Semester-V

		L	T	P	C
17PAU501A	**DSE: 1 - RESEARCH METHODOLOGY**	**6**	**2**	**-**	**6**

SCOPE

Research Methodology represents the concept of the Research Process, Research Design, Sampling Techniques and testing of Hypothesis by applying Parametric and Non Parametric test. This paper presents the various research tools and techniques in order to facilitate the research work.

OBJECTIVES

- To make the students understand the concept of Research Methodology.
- Enlighten the students knowledge in Sampling Techniques
- Impart student's knowledge in writing a good research report.

UNIT I

Research - Meaning - Scope and Significance - Utility of Research - Qualities of Good Researcher - Types of Research - Research Process - Identification - Selection and Formulation of Research Problems - Hypothesis - Research Design

UNIT II

Sampling - Methods and techniques - Sample size - Sampling Error - Field work and Data Collection - Tools of Data Collection - Interview Schedule - Questionnaire - Observation - Interview and Mailed Questionnaire - Pilot Study and Final Collection of Data - Secondary Data

UNIT III

Measurement and Scaling Techniques - Processing and Analysis of Data - Editing and Coding - Transcription and Tabulation - Statistical tools used in Research - Measures of Central Tendency - Median - Mode - Standard Deviation - Correlation Analysis - Regression Analysis.

UNIT IV

Hypothesis - Meaning - Sources -Types - Formulation - Data Analysis - Z test (mean, diff. of mean, diff. of proportion) - t-test (mean) - Paired t-test - Chi square test - Introduction to theoretical concept of ANOVA - Factor Analysis and Discriminant Analysis.

UNIT V

Interpretation - Meaning - Techniques of Interpretation - Report writing - Significance - Report Writing - Steps in Report Writing - Layout of report - Types of Reports - Oral Presentation - Executive Summary - Mechanics of Writing Research Report - Precautions for Writing Report - Norms for using Tables - Charts - and Diagrams - Appendix - Norms for using Index and Bibliography.

Note: The question paper shall cover 60% theory and 40% problem

SUGGESTED READINGS:

TEXT BOOKS

1. Kothari, C.R**.** (2009). *Research Methodology*. New Delhi: Wishwa Prakashan, Publications.

REFERENCES

1. Zikmund, Babin & Carr. (2009). *Business Research Methods* (8th ed.) New Delhi: South-Western.
2. NareshMalhotra. (2012*). Basic Marketing Research: Integration of Social Media*. New Delhi: Pearson Publisher.
3. Mark N.K. Saunders, Philip Lewis, & Adrian Thornhill. (2015). *Research Methods for Business Students* (7th ed.). New Delhi: Vikas Publisher.
4. Rao, K.V. (2012), *Research Methods for Management and Commerce*. Mumbai: Sterling Publishers Pvt., Ltd., Himalaya Publishing house.
5. Donald R.Cooper, & Pamela S.Schindler. (2008), *Business Research Methods*. New Delhi: Tata McGraw Hill.
6. Uma Sekaran. (2007). *Research Methods for Business***.** New Delhi: Wiley Publications.

Unit I – Introduction to Research

Research Methodology: Meaning – Objectives - Scope and Significance- Types of research- Characteristics of good research - Criteria for good research - Objectivity in research – Research strategies - Research process – Research in management decisions - Problems encountered by Researchers in India.

MEANING OF RESEARCH

Research is an endeavour to discover, develop and verify knowledge. It is an intellectual act that begins with the asking of questions and progressiveness through the critical examination of evidence that is both relevant and reliable to the revelation of truth. Research can be defined as the search for knowledge, or as any systematic investigation, with an open mind, to establish novel facts, solve new or existing problems, prove new ideas, or develop new theories, usually using a scientific method. The primary purpose of research is discovering, interpreting, and the development of methods and systems for the advancement of human knowledge on a wide variety of scientific matters of our world and the universe.

DEFINITIONS OF RESEARCH

Webster's New International Dictionary: "Research is careful critical enquiry or examination in seeking facts or principles, diligent investigation in order to ascertain something.

John W. Best: Research may be defined as the systematic and objective analysis and recording to controlled observations that may lead to the development of generalization, principles of theories resulting in prediction and possible ultimate control of events

Robert Ross: Research is essentially an investigation, a recording and an analysis of evidence too the purpose of gaining knowledge

Clifford Woody: Research comprises of defining and redefining problems, formulating hypothesis or suggested solutions, collecting, organizing and evaluating data making deduction and reaching conclusion and at last carefully testing conclusions to determine whether they fit in formulating hypothesis.

John Dewey: Research is considered to be the formal, systematic, intensive process of carrying on the scientific method of analysis. It involves a more systematic structure of investigation, usually in some sort of formal record of procedures and a report of result or conclusions.

Fred Kerlinger: Research is an organized enquiry designed and carried out to provide information for solving a problem.

Redman and Mory: Systematized effort to gain new knowledge

OBJECTIVES OF RESEARCH

- The purpose of research is to discover answers to questions through the application of scientific procedures. The main aim of research is to find out the truth which is hidden and which has not been discovered as yet.
- To gain familiarity with a phenomenon or to achieve new insights into it (studies with this object in view are termed as exploratory or formulative research studies);
- To portray accurately the characteristics of a particular individual, situation or a group (studies with this object in view are known as descriptive research studies);
- To determine the frequency with which something occurs or with which it is associated with something else (studies with this object in view are known as diagnostic research studies);

- To test a hypothesis of a causal relationship between variables (such studies are known as hypothesis-testing research studies).

