal to zero. Sampling error decreases with the increase in the size of the sample, and it happens to be of a smaller magnitude in case of homogeneous population.

Sampling error can be measured for a given sample design and size. The measurement of sampling error is usually called the 'precision of the sampling plan'. If we increase the sample size, the precision can be improved. But increasing the size of the sample has its own limitations viz., a large sized sample increases the cost of collecting data and also enhances the systematic bias. Thus the effective way to increase precision is usually to select a better sampling design which has a smaller sampling error for a given sample size at a given cost. In practice, however, people prefer a less precise design because it is easier to adopt the same and also because of the fact that systematic bias can be controlled in a better way in such a design.

In brief, *while selecting a sampling procedure, researcher must ensure that the procedure causes a relatively small sampling error and helps to control the systematic bias in a better way.*

Characteristics Of A Good Sample Design

From what has been stated above, we can list down the characteristics of a good sample design as under:

- (a) Sample design must result in a truly representative sample.
- (b) Sample design must be such which results in a small sampling error.
- (c) Sample design must be viable in the context of funds available for the research study.
- (d) Sample design must be such so that systematic bias can be controlled in a better way.
- (e) Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

Different Types Of Sample Designs

There are different types of sample designs based on two factors viz., the representation basis and the element selection technique. On the representation basis, the sample may be probability sampling or it may be non-probability sampling. Probability sampling is based on the concept of random selection, whereas non-probability sampling is 'non-random' sampling. On element selection basis, the sample may be either unrestricted or restricted. When each sample element is drawn individually from the population at large, then the sample so drawn is known as 'unrestricted sample', whereas all other forms of sampling are covered under the term 'restricted sampling'. The following chart exhibits the sample designs as explained above. Thus, sample designs are basically of two types viz., non-probability sampling and probability sampling. We take up these two designs separately.

Non-probability sampling: Non-probability sampling is that sampling procedure which does not afford any basis for estimating the probability that each item in the population has of being included in the sample. Non-probability sampling is also known by different names such as deliberate sampling, purposive sampling and judgement sampling. In this type of sampling, items for the sample are selected deliberately by the researcher; his choice concerning the items remains supreme. In other words, under non-probability sampling the organisers of the inquiry purposively choose the particular units of the universe for constituting a sample on the basis that the small mass that they so select out of a huge one will be typical or representative of the whole. For instance, if economic conditions of people living in a state are to be studied, a few towns and villages may be purposively selected for intensive study on the principle that they can be representative of the entire state. Thus, the judgement of the organisers of the study plays an important part in this sampling design. In such a design, personal element has a great chance of entering into the selection of the sample. The investigator may select a sample which shall yield results favourable to his point of view and if that happens, the entire inquiry may get vitiated. Thus, there is always the danger of bias entering into this type of sampling technique. But in the investigators are impartial, work without bias and have the necessary experience so as to take sound judgement, the results obtained from an analysis of deliberately selected sample may be tolerably reliable. However, in such a sampling, there is no assurance that every element has some specifiable chance of being included. Sampling error in this type of sampling cannot be estimated and the element of bias, great or small, is always there. As such this sampling design is rarely adopted in large inquiries of importance. However, in small inquiries and researches by individuals, this design may be adopted because of the relative advantage of time and money inherent in this method of sampling. *Quota sampling* is also an example of non-probability sampling. Under quota sampling the interviewers are simply given quotas to be filled from the different strata, with some restrictions on how they are to be filled. In other words, the actual selection of the items for the sample is left to the interviewer's discretion. This type of sampling is very convenient and is relatively inexpensive. But the samples so selected certainly do not possess the characteristic of random samples. Quota

samples are essentially judgement samples and inferences drawn on their basis are not amenable to statistical treatment in a formal way. sampling'. Under this sampling design, every item of the universe has an equal chance of inclusion in the sample. It is, so to say, a lottery method in which individual units are picked up from the whole group not deliberately but by some mechanical process. Here it is blind chance alone that determines whether one item or the other is selected. The results obtained from probability or random sampling can be assured in terms of probability i.e., we can measure the errors of estimation or the significance of results obtained from a random sample, and this fact brings out the superiority of random sampling design over the deliberate sampling design. Random sampling ensures the law of Statistical Regularity which states that if on an average the sample chosen is a random one, the sample will have the same composition and characteristics as the universe. This is the reason why random sampling is considered as the best technique of selecting a representative sample.

Random sampling from a finite population refers to that method of sample selection which gives each possible sample combination an equal probability of being picked up and each item in the entire population to have an equal chance of being included in the sample. This applies to sampling without replacement i.e., once an item is selected for the sample, it cannot appear in the sample again (Sampling with replacement is used less frequently in which procedure the element selected for the sample is returned to the population before the next element is selected. In such a situation the same element could appear twice in the same sample before the second element is chosen). In brief, the implications of random sampling (or simple random sampling) are:

(a) It gives each element in the population an equal probability of getting into the sample; and all choices are independent of one another.

(b) It gives each possible sample combination an equal probability of being chosen. Keeping this in view we can define a simple random sample (or simply a random sample) from a finite population as a sample which is chosen in such a way that each of the NC_n possible samples has the same probability, $1/NC_n$, of being selected. To make it more clear we take a certain finite population consisting of six elements (say a, b, c, d, e, f) i.e., $N = 6$. Suppose that we want to take a sample of size $n = 3$ from it. Then there are $6C_3 = 20$ possible distinct samples of the required size, and they consist of the elements $abc, abd, abe, abf, acd, ace, acf, ade, adf, aef, bcd, bce, bcf, bde, bdf, bef, cde, cdf, cef,$ and def . If we choose one of these samples in such a way that each has the probability $1/20$ of being chosen, we will then call this a random sample.

How To Select A Random Sample ?

With regard to the question of how to take a random sample in actual practice, we could, in simple cases like the one above, write each of the possible samples on a slip of paper, mix these slips thoroughly in a container and then draw as a lottery either blindfolded or by rotating a drum or by any other similar device. Such a procedure is obviously impractical, if not altogether impossible in complex problems of sampling. In fact, the practical utility of such a method is very much limited. Fortunately, we can take a random sample in a relatively easier way without taking the trouble of enlisting all possible samples on paper-slips as explained above. Instead of this, we can write the name of each element of a finite population on a slip of paper, put the slips of paper so prepared into a box or a bag and mix them thoroughly and then draw (without looking) the required number of slips for the sample one after the other without replacement. In doing so we must make sure that in successive drawings each of the remaining elements of the population has the same chance of being selected. This procedure will also result in the same probability for each possible sample. We can verify this by taking the above example. Since we have a finite population of 6 elements and we want to select a sample of size 3, the probability of drawing any one element for our sample in the first draw is $3/6$, the probability of drawing one more element in the second draw is $2/5$, (the first element drawn is not replaced) and similarly the probability of drawing one more element in the third draw is $1/4$. Since these draws are independent, the joint probability of the three elements which constitute our sample is the product of their individual probabilities and this works out to $3/6 \times 2/5 \times 1/4 = 1/20$.

This verifies our earlier calculation.

Even this relatively easy method of obtaining a random sample can be simplified in actual practice by the use of random number tables. Various statisticians like Tippett, Yates, Fisher have prepared tables of random numbers which can be used for selecting a random sample. Generally, Tippett's random number tables are used for the purpose. Tippett gave 10400 four figure numbers. He selected 41600 digits from the census reports and combined them into fours to give his random numbers which may be used to obtain a random sample.

We can illustrate the procedure by an example. First of all we reproduce the first thirty sets of Tippett's numbers

2952 6641 3992 9792 7979 5911
3170 5624 4167 9525 1545 1396
7203 5356 1300 2693 2370 7483
3408 2769 3563 6107 6913 7691
0560 5246 1112 9025 6008 8126

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

The units bearing the above serial numbers would then constitute our required random sample. One may note that it is easy to draw random samples from finite populations with the aid of random number tables only when lists are available and items are readily numbered. But in some situations it is often impossible to proceed in the way we have narrated above. For example, if we want to estimate the mean height of trees in a forest, it would not be possible to number the trees, and choose random numbers to select a random sample. In such situations what we should do is to select some trees for the sample haphazardly without aim or purpose, and should treat the sample as a random sample for study purposes.

Random Sample From An Infinite Universe

So far we have talked about random sampling, keeping in view only the finite populations. But what about random sampling in context of infinite populations? It is relatively difficult to explain the concept of random sample from an infinite population. However, a few examples will show the basic characteristic of such a sample. Suppose we consider the 20 throws of a fair dice as a sample from the hypothetically infinite population which consists of the results of all possible throws of the dice. If the probability of getting a particular number, say 1, is the same for each throw and the 20 throws are all independent, then we say that the sample is random. Similarly, it would be said to be sampling from an infinite population if we sample with replacement from a finite population and our sample would be considered as a random sample if in each draw all elements of the population have the same probability of being selected and successive draws happen to be independent. In brief, one can say that the selection of each item in a random sample from an infinite population is controlled by the same probabilities and that successive selections are independent of one another.

Complex Random Sampling Designs

Probability sampling under restricted sampling techniques, as stated above, may result in complex random sampling designs. Such designs may as well be called 'mixed sampling designs' for many of such designs may represent a combination of probability and non-probability sampling procedures in selecting a sample. Some of the popular complex random sampling designs are as follows:

(i) Systematic sampling: In some instances, the most practical way of sampling is to select every i th item on a list. Sampling of this type is known as systematic sampling. An element of randomness is introduced into this kind of sampling by using random numbers to pick up the unit with which to start. For instance, if a 4 per cent sample is desired, the first item would be selected randomly from the first twenty-five and thereafter every 25th item would automatically be included in the sample. Thus, in systematic sampling only the first unit is selected randomly and the remaining units of the sample are selected at fixed intervals. Although a systematic sample is not a random sample in the strict sense of the term, but it is often considered reasonable to treat systematic sample as if it were a random sample. Systematic sampling has certain plus points. It can be taken as an improvement over a simple random sample in as much as the systematic sample is spread more evenly over the entire population.

It is an easier and less costlier method of sampling and can be conveniently used even in case of large populations. But there are certain dangers too in using this type of sampling. If there is a hidden periodicity in the population, systematic sampling will prove to be an inefficient method of sampling.

For instance, every 25th item produced by a certain production process is defective. If we are to select a 4% sample of the items of this process in a systematic manner, we would either get all defective items or all good items in our sample depending upon the random starting position. If all elements of the universe are ordered in

a manner representative of the total population, i.e., the population list is in random order, systematic sampling is considered equivalent to random sampling. But if this is not so, then the results of such sampling may, at times, not be very reliable. In practice, systematic sampling is used when lists of population are available and they are of considerable length.

(ii) Stratified sampling: If a population from which a sample is to be drawn does not constitute a homogeneous group, stratified sampling technique is generally applied in order to obtain a representative sample. Under stratified sampling the population is divided into several sub-populations that are individually more homogeneous than the total population (the different sub-populations are called 'strata') and then we select items from each stratum to constitute a sample. Since each stratum is more homogeneous than the total population, we are able to get more precise estimates for each stratum and by estimating more accurately each of the component parts, we get a better estimate of the whole. In brief, stratified sampling results in more reliable and detailed information.

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

(a) How to form strata?

(b) How should items be selected from each stratum?

(c) How many items be selected from each stratum or how to allocate the sample size of each stratum?

Regarding the first question, we can say that the strata be formed on the basis of common characteristic(s) of the items to be put in each stratum. This means that various strata be formed in such a way as to ensure elements being most homogeneous within each stratum and most heterogeneous between the different strata. Thus, strata are purposively formed and are usually based on past experience and personal judgement of the researcher. One should always remember that careful consideration of the relationship between the characteristics of the population and the characteristics to be estimated are normally used to define the strata. At times, pilot study may be conducted for determining a more appropriate and efficient stratification plan. We can do so by taking small samples of equal size from each of the proposed strata and then examining the variances within and among the possible stratifications, we can decide an appropriate stratification plan for our inquiry. In respect of the second question, we can say that the usual method, for selection of items for the sample from each stratum, resorted to is that of simple random sampling. Systematic sampling can be used if it is considered more appropriate in certain situations.

Regarding the third question, we usually follow the method of proportional allocation under which the sizes of the samples from the different strata are kept proportional to the sizes of the strata. That is, if P_i represents the proportion of population included in stratum i , and n represents the total sample size, the number of elements selected from stratum i is $n \cdot P_i$. To illustrate it, let us suppose that we want a sample of size $n = 30$ to be drawn from a population of size $N = 8000$ which is divided into three strata of size $N_1 = 4000$, $N_2 = 2400$ and $N_3 = 1600$. Adopting proportional allocation, we shall get the sample sizes as under for the different strata:

For strata with $N_1 = 4000$, we have $P_1 = 4000/8000$

and hence $n_1 = n \cdot P_1 = 30 (4000/8000) = 15$

Similarly, for strata with $N_2 = 2400$, we have

$n_2 = n \cdot P_2 = 30 (2400/8000) = 9$, and

for strata with $N_3 = 1600$, we have

$n_3 = n \cdot P_3 = 30 (1600/8000) = 6$.

Thus, using proportional allocation, the sample sizes for different strata are 15, 9 and 6 respectively which is in proportion to the sizes of the strata viz., 4000 : 2400 : 1600. Proportional allocation is considered most efficient and an optimal design when the cost of selecting an item is equal for each stratum, there is no difference in within-stratum variances, and the purpose of sampling happens to be to estimate the population value of some characteristic. But in case the purpose happens to be to compare the differences among the strata, then equal sample selection from each stratum would be more efficient even if the strata differ in sizes. In cases where strata differ not only in size but also in variability and it is considered reasonable to take larger samples from the more variable strata and smaller samples from the less variable strata, we can then account for both (differences in stratum size and differences in stratum variability) by using disproportionate sampling design by requiring:

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

- (a) How to form strata?
- (b) How should items be selected from each stratum?
- (c) How many items be selected from each stratum or how to allocate the sample size of each stratum?

Regarding the first question, we can say that the strata be formed on the basis of common characteristic(s) of the items to be put in each stratum. This means that various strata be formed in such a way as to ensure elements being most homogeneous within each stratum and most heterogeneous between the different strata. Thus, strata are purposively formed and are usually based on past experience and personal judgement of the researcher. One should always remember that careful consideration of the relationship between the characteristics of the population and the characteristics to be estimated are normally used to define the strata. At times, pilot study may be conducted for determining a more appropriate and efficient stratification plan. We can do so by taking small samples of equal size from each of the proposed strata and then examining the variances within and among the possible stratifications, we can decide an appropriate stratification plan for our inquiry.