SIGNIFICANCE OF RESEARCH

- "All progress is born of inquiry. Doubt is often better than overconfidence, for it leads to inquiry, and inquiry leads to invention" is a famous Hudson Maxim in context of which the significance of research can well be understood. Increased amounts of research make progress possible. Research inculcates scientific and inductive thinking and it promotes the development of logical habits of thinking and organisation.
- The role of research in several fields of applied economics, whether related to business or to the economy as a whole, has greatly increased in modern times. The increasingly complex nature of business and government has focused attention on the use of research in solving operational problems. Research, as an aid to economic policy, has gained added importance, both for government and business.
- **1) Research provides the basis for nearly all Government Policies in our Economic System**

For instance, government's budgets rest in part on an analysis of the needs and desires of the people and on the availability of revenues to meet these needs. The cost of needs has to be equated to probable revenues and this is a field where research is most needed. Through research we can devise alternative policies and can as well examine the consequences of each of these alternatives. Decision-making may not be a part of research, but research certainly facilitates the decisions of the policy maker. Government has also to chalk out programmes for dealing with all facets of the country's existence and most of these will be related directly

or indirectly to economic conditions. The plight of cultivators, the problems of big and small business and industry, working conditions, trade union activities, the problems of distribution, even the size and nature of defense services are matters requiring research. Thus, research is considered necessary with regard to the allocation of nation's resources. Another area in government, where research is necessary, is collecting information on the economic and social structure of the nation. Such information indicates what is happening in the economy and what changes are taking place. Collecting such statistical information is by no means a routine task, but it involves a variety of research problems. These days nearly all governments maintain large staff of research technicians or experts to carry on this work. Thus, in the context of government, research as a tool to economic policy has three distinct phases of operation, viz., (i) investigation of economic structure through continual compilation of facts; (ii) diagnosis of events that are taking place and the analysis of the forces underlying them; and (iii) the prognosis, i.e., the prediction of future developments.

- **2) Research has its Special Significance in Solving various Operational and Planning Problems of Business and Industry**

Operations research and market research, along with motivational research, are considered crucial and their results assist, in more than one way, in taking business decisions. Market research is the investigation of the structure and development of a market for the purpose of formulating efficient policies for purchasing, production and sales. Operations research refers to the application of mathematical, logical and analytical techniques to the solution of business problems of cost minimization or of profit maximization or what can be termed as optimization problems. Motivational research of determining why people behave as

they do is mainly concerned with market characteristics. In other words, it is concerned with the determination of motivations underlying the consumer (market) behaviour. All these are of great help to people in business and industry who are responsible for taking business decisions. Research with regard to demand and market factors has great utility in business. Given knowledge of future demand, it is generally not difficult for a firm, or for an industry to adjust its supply schedule within the limits of its projected capacity. Market analysis has become an integral tool of business policy these days. Business budgeting, which ultimately results in a projected profit and loss account, is based mainly on sales estimates which in turn depend on business research. Once sales forecasting is done, efficient production and investment programmes can be set up around which are grouped the purchasing and financing plans. Research, thus, replaces intuitive business decisions by more logical and scientific decisions.

- **3) Research is equally important for social scientists in studying social relationships and in seeking answers to various social problems.**

It provides the intellectual satisfaction of knowing a few things just for the sake of knowledge and also has practical utility for the social scientist to know for the sake of being able to do something better or in a more efficient manner. Research in social sciences is concerned both with knowledge for its own sake and with knowledge for what it can contribute to practical concerns. “This double emphasis is perhaps especially appropriate in the case of social science. On the one hand, its responsibility as a science is to develop a body of principles that make possible the understanding and prediction of the whole range of

human interactions. On the other hand, because of its social orientation, it is increasingly being looked to for practical guidance

- In addition to what has been stated above, the significance of research can also be understood keeping in view the following points:
- To those students who are to write a master's or Ph.D. thesis, research may mean careerism or a way to attain a high position in the social structure;
- To professionals in research methodology, research may mean a source of livelihood;
- To philosophers and thinkers, research may mean the outlet for new ideas and insights;
- To literary men and women, research may mean the development of new styles and creative work;
- To analysts and intellectuals, research may mean the generalizations of new theories.
- Thus, research is the fountain of knowledge for the sake of knowledge and an important source of providing guidelines for solving different business, governmental and social problems. It is a sort of formal training which enables one to understand the new developments in one's field in a better way.

TYPES OF RESEARCH

1) Descriptive Research

A descriptive study may be simple or complex. It determines who, what, where and how of a topic. It is concerned with describing the characteristics (e.g., the extent to which libraries are used) estimating the proportion of the people in a specified population who hold certain views or attitudes (e.g., how many favour the abolition of capital punishment?) predicting specifically (e.g., how may will cash their government bonds during a given period?) and discovering or testing whether certain

variables are associated (e.g., people who spend a good deal of time for reading, go to movies often with each other)

Descriptive study may employ any of or all the methods of data collection such as interview, questionnaire, observation, tests and cumulative record cards. In the descriptive study the researcher must be careful to make a note of the bias and extravagance that may creep in at every stage of the study – formulating the objectives of the study; designing the methods of data collection; selecting the sample; collecting, processing and analyzing the data; and reporting the findings.

2) Analytical Research

Analytical study makes use of available information by analyzing and doing critical evaluation. Analytical study makes use of higher level statistical tools which are not commonly used.

3) Applied Research

Applied research aims at finding a solution for an immediate problem faced by any business organization. This research deals with real life situations. Example: "Why have sales decreased during the last quarter"? Market research is an example of applied research. Applied research has a practical problem-solving emphasis. It brings out many new facts.

Examples:

1. Use of fibre glass body for cars instead of metal.

2. To develop a new market for the product.

4) Fundamental Research

This is otherwise known as basic research or fundamental research. Gathering knowledge for knowledge's sake is known as fundamental research. It does not have any commercial potential. It is

not connected to any practical problem. e.g. Theory of Relativity. It is only for the enrichment of the knowledge.

5) Quantitative Research

Quantitative researches are based on the measurements of quantity or amounts. It means that these type of researches deals with items which are expressed in numbers.

6) Qualitative Research

Qualitative researches deals with the qualitative phenomena. i.e. anything which cannot be expressed in numerical terms. Motivation research is an example of qualitative research.

7) Conceptual Research

Conceptual research is that related to some abstract idea(s) or theory. It is generally used by philosophers and thinkers to develop new concepts or to reinterpret existing ones.