In respect of the second question, we can say that the usual method, for selection of items for the sample from each stratum, resorted to is that of simple random sampling. Systematic sampling can be used if it is considered more appropriate in certain situations.

Regarding the third question, we usually follow the method of proportional allocation under which the sizes of the samples from the different strata are kept proportional to the sizes of the strata. That is, if P_i represents the proportion of population included in stratum i , and n represents the total sample size, the number of elements selected from stratum i is $n \cdot P_i$. To illustrate it, let us suppose that we want a sample of size $n = 30$ to be drawn from a population of size $N = 8000$ which is divided into three strata of size $N_1 = 4000$, $N_2 = 2400$ and $N_3 = 1600$. Adopting proportional allocation, we shall get the sample sizes as under for the different strata:

For strata with $N_1 = 4000$, we have $P_1 = 4000/8000$

and hence $n_1 = n \cdot P_1 = 30 (4000/8000) = 15$

Similarly, for strata with $N_2 = 2400$, we have

$n_2 = n \cdot P_2 = 30 (2400/8000) = 9$, and

for strata with $N_3 = 1600$, we have

$n_3 = n \cdot P_3 = 30 (1600/8000) = 6$.

Thus, using proportional allocation, the sample sizes for different strata are 15, 9 and 6 respectively which is in proportion to the sizes of the strata viz., 4000 : 2400 : 1600.

Proportional allocation is considered most efficient and an optimal design when the cost of selecting an item is equal for each stratum, there is no difference in within-stratum variances, and the purpose of sampling happens to be to estimate the population value of some characteristic. But in case the purpose happens to be to compare the differences among the strata, then equal sample selection from each stratum would be more efficient even if the strata differ in sizes. In cases where strata differ not only in size but also in variability and it is considered reasonable to take larger samples from the more variable strata and smaller samples from the less variable strata, we can then account for both (differences in stratum size and differences in stratum variability) by using disproportionate sampling design by requiring.

Data collection

Meaning and importance of data

Data and facts and other relevant materials, past and present, serving as bases for study and analyses. Data are the facts and figures collected for statistical investigation.

There are two types of data:

1. Primary data,
2. Secondary data (desk research)

Method of Collecting primary data

- a. Observation method
- b. Interview method
- c. Local correspondences
- d. Questionnaire and schedule method

Sources of secondary data

1. Personal sources: auto-biography, life history, diaries, letters, memoirs
2. Public source: I. published source – books, journals, reports, newspapers etc.
II. Unpublished source

Tool For Data Collection

1. Inquiry forms:
 - a) Schedule
 - b) Questionnaire
 - c) checklist
 - d) rating scale
 - e) Score board
2. Observation
3. Interview
4. Socio-metric techniques
5. Psychological test

Types Of Data

Quantitative data:

It applies various scales of measurement. The experiences of people are fit into standardized responses to which numerical values are attached.

Qualitative data:

They are verbal or other symbolic materials. The responses to open ended questions of a questionnaire or a schedule, first hand information from people about their experiences, ideas, beliefs, etc. and selected from content or accepts from documents, case history, personal diaries, and letters are other examples of qualitative data.

Importance Of Data

- It serve as the bases or raw materials for analysis
- It provide correct answers for analysis
- It serves the basis for testing the hypothesis
- It helps for constructing measurement scales and tables
- It determine the quality of the findings of the study

Sources Of Data

A significant and distinctive stage of research in any science is the collection of necessary information to prove their hypothesis. For this purpose, the researcher should look to diverse sources which provide the necessary information. The sources of information are generally classified as primary and secondary, while P.V. Young feels that sources of data can be divided documentary and field sources. There are two sources of data:

1. Primary sources (primary data)
2. Secondary sources-desk research (secondary data)

The primary data are those which are collected afresh and for the first time, and thus happen to be original in character or information collected or generated by the researcher for the purpose of the project immediately at hand.

Advantages of primary data

- Primary data are the first –hand account of the situation.
- There is a greater scope for reliability of the information.
- Primary data are the logical starting point for research in several disciplines.
- Primary data are the only source to understand one’s opinions, personal qualities, attitudes, etc.

Secondary Sources Of Data

- The secondary data are those which have already been collected by someone else and which have already been passed through the statistical process.
- Secondary data refer to the information that have been collected by someone other than researcher for purposes other than those involved in the research project at hand.

- Books, journals, manuscripts, diaries, letters, etc., all become secondary sources of data as they are written or compiled for a separate purpose.
- As a matter of fact, the difference between primary and secondary sources is a matter of relativity. Data which are primary in the hands of one, becomes secondary in the hands of the other.

Advantages of secondary data.

- a. It saves time, energy and money
- b. It provide information that may not be secured by the individual researcher.

Methods/techniques of data collection

□ This is a very important aspect of research design and the ability to achieve the research aims and answer the research question depends on the effectiveness of data collection.

Method of collecting data

1. Observation method
2. Interview method
3. Survey method
4. Experimentation
5. Panel method
6. Projective technique
7. Socio metry
8. Content analysis

Schedule

Schedule is the tool or instrument used to collect data from the respondents while interview is conducted. **Schedule** contains questions, statements (on which opinions are elicited) and blank spaces/tables for filling up the respondents. The **features of schedules** are :

- The schedule is presented by the interviewer. The questions are asked and the answers are noted down by him.
- The list of questions is a more formal document, it need not be attractive.
- The schedule can be used in a very narrow sphere of social research.

The main **purposes of schedule** are three fold :

- To provide a standardized tool for observation or interview in order to attain objectivity,
- To act as memory tickler i.e., the schedule keeps the memory of the interviewer/ observer refreshed and keeps him reminded of the different aspects that are to be particularly observed, and
- To facilitate the work of tabulation and analysis.

Types of Schedule

There are several kinds of schedule. **Rating Schedules** is a schedule used to obtain opinions, preferences etc, respondents over statements on the phenomenon studied. The schedule consists of positive and negative statements of opinion on the phenomenon. **Documents Schedules** are used to collect data/information from recorded evidences and/or case histories. Here the blanks, functional issues related blanks and the like to be filled up from records and documents are present. **Survey Schedules** are like questionnaires. **Observation Schedules** are schedules used when observational method of data collection is used. These could be structured or unstructured interview schedules are used for collecting data when interview method of communication with the respondents is used.

Essentials of a Good Schedule

A good schedule must have the following features

- **Content:** Should cover questions or statements relating to all significant aspects of the study.
- **Dissectional:** Should look into the problem analytically, dissecting every, major and significant components of the problem.
- **Context:** Should suit the context in which it is applied. Different types of studies need different schedules.
- **Criterion:** Should use sound logic in classifying respondents based opinions expressed.
- **Construction:** Should be constructed in such a way that questions statements progress gradually and in order. Better it is sub-divided into parts, each part dealing with a certain sub topic of the issue studied. For each objective, a separate part may be devoted.
- **Language:** Should be linguistically superbly designed. Clear and straight forward language be used.
- **Reliable:** Should be reliable such that same results are obtained whenever the schedule is used when everything else remains same.
- **Mechanical Aspects:** Paper used, margin space given, spacing, printing, size of letters, etc. should be normal.
- **Size:** Should not too length nor too short. Should give fair coverage to the topic.
- **Qualities to be Avoided:** Long, complex, presumptuous, personal, embarrassing, hypothetical issues, morality oriented, upsetting type and necessary questions must be avoided.

To sum up, accurate information and accurate response are the two essential conditions of a good schedule. Accurate communication is effected by proper wording of questions so as to produce desired sense without any ambiguity. Accurate response is said to have been achieved when replies contain the information sought for. The response is achieved by stimulating the respondents to fill the schedule. Besides, the physical structure of the schedule should be attractive; the questions asked or information sought should be adequate and relevant to the enquiry, so that final generalization may be based upon it. The information sought should not only be valid, it should also be capable of being tabulated and if possible being subjected to statistical analysis.

Procedure for Formulating a Schedule

- Study the different aspects of the problem. The problem under study should first of all be split up into various aspects. The determination of these aspects will depend upon clear understanding of the problem under study.
- Sub-divide the problem to get necessary information. Each aspect has again to be broken up into a number of sub-parts. These sub-parts should be quite exhaustive to give a full and complete picture of the aspect under study.
- Class questions. Care should be taken to see that the questions convey the exact sense. Respondents will be willing to supply information without any hesitation, bias or distortion of facts, if questions are exact and clear.
- Serialization of Questions. In order to obtain well-organised information, it is necessary that the questions should be presented to the respondents in a well-ordered serial. It has been experienced to various field studies that the change in the order of questions affects the answers adversely.
- Testing the validity of schedule Whatever may be the degree of precaution taken, some slips are based to be left out and these can be located when the schedule is put into a reliability and validity test.
- Division. The schedule be divided into adequate number of divisions. Introductory part, instructional part, issues related parts, etc. are certain parts by which the schedule is divided into parts.
- Appropriate form of questions. Use appropriate forms of questions at appropriate places. Open ended, close ended, pictorial, Yes or No (Questions), multiple choice questions, etc. can be used.

Process of data collection through schedules:

Data collection under this method proceeds in a systematic manner. The investigators or enumerators proceed to the field with the schedules and administer them on the sample, selected by them. They go on asking the questions incorporated in the schedule and note down the responses of the respondents. The quality of the data depends on the people who go to the field and collect the data. Investigators or enumerator should be trained sufficiently. They should be intelligent and must possess the capacity of cross examination in order to find out the facts. Above all they should be honest, sincere, hardworking and should have patience and

perseverance since the quality of data affects the validity of the conclusion, every care should be taken to collect as accurately as possible. The procedure of constructing a schedule, the analysis and interpretation of data gathered through it are not very different from those of a questionnaire.

1. Questionnaire

Questionnaire is the most common instrument of data collection. A questionnaire consists of number of questions printed or typed in a definite order on a form or set of forms. In a questionnaire respondents read the questions, interpret what is expected then write down the answers.

Mailed Questionnaire

A questionnaire consists of a schedule of a questions sent by mail to the persons on a list or in a sample survey. Questionnaire can be further classified as follows:

- Structured questionnaire
- Non-structured questionnaire
- Disguised questionnaire
- Non-disguised questionnaire
- Structured disguised questionnaire
- Structured non-disguised questionnaire

The following are some of the important merits of the use of questionnaire method:

1. **Low cost:** The primary advantage of questionnaire method is that it is less expensive to administer. This is because questionnaires are most often mailed or handed over to a large number of interviewees simultaneously.
2. **Avoid bias:** The questionnaire method does not give scope to the investigator to manipulate the data or respondent. It is free from the bias of the interviewer and the answers are in own language of the respondents.
3. **Anonymity:** It offers greater anonymity. Since the investigator is interested in the answers to the questions, rather than the persons, the respondents may feel free to express their opinion without ambiguity or fear.
4. **Less pressure:** This method gives less pressure on the respondents for immediate responses. Sometimes. This may be necessary, if the respondent is required to report information which he needs to check up.
5. **Wide Coverage:** The questionnaire is the best method to reach far off places with very low cost.
6. **Dependable and reliable result:** Large samples can be made use of and thus the results can be made more dependable and reliable.

Formulation Of A Questionnaire

The success of a questionnaire depends upon the skills and insights with which the lists of questions are formulated along with the type of questions used. The following considerations in mind while formulating questionnaire.

Appeal

The appeal should be short, clear and direct establishing the genuineness of the research and its utility for all concerned.

Instruction For Filling Up The Questionnaire

The questionnaire must carry a list of instructions for filling it up and dispatching it.

Form Of The Questionnaire

The outlook and appearance of the questionnaire should be attractive. It must be printed in an appealing style on high quality paper.

Clarity Of Questions

The best method is to ensure that our questionnaire does not have any misleading or confusing questionnaire. It should first be tried on a selected group of individuals and suitable modifications should be made in questions in the light of the experience with the selected group.

Sequence Of Questions

The order of framing questions is also important. The sequence must be logical and arouse interest in the questions. The disorderly sequence of the questions disturbs the mind of the respondent and he may fail to answer the questions, adequately.

General Form

So far as the general form of a questionnaire is concerned, it can either be structured or unstructured questionnaire. Structured questionnaires are those questionnaires in which there are definite, concrete and pre-determined questions. The questionnaires which do not have definite, concrete and pre-determined questions are termed as unstructured questionnaires. In this type, the researcher/ interviewer is provided with a general guide on the type of information to be obtained.

Question Formulation And Wording

In general, all questions should meet the following standards –

- (i) should be easily understood;
- (ii) should be simple that is, should convey only one thought at a time;
- (iii) should be concrete and should conform as much as possible to the respondent's way of Thinking.

Construction Of Questionnaire

The process of drafting a questionnaire should have the following aspects:

Information required: The first step in the formulation of a questionnaire is to decide in advance what type of information is needed for the study.

Type of questionnaire to be used: Questionnaires may be categorized according to structure and directness. Structure refers to the degree to which the questions and responses are formal and standardized.

First draft: Before finalizing the questionnaire the researcher should prepare a preliminary draft of the questionnaire based on the aim and objectives of the study. The researcher can check, revise and prepare the final draft of the questionnaire based on pre-testing results of the questionnaire. Revising questions: the researcher should revise the questionnaire before the final edition. This will help the researcher to understand unforeseen problems related to wording, format, sequence etc.

Editing of questionnaire: Researcher must pay proper care and attention to the editing of the questionnaire. Simple words, which are familiar to all respondents should be employed. Words with ambiguous meanings, danger words, words with emotional connotations should be avoided.

Explaining the procedure for its use: the researcher should give or specify the procedure or instructions for its use. If the instrument is meant for mail survey, instructions regarding the mode of answering should be specified at the top of the first page. The anonymity should be assured.

Pre-testing of questionnaire: Pretest is a try out of the questionnaire to see how it works and whether changes are necessary before the start of full scale study. It provides a means of catching and solving unforeseen problems in the administration of the questionnaire, such as the wording, sequence of questions or even length. It may also indicate the need for additional questions or the elimination of others. Pretests are best done by personal interview even if the survey is to be handled by mail or telephone.

Final draft of the questionnaire: After pretesting the questionnaire if any mistakes are found, they must be corrected and the draft revised. In fact, there is no final word about the construction of a questionnaire. Improvements can always be thought of and the lacunae existing in the questionnaire would be coming to light as the work is progressing.

Various methods/ techniques for getting the response

There are several methods to get the responses through a questionnaire. They can be listed as:

1. Self addressed envelope
2. Incentives/schemes/prizes
3. Reminders
4. Using mediator
5. Sufficient interval
6. Retention
7. Good and interest creating questions

Characteristics Of Good Questions

Good questions should

1. be simple and clear
2. Be suitable to respondent's intelligence level
3. Be relevant
4. Be indirect
5. Have serial order

6. Have classifiability
7. Have verifiable questions
8. Have ambiguous questions
9. Be answerable in few words

Scaling

Definition: Scaling is the process of generating the continuum, a continuous sequence of values, upon which the measured objects are placed.

In Marketing Research, several scaling techniques are employed to study the relationship between the objects. The most commonly used techniques can be classified as:

Comparative Scales: In comparative scaling there is a **direct comparison of stimulus object**. For example, the respondent might be asked directly about his preference between the ink pen and gel pen. The comparative data can only be interpreted in relative terms and hence possess the ordinal or rank-order properties. This is the reason why the comparative scaling is also called as **nonmetric scaling**. The Comparative Scaling includes the following techniques:

- **Paired Comparison Scaling**
- **Rank Order Scaling**
- **Constant Sum Scaling**
- **Q-Sort Scaling**

2.

Non comparative Scales: The non comparative scale, also called as **monadic or metric scale** is a scale in which each object is scaled independently of the other objects in the stimulus set under study. Generally, the resulting data are assumed to be an interval and ratio scaled. For example, a respondent may be asked to rate their preference for the gel pen on a preference scale (1 = not at all preferred, 6 = greatly preferred). The noncomparative scale includes the following techniques:

- **Continuous Rating Scale**
- **Itemized Rating Scale**

Thus, the researcher can apply any of the scaling techniques to determine the characteristic of an individual and then locating him on the scale that best fits the defined characteristics.

Narrative interviews

Narratives are stories that are based on the unfolding of events or actions from the perspective of a participant's life experience. Narration is not new; in fact, it is one of the oldest human activities.^[19] In diabetes, patients tell their stories of illness and how they live with illness over time. The story of the individual patient (the case) is still, despite the reliance of medicine on scientific theory and generalizable results, an important mechanism for understanding how general scientific knowledge is applied.^[20] In recent years, more formal study of narratives in social health has become a method to represent and interpret an individual's lived experience.

Certain questions or concerns in diabetes social health research lend themselves to a narrative interview approach. It is an approach to use when little is known about the research topic, for instance how religious beliefs might affect diabetes self-management. The researcher could begin the narrative interview with a wide net, such as "what are your religious beliefs?" with one follow-up question of "how do you these beliefs impact your diabetes management?"

Researchers who conduct narrative research assume that a narrative of chronic illness, such as diabetes, is not simply the story of an illness, but the story of a life that is altered by illness.^[21] Researchers interested in

narrative and diabetes have conducted studies in a variety of topics ^{[22],[23],[24],[25]} , which are displayed below in [Table 1]. After asking the narrative question, researchers encourage participants to tell their illness stories.

The benefit of narrative interviews is that the participant guides the interview and may tell you information that could not have been predicted. The downside of these types of interviews is that they are often lengthy, lasting often 1 h. They are also more difficult to analysis than other types of interviews because it is an unstructured approach to interviewing that yields wide and deep themes.

The Purpose of Narrative Writing

Narration means the art of storytelling, and the purpose of narrative writing is to tell stories. Any time you tell a story to a friend or family member about an event or incident in your day, you engage in a form of narration. In addition, a narrative can be factual or fictional. A factual story is one that is based on, and tries to be faithful to, actual events as they unfolded in real life. A fictional story is a made-up, or imagined, story; the writer of a fictional story can create characters and events as he or she sees fit.

The big distinction between factual and fictional narratives is based on a writer’s purpose. The writers of factual stories try to recount events as they actually happened, but writers of fictional stories can depart from real people and events because the writers’ intents are not to retell a real-life event. Biographies and memoirs are examples of factual stories, whereas novels and short stories are examples of fictional stories.

Know Your Purpose

Because the line between fact and fiction can often blur, it is helpful to understand what your purpose is from the beginning. Is it important that you recount history, either your own or someone else’s? Or does your interest lie in reshaping the world in your own image—either how you would like to see it or how you imagine it could be? Your answers will go a long way in shaping the stories you tell.

Ultimately, whether the story is fact or fiction, narrative writing tries to relay a series of events in an emotionally engaging way. You want your audience to be moved by your story, which could mean through laughter, sympathy, fear, anger, and so on. The more clearly you tell your story, the more emotionally engaged your audience is likely to be.

The Structure of a Narrative Essay

Major narrative events are most often conveyed in chronological order, the order in which events unfold from first to last. Stories typically have a beginning, a middle, and an end, and these events are typically organized by time. Certain transitional words and phrases [aid in keeping the reader oriented in the sequencing of a story](#). Some of these phrases are listed here:

Chronological Transitional Words

after/afterward	as soon as	at last	before
currently	during	eventually	meanwhile
next	now	since	soon
finally	later	still	then
until	when/whenever	while	first, second, third

Other basic components of a narrative are:

- Plot - The events as they unfold in sequence.
- Characters - The people who inhabit the story and move it forward. Typically, there are minor characters and main characters. The minor characters generally play supporting roles to the main character, or the protagonist.
- Conflict - The primary problem or obstacle that unfolds in the plot that the protagonist must solve or overcome by the end of the narrative. The way in which the protagonist resolves the conflict of the plot results in the theme of the narrative.
- Theme - The ultimate message the narrative is trying to express; it can be either explicit or implicit.

Writing a Narrative Essay

When writing a narrative essay, start by asking yourself if you want to write a factual or fictional story. Then free write, brainstorm, or mind map about topics that are of general interest to you. For more information about pre-writing, review My Writing Process.

Once you have a general idea of what you will be writing about, you should sketch out the major events of the story that will compose your plot. Typically, these events will be revealed chronologically and climax at a central conflict that must be resolved by the end of the story. The use of strong details is crucial as you describe the events and characters in your narrative. You want the reader to emotionally engage with the world that you create in writing.

Keep the Senses in Mind

To create strong details, keep the human senses in mind. You want your reader to be immersed in the world that you create, so focus on details related to sight, sound, smell, taste, and touch as you describe people, places, and events in your narrative.

As always, it is important to start with a strong introduction to hook your reader into wanting to read more. Try opening the essay with an event that is interesting to introduce the story and get it going. Finally, your conclusion should help resolve the central conflict of the story and impress upon your reader the ultimate theme of the piece.

Narrative Essay Example

My College Education

The first class I went to in college was philosophy, and it changed my life forever. Our first assignment was to write a short response paper to the Albert Camus essay “The Myth of Sisyphus.” I was extremely nervous about the assignment as well as college. However, through all the confusion in philosophy class, many of my questions about life were answered.

I entered college intending to earn a degree in engineering. I always liked the way mathematics had right and wrong answers. I understood the logic and was very good at it. So when I received my first philosophy assignment that asked me to write my interpretation of the Camus essay, I was instantly confused. What is the right way to do this assignment, I wondered? I was nervous about writing an incorrect interpretation and did not want to get my first assignment wrong. Even more troubling was that the professor refused to give us any guidelines on what he was looking for; he gave us total freedom. He simply said, “I want to see what you come up with.”

Full of anxiety, I first set out to read Camus’s essay several times to make sure I really knew what it was about. I did my best to take careful notes. Yet even after I took all these notes and knew the essay inside and out, I still did not know the right answer. What was my interpretation? I could think of a million different ways to interpret the essay, but which one was my professor looking for? In math class, I was used to examples and explanations of solutions. This assignment gave me nothing; I was completely on my own to come up with my individual interpretation.

Next, when I sat down to write, the words just did not come to me. My notes and ideas were all present, but the words were lost. I decided to try every prewriting strategy I could find. I brainstormed, made idea maps, and even wrote an outline. Eventually, after a lot of stress, my ideas became more organized and the words fell on the page. I had my interpretation of “The Myth of Sisyphus,” and I had my main reasons for interpreting the essay. I remember being unsure of myself, wondering if what I was saying made sense, or if I was even on the right track. Through all the uncertainty, I continued writing the best I could. I finished the conclusion paragraph, had my spouse proofread it for errors, and turned it in the next day simply hoping for the best.

Then, a week or two later, came judgment day. The professor gave our papers back to us with grades and comments. I remember feeling simultaneously afraid and eager to get the paper back in my hands. It turned out, however, that I had nothing to worry about. The professor gave me an A on the paper, and his notes suggested that I wrote an effective essay overall. He wrote that my reading of the essay was very original and that my thoughts were well organized. My relief and newfound confidence upon reading his comments could not be overstated.

What I learned through this process extended well beyond how to write a college paper. I learned to be open to new challenges. I never expected to enjoy a philosophy class and always expected to be a math and science person. This class and assignment, however, gave me the self-confidence, critical-thinking skills, and courage to try a new career path. I left engineering and went on to study law and eventually became a lawyer. More important, that class and paper helped me understand education differently. Instead of seeing college as a direct stepping stone to a career, I learned to see college as a place to first learn and then seek a career or enhance an existing career. By giving me the space to express my own interpretation and to argue for my own values, my philosophy class taught me the importance of education for education’s sake. That realization continues to pay dividends every day.

Interview Schedule

Meaning of Interview Schedule :

Interview is a face to face or one to one situation in which the interviewer gathers information about the behaviour, problems and future plans of the pupil. It is also designed to assist the pupil to understand himself and his environment, so as to be able to solve his problems or modify his plans. There are various types of interviews depending upon their purpose and design.

When interview is used as a tool for gathering data for research purpose it is called “research interview” When interview is used for clinical purpose or to secure information about a pupil’s problems, his past history, adjustment patterns etc. it is called clinical interview. Similarly, interview can cover only one pupil at a time or a group of pupils. Accordingly, interview can be “individual interview or group interview”. Interviews are also classified as “structured interview” and “unstructured interview.” But our concern is to have a clear-cut look on interview schedule.

Interview schedule is another procedure under self-reporting technique of individual data collection. In the interview the individual is made to answer several questions put to him related to a specific aspect in a face to face situation. In conducting interview the interviewer (guidance personnel) may use questions specified before hand. This is called structured interview. If the interviewer doesn’t have any pre-specified questions while conducting interview, it is called unstructured interview.

Besides the above types, interviews may be counselling interview, diagnostic interview, non-directive interview, authoritarian interview and non-authoritarian interview. For guidance purpose non directive interview seems to be most useful and helpful. The counsellor follows the student’s needs, reflex and helps to clarify his feeling. He doesn’t inject his own ideas into the conversation by questions or suggestions or by giving information or advice.

Uses of Interview Schedule:

The interview schedule has the following uses:

1. It is self-reporting technique which provides considerable flexibility to the interviewer.
2. Questions can be clarified, if necessary the interview can be given an opportunity to qualify or modify his answer and the interviewer can carefully observe the individual during the session, noting down the feeling attached to his answer the topic or cases when he seems to be evasive and areas on which he is most vocal.
3. If the interviewer is experienced and trained, he can go beyond external purposes of the interview and he understands his inner feelings, wishes, desires, likes and dislikes.
4. While interviewing, the behaviour of the subject can be observed and information with regard to his emotional complex can be observed which will be greatly helpful for individual guidance.
5. It is a potent and indispensable tool for getting data that no other research tool can do.
6. It can be adaptable, capable of being used with all types of individuals.

Limitations of Interview Schedule:

The Interview schedule has the following limitations so far as its application in the field of guidance is concerned:

1. It is extremely time consuming process.
2. Information obtained is not standardized from one person to another.
3. It suffers from bias of the interviewer.
4. Sometimes interviewer's own Masses influence the questions that are asked.
5. Some of the interviewers are becoming too rigid in taking judgments, while some others get influenced by others' judgments.
6. A great deal of differences are found among the interviewers. As a result, the results obtained can't be called reliable.
7. Uniformity can't be maintained from one interview situation to another.

In spite of the above limitations, interview technique may be quite useful for guidance. If the interviewer is trained and if he possesses professional knowledge manners, maturity, objectivity, adequacy, well-defined social values, the interview technique can get a high achieving situation in individual guidance service.

Observation Method

Observation is one of the cheaper and more effective techniques of data collection. Observation, in simple terms, is defined as watching the things with some purpose in view. However, in research activity the term has a wider meaning than simple watching. Observation, is a systematic and deliberate study through eye of spontaneous occurrence at the time, they occur. Observation may serve a variety of research purposes, it may be used to explore the given area of subject matter or to gain insight in to the research problem and provide a basis for development of hypotheses.

Observation may also be used as the primary technique of data collection in descriptive studies and also in the experimental studies designed for testing casual hypotheses. Observation many times is a perception. Observation has mainly three components-Sensation, attention and perception. The accuracy of observation depends on knowledge and experience. Generally, the intellectual, physical and moral conditions are very important in observation.

General characteristics of observation method

- a. It is a physical and mental activity.
- b. It is selective and purposeful.
- c. It is a scientific tool of research.
- d. It is a direct study of the situation or phenomenon.
- e. It tries to establish cause and effect relationship in the observed phenomenon.

Process Of Observation

There are five sequential steps in the observation method.

- a. Preparation and training.
- b. Entry in to the study environment.
- c. Initial interaction.
- d. Observation and training.
- e. Termination of field work.

Aids in observation process

In order to make the process of observation effective and reduce the faults of the observer, a researcher may use a range of tools for systematising and recording data. Diaries, field notes, maps, check lists, cameras, audio, video tape recorders, maps ,analogy, checklist, socio metric scales, mechanical devices are the major tools adopted by the researcher to make the observation process as accurate as possible.

Types of Observation

Observation, which is the most classical method of scientific enquiry, may take many forms. With reference to investigators role, it may be classified into

- a. **Participant observation:** In this observation, the observer is a part of the phenomenon or group which is observed and he acts as both an observer and a participant. The persons who are observed group should not be aware of the researcher's purpose. Then only their behaviour will be natural. The observer can understand the emotional reactions of the observe group, and get a deeper insight of their experiences.
- b. **Non-Participant observation:** In this type of observation, the researcher does not actually participate in the activities of the group to be studied. There is no emotional involvement on the part of the observer. Observer would be simply present in the group to note down the behaviour of the respondents.
- c. **Controlled observation:** This type of observation is found quite useful in either in the laboratory origin the field. This involves standardization of the fields like psychology and sociology. Controlled observation is carried out observational techniques and exercise of maximum control over extrinsic and intrinsic variables
- d. **Uncontrolled observation:** If the observation takes place in the natural settings, it may be termed as uncontrolled observation. The main aim of this observation is get spontaneous picture of life. This does not involve control over any extrinsic or intrinsic variables.
- e. **Direct observation:** In this type of observation, the event or the behaviour of the person is observed as it occurs. This method is flexible and allows the observer to see and record subtle aspects of events and behaviour as they occur.
- f. **Indirect observation;** This does not involve the physical presence of the observer , and the recording is done by mechanical, photographic or electronic devices. This method is less flexible than direct observation. In other words, the behaviour of the person is not observed, rather its effects are observed.