8) Empirical Research

Empirical research relies on experience or observation alone, often without due regard for system and theory. It is data-based research, coming up with conclusions which are capable of being verified by observation or experiment. We can also call it as experimental type of research. In such a research it is necessary to get at facts firsthand, at their source, and actively to go about doing certain things to stimulate the production of desired information. In such a research, the researcher must first provide himself with a working hypothesis or guess as to the probable results. He then works to get enough facts (data) to prove or disprove his hypothesis. He then sets up experimental designs which he thinks will manipulate the persons or the materials concerned so as to bring forth the desired information. Such research is thus characterized by the experimenter's control over the variables

under study and his deliberate manipulation of one of them to study its effects. Empirical research is appropriate when proof is sought that certain variables affect other variables in some way. Evidence gathered through experiments or empirical studies is today considered to be the most powerful support possible for a given hypothesis.

9) One-time research or Longitudinal Research

the former case the research is confined to a single time-period, whereas in the latter case the research is carried on over several time-periods.

10) Field Method

Field study is a scientific enquiry aimed at discovering the relations and interactions among sociological, physiological and educational variables in real social structures and life situations like communities, schools, factories, organizations and institutions. Hence, it is called field study.

11) Exploratory Research

Explanatory research is carried, when the reason for a problem is not clear. In exploratory research, all possible reasons which are very obvious are eliminated, thereby directing the research to proceed further with limited options.

Example for Exploratory Research

Sales decline in a company may be due to:

- Inefficient service
- Improper price

- Inefficient sales force
- Ineffective promotion
- Improper quality

12) Formalized Research

Formalized research studies are those with substantial structure and with specific hypotheses to be tested.

13) Historical Research

This research is the induction of principles through research into the past and social forces which have shaped the present. Its aim is to apply reflective thinking to unsolved social problems by discovering past trends of events, facts and attitudes, and by tracing lines of development in human thought and action.

14) Decision Oriented Research

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

15) Individual and Group Research

The research undertaken by an individual is called individual research. The bulk of research activities in universities, and colleges are made by the individual. The individual research is done on the basis of one's own judgement, interest and capacity.

Group research is undertaken by several researchers. Their activities are coordinated by a director, Research conducted by a firm, trade association and government agency is performed by a team of researchers under a project director. Research in colleges and universities financed by grants is done on a group basis.

16) Operations Research

This method of research has been done for solving problems by using scientific methods and quantitative techniques. While the researchers care to study the development of methods, the industrial operations researcher evinces interest in the applications of methods to solve the pressing or critical problems of their firm.

Research can also be classified as conclusion-oriented and decision-oriented. While doing conclusion-oriented research a researcher is free to pick up a problem, redesign the enquiry and is free to conceptualize as he wishes. Decision-oriented research always implies taking a rational decision. Operational research is an example of decision-oriented research.

CRITERIA OF GOOD RESEARCH

- Whatever may be the types of research works and studies, one thing that is important is that they all meet on the common ground of scientific method employed by them. One expects scientific research to satisfy the following criteria:
- The purpose of the research should be clearly defined and common concepts be used.
- The research procedure used should be described in sufficient detail to permit another researcher to repeat the research for further advancement, keeping the continuity of what has already been attained.
- The procedural design of the research should be carefully planned to yield results that are as objective as possible.

- The researcher should report with complete frankness, flaws in procedural design and estimate their effects upon the findings.
- The analysis of data should be sufficiently adequate to reveal its significance and the methods of analysis used should be appropriate. The validity and reliability of the data should be checked carefully.
- Conclusions should be confined to those justified by the data of the research and limited to those for which the data provide an adequate basis.
- Greater confidence in research is warranted if the researcher is experienced, has a good reputation in research and is a person of integrity.

QUALITIES OF GOOD RESEARCH

1) Good research is Systematic

It means that research is structured with specified steps to be taken in a specified sequence in accordance with the well defined set of rules. Systematic characteristic of the research does not rule out creative thinking but it certainly does reject the use of guessing and intuition in arriving at conclusions.

2) Good research is Logical

This implies that research is guided by the rules of logical reasoning and the logical process of induction and deduction are of great value in carrying out research. Induction is the process of reasoning from a part to the whole whereas deduction is the process of reasoning from some premise to a conclusion which follows from that very premise.

3) Good research is Empirical

It implies that research is related basically to one or more aspects of a real situation and deals with concrete data that provides a basis for external validity to research results.

4) Good research is Replicable

This characteristic allows research results to be verified by replicating the study and thereby building a sound basis for decisions.

QUALITIES OF A GOOD RESEARCHER

A) General Qualities

1) Scientific Attitude

The first essential Quality of a successful research worker is that he must possess a scientific (systematic) frame (structure) of mind. He must have the determination (willpower / strength of mind) and ability to get the naked (hidden) facts and not to be influenced by one's own wishes.

As human beings he has certain praises (admiration) and prejudices (bias). He has also certain precarceived notions (ideas) about the problems being researched. He should keep all these things with him.

2) Imagination and Insight

Researcher must possess high degree of imagination. He should be able to go deeper and deeper into the realm (area) of abstract social phenomena (fact / event) and visualize the intangible aspects (features) of the society

3) Perseverance

Work of scientific research requires steady of mind. Researcher should not get easily discouraged. It is equally possible that he might subsequently feel that the choice of the problem was

wrong. Inspite of all this he must have more courage to face the difficulties and work patiently and continuously over long periods

4) Quick Grasping Power

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

5) Clarity of Thinking

A good researcher should have clear idea about the terminology that he is going to use.

B) Specific Qualities

1) Knowledge of the Subject

The researcher should be enough knowledge in his area of research. Such knowledge helps him in preparing questionnaire and schedule to get proper information. He can enter into face to face discussion and remove any doubts arising the minds of the people regarding the study

2) Knowledge of the technique of Research

Researcher should have basic idea on tools used in his research

3) Personal Taste in the Study

A personal taste in the study will inspire him and keep his morale (confidence) in times of difficulties. A forced work is often monotonous and very tiresome

4) Familiarity about the Information

The researcher should be familiar with the people whom he is studying. Familiarity will help him to get intimate (close) information

5) Unbiased Attitude

The researcher should have no preconceptions (idea / bias) about the subject under study. He should go to his research with absolutely a clean state. He should maintain an open mind and look for data which would substantiate (validate / verify) and give his theory a new meaning.