Advantages of observation method.

- a. It is the most direct means of studying a wide variety of phenomena based on actual and first-hand experience.
- b. It enables the observer to code and record behavior at the time of its occurrence.
- c. The behavior of human beings can be best studied.
- d. It is the basis for formulating hypothesis.
- e. Data collected under this method is more accurate and reliable, as it is based on the first hand perception of the eyes.

Field observation studies

Participant observers may use multiple methods to gather data. One primary approach involves writing field notes. There are several guides for learning how to prepare field notes.

Researchers may be interested in creating or using a template to guide a researchers' observations.

- Templates or observational coding sheets can be useful when data is collected by inexperienced observers
- Templates or observational coding sheets should only be developed after observation in the field that is not inhibited by such a template
- Theories and concepts can be driven by templates and result in focused data collection
- Templates can deflect attention from unnamed categories, unimagined and unanticipated activities that can be very important to understanding a phenomenon and a setting .

Advantages of Observation:

(1) Simplest Method:

Observation is probably the most common and the simplest method of data collection. It does not require much technical knowledge. Although scientific controlled observation requires some technical skill of the researcher, still it is easier than other methods. Everybody in this world observes many things in their daily life. A little training can make a person perfect, to observe his surroundings.

(2) Useful for Framing Hypothesis:

Observation is one of the main bases of formulating hypothesis. By observing a phenomenon continuously, the researcher may get well acquainted with the observed. He came to know about their habits, likes, dislikes, problems, perception, different activities and so many other things. All these help him a lot to form a hypothesis on them. Any researcher, therefore, has to be a good observer.

(3) Greater Accuracy:

In other methods like interview, questionnaire etc., the researcher has to depend on information provided by the respondents. So these are indirect methods and here the investigator does not have any means to examine the accuracy of the data supplied by them. But in observation the observer can directly check the accuracy from the observed. He can apply various devices to test the reliability of their behaviour. So very often the data collected through observation is more reliable than these collected through interview or questionnaire.

(4) An Universal Method:

Observation is a common method used in all sciences, whether physical or social. So it has greater universality of practice. As a common method, it is very easily followed and accepted.

(5) Observation is the Only Appropriate Tool for Certain Cases:

Observation can deal with phenomena which are not capable of giving verbal information about their behaviour, feeling and activities simply for the reason that they cannot speak e.g. infants or animals. Observation is indispensable for studies on infants who can neither understand the queries of the researcher nor express themselves clearly. In the case of animals observation is the only way out. For deaf and dumb persons, for serious cases of abnormality or mad persons, for non-cooperative persons, for too shy persons and for persons who do not understand the language of researcher, observation will be the only appropriate tool.

(6) Independent of People's Willingness to Report:

Observation does not require the willingness of the people to provide various information about them. Often some respondents do not like to speak about themselves to an outsider. Some people do not have time or

required skill to provide important information to the researcher. Although observation cannot always overcome such problems, still relatively speaking it requires less active co-operation and willingness of respondents. Observation is ever possible without the knowledge of the respondents.

Limitations of Observation:

(1) Some of the Occurrences may not be Open to Observation:

There are many personal behaviours or secret activities which are not open for observation. For example, no couple will allow the researcher to observe their sexual activities. In most of the cases people do not allow the outsider to study their activities.

(2) Not all Occurrences Open to Observation can be Observed when Observer is at Hand:

Such problems arise because of the uncertainty of the event. Many social events are very much uncertain in nature. It is a difficult task on the part of the researcher to determine their time and place. The event may take place in the absence of the observer. On the other hand, it may not occur in the constant presence of the observer. For example, the quarrel and fight between two individuals or groups is never certain. Nobody knows when such an event will take place.

(3) Not all Occurrences Lend Themselves to Observational Study:

Most of the social phenomenon is abstract in nature. For example, love, affection, feeling and emotion of parents towards their children are not open to our senses and also cannot be quantified by observational techniques. The researcher may employ other methods like case study; interview etc. to study such phenomena.

(4) Lack of Reliability:

Because social phenomena cannot be controlled or used for laboratory experiments, generalizations made by observation method are not very reliable. The relative-ness of the social phenomena and the personal bias of the observer again create difficulty for making valid generalization in observation. P.V. Young remarks that in observation, no attempt is made to use instruments of precision to check the accuracy of the phenomenon.

(5) Faulty Perception:

Observation is a highly technical job. One is never sure that what he is observing is the same as it appears to his eyes. Two persons may judge the same phenomena differently. One person may find something meaningful and useful from a situation but the other may find nothing from it. Only those observers who are having the technical knowledge about the observation can make scientific observation.

(6) Personal Bias of the Observer:

The personal bias, personal view or looking at things in a particular way often creates obstacle for making valid generalization. The observer may have his own ideas of right and wrong or he may have different pre-conceptions regarding an event which kills the objectivity in social research.

(7) Slow Investigation:

Observation is a time taking process. P.V. Young rightly remarks that the valid observation cannot be hurried; we cannot complete our investigation in a short period through observation. It sometimes reduces the interest of both observer and observed to continue their observation process.

(8) Expensive:

Observation is a costly affair. It requires high cost, plenty of time and hard effort. Observation involves travelling, staying at the place of phenomena and purchasing of sophisticated equipment's. Because of this it is called as one of the most expensive methods of data collection.

(9) Inadequate Method:

According to P.V. Young, "the full answers cannot be collected by observation alone". Therefore many suggested that observation must be supplemented by other methods also.

(10) Difficulty in Checking Validity:

Checking the validity of observation is always difficult. Many of the phenomena of observation cannot be defined with sufficient precision and does not help in drawing a valid generalization. The lack of competence of the observer may hamper validity and reliability of observation.

Sources of Documentary Information

A document constitutes embodied thought which is a record of work on paper or other material fit for physical handling, transport across space and preservation through time. It may include manuscripts, handwritten and engraved materials including printed books, periodical, microform, photograph, gramophone records, tape records, etc. The recent advances in science and technology helps originate another kind of document i.e. computer readable forms that includes C.D., DVD, pen drive, hard disk, web resources etc. All documents are the records of human observation and thought and in its creation direct human intervention is necessary. They provide some information to its readers or users. A library as a gateway of knowledge provides access to a variety of such documentary sources of information.

The sources of documentary information can also be termed as an *information product*. It is generated out of a service to be provided to the user. It is a kind of consolidation and presentation process giving tangibility to information.

1. Classification of Documentary Sources of Information: Different authors classified the documentary sources of information into different categories. Some popular classifications are listed bellow

a) C. W. Hanson Classification: C. W. Hanson (1971) in the article "Introduction to science Information work" published in ASLIB (previously Association of Special Libraries and Information Bureau but now known as Association for Information Management) divides documentary sources of information into two categories i.e. primary and secondary.

i) Primary: The primary documents exist of their own and usually contain original information on the first formulation of any new observation, experiment, ideas, etc. Thus, according to C. W. Hanson, a monograph, an article in periodical, text book, and encyclopaedia are all primary documents. An article in encyclopedia or text book may not contain any new information on the subject but it presents the information in the particular form for the first time. The articles concerned are not a condensation or rewriting in any way of any existing document but has been written specifically for the text book or the encyclopedia.

ii) Secondary: All secondary publications present the contents of primary document in a condensed form or list them in a helpful way so that the existence of a primary document can be known and access to it can be made.

iii) Primary / Secondary Sources of Information: Conference proceedings, theses and dissertations, monographs, etc. have the characteristics of both primary and secondary sources of information. Those of documents representing new facts can be regarded as primary publication and those having the character of reviews can be grouped as secondary publication. As a result of such mixing of primary and secondary sources of information some expert doesn't consider this division to be much practical utility.

b) Denis Grogan Classification: Denis Grogan, on the basis of level of reorganization, has classified the documents into three categories. They are: *primary*, *secondary* and *tertiary*.

i) Primary Sources: Primary publications are those in which the author for the first time supplies evidence, describes a discovery, makes or drives a new proposition or brings forward new evidence about previous

proposition. It was created at or near the time being studied, often by the people being studied. It is a fundamental, authoritative document related to a subject of inquiry, used in the preparation of a later derivative work. Thus, the primary sources of information are basic sources of new information which are not passed through any filtering mechanism like condensation, interpretation or evaluation and are the original work of the author.

UNESCO (1968) defines a primary publication as “original scientific paper describing new research, techniques or apparatus.” Primary does not mean superior. It refers to the creation by the primary players, and is distinguished from a secondary source, which is a historical work, like a scholarly book or article, built up from primary sources.

Primary sources may include periodicals, patent, standard, report, reprint, trade journal, classic book, letters, diaries, and other personal papers, photographs, interviews and transcripts, government and historical records, newspaper clippings, and other original sources. The significance of primary publications is as follows:

- a) A subject becomes a discipline in its own right when independent primary sources begin to be produced in that area. The progress and development of a country directly depend on the primary literature that reports a new discovery.
- b) The information published in primary sources are newly generated, recent, current, full and up-to-date for all other investigators working in the same field.
- c) When any research or investigation or its any concept is first published in primary sources only it becomes the basic and original sources of communication of information and reports quickly to be used by other users.
- d) Publication of primary sources of information avoids doubling and duplication in the research, thus saves time, money and labour to be spent on it. It also acts as a guide to the researcher engaged in the same field by pointing out what has been done? And up to what level? etc.
- e) The primary sources of information help in the compilation of secondary and tertiary sources of information. Often primary sources of information may be the only sources of information in existence.

There are certain primary sources of information, which remain unpublished. Very often these may be consulted for historical interest. Such materials include laboratory note book, memoranda, diaries, letters to and from a particular individual, company, etc. The library also tries to procure such type of material if it comes within its scope of area or is relevant to its purpose.

ii) Secondary Sources: A document concerning a particular subject of inquiry which is derived from or based on the study and analysis of the primary source of information is called the secondary source of information. In the secondary source of information the original information is selected, modified and arranged in a suitable format for the purpose of easy location by the user. The secondary sources of information thus provide digested information and also serve as bibliographical key to primary sources of information. Secondary publication includes text book, reference book, review of the literature, etc.

iii) Tertiary Sources: The tertiary sources of information are last to appear and mostly do not contain subject knowledge. It is designed to provide information about information and so acts as a guide to the primary and secondary sources of information. The main function of tertiary sources is to aid the user in using primary and secondary sources of information. The tertiary sources of information are bibliography of bibliographies, guides to libraries, other organizations, indexing and abstracting periodicals, list of accession, list of research in progress, directories, etc.

Eventually there is no rigid line of demarcation between primary, secondary and tertiary sources of information.

c) S. R. Ranganathan Classification: Based on the physical characteristics of documents S. R. Ranganathan classified documentary sources of information into four categories. These also reflect the chronological order of their development. They are:

- i) **Conventional:** Books, periodicals, Map etc.;
- ii) **Neo Conventional:** Standards, specification, patent etc.;
- iii) **Non Conventional:** Audio visual, microcopy etc.;
- iv) **Meta Document:** Direct records unmediated by human mind.

2. Types of Documentary Sources of Information: The documentary sources of information can be of the following types

a) Newspaper: Newspapers are usually published as dailies or weeklies. The type of paper they are printed on, called newsprint is not meant to last. They are usually preserved on microfilm for this reason. Libraries usually keep paper copies of newspapers until the microfilm copies arrive. Nowadays many newspapers are available on the Internet, some for free, and others by subscription.

b) Periodical: Periodicals are issued at intervals and numbered consecutively. They are given volume designations, several issues making up a volume. Periodicals include journals and magazines.

i) Journal: Journal is a scholarly publication devoted to disseminating current information about research and developments in a specific field or subfield of human knowledge. Journal is usually regularly published at interval. Most journal articles are long and include a paragraph at the beginning, called an abstract which summarizes the main points of the article and at the end a bibliography or list of works cited. The writings of the journals are most often peer-reviewed.

ii) Magazine: The magazine usually refers to the non-scholarly publications written for an educated audience and contains popular reading.

c) Reprint: Once an article is published in a journal additional copies are taken out separately and provided to the author. A fixed number is generally supplied free of charge. Additional copies are supplied at a cost; these copies are known as reprints and used for exchange with other scientists working in the same field.

d) House Journal: It is a publication issued by an organization to inform the public of its performance and style of function and also to know the reaction, opinions of its public. Generally house journals are of two types:

i) External House Journals: The external house journal is meant for the external audience of an organization. The external audience of an organization refers to those who do not work under the roof of the organization, but are interested in it.

ii) Internal House Journal: Internal house journals are meant for the employees under the roof of an organization. Broadly speaking, it aims to inform and educate the employees of all levels about the organization's activities, functions, etc.

e) Newsletter: Newsletter is a publication issued by an organization often simple in format and crisp in style to provide speedy information for a definite audience. Newsletters are always issued regularly and have a short life span. It is a modest publication containing limited pages say four to eight and a few pictures and illustrations. Generally, the organizations that do not go for house journals find a good substitute in newsletters. While some newsletters are intended for the employees, others are meant for the external public.

f) Patents: A patent presents a detailed account of a new manufacturing process or improvement of an existing process, a new product, a new method of testing and control etc. Generally, when some kind of invention is made the manufacturer wants to protect his invention and the patent offices in various countries on the request of the manufacturer generally issue the patent, which provides an exclusive right to the manufacturer on the invention. It takes the form of an official document having the seal of the government attached to it, which confers an exclusive privilege or right over a period of time to the proceeds of an invention.

g) Standards: Standards are units or measures in terms of weight, size, length, quality, composition, process of production, etc., established by National and International Organizations. Standards are often finalized through testing, research, and study and prescribe the accepted quality or performance value of a product.

h) Research Report: Research reports are published as part of the annual report of an organization or as a separate report published at periodical intervals by individual and agencies that obtain research grants and have to produce them as a condition of such grants. The research reports are generally produced in limited number of copies and the distribution is also restricted and controlled.

i) Trade and Product Bulletin / Journal: Trade journals contain primary articles but of the nature of applied research. It contains the particulars of goods manufactured by or sold by a firm. Frequently illustrated and containing prices, it also often contains application oriented description rather than theoretical description. These are published by Research and Development Organizations, Trade Associations etc. The original objectives of all trade journals are product advertisement. The complete description, principles and working of a newly developed and highly sophisticated instrument may for quite time be available only in the manufacturer trade journals. Eg.: International Product Finder. Bombay: Business press.

j) Conference Proceedings: Many conference proceedings present new findings or results of work for the first time or at least months before they are published in scientific journals. Some times, conference proceedings also include questions from participants and answers and clarifications from the authors of the papers. The conference proceedings generally contain the statement of objectives, opening address or presidential address, list of participants or conference's who's who, resolutions or recommendations, etc.

k) Thesis and Dissertation: Thesis and dissertation are the results of purely academic pursuit. It reports some original work in a specific field. Among all the primary sources of information thesis and dissertation are probably least used mainly because their existence is not known in many cases and also due to the limited number of copies of the document.

l) Treatise: A treatise provides an exhaustive treatment of a broad subject. It is encyclopaedic in coverage of the subject but different in its treatment. It presents in a systematic and consolidated manner the result of work and research in the field with full reference to the primary sources.

m) Monograph: The scope of a monograph is narrower than that of a treatise. Monograph is on a single topic whereas a treatise is on a broad subject. Research monographs are separately published reports on an original research that is too long, too specialized or otherwise unsuitable for publication in one of the standard journals. Each monograph is self contained which frequently summarizes the particular existing theory or practice along with the author's original work.

n) Review: A review is actually a narrative account or critical synthesis of the progress of a particular field of study prepared by an expert in the field. It shifts, evaluates and puts each significant contribution into its proper perspective. It indicates interrelationship of ideas, significance and possible areas of application and so on, so that one can easily get an expert view of the subject without having to go through the mass of literature.

o) Text Book: A text book is made of continuous exposition, sentences mount into paragraph, paragraphs into chapter, chapters get woven into a single swelling exposition in the continuous pursuit of a single idea, simple or complex, and text books are read consecutively for inspiration, enjoyment or information. There is a link at each stage. There is an element of continuity. According to Grogan, "a text book is a teaching instrument; its primary aim is not to import information about its subject but to develop understanding of it. It concentrates on demonstrating principles rather than recounting detail".

p) State of the Art Report: These are types of reviews which do not have all embracing scope and historical orientation. These present information assembled from various sources and subjects to the operation of analysis, consolidation, extraction and evaluation in a formal presentation representing the most advanced degree of technical achievement in its field at the present time. Some owe their existence to a specific query while others are issued on a regular basis, in many cases once in a year. State of the Art report emphasizes on the recent and up-to-date ideas.

q) Trend Report: Trend report gives an account of the general direction of research in the subject based on a review of the documents on current development.

r) Technical Digests: A digest service is directed to executives, engineers, technical worker, etc. working in industries. It provides up to date technical information. It presents descriptive text of information in a condensed form and on the core ideas in brief and orderly form.

Unit -4

ANALYSIS AND PRESENTATION OF DATA

Analysis of data is considered to be highly skilled and technical job which should be carried out .Only by the researcher himself or under his close supervision. Analysis of data means critical examination of the data for studying the characteristics of the object under study and for determining the patterns of relationship among the variables relating to it's using both quantitative and qualitative methods.

Purpose of Analysis

Statistical analysis of data saves several major purposes.

1. It summarizes large mass of data in to understandable and meaningful form.
2. It makes descriptions to be exact.
3. It aids the drawing of reliable inferences from observational data.
4. It facilitates identification of the casual factors underlying complex phenomena
5. It helps making estimations or generalizations from the results of sample surveys.
6. Inferential analysis is useful for assessing the significance of specific sample results under assumed population conditions.

Steps in Analysis

Different steps in research analysis consist of the following.

1. The first step involves construction of statistical distributions and calculation of simple measures like averages, percentages, etc.
2. The second step is to compare two or more distributions or two or more subgroups within a distribution.
3. Third step is to study the nature of relationships among variables.
4. Next step is to find out the factors which affect the relationship between a set of variables
5. Testing the validity of inferences drawn from sample survey by using parametric tests of significance.

Types of Analysis

Statistical analysis may broadly classified as descriptive analysis and inferential analysis

Descriptive Analysis

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Descriptive statistics is the discipline of quantitatively describing the main features of a collection of data or the quantitative description itself. In such analysis there are univariate analysis bivariate analysis and multivariate analysis.

• Univariate analysis

• Univariate analysis involves describing the distribution of a single variable, including its central tendency (including the mean, median, and mode) and dispersion (including the range and quartiles of the data-set, and measures of spread such as the variance and standard deviation). The shape of the distribution may also be described via indices such as skewness and kurtosis. Characteristics of a variable's distribution may also be depicted in graphical or tabular format, including histograms and stem-and-leaf display.

• Bivariate analysis

Bivariate analysis is one of the simplest forms of the quantitative (statistical) analysis. It involves the analysis of two variables (often denoted as X , Y), for the purpose of determining the empirical relationship between them. Common forms of bivariate analysis involve creating a percentage table or a scatter plot graph and computing a simple correlation coefficient

• Multivariate analysis.

In multivariate analysis multiple relations between multiple variables are examined simultaneously. Multivariate analysis (MVA) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest

Inferential Analysis

Inferential statistics is concerned with making predictions or inferences about a population from observations and analyses of a sample. That is, we can take the results of an analysis using a sample and can generalize it to the larger population that the sample represents. There are two areas of statistical inferences (a) statistical estimation and (b) the testing of hypothesis.

Tools and Statistical Methods For Analysis

The tools and technique of statistics can be studied under two divisions of statistics.

(A) Descriptive Statistics

In descriptive statistics we develop certain indices and measures of raw data. They are;

1. Measures of Central Tendency
2. Measures of Dispersion
3. Measures of skewness and kurtosis
4. Measures of correlation
5. Regression analysis
6. Index numbers
7. Time series analysis
8. Coefficient of association

1. Measures of Central Tendency.

The central tendency of a distribution is an estimate of the "center" of a distribution of values. There are different types of estimates of central tendency such as mean, median, mode, geometric mean, and harmonic mean.

2. Measures of Dispersion.

Dispersion refers to the spread of the values around the central tendency. There are two common measures of dispersion, the range and the standard deviation. It can be used to compare the variability in two statistical series

3. Measures of skewness and kurtosis

A fundamental task in many statistical analyses is to characterize the *location* and *variability* of a data set. A further characterization of the data includes skewness and kurtosis. Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. That is, data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails.

4. Measures of correlation

Correlation refers to any of a broad class of statistical relationships involving dependence. When there are two variables, the correlation between them is called simple correlation. When there are more than two variables and we want to study relation between two of them only, treating the others as constant, the relation is called partial correlation. When there are more than two variables and we want to study relation of one variable with all other variables together, the relation is called multiple correlations.

5. Regression analysis

Regression analysis is a statistical process for estimating the relationships among variables. It includes many techniques for modelling and analysing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables.

6. Index numbers

An **index** is a statistical measure of changes in a representative group of individual data points. Index numbers are designed to measure the magnitude of economic changes over time. Because they work in a similar way to percentages they make such changes easier to compare.

7. Time series analysis

A time series is a sequence of data points, measured typically at successive points in time spaced at uniform time intervals. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.

8. Coefficient of association

Coefficient of association like, Yule's coefficient, measures the extent of association between two attributes.

(B) Inferential Statistics

Inferential statistics deals with forecasting, estimating or judging some results of the universe based on some units selected from the universe. This process is called Sampling. It facilitates estimation of some population values known as parameters. It also deals with testing of hypothesis to determine with what validity the conclusions are drawn.

Ratios, percentages and averages

In statistical analysis Ratios, percentages and weighted averages play a very important role. Ratios show the relation of one figure to another. For example, if the total number of students in a school is 2000, and total number of teachers is 250, then the ratio between teachers and students is 250:2000. To make it percentage, multiply by 100.

Measures of central tendency (averages)

An average is a single significant figure which sums up characteristic of a group of figures. The various measures of central tendency are;

- (1) Arithmetic mean
- (2) Median
- (3) Mode
- (4) Geometric mean
- (5) Harmonic mean

Arithmetic Mean

The Mean or average is probably the most commonly used method of describing central tendency. To compute the mean all you do is add up all the values and divide by the number of value.

$$\bar{x} = \frac{\sum x}{n}$$

Arithmetic mean =

Where x stands for an observed value,
n stands for the number of observations in the data set
 $\sum x$ stands for all observed x values, and
stands for the mean value of x

For example, consider the test score values:

15, 20, 21, 20, 36, 15, 25, 15

The sum of these 8 values is 167, so the mean is $167/8 = 20.875$.

Ex. 1 calculate mean from the following data

Value: 5 15 25 35 45 55 65 75

Freq: 1 20 25 24 12 31 71 52

Values	frequency	Fx
5	15	75
15	20	300
25	25	625
65	24	840
45	12	540
55	31	1705
65	71	4615
75	52	3900
	250	12600

$$\bar{x} = \frac{\sum x}{n} = 12600/250 = 50.4$$

Arithmetic mean =

Median

The Median is the score found at the exact middle of the set of values. One way to compute the median is to list all scores in numerical order, and then locate the score in the center of the sample.

For example, if there are 500 scores in the list, score #250 would be the median. It is also, called $\{(n + 1) \div 2\}$ **th value**, where **n** is the number of values in a set of data.

Example

Imagine that a top running athlete in a typical 200-metre training session runs in the following times: 26.1, 25.6, 25.7, 25.2 et 25.0 seconds.

First, the values are put in ascending order: 25.0, 25.2, 25.6, 25.7, and 26.1. Then, using the following formula, figure out which value is the middle value. Remember that n represents the number of values in the data set.

$$\text{Median} = \{(n+1) \div 2\} \text{th value} = (5+1) \div 2 = 3$$

The third value in the data set will be the median. Since 25.6 is the third value, 25.6 seconds would be the median time.

$$= 25.6 \text{ seconds}$$

Example

Now, if the runner sprints the sixth 200-metre race in 24.7 seconds, what is the median value now? Again, you first put the data in ascending order: 24.7, 25.0, 25.2, 25.6, 25.7, 26.1. Then, you use the same formula to calculate the median time.

$$\text{Median} = \{(n+1) \div 2\} \text{th value}$$

$$(6+1) \div 2$$

$$7 \div 2$$

$$= 3.5$$

Since there is an even number of observations in this data set, there is no longer a distinct middle value. The median is the 3.5th value in the data set meaning that it lies between the third and fourth values. Thus, the median is calculated by averaging the two middle values of 25.2 and 25.6. Use the formula below to get the average value.

$$\text{Average} = (\text{value below median} + \text{value above median}) \div 2$$

$$= (\text{third value} + \text{fourth value}) \div 2$$

$$= (25.2 + 25.6) \div 2$$

$$= 50.82$$

$$= 25.4$$

The value 25.4 falls directly between the third and fourth values in this data set, so 25.4 seconds would be the median

The various steps in the computations of median in a discrete series are as follows:

(i) Arrange the values in ascending or descending order of magnitude.

(ii) Find out the cumulative frequencies.

(iii) Find out the middle item by the formula $N + 1/2$

(iv) Now find out the value of $(N + 1/2)$ th item. It can be found by first locating the cumulative frequency which is equal to or $(N + 1/2)$ next higher to it, and then determining the value corresponding to it. This will be the value of the median.

Finding the Value of Median

Find out the value of median from the following data

Daily wages	10	5	7	11	8
Number of Workers	15	20	15	18	12

Solution: Calculation of median

Wages in ascending order	Number of persons (f)	Cumulative Frequency (c.f.)
5	20	20
7	15	35
8	12	47
10	15	62
11	18	80

Median is the value of $(N+1)/2$ th or $((80+1)/2)$ th or 40.5th item.

All items from 35 onwards up to 47 have a value of 8. Thus the median value would be 8. In the case of continuous frequency distribution, median class corresponds to the cumulative frequency which includes $N/2$. After getting median class find median by using the following interpolation formula.

$$\text{Median, } m = L1 + \left[\frac{(N/2 - CF)}{f} \right] C$$

L1 means lower boundary of the median class

N means sum of frequencies

CF means cumulative frequency before the median class. Meaning that the class before the median class what is the frequency

f means frequency of the median class

C means the size of the median class

Find out the value of median from the following data

Class : 0-10 10-20 20-30 30-40 40-50 50-60 60-70

Frequency: 8 12 20 23 18 7 2

Class	Frequency	Cumulative frequency
0-10	8	8
10-20	12	20
20-30	20	40
30-40	23	63 $n/2$
40-50	18	81
50-60	7	88
60-70	2	90
	90	

Median = $(N/2)$ th item = size of $(90/2)$ th item = size of 45th item

45 is included in the cumulative frequency 63. The class having cf 63 is 30-40

Therefore 30-40 is the median class

Applying interpolation formula

$$\text{Median, } m = L1 + \left[\frac{(N/2 - CF)}{f} \right] C$$

Here $L1=30$, $N/2=45$, $cf=40$, $f=23$, $c=10$

$$\text{Median, } = 30 + \left[\frac{(45 - 40)}{23} \right] 10 = 30 + 50/23 = 32.17$$

Mode

Mode is the value of the item of a series which occurs most frequently. According to Kenny 'the value of the variable which occurs most frequently in a distribution is called a mode'.

In the case of individual series, the value which occurs more number of times is mode. For

example, a set of students of a class report the following number of video movies they see in a month.

No of movies: 10, 15, 20, 15, 15, 8

Mostly the students see 15 movies in a month. Therefore mode = 15

When no item appears more number of items than others we say mode is ill defined. In that case, mode is obtained by the formula, mode = $3\text{median} - 2\text{mean}$

Ex: find mode from the values 40, 25, 60, 35, 81, 75, 90, 10

Ans: all items appear equal number of items. So mode is ill defined.

Therefore, mode = $3\text{median} - 2\text{mean}$

$$\text{Mean} = 416/8 = 52$$

$$\text{Median} = \{(n + 1) \div 2\}\text{th value} = \text{size of } 4.5\text{th item} = (40 + 60) / 2 = 50$$

$$\text{Mode} = 3\text{median} - 2\text{mean}$$

$$= (3 * 50) - (2 * 52) = 150 - 104 = 46$$

In the case of discrete frequency distribution, the value having highest frequency is taken as mode

Ex: find mode

Size :	5	8	10	12	29	35	40	46
No of items:	3	12	25	40	31	20	18	7

Ans: the value 12 has the highest frequency. Therefore 12 is the mode.

In the case of continuous frequency distribution, mode lies in the class having highest frequency.