RESEARCH PROCESS

Research process consists of series of actions or steps necessary to effectively carry out research and the desired sequencing of these steps.

1. Formulating the Research Problem;
2. Extensive literature survey;
3. Developing the Hypothesis;
4. Preparing the Research Design;
5. Determining Sample Design;
6. Collecting the Data;
7. Execution of the Project;
8. Analysis of Data;
9. Hypothesis testing;
10. Generalizations and Interpretation, and
11. Preparation of the Report or Presentation of the Results

1) Formulating the Research Problem

There are two types of research problems, viz., those which relate to states of nature and those which relate to relationships between variables. At the very outset the researcher must single out the

problem he wants to study, i.e., he must decide the general area of interest or aspect of a subject-matter that he would like to inquire into. Initially the problem may be stated in a broad general way and then the ambiguities, if any, relating to the problem be resolved. Then, the feasibility of a particular solution has to be considered before a working formulation of the problem can be set up. The formulation of a general topic into a specific research problem, thus, constitutes the first step in a scientific enquiry. Essentially two steps are involved in formulating the research problem, viz., understanding the problem thoroughly, and rephrasing the same into meaningful terms from an analytical point of view.

The best way of understanding the problem is to discuss it with one's own colleagues or with those having some expertise in the matter. In an academic institution the researcher can seek the help from a guide who is usually an experienced man and has several research problems in mind. Often, the guide puts forth the problem in general terms and it is up to the researcher to narrow it down and phrase the problem in operational terms. In private business units or in governmental organisations, the problem is usually earmarked by the administrative agencies with whom the researcher can discuss as to how the problem originally came about and what considerations are involved in its possible solutions.

The researcher must at the same time examine all available literature to get himself acquainted with the selected problem. He may review two types of literature the conceptual literature concerning the concepts and theories, and the empirical literature consisting of studies made earlier which are similar to the one proposed. The basic outcome of this review will be the knowledge as to what data and other materials are available for operational purposes which will enable the researcher to specify his own research problem in a meaningful context. After this the researcher rephrases the problem

into analytical or operational terms i.e., to put the problem in as specific terms as possible. This task of formulating, or defining, a research problem is a step of greatest importance in the entire research process. The problem to be investigated must be defined unambiguously for that will help discriminating relevant data from irrelevant ones. Care must, however, be taken to verify the objectivity and validity of the background facts concerning the problem.

2) Extensive Literature Survey

Once the problem is formulated, a brief summary of it should be written down. It is compulsory for a research worker writing a thesis for a Ph.D. degree to write a synopsis of the topic and submit it to the necessary Committee or the Research Board for approval. At this juncture the researcher should undertake extensive literature survey connected with the problem. For this purpose, the abstracting and indexing journals and published or unpublished bibliographies are the first place to go to. Academic journals, conference proceedings, government reports, books etc., must be tapped depending on the nature of the problem. In this process, it should be remembered that one source will lead to another. The earlier studies, if any, which are similar to the study in hand should be carefully studied. A good library will be a great help to the researcher at this stage.

3) Developing the Hypothesis

After extensive literature survey, researcher should state in clear terms the working hypothesis or hypotheses. Working hypothesis is tentative assumption made in order to draw out and test its logical or empirical consequences. As such the manner in which research hypotheses are developed is particularly important since they provide the focal point for research. They also affect the manner in which tests must be conducted in the analysis of data and indirectly the quality of data which is required for the analysis. In most types of research, the development of working hypothesis plays an

important role. Hypothesis should be very specific and limited to the piece of research in hand because it has to be tested. The role of the hypothesis is to guide the researcher by delimiting the area of research and to keep him on the right track. It sharpens his thinking and focuses attention on the more important facets of the problem. It also indicates the type of data required and the type of methods of data analysis to be used.

How does one go about developing working hypotheses? The answer is by using the following approach:

- Discussions with colleagues and experts about the problem, its origin and the objectives in seeking a solution;
- Examination of data and records, if available, concerning the problem for possible trends, peculiarities and other clues;
- Review of similar studies in the area or of the studies on similar problems; and
- Exploratory personal investigation which involves original field interviews on a limited scale with interested parties and individuals with a view to secure greater insight into the practical aspects of the problem.

Thus, working hypotheses arise as a result of a-priori thinking about the subject, examination of the available data and material including related studies and the counsel of experts and interested parties. Working hypotheses are more useful when stated in precise and clearly defined terms. It may as well be remembered that occasionally we may encounter a problem where we do not need working hypotheses, especially in the case of exploratory or formulative researches which do not aim at testing the hypothesis. But as a general rule, specification of working hypotheses in another basic step of the research process in most research problems.

4) Preparing the Research Design

The research problem having been formulated in clear cut terms, the researcher will be required to prepare a research design, i.e., he will have to state the conceptual structure within which research would be conducted. The preparation of such a design facilitates research to be as efficient as possible yielding maximal information. In other words, the function of research design is to provide for the collection of relevant evidence with minimal expenditure of effort, time and money. But how all these can be achieved depends mainly on the research purpose. Research purposes may be grouped into four categories, viz., (i) Exploration, (ii) Description, (iii) Diagnosis, and (iv) Experimentation. A flexible research design which provides opportunity for considering many different aspects of a problem is considered appropriate if the purpose of the research study is that of exploration. But when the purpose happens to be an accurate description of a situation or of an association between variables, the suitable design will be one that minimises bias and maximises the reliability of the data collected and analysed.

There are several research designs, such as, experimental and non-experimental hypothesis testing. Experimental designs can be either informal designs (such as before-and-after without control, after-only with control, before-and-after with control) or formal designs (such as completely randomized design, randomized block design, Latin square design, simple and complex factorial designs), out of which the researcher must select one for his own project.

The preparation of the research design, appropriate for a particular research problem, involves usually the consideration of the following:

a) The means of obtaining the information;

b) The availability and skills of the researcher and his staff (if any);

c) Explanation of the way in which selected means of obtaining information will be organized and the reasoning leading to the selection;

d) The time available for research; and

e) The cost factor relating to research, i.e., the finance available for the purpose.

5) Determining Sample Design

All the items under consideration in any field of inquiry constitute a 'universe' or 'population'. A complete enumeration of all the items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry when all the 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 observations 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. Not only this, census inquiry is not possible in practice under many circumstances. For instance, blood testing is done only on sample basis. Hence, quite often we select only a few items from the universe for our study purposes. The items so selected constitute what is technically called a sample.

The researcher must decide the way of selecting a sample or what is popularly known as the sample design. In other words, a sample design is a definite plan determined before any data are actually collected for obtaining a sample from a given population. Thus, the plan to select 12 of a city's 200 drugstores in a certain way constitutes a sample design. Samples can be either probability samples or non-probability samples. With probability samples each element has a known probability of being included in the sample but the non-probability samples do not allow the researcher to determine this probability. Probability samples are those based on simple random sampling,

systematic sampling, stratified sampling, cluster/area sampling whereas non-probability samples are those based on convenience sampling, judgement sampling and quota sampling techniques. A brief mention of the important sample designs is as follows:

a) Deliberate Sampling

b) Simple Random Sampling

c) Systematic Sampling

d) Stratified Sampling

e) Quota Sampling

f) Cluster Sampling and Area Sampling

g) Multi-stage Sampling

h) Sequential Sampling

6) Collecting the Data

a) By Observation

This method implies the collection of information by way of investigator's own observation, without interviewing the respondents. The information obtained relates to what is currently happening and is not complicated by either the past behaviour or future intentions or attitudes of respondents. This method is no doubt an expensive method and the information provided by this method is also very limited. As such this method is not suitable in inquiries where large samples are concerned.

b) Through Personal Interview

The investigator follows a rigid procedure and seeks answers to a set of pre-conceived questions through personal interviews. This method of collecting data is usually carried out in a structured way where output depends upon the ability of the interviewer to a large extent.

c) Through Telephone Interview

This method of collecting information involves contacting the respondents on telephone itself. This is not a very widely used method but it plays an important role in industrial surveys in developed regions, particularly, when the survey has to be accomplished in a very limited time.

d) By Mailing of Questionnaire

The researcher and the respondents do come in contact with each other if this method of survey is adopted. Questionnaires are mailed to the respondents with a request to return after completing the same. It is the most extensively used method in various economic and business surveys. Before applying this method, usually a Pilot Study for testing the questionnaire is conduced which reveals the weaknesses, if any, of the questionnaire. Questionnaire to be used must be prepared very carefully so that it may prove to be effective in collecting the relevant information.

e) Through Schedules

Under this method the enumerators are appointed and given training. They are provided with schedules containing relevant questions. These enumerators go to respondents with these schedules. Data are collected by filling up the schedules by enumerators on the basis of replies given by respondents. Much depends upon the capability of enumerators so far as this method is concerned. Some occasional field checks on the work of the enumerators may ensure sincere work.

The researcher should select one of these methods of collecting the data taking into consideration the nature of investigation, objective and scope of the inquiry, finanical resources, available time and the desired degree of accuracy. Though he should pay attention to all these factors

but much depends upon the ability and experience of the researcher. In this context Dr A.L Bowley very aptly remarks that in collection of statistical data commonsense is the chief requisite and experience the chief teacher.

7) Execution of the Project

Execution of the project is a very important step in the research process. If the execution of the project proceeds on correct lines, the data to be collected would be adequate and dependable. The researcher should see that the project is executed in a systematic manner and in time. If the survey is to be conducted by means of structured questionnaires, data can be readily machine-processed. In such a situation, questions as well as the possible answers may be coded. If the data are to be collected through interviewers, arrangements should be made for proper selection and training of the interviewers. The training may be given with the help of instruction manuals which explain clearly the job of the interviewers at each step. Occasional field checks should be made to ensure that the interviewers are doing their assigned job sincerely and efficiently. A careful watch should be kept for unanticipated factors in order to keep the survey as much realistic as possible. This, in other words, means that steps should be taken to ensure that the survey is under statistical control so that the collected information is in accordance with the pre-defined standard of accuracy. If some of the respondents do not cooperate, some suitable methods should be designed to tackle this problem. One method of dealing with the non-response problem is to make a list of the non-respondents and take a small sub-sample of them, and then with the help of experts vigorous efforts can be made for securing response.

8) Analysis of Data

After the data have been collected, the researcher turns to the task of analysing them. The analysis of data requires a number of closely related operations such as establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences. The unwieldy data should necessarily be condensed into a few manageable groups and tables for further analysis. Thus, researcher should classify the raw data into some purposeful and usable categories. Coding operation is usually done at this stage through which the categories of data are transformed into symbols that may be tabulated and counted. Editing is the procedure that improves the quality of the data for coding. With coding the stage is ready for tabulation. Tabulation is a part of the technical procedure wherein the classified data are put in the form of tables. The mechanical devices can be made use of at this juncture. A great deal of data, specially in large inquiries, is tabulated by computers. Computers not only save time but also make it possible to study large number of variables affecting a problem simultaneously.

Analysis work after tabulation is generally based on the computation of various percentages, coefficients, etc., by applying various well defined statistical formulae. In the process of analysis, relationships or differences supporting or conflicting with original or new hypotheses should be subjected to tests of significance to determine with what validity data can be said to indicate any conclusion(s). For instance, if there are two samples of weekly wages, each sample being drawn from factories in different parts of the same city, giving two different mean values, then our problem may be whether the two mean values are significantly different or the difference is just a matter of chance. Through the use of statistical tests we can establish whether such a difference is a real one or is the result of random fluctuations. If the difference happens to be real, the inference will be that the two samples come from different universes and if the difference is due to chance, the conclusion would be

that the two samples belong to the same universe. Similarly, the technique of analysis of variance can help us in analysing whether three or more varieties of seeds grown on certain fields yield significantly different results or not. In brief, the researcher can analyse the collected data with the help of various statistical measures.