From the modal class, mode is calculated using interpolation formula.

$$\text{Mode} = L1 + \left[\frac{(f1 - f0) c}{2f1 - f0 - f2} \right]$$

Where, L1 is the lower limit of model class. f0 and f2 are respectively the frequencies of class just preceding and succeeding model class, f1 is the frequency of the model class.

Ex: calculate mode from the following data.

Size:	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50
Freq:	4	8	18	30	20	10	5	2

Ans: Modal class is 25-30 since it has highest frequency.

$$\text{Mode} = L1 + \left[\frac{(f1 - f0) \cdot c}{2f1 - f0 - f2} \right]$$

$$= 25 + \left[\frac{(30 - 18) \cdot 5}{2 \cdot 30 - 18 - 20} \right] = 25 + \frac{60}{22} = 27.73$$

Index Numbers

Index numbers are designed to measure the magnitude of economic changes over time. A statistic which assigns a single number to several individual statistics in order to quantify trends. Index numbers are the indicators of the various trends in an economy. Price index numbers indicate the position of prices whether they are rising or falling and at what rate. Similarly, index numbers regarding agricultural production indicates the trend of change whether it is rising or falling at what rate over a period of time. An index number is an economic data figure reflecting price or quantity compared with a standard or base value. The base usually equals 100 and the index number is usually expressed as 100 times the ratio to the base value. For example, if a commodity costs twice as much in 1970 as it did in 1960, its index number would be 200 relative to 1960. Index numbers are used especially to compare business activity, the cost of living, and employment. An index number is specialized average. Index numbers may be simple or weighted depending on whether we assign equal importance to every commodities or different importance to different commodities according to the percentage of income spent on them or on the basis of some other criteria. In this chapter, we shall discuss both simple and weighted index numbers.

Simple and weighted index numbers

Simple index numbers are those in the calculation of which all the items are treated as equally important. Here items are not given any weight. Weighted index numbers are those in the calculation of which each item is assigned a particular weight.

Price Index Numbers

Price index numbers measure changes in the price of a commodity for a given period in comparison with another period.

Various methods used for construction of Price index numbers

1) Simple Aggressive Method

This is the simplest method. The prices for base year and current year are only required. The aggregate of current year price is divided by aggregate of base year price and multiplied by 100.

i.e. $\frac{\sum p1}{\sum p0} \cdot 100$ where, p1 is the aggregate of price in the current year and p0 is the aggregate no of prices in the base year.

Ex: for the data given below calculate simple index number

Commodities: A B C D E

Price in 2008 5 8 12 25 3

Price in 2010 7 9 15 24 4

Ans: we take 2008 as base year and 2010 as current year, since 2008 is the back period

Commodities	Price in 2008(p1)	Price in 2010 (p0)
A	5	7
B	8	9
C	12	15
D	25	24
E	3	4
	53	59

$$\text{Simple index number} = \frac{\sum p1}{\sum p0} \cdot 100 = \frac{59}{53} \cdot 100 = 111.3$$

2) Simple Average Relative Method

In this method, price relative for each item is found out.

Price relative is $I = \text{current year price} \div \text{base year price} * 100$

The average of these relatives is found out. ie price index number $= \Sigma I/n$

Ex: for the data given below calculate simple index number by average relative method

Items : 1 2 3 4 5

Price in base year : 5 10 15 20 8

price in current year : 7 12 25 18 9

Items	price in base year	price in current year	$I = \Sigma p_1 \div \Sigma p_0 * 100$
1	5	7	140.0
2	10	12	120.0
3	15	25	166.7
4	20	18	90.0
5	8	9	112.5
			629.2

Simple index number $= \Sigma I/n = 629.2/5 = 125.84$

3) weighted aggressive method

in this method weights are assigned to each item. The two well known methods used for assigning weights are known as Laspeyer's method and Paasche's method.

Laspeyer's method: base year quantity is taken as weight.

Laspeyer's index number $= \Sigma p_1 q_0 / \Sigma p_0 q_0 * 100$

Paasche's method. : current year quantity is taken as weight.

Paasche's index number $= \Sigma p_1 q_1 / \Sigma p_0 q_1 * 100$

Prof. Irving Fisher has suggested a formula for the construction of index numbers.

Fisher's index number $= \Sigma p_1 q_0 \Sigma p_1 q_1$

$\Sigma p_0 q_0 \times \Sigma p_0 q_1$

Ex: calculate (i) Laspeyer's (ii) Paasche's (iii) Fisher's index numbers from the following data.

Commodity	price		quantity assumed	
	2009	2010	2009	2010
A	0.80	0.70	10	11.0
B	0.85	0.90	8	9.0
C	1.30	0.80	5	5.5

Commodity	p0	p1	q0	q1	p1q1	p0q1	p1q0	p0q0
A	0.80	0.70	10	11	7.7	8.80	7.0	8.00
B	0.85	0.90	8	9	8.1	7.65	7.2	6.8
C	1.30	0.80	5	5.5	4.4	7.15	4.0	6.5
					202	23.6	18.2	21.3

Laspeyer's index number $= \Sigma p_1 q_0 / \Sigma p_0 q_0 * 100$

$= (18.2/21.3) * 100 = 85.45$

Paasche's index number $= \Sigma p_1 q_1 / \Sigma p_0 q_1 * 100$

$= (20.2/23.6) * 100 = 85.59$

Fisher's index number $= \Sigma p_1 q_0 \Sigma p_1 q_1$

$\Sigma p_0 q_0 \times \Sigma p_0 q_1$

SQUARE ROOT $\frac{18.2}{21.3} \frac{20.2}{23.6} = 85.5$

21.3×23.6

4) Weighted Average Of Price Relative Method

In this method, we are using some arbitrary numbers as weight.

The formula is $\Sigma IV / \Sigma V$ where, 'V' is the weight and $I = (p_1 / p_0) * 100$

Calculate index number of price for 2009 on the basis of 2008

Commodity weight price (2008) price(2009)

A	40	16	20
B	25	40	60
C	5	2	2
D	20	5	6
E	10	2	1

Ans:

Commodity	V	P0	P1	I	IV
A	40	16	20	125	5000
B	25	40	60	150	3750
C	5	2	2	100	500
D	20	5	6	120	2400
E	10	2	1	50	500
	100				12150

Index number for 2009 = $\frac{\Sigma IV}{\Sigma V} = \frac{12150}{100} = 121.5$

Interpretation

Interpretation refers to the technique of drawing inference from the collected facts and explaining the significance of those inferences after an analytical and experimental study. It is a search for broader and more abstract means of the research findings. If the interpretation is not done very carefully, misleading conclusions may be drawn. The interpreter must be creative of ideas he should be free from bias and prejudice.

Fundamental principles of interpretation

1. Sound interpretation involves willingness on the part of the interpreter to see what is in the data.
2. Sound interpretation requires that the interpreter knows something more than the mere figures.
3. Sound interpretation demands logical thinking.
4. Clear and simple language is necessary for communicating the interpretation

Need for interpretation (importance of interpretation.)

1. It is through interpretation that the interpreter is able to know the abstract principles lying in his conclusions.
2. On the basis of the principles underlying his findings, a researcher can make various predictions about the various other events which are unrelated to his area of findings.
3. Interpretation leads to the establishment of explaining concepts.
4. A researcher can appreciate only through interpretation, why his findings are and what they are.
5. The interpretation of the findings of exploratory research study usually results in to hypothesis for experimental research.

Steps involved in the technique of interpretation

1. Researcher must give reasonable explanations of the relations he have found. He must be able to see uniformity in diversified research findings so that generalization of findings is possible.
2. If any extraneous information is collected during the study, it must be considered while interpreting the final result of research study.
3. The researcher can consult with those having insight in to the study who can point out the omission and errors in logical arguments.
4. The researcher must consider all relevant factors affecting the problem at the time of interpretation.
5. The conclusions appearing correct at the beginning may prove to be inaccurate later. So researcher must not be in a hurry while interpreting.

Errors of interpretation

The errors of interpretation can be classified into two groups.

1. Errors due to false generalizations

Errors occur when (i) unwarranted conclusions are drawn from the facts available. (ii) Drawing conclusions from an argument running from effect to cause. (iii) Comparing between two sets of data with unequal base. (iv) Conclusions are drawn from data irrelevant to the problem. (v) False generalizations and faulty statistical methods are made.

2. Errors due to misuse of statistical measures

When (i) conclusions are based on what is true, on an average. (ii) Percentages are used for comparisons, when total numbers are different. (iii) Index numbers are used without proper care. (iv) Casual correlation is used as real correlation.

Test of significance for small samples

So far we have discussed problems belonging to large samples. When a small sample (size < 30) is considered, the above tests are inapplicable because the assumptions we made for large sample tests, do not hold good for small samples. In case of small samples it is not possible to assume (i) that the random sampling distribution of a statistics normal and (ii) the sample values are sufficiently close to population values to calculate the S.E. of estimate.

Thus an entirely new approach is required to deal with problems of small samples. But one should note that the methods and theory of small samples are applicable to large samples but its converse is not true.

Degree of freedom (df): By degree of freedom we mean the number of classes to which the value can be assigned arbitrarily or at will without voicing the restrictions or limitations placed.

For example, we are asked to choose any 4 numbers whose total is 50. Clearly we are at freedom to choose any 3 numbers say 10, 23, 7 but the fourth number, 10 is fixed since the total is 50 [$50 - (10 + 23 + 7) = 10$]. Thus we are given a restriction; hence the freedom of selection of number is $4 - 1 = 3$.

The degree of freedom (df) is denoted by ν (nu) or df and it is given by $\nu = n - k$, where n = number of classes and k = number of independent constrains (or restrictions).

In general for a Binomial distribution, $\nu = n - 1$

For Poisson distribution, $\nu = n - 2$ (since we use total frequency and arithmetic mean).

For normal distribution, $\nu = n - 3$ (since we use total frequency, mean and standard deviation) etc.

Student's t-distribution

This concept has been introduced by W. S. Gosset (1876 - 1937). He adopted the pen name "student." Therefore, the distribution is known as 'student's t-distribution'.

It is used to establish confidence limits and test the hypothesis when the population variance is not known and sample size is small (< 30).

If a random sample x_1, x_2, \dots, x_n of n values be drawn from a normal population with mean μ and standard deviations then the mean of sample

$$\bar{x} = \frac{\sum x_i}{n}$$

Estimate of the variance : Let s^2 be the estimate of the variance of the sample then s^2 given by

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

($n - 1$) as denominator in place 'n'.

(I) The statistic 't' is defined as

$$t = \frac{|\bar{x} - \mu|}{\sqrt{\frac{s^2}{n}}} \text{ or } \frac{|\bar{x} - \mu|}{s} \sqrt{n}$$

Where \bar{x} = sample mean, μ = actual or hypothetical mean of population, n = sample size, s = standard deviation of sample.

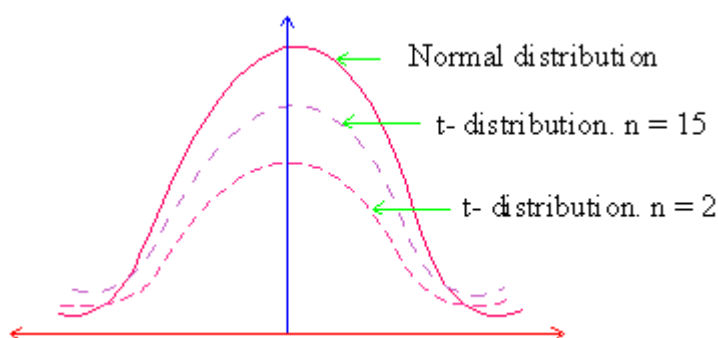
$$\text{Where } s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Note: 't' is distributed as the student distribution with $(n - 1)$ degree of freedom (df).

(II) 1) The variable 't' distribution ranges from minus infinity to plus infinity.

2) Like standard normal distribution, it is also symmetrical and has mean zero

3) σ^2 of t-distribution is greater than 1, but becomes 1 as 'df' increases and thus the sample size becomes large. Thus the variance of t-distribution approaches the variance of the normal distribution as the sample size increases for $\sigma^2(\text{df}) = 1$, the t-distribution matches with the normal distribution. (observe the adjoining figure).



Also note that the t-distribution is lower at the mean and higher at the tails than the normal distribution. The t-distribution has proportionally greater area at its tails than the normal distribution.

(III) 1) If $|t|$ exceeds $t_{0.05}$ then difference between \bar{x} and μ is significant at 0.05 level of significance.

2) If $|t|$ exceeds $t_{0.01}$, then difference is said to highly significant at 0.01 level of significance.

3) If $|t| < t_{0.05}$ we conclude that the difference between \bar{x} and μ is not significant and the sample might have been drawn from a population with mean = μ i.e. the data is consistent with the hypothesis.

(IV) Fiducial limits of population mean

$$\text{For } 95\%, \bar{x} \pm \frac{S}{\sqrt{n}} t_{0.05}$$

$$\text{For } 99\%, \bar{x} \pm \frac{S}{\sqrt{n}} t_{0.01}$$

Example A random sample of 16 values from a normal population is found to have a mean of 41.5 and standard deviation of 2.795. On this basis is there any reason to reject the hypothesis that the population mean $\mu = 43$? Also find the confidence limits for μ .

Solution: Here $n = 16 - 1 = 15$, $\bar{x} = 41.5$, $s = 2.795$ and $\mu = 43$.

$$\text{Now } t = \frac{|\bar{x} - \mu|}{s} \sqrt{n} = \frac{1.5 \times \sqrt{15}}{2.795} = 2.078$$

From the t-table for 15 degree of freedom, the probability of t being 0.05, the value of t = 2.13. Since $2.078 < 2.13$. The difference between \bar{x} and μ is not significant.

Now, null hypothesis : $H_0 : \mu = 43$ and

Alternative hypothesis : $H_1 : \mu \neq 43$.

Chi-Square

Tests like z-score, t, F are based on the assumption that the samples were drawn from normally distributed populations or more accurately that the sample means were normally distributed. As these tests require assumption about the type of population or parameters, these tests are known as 'parametric tests.'

There are many situations in which it is impossible to make any rigid assumption about the distribution of the population from which samples are drawn. This limitation led to search for non-parametric tests. Chi-square (Read as Ki - square) test of independence and goodness of fit is a prominent example of a non-parametric test. The chi-square (χ^2) test can be used to evaluate a relationship between two nominal or ordinal variables.

χ^2 (chi-square) is measure of actual divergence of the observed and expected frequencies. In sampling studies we never expect that there will be a perfect coincidence between actual and observed frequencies and the question that we have to tackle is about the degree to which the difference between actual and observed frequencies can be ignored as arising due to fluctuations of sampling. If there is no difference between actual and observed frequencies then $\chi^2 = 0$. If there is a difference, then χ^2 would be more than 0. But the difference may also be due to sample fluctuation and thus the value of χ^2 should be ignored in drawing the inference. Such values of χ^2 under different conditions are given in the form of tables and if the actual value is greater than the table value, it indicates that the difference is not solely due to sample fluctuation and that there is some other reason.

On the other hand, if the calculated χ^2 is less than the table value, it indicates that the difference may have arisen due to chance fluctuations and can be ignored. Thus χ^2 -test enable us to find out the divergence between theory and fact or between expected and actual frequencies is significant or not.

If the calculated value of χ^2 is very small, compared to table value then expected frequencies are very little and the fit is good.

If the calculated value of χ^2 is very large as compared to table value then divergence between the expected and the observed frequencies is very big and the fit is poor.

We know that the degree of freedom r (df) is the number of independent constraints in a set of data.

Suppose there is a 2×2 association table and actual frequencies of the various classes are as follows :

	A	a	
B	AB -22	aB -38	60
b	Ab -8	ab -32	40
	30	70	100

Now the formula for calculating expected frequency of any class (cell)

$$= \frac{\text{Row total for row containing the cell} \times \text{column total for column containing the cell}}{\text{The total number of observations}}$$

In notations : Expected frequency = $\frac{R \times C}{N}$

For example, if we have two attributes A and B that are independent then the expected frequency

of the class (cell) AB would be $\frac{30 \times 60}{100} = 18$

Once the expected frequency of cell (AB) is decided the expected frequencies of remaining three classes are automatically fixed.

Thus for class (aB) it would be $60 - 18 = 42$

for class (Ab) it would be $30 - 18 = 12$

for class (ab) it would be $70 - 42 = 28$

This means that so far as 2×2 association (contingency). Table is concerned, there is 1 degree of freedom.

In such tables the degrees of freedom are given by a formula $n = (c - 1) (r - 1)$

where c = Number of columns

r = Number of rows

Thus in 2×2 table $df = (2 - 1) (2 - 1) = 1$

3×3 table $df = (3 - 1) (3 - 1) = 4$

4×4 table $df = (4 - 1) (4 - 1) = 9$ etc.

If the data is not in the form of contingency tables but as a series of individual observations or discrete or continuous series then it is calculated by $n = n - 1$ where n is the number of frequencies or values of number of independent individuals.

Formula :

$$\chi^2 = \sum \left[\frac{(\text{observed frequency} - \text{expected frequency})^2}{\text{expected frequency}} \right]$$

$$\chi^2 = \sum \left[\frac{(O - E)^2}{E} \right]$$

where O = observed frequency and E = expected frequency.

Example The following table shows the age groups of people interviewed according to their age-group and the number in each group estimated to have T. B.

Age group	Nos. interviewed	T. B. cases
15 - 20	199	1
20 - 25	300	8
25 - 35	1128	38
35 - 45	1375	96
45 - 55	1089	105
55 - 65	625	56
65 - 73	155	12
Total	4871	316

Do these figures justify the hypothesis that T. B. is equally popular in all age groups?

Solution: If T.B. equally popular in all groups then in each age group

$$\left[\frac{316}{4871} \times 100 = 6.5\% \text{ of the people suffer from it} \right]$$

On this basis the observed and expected frequencies would be as

Age group Observed Cases Expected cases

15 - 20	1	13
20 - 25	8	19.5
25 - 35	38	73
35 - 45	96	89
45 - 55	105	71
55 - 65	56	40.5
65 - 75	12	10

Using formula, we get

$$\chi^2 = \frac{(1-13)^2}{13} + \frac{(8-19.5)^2}{19.5} + \frac{(38-73)^2}{73} + \frac{(96-89)^2}{89} \\ + \frac{(105-71)^2}{71} + \frac{(56-40.5)^2}{40.5} + \frac{(12-10)^2}{10}$$

Therefore, $\chi^2 = 57.6$

$$\text{Also } n(\text{df}) = (c - 1)(n - 1) = (7 - 1)(2 - 1) = 6$$

From the χ^2 - table, $\chi^2_{0.05, n=6} = 12.59$

Since $57.6 > 12.59$ at 0.05 level of significance for 6 degree of freedom, the difference is significant and the hypothesis is not justified.

Unit -5
Presentation of Data

Diagrams

A diagram is a visual form of presentation of statistical data, highlighting their basic facts and relationship. There are geometrical figures like lines, bars, squares, rectangles, circles, curves, etc. Diagrams are used with great effectiveness in the presentation of all types of data.

When properly constructed, they readily show information that might otherwise be lost amid the details of numerical tabulation.

Importance of Diagrams :

A properly constructed diagram appeals to the eye as well as the mind since it is practical, clear and easily understandable even by those who are unacquainted with the methods of presentation. Utility or importance of diagrams will become clearer from the following points

(i) **Attractive and Effective Means of Presentation** : Beautiful lines ;full of various colours and signs attract human sight, and do not strain the mind of the observer. A common man who does not wish to indulge in figures, get message from a well prepared diagram.

(ii) **Make Data Simple and Understandable** : The mass of complex data, when prepared through diagram, can be understood easily. According to Shri Morane, *“Diagrams help us to understand the complete meaning of a complex numerical situation at one sight only”*.

(iii) **Facilitate Comparison** : Diagrams make comparison possible between two sets of data of different periods, regions or other facts by putting side by side through diagrammatic presentation.

(iv) **Save Time and Energy** : The data which will take hours to understand, becomes clear by just having a look at total facts represented through diagrams.

(v) **Universal Utility** : Because of its merits, the diagrams are used for presentation of statistical data in different areas. It is widely used technique in economic, business, administration, social and other areas.

(vi) **Helpful in Information Communication** : A diagram depicts more information than the data shown in a table. Information concerning data to general public becomes more easy through diagrams and gets into the mind of a person with ordinary knowledge.

Types of diagrams

The different types of diagrams can be divided into following heads -

- (1) One dimensional diagrams
- (2) Two dimensional diagrams
- (3) Three dimensional diagrams
- (4) Pictograms
- (5) Cartograms

One Dimensional Diagrams or Bar Diagrams :

Bar diagrams are the most common types of diagram. A bar is a thick line whose width is shown merely for attention. They are called one dimensional because it is only the length of the bar that matters and not the width.

Kinds of Bar Diagrams :

Line Diagrams : When the number of items is large, but the proportion between the maximum and minimum is low, lines may be drawn to economise space. Only individual or time series are represented by these diagrams

Simple Bar Diagrams : A simple bar diagram is used to represent only one variable. For example the figures of sales, production etc. of various years may be shown by means of simple bar diagram. The bars are of equal width only the length varies.

These diagrams are appropriate in case of individual series, discrete series and time series.

Multiple Bar Diagrams : In a multiple bar diagram, two or more sets of interrelated data are represented. Different shades, colors or dots are used to distinguish between the bars. These are used to compare two or more related variables based on time and place.

Sub-divided Bar Diagrams : If a bar is divided into more than one parts, it will be called sub-divided bar diagram. Each component occupies a part of the bar proportional to its share in the total. For example total

expenditure incurred by a family on various items such as food, clothing, education, house rent etc can be represented by means of sub-divided bar diagram.

Percentage Sub-Divided Bar Diagrams : Percentage sub-divided bars are particularly useful to measure relative changes of data. When such diagrams are prepared, the length of the bar is kept equal to 100 and segments are cut in these bars to represent the components of an aggregate.

Profit – Loss Diagrams : If relative change of cost & sales or profit or loss are to be represented with the help of bars, then profit – loss diagram are constructed. These diagrams are similar to percentage sub-divided bars and are prepared in the same way.

Duo-Directional Bar Diagrams : In duo-directional bar diagram comparative study of two major parts of data is represented in a single bar. Such duo directional diagrams are represented on both sides of the horizontal axis, i.e. above and below the base line.

Paired Bars : If two different information which are in different units are to be presented then paired bar diagram are used. These bars are not vertical but horizontal and the first scale is in the first half and second scale is in the second half.

Deviation Bar Diagram : Deviation bars are popularly used for representing net quantities i.e. net profit, net loss, net exports or net imports etc. Such bars can have both positive or negative values. Positive values are shown above the base line and negative values below it.

Progress Chart or Gant Chart : These charts are mainly used in factories for comparing the actual production with targeted production. By looking at it, it can be known how much production has been achieved and how much they are lacking behind the capacity.

Pyramid Bar Diagram : These diagrams are constructed to show population distribution. The distribution of population according to sex, age, education etc are represented by this diagram. In this diagram, the base line is in the middle and its shape is like a pyramid.

Sliding Bar Diagrams : These bars are like duo-directional bars but instead of absolute figures percentages of two variables are shown. One of them is shown on the right side of the base and the other on its left.

Two Dimensional Diagrams :

In two dimensional diagrams, the height as well as the width of the bars will be considered. The area of the bars represents the magnitude of data. Such diagrams are also known as area diagrams or surface diagrams.

The important types are –

- (i) Rectangle diagram
- (ii) Square diagram
- (iii) Circle or Pie diagram

Rectangle Diagram : Rectangles are often used to represent the relative magnitude of two or more values. The area of rectangles is kept in proportion to the values. They are placed side by side like bars and uniform space is left between different rectangles.

The rectangles may be of different types -

- a) Simple Rectangles
- b) Sub-divided Rectangles
- c) Percentage sub-divided Rectangles

Square Diagrams : When there is a large difference between the extreme values (example the smallest value is 4 and the biggest value is 800) is such a case square diagram is more appropriate.

First of all square roots of the given values are calculated and the sides are taken in the proportion of square roots. The squares are drawn on the common base line, serially either in increasing or decreasing heights to have beautiful and attractive appearance.

For calculating the scale, the area is calculated by squaring its side, on the basis of that value of 1 sq.cm. is calculated.

Circle or Pie Diagram : These diagrams are more attractive, therefore, pie diagrams are preferred to square diagrams. These diagrams are used to represent data of population, foreign trade, production etc.

The square roots of the given values are calculated and then it is divided by some common factor so as to attain the radii for the circles. The area of the circle is calculated by the formula - (πr^2). The sides for squares are taken as the radii for different circles.

A circle can also be sub-divided on the basis of angles to be calculated for each component. There is 360 degree at the centre of the circle and proportionate sectors are cut taking the whole data equal to 360 degrees. Such a circle is known as sub-divided circle or Angular diagram.

Three Dimensional Diagrams :

In three-dimensional diagrams length, width and height (depth) are taken into consideration. If the difference between the minimum and maximum value is so wide as it is difficult to represent them by square or circle diagram, then three dimensional diagram is used. For this cubic roots of the given numbers is calculated. Three dimensional diagrams include cubes, blocks, spheres and cylinders etc.

Pictograms' : Pictograms are used by government and non-government organizations for the purpose of advertisement and publicity through appropriate pictures. It is a popular technique particularly when statistical facts are to be presented for a layman having no background of mathematics or statistics. For representing data relating to social, business and economic phenomena for general masses in fairs and exhibitions, this method is used.

Cartograms : Cartograms or statistical maps are also used to represent data. Cartograms are simple and elementary form of visual presentation and are very easy to understand. While highlighting the regional or geographical comparisons, mapographs or cartograms are generally used.

Graphs: construction of graphs

According to M.M. Blair – “ *The simplest to understand, the easiest to make, the most variable and the most widely used type of chart is graph.* ”.

Graphic presentation is a visual form of presenting statistical data. Graph acts as a tool of analysis and makes complex data simple and intelligible.

Graphs are more appropriate in the following cases -

- a) If tendency instead of real measurement is important.
- b) When comparative study of many data series is required on one graph.
- c) If estimation and interpolation are to be presented by graph.
- d) If frequency distribution is presented by two or more curves.

In research, the data collected may be of complex nature. Diagrams and graphs is one of the methods which simplifies the complexity of quantitative data and make them easily intelligible. They present dry and uninteresting statistical facts in the shape of attracting and appealing pictures. They have a lasting effect on the human mind than the conventional numbers.

Uses of Graphs and Diagrams

1. They help in presenting quantitative facts in simple, clear and effective pictures.
2. They make the whole data readily intelligible.
3. They can be used for comparison purpose.
4. They are useful in analysing complex economic theories.
5. They save much time in understanding data.
6. Facts can be understood without doing mathematical calculations.
7. They help in locating statistical measures such as median, quartile, mode etc

Types of Graphs

The following graphs are commonly used to represent data

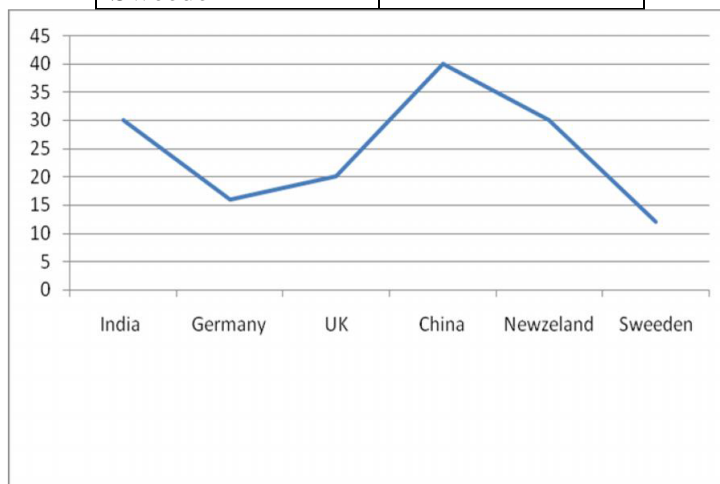
1. Charts or line graphs
2. Bar charts
3. Circle charts or pie diagram
4. Pictograms

1. Line Graphs

A line graph displays information in a series of data points that each represents an individual measurement or piece of data. The series of points are then connected by a line to show a visual trend in data over a period of time. The line is connected through each piece chronologically.

For eg; following data show birth rate per thousands of six countries over a period.