9) Hypothesis Testing

After analysing the data as stated above, the researcher is in a position to test the hypotheses, if any, he had formulated earlier. Do the facts support the hypotheses or they happen to be contrary? This is the usual question which should be answered while testing hypotheses. Various tests, such as Chi square test, t-test, F-test, have been developed by statisticians for the purpose. The hypotheses may be tested through the use of one or more of such tests, depending upon the nature and object of research inquiry. Hypothesis-testing will result in either accepting the hypothesis or in rejecting it. If the researcher had no hypotheses to start with, generalisations established on the basis of data may be stated as hypotheses to be tested by subsequent researches in times to come.

10) Generalizations and Interpretation

If a hypothesis is tested and upheld several times, it may be possible for the researcher to arrive at generalisation, i.e., to build a theory. As a matter of fact, the real value of research lies in its ability to arrive at certain generalisations. If the researcher had no hypothesis to start with, he might seek to explain his findings on the basis of some theory. It is known as interpretation. The process of interpretation may quite often trigger off new questions which in turn may lead to further researches.

11) Preparation of the Report or Presentation of the Results

Finally, the researcher has to prepare the report of what has been done by him. Writing of report must be done with great care keeping in view the following:

1) The layout of the report should be as follows: (i) the preliminary pages; (ii) the main text, and (iii) the end matter.

In its preliminary pages the report should carry title and date followed by acknowledgements and foreword. Then there should be a table of contents followed by a list of tables and list of graphs and charts, if any, given in the report.

The main text of the report should have the following parts:

a) Introduction: It should contain a clear statement of the objective of the research and an explanation of the methodology adopted in accomplishing the research. The scope of the study along with various limitations should as well be stated in this part.

b) Summary of Findings: After introduction there would appear a statement of findings and recommendations in non-technical language. If the findings are extensive, they should be summarised.

c) Main Report: The main body of the report should be presented in logical sequence and broken-down into readily identifiable sections.

d) Conclusion: Towards the end of the main text, researcher should again put down the results of his research clearly and precisely. In fact, it is the final summing up.

At the end of the report, appendices should be enlisted in respect of all technical data. Bibliography, i.e., list of books, journals, reports, etc., consulted, should also be given in the end. Index should also be given specially in a published research report.

2) Report should be written in a concise and objective style in simple language avoiding vague expressions such as 'it seems,' 'there may be', and the like.

3) Charts and illustrations in the main report should be used only if they present the information more clearly and forcibly.

4) Calculated 'confidence limits' must be mentioned and the various constraints experienced in conducting research operations may as well be stated.

PROBLEMS ENCOUNTERED BY RESEARCHERS IN INDIA

Researchers in India, particularly those engaged in empirical research, are facing several problems.Some of the important problems are as follows:

1. The lack of a scientific training in the methodology of research is a great impedimentfor researchers in our country. There is paucity of competent researchers. Many researchers take a leap in the dark without knowing research methods. Most of the work, which goes in the name of research is not methodologically sound. Research to many researchers and even to their guides, is mostly a scissor and paste job without any insight shed on the collated materials. The consequence is obvious, viz., the research results, quite often, donot reflect the reality or realities. Thus, a systematic study of research methodology is anurgent necessity. Before undertaking research projects, researchers should be well equippedwith all the methodological aspects. As such, efforts should be made to provide shortduration intensive courses for meeting this requirement.

2. There is insufficient interaction between the university research departments on one sideand business establishments, government departments and research institutions on the other side. A great deal of primary data of non-confidential nature remains untouched/untreated by the researchers for want of proper contacts. Efforts should be made to develop satisfactory liaison among all concerned for better and realistic researches. There is need for developing some mechanisms of a university—industry interaction programme so that academics can get ideas from practitioners on what needs to be researched and practitioners can apply the research done by the academics.

3. Most of the business units in our country do not have the confidence that the materialsupplied by them to researchers will not be misused and as such they are often reluctant insupplying the needed information to researchers. The concept of secrecy seems to besacrosanct to business organisations in the country so much so that it proves an impermeable barrier to researchers. Thus, there is the need for generating the confidence that the information/data obtained from a business unit will not be misused.

4. Research studies overlapping one another are undertaken quite often for want of adequate information. This results in duplication and fritters away resources. This problem can be solved by proper compilation and revision, at regular intervals, of a list of subjects on which and the places where the research is going on. Due attention should be given toward identification of research problems in various disciplines of applied science which are of immediate concern to the industries.

5. There does not exist a code of conduct for researchers and inter-university and interdepartmental rivalries are also quite common. Hence, there is need for developing a code of conduct for researchers which, if adhered sincerely, can win over this problem.

6. Many researchers in our country also face the difficulty of adequate and timely secretarial assistance, including computerial assistance. This causes unnecessary delays in the completion of research studies. All possible efforts be made in this direction so that efficient secretarial assistance is made available to researchers and that too well in time. University Grants Commission must play a dynamic role in solving this difficulty.

7. Library management and functioning is not satisfactory at many places and much ofthe time and energy of researchers are spent in tracing out the books, journals, reports, etc., rather than in tracing out relevant material from them.

8. There is also the problem that many of our libraries are not able to get copies of oldand new Acts/Rules, reports and other government publications in time. This problemis felt more in libraries which are away in places from Delhi and/or the state capitals. Thus,efforts should be made for the regular and speedy supply of all governmental publicationsto reach our libraries.

9. There is also the difficulty of timely availability of published data from variousgovernment and other agencies doing this job in our country. Researcher also faces theproblem on account of the fact that the published data vary quite significantly because ofdifferences in coverage by the concerning agencies.

10. There may, at times, take place the problem of conceptualization and also problemsrelating to the process of data collection and related things.

RESEARCH PROBLEM

SELECTING THE PROBLEM

The research problem undertaken for study must be carefully selected. The task is a difficult one, although it may not appear to be so. Help may be taken from a research guide in this connection.

A problem must spring from the researcher's mind like a plant springing from its own seed. If our eyes need glasses, it is not the optician alone who decides about the number of the lens we require. We have to see ourselves and enable him to prescribe for us the right number by cooperating with him. Thus, a research guide can at the most only help a researcher choose a subject. However, the following points may be observed by a researcher in selecting a research

problem or a subject for research:

(i) Subject which is overdone should not be normally chosen, for it will be a difficult task to throw any new light in such a case.

(ii) Controversial subject should not become the choice of an average researcher.

(iii) Too narrow or too vague problems should be avoided.

(iv) The subject selected for research should be familiar and feasible so that the related research material or sources of research are within one's reach. Even then it is quite difficult to supply definitive ideas concerning how a researcher should obtain ideas for his research.

For this purpose, a researcher should contact an expert or a professor in the University who is already engaged in research. He may as well read articles published in current literature available on the subject and may think how the techniques and ideas discussed therein might be applied to the solution of other problems. He may discuss with others what he has in mind concerning a problem. In this way he should make all possible efforts in selecting a problem.

(v) The importance of the subject, the qualifications and the training of a researcher, the costs involved, the time factor are few other criteria that must also be considered in selecting a problem. In other words, before the final selection of a problem is done, a researcher must ask himself the following questions:

(a) Whether he is well equipped in terms of his background to carry out the research?

(b) Whether the study falls within the budget he can afford?

(c) Whether the necessary cooperation can be obtained from those who must participate in research as subjects?

If the answers to all these questions are in the affirmative, one may become sure so far as the practicability of the study is concerned.

(vi) The selection of a problem must be preceded by a preliminary study. This may not be necessary when the problem requires the conduct of a research closely similar to one that has already been done. But when the field of inquiry is relatively new and does not have available a set of well developed techniques, a brief feasibility study must always be undertaken.

Technique involved in Defining a Problem

(i) Statement of the problem in a general way: First of all the problem should be stated in a broad general way, keeping in view either some practical concern or some scientific or intellectual interest. For this purpose, the researcher must immerse himself thoroughly in the subject matter concerning which he wishes to pose a problem. In case of social research, it is considered advisable to do some field observation and as such the researcher may undertake some sort of preliminary survey or what is often called *pilot survey*. Then the researcher can himself state the problem or he can seek the guidance of the guide or the subject expert in accomplishing this task. Often, the guide puts forth the problem in general terms, and it is then up to the researcher to narrow it down and phrase the problem in operational terms. In case there is some directive from an organizational authority, the problem then can be stated accordingly. The problem stated in a broad general way may contain various ambiguities which must be resolved by cool thinking and rethinking over the problem. At the

same time the feasibility of a particular solution has to be considered and the same should be kept in view while stating the problem.

(ii) Understanding the nature of the problem: The next step in defining the problem is to understand its origin and nature clearly. The best way of understanding the problem is to discuss it with those who first raised it in order to find out how the problem originally came about and with what objectives in view. If the researcher has stated the problem himself, he should consider once again all those points that induced him to make a general statement concerning the problem. For better understanding of the nature of the problem involved, he can enter into discussion with those who have a good knowledge of the problem concerned or similar other problems. The researcher should also keep in view the environment within which the problem is to be studied and understood.

(iii) Surveying the available literature: All available literature concerning the problem at hand must necessarily be surveyed and examined before a definition of the research problem is given. This means that the researcher must be well-conversant with relevant theories in the field, reports and records as also all other relevant literature. He must devote sufficient time in reviewing of research already undertaken on related problems. This is done to find out what data and other materials, if any, are available for operational purposes.

(v) Rephrasing the research problem: Finally, the researcher must sit to rephrase the research problem into a working proposition. Once the nature of the problem has been clearly understood, the environment (within which the problem has got to be studied) has been defined, discussions over the problem have taken place and the available literature has been surveyed and examined, rephrasing the problem into analytical or operational terms is not a difficult task. Through rephrasing, the researcher

puts the research problem in as specific terms as possible so that it may become operationally viable and may help in the development of working hypotheses.

(a) Technical terms and words or phrases, with special meanings used in the statement of the problem, should be clearly defined.

(b) Basic assumptions or postulates (if any) relating to the research problem should be clearly stated.

(c) A straight forward statement of the value of the investigation (i.e., the criteria for the selection of the problem) should be provided.

(d) The suitability of the time-period and the sources of data available must also be considered by the researcher in defining the problem.

(e) The scope of the investigation or the limits within which the problem is to be studied must mentioned explicitly in defining a research problem

RESEARCH DESIGN

The formidable problem that follows the task of defining the research problem is the preparation of the design of the research project, popularly known as the "research design". Decisions regarding what, where, when, how much, by what means concerning an inquiry or a research study constitute a research design. "A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in

procedure."1 In fact, the research design is the conceptual structure within which research is conducted; it constitutes the blueprint for the collection, measurement and analysis of data. As such the design includes an outline of what the researcher will do from writing the hypothesis and its operational implications to the final analysis of data. More explicitly, the desing decisions happen to be in respect of:

1. What is the study about?
2. Why is the study being made?
3. Where will the study be carried out?
4. What type of data is required?
5. Where can the required data be found?
6. What periods of time will the study include?
7. What will be the sample design?
8. What techniques of data collection will be used?
9. How will the data be analyzed?
10. In what style will the report be prepared?

Keeping in view the above stated design decisions, one may split the overall research design into the following parts:

a) The sampling design which deals with the method of selecting items to be observed for the given study;

b) The observational design which relates to the conditions under which the observations are to be made;

c) The statistical design which concerns with the question of how many items are to be observed and how the information and data gathered are to be analysed; and

d) The operational design which deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

From what has been stated above, we can state the important features of a research design as under:

i) It is a plan that specifies the sources and types of information relevant to the research problem.

ii) It is a strategy specifying which approach will be used for gathering and analysing the data.

iii) It also includes the time and cost budgets since most studies are done under these two constraints.

In brief, research design must, at least, contain—(a) a clear statement of the research problem; (b) procedures and techniques to be used for gathering information;

(c) The population to be studied; and (d) methods to be used in processing and analyzing data.

METHODS OF RESEARCH DESIGN

Different research designs can be conveniently described if we categorize them as: (1) research design in case of exploratory research studies; (2) research design in case of descriptive and diagnostic research studies, and (3) research design in case of hypothesis-testing research studies.

1) Research design in case of exploratory research studies

Exploratory research studies are also termed as formulative research studies. The main purpose of such studies is that of formulating a problem for more precise (accurate) investigation or of developing the working hypotheses from an operational point of view. The major emphasis in such studies is on the discovery of ideas and insights. As such the research design appropriate for

such studies must be flexible enough to provide opportunity for considering different aspects of a problem under study. Inbuilt flexibility in research design is needed because the research problem, broadly defined initially, is transformed into one with more precise meaning in exploratory studies, which fact may necessitate changes in the research procedure for gathering relevant data. Generally, the following three methods in the context of research design for such studies are talked about: (a) the survey of concerning literature; (b) the experience survey and (c) the analysis of 'insight-stimulating' examples.

The survey of concerning literature happens to be the most simple and fruitful method of formulating precisely the research problem or developing hypothesis. Hypotheses stated by earlier workers may be reviewed and their usefulness be evaluated as a basis for further research. It may also be considered whether the already stated hypotheses suggest new hypothesis. In this way the researcher should review and build upon the work already done by others, but in cases where hypotheses have not yet been formulated, his task is to review the available material for deriving the relevant hypotheses from it.

Besides, the bibliographical survey of studies, already made in one's area of interest may as well as made by the researcher for precisely formulating the problem. He should also make an attempt to apply concepts and theories developed in different research contexts to the area in which he is himself working. Sometimes the works of creative writers also provide a fertile ground for hypothesis-formulation and as such may be looked into by the researcher.

Experience survey means the survey of people who have had practical experience with the problem to be studied. The object of such a survey is to obtain insight into the relationships between variables and new ideas relating to the research problem. For such a survey people who are competent

and can contribute new ideas may be carefully selected as respondents to ensure a representation of different types of experience. The respondents so selected may then be interviewed by the investigator. The researcher must prepare an interview schedule for the systematic questioning of informants. But the interview must ensure flexibility in the sense that the respondents should be allowed to raise issues and questions which the investigator has not previously considered. Generally, the experience-collecting interview is likely to be long and may last for few hours. Hence, it is often considered desirable to send a copy of the questions to be discussed to the respondents well in advance. This will also give an opportunity to the respondents for doing some advance thinking over the various issues involved so that, at the time of interview, they may be able to contribute effectively. Thus, an experience survey may enable the researcher to define the problem more concisely (briefly) and help in the formulation of the research hypothesis. This survey may as well provide information about the practical possibilities for doing different types of research.

Analysis of 'insight-stimulating' examples is also a fruitful method for suggesting hypotheses for research. It is particularly suitable in areas where there is little experience to serve as a guide. This method consists of the intensive study of selected instances of the phenomenon in which one is interested. For this purpose the existing records, if any, may be examined, the unstructured interviewing may take place, or some other approach may be adopted. Attitude of the investigator, the intensity of the study and the ability of the researcher to draw together diverse information into a unified interpretation are the main features which make this method an appropriate procedure for evoking insights.

Now, what sort of examples is to be selected and studied? There is no clear cut answer to it. Experience indicates that for particular problems certain types of instances are more appropriate than

others. One can mention few examples of 'insight-stimulating' cases such as the reactions of strangers, the reactions of marginal individuals, the study of individuals who are in transition from one stage to another, the reactions of individuals from different social strata and the like. In general, cases that provide sharp contrasts or have striking features are considered relatively more useful while adopting this method of hypotheses formulation.

Thus, in an exploratory of formulative research study which merely leads to insights or hypotheses, whatever method or research design outlined above is adopted, the only thing essential is that it must continue to remain flexible so that many different facets of a problem may be considered as and when they arise and come to the notice of the researcher.

2) Research design in case of descriptive and diagnostic research studies

Descriptive research studies are those studies, which are concerned with describing the characteristics of a particular individual, or of a group, whereas diagnostic research studies determine the frequency with which something occurs or its association with something else. The studies concerning whether certain variables are associated are examples of diagnostic research studies. As against this, studies concerned with specific predictions, with narration of facts and characteristics concerning individual, group or situation are all examples of descriptive research studies. Most of the social research comes under this category. From the point of view of the research design, the descriptive as well as diagnostic studies share common requirements and as such we may group together these two types of research studies. In descriptive as well as in diagnostic studies, the researcher must be able to define clearly, what he wants to measure and must find adequate methods for measuring it along with a clear cut definition of 'population' he wants to study. Since the aim is to obtain complete and accurate information in the said studies, the procedure to be used must be

carefully planned. The research design must make enough provision for protection against bias and must maximize reliability, with due concern for the economical completion of the research study. The design in such studies must be rigid and not flexible and must focus attention on the following:

- Formulating the objective of the study (what the study is about and why is it being made?)
- Designing the methods of data collection (what techniques of gathering data will be adopted?)
- Selecting the sample (how much material will be needed?)
- Collecting the data (where can the required data be found and with what time period should the data be related?)
- Processing and analyzing the data.
- Reporting the findings.

In a descriptive/diagnostic study the first step is to specify the objectives with sufficient precision (accuracy) to ensure that the data collected are relevant. If this is not done carefully, the study may not provide the desired information.

Then comes the question of selecting the methods by which the data are to be obtained. In other words, techniques for collecting the information must be devised. Several methods (viz., observation, questionnaires, interviewing, examination of records, etc.), with their merits and limitations, are available for the purpose and the researcher may user one or more of these methods which have been discussed in detail in later chapters. While designing data-collection procedure, adequate safeguards against bias and unreliability must be ensured. Whichever method is selected, questions must be well examined and be made unambiguous; interviewers must be instructed not to express their own opinion; observers must be trained so that they uniformly record a given item of behaviour. It is

always desirable to pretest the data collection instruments before they are finally used for the study purposes. In other words, we can say that "structured instruments" are used in such studies.

In most of the descriptive/diagnostic studies the researcher takes out sample