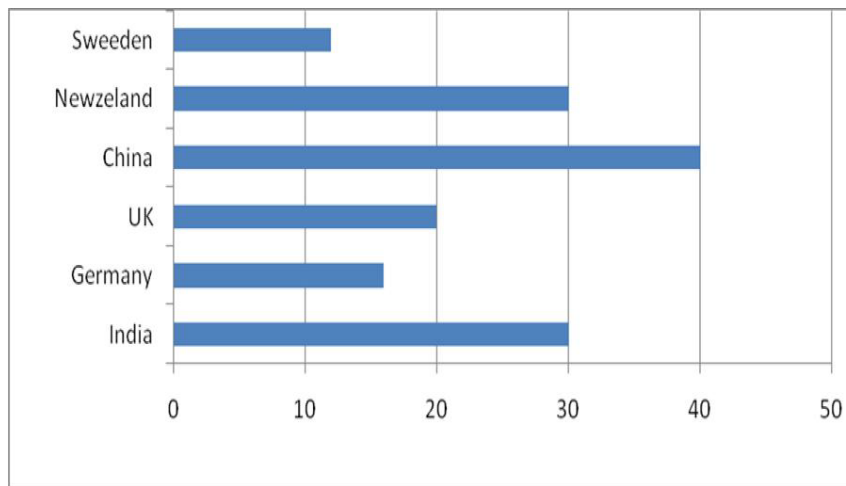
Country	Birth Rate
India	30
Germany	16
UK	20
China	40
Newzeland	30
Sweedden	12



2. BAR CHARTS

The bar graph is a common type of graph which consists of parallel bars or rectangles with lengths that are equal to the quantities that occur in a given data set. The bars can be presented vertically or horizontally to show the contrast and record information. Bar graphs are used for plotting discontinuous (discrete) data. Discrete data contains discrete values and are not continuous.

Country	Birth Rate
India	30
Germany	16
UK	20
China	40
Newzeland	30
Sweedden	12



Histogram

A histogram is a graph of frequency distributions. It is a set of vertical bars whose are proportional to the frequencies. While constructing histogram, the variable is always taken on the x- axis and the frequencies on y- axis.

Frequency Polygon

The frequency polygon is a graph of frequency distribution. Here we draw histogram of the data and then join by straight line and mid points of upper horizontal sides of these bars. Join both ends of the frequency polygon with the x- Axis.

Frequency Curves

A continuous frequency distribution can be represented by a smoothed curve known as Frequency curves

Ogive or Cumulative Frequency Curve

A frequency distribution can be cumulated in two ways, less than cumulative series and more than cumulative series. Smoothed frequency curves drawn for these two cumulative series are called cumulative frequency curves or ogives.

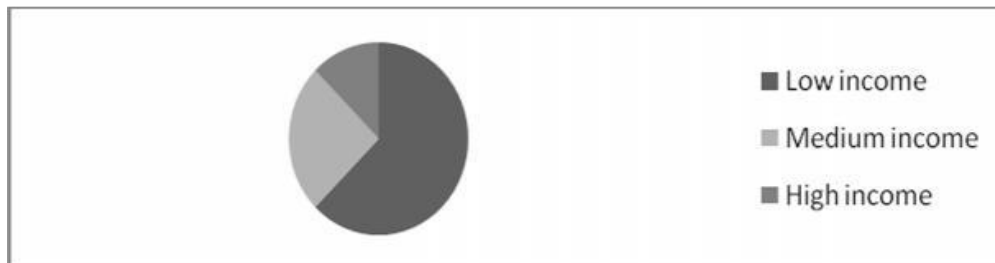
a. **Less than ogive curve:** In less than ogive curve the upper limit per limit of each class interval is taken on x- axis in increasing order. For each such upper limit on x-axis, the cumulative frequency of all the class intervals from the first class interval to last class interval are taken on the y-axis.

b. **More than ogive curve:** In more than ogive curve the lower limit of each class interval is taken on x- axis in increasing order. For each such lower limit on x- axis the cumulative frequency of all the class interval from that class interval to the last class interval are taken on y-axis.

3. Circle Charts or Pie Diagram

A pie graph is a circle divided into sections which each display the size of a relative piece of information. Each section of the graph comes together to form a whole. In a pie graph, the length of each sector is proportional to the percentage it represents. Pie graphs work particularly well when ach slice of the pie represents 25 to 50 percent of the given data. The pie chart shows that 62.5% belong to low income group, 25% to middle income group and 12.5%to high income group.

Income category	Percentage
Low income	62.5
Medium income	25.0
High income	12.5



The pie chart shows that 62.5% belong to low income group, 25% to middle income group and 12.5% to high income group.

4. Pictograms

A pictogram, also called a pictogram or pictograph, is an ideogram that conveys its meaning through its pictorial resemblance to a physical object. Pictographs are often used in writing and graphic systems in which the characters are to a considerable extent pictorial in appearance. Pictography is a form of writing which uses representational, pictorial drawings. It is a basis of cuneiform and, to some extent, hieroglyphic writing, which also uses drawings as phonetic letters or determinative rhymes.

Kerala	20
Karnataka	28
Tamil Nadu	39
Andhra Pradesh	42

False Base Line

While making the graph the vertical scale (y-axis) must start from zero origin. But in some cases adjustment of scale is not possible on (y-axis) starting with zero since the values are big. If the vertical scale begins with zero, the curve will be concentrated mostly on the top of the graph paper.

In such cases that portion of the scale from zero to the minimum value of items to be depicted is omitted. This fact is indicated in the graph paper itself either by drawing a double saw tooth lines in the empty space or by a cut in y-axis. This is known as using false base line. It is drawn in this way to show that Y axis is lost in between

Cumulative Frequency

Cumulative frequency is defined as a running total of frequencies. The frequency of an element in a set refers to how many of that element there are in the set. Cumulative frequency can also defined as the sum of all previous frequencies up to the current point.

The cumulative frequency is important when analyzing data, where the value of the cumulative frequency indicates the number of elements in the data set that lie below the current value. The cumulative frequency is also useful when representing data using diagrams like histograms.

Cumulative Frequency Table

The cumulative frequency is usually observed by constructing a cumulative frequency table. The cumulative frequency table takes the form as in the example below.

Example 1

The set of data below shows the ages of participants in a certain summer camp. Draw a cumulative frequency table for the data.

Age (years)	Frequency
10	3
11	18
12	13
13	12
14	7
15	27

Solution: The cumulative frequency at a certain point is found by adding the frequency at the present point to the cumulative frequency of the previous point.

The cumulative frequency for the first data point is the same as its frequency since there is no cumulative frequency before it.

age (years)	Frequency	Cumulative Frequency
10	3	3
11	18	$3+18 = 21$
12	13	$21+13 = 34$
13	12	$34+12 = 46$
14	7	$46+7 = 53$
15	27	$53+27 = 80$

Cumulative Frequency Graph (Ogive)

A cumulative frequency graph, also known as an Ogive, is a curve showing the cumulative frequency for a given set of data. The cumulative frequency is plotted on the y-axis against the data which is on the x-axis for un-grouped data. When dealing with grouped data, the Ogive is formed by plotting the cumulative frequency against the upper boundary of the class. An Ogive is used to study the growth rate of data as it shows the accumulation of frequency and hence its growth rate.

Example 2

Plot the cumulative frequency curve for the data set below

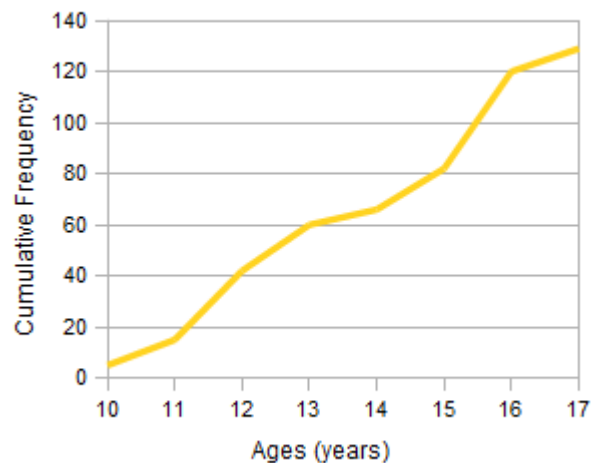
Age (years)	Frequency
10	5
11	10
12	27

13	18
14	6
15	16
16	38
17	9

Solution:

Age (years)	Frequency	Cumulative Frequency
10	5	5
11	10	$5+10 = 15$
12	27	$15+27 = 42$
13	18	$42+18 = 60$
14	6	$60+6 = 66$
15	16	$66+16 = 82$
16	38	$82+38 = 120$
17	9	$120+9 = 129$

Cumulative Frequency Graph (Ogive)



Percentiles

A percentile is a certain percentage of a set of data. Percentiles are used to observe how many of a given set of data fall within a certain percentage range; for example; a thirtieth percentile indicates data that lies the 13% mark of the entire data set.

Calculating Percentiles

Let designate a percentile as P_m where m represents the percentile we're finding, for example for the tenth percentile, m would be 10. Given that the total number of elements in the data set is N

$$P_m = \frac{m}{100} \times N$$

REPORT WRITING

TITLE

Titles are designed to **inform readers on the content of your work**. Aim for a title that is informative and **specific** to your research. Make sure that your title clearly indicates and reflects the contents of the report.

Effective titles

- Titles should be concise, descriptive and specific. Specify your topic in a subtitle if possible.
- Do not use abbreviations in titles.
- The initial working title may not adequately reflect your actual type of study. Adjust your title at the end of your project or research if possible.

ABSTRACT

The **abstract** is a brief summary of the study rationale, methodology and the main findings and conclusions of the report. Make sure to highlight in which way your paper is making a contribution to the field.

Writing an effective abstract

- An abstract should be self-contained. Do not copy your introduction in your abstract. An introduction is merely introducing your paper, while the abstract is summarizing the whole paper and its conclusions.
- Although some terminology will be inevitable, ensure that your abstract is written **accessibly**.
- Do not include general background information in your abstract. As abstracts should contain a limited number of words (400 or less), they should convey the essential information found in your paper.
- Do not include tables and figures in an abstract, unless stated otherwise.
- Do not cite references in an abstract, unless stated otherwise.
- Leave writing your abstract to the end of your project or research, as you will have a clearer picture of your main findings and conclusions at that point.

TABLE OF CONTENTS

The purpose of the table of contents is to **give an overview of the subject matter and the structure of the report**, so that readers can easily jump to a specific part of the text containing the information they need. The structure of the table of contents needs to be logical and transparent.

Formatting a table of contents

- Use "Contents" as a header for the table of contents.
- Use the correct **indentation**: main titles (Contents, Introduction, Conclusion and Appendices) should be left-aligned and titles of sections should be indented from the left margin. For each additional level, you should set an additional indent.
- Ensure your table of contents is structured in an **orderly fashion**. Make sure that titles or headings are as significant as possible.
 - Divide all subjects in subtopics, but take into account that each subtopic could belong to only one subject of a higher level. Please note that each subtopic should have at least one other subtopic at the same level.
 - Arrange all subtopics in a systematic manner, using a unique criterion.
 - Use parallel structure for headings at the same level:

LISTS OF TABLES, FIGURES AND GRAPHS

Lists of tables, figures and graphs guide readers to find the information they are looking for in the body of the text. A list of symbols and abbreviations enables readers to quickly find the meaning of each symbol and

abbreviation. For the author, writing and updating lists of symbols and abbreviations contributes to systematic notation and avoidance of double use of symbols.

Lists of tables, figures, graphs, symbols, formulas and abbreviations are generally provided after the table of contents.

INTRODUCTION

An introduction should **capture the audience's attention**. Introductions generally start by identifying and situating a problem in the existing literature. Next, introductions describe how the project or research was conducted, formulate the purpose of the research or paper and highlight in which way it is making a new contribution to the field. Finally, the introduction indicates the main points as well as the outline of the report. Do not forget to mention the relevance of the work done.

Writing a strong introduction

- Describe the **problem statement** and situate the problem in its wider context. If appropriate, the introduction defines key concepts and explains new concepts.
- Frame your research within the existing **literature** and refer to previous work. Present a comprehensive yet brief literature review and cite the sources you have used both in the text and in a reference list.
- Provide a brief overview of your **methodology** or the procedures followed.
- Finally, indicate the **outline of the report** with explicit reference to the different chapters and/or sections.

BODY

The body of the text explains in detail how the study was conducted, reports key findings and provides evidence supporting your conclusion.

Organizing the body of the text

- Divide the main body of the text in chapters. Chapters should follow a logical outline and usually include the following three main parts: methodology, results and discussion and interpretation:
- **Methods:** describing the project (elaboration), while explaining the key working principle applied. The method section elaborates on the methodology used and makes objective arguments to justify the approach taken.
- **Results:** reporting results and analyses, focusing on key results and interpretations, acknowledging limitations and implications for the interpretation of results. The line of reasoning should be clear and well-supported and assumptions should be justified.
- **Discussion:** discussing main points in relation to the problem statement, analyzing and interpreting main findings through consistent reasoning and argumentation, eliminating alternative explanations and pointing to the significance of the results.
- Check whether the body of the text is balanced. Long chapters could perhaps be split up in smaller chapters. If you have a few short chapters, verify whether they or not could be combined with other chapters.

Organizing chapters in sections

- Subdivide chapters in sections and do so in a systematic manner. Sections consist of associated paragraphs: lines of thought with each developing only one main idea. Check whether ideas are equally developed in each paragraph and that the length of your paragraphs is proportional to your paper.
- Ensure appropriate titles or headings. Use parallel structures for headings at the same level, to make your text easier to understand.
- Open each chapter with an introduction describing the procedure, main insights developed and the outline of the chapter.

- Close each chapter with a conclusion shortly summarizing the main results. Indicate and explain the connection with other chapters in the body of the text as well as what contribution the chapter has made to the whole of the paper.

CONCLUSION The main text ends with a concluding section. Remember that this section will be read by prospective readers first, therefore it should be independent of the main body of the text.

Writing your conclusion

- Recapitulate your main findings, general conclusions and contributions.
- Briefly discuss your results if appropriate and provide an answer to the problem statement.
- Conclude with recommendations for improvement and/or suggestions for further research.

Pay attention

- Do not add new information to your general conclusion.
- Relate your results to previous work, but do not include an extensive discussion.
- Identify the limitations of the study and explain their implications for the interpretation of the results.

REFERENCES

The reference section comprises **a list of all sources that were cited in the text.**

Compiling a reference list

The reference list should be **alphabetical** according to the first author's last name. If you use the **number system**, then references are listed in the order that they have been cited in the text.

Pay attention

- Compile the reference list preferably **as you write the body of the text.**
- Note that the reference list should always be **accurate** and **complete.**
- If you are listing your references alphabetically, make sure that the first line of your reference is at the margin and subsequent lines are indented (e.g. 0,5 cm).

APPENDICES

Appendices convey important but non-essential information, such as raw data, long proofs, calculations and extensive experimental results. If they would be included in the body of the text, these could harm the readability of the text.

Pay attention

- Appendices should always be numbered, captioned and referred to from the text.
- Appendices should be clear and self-contained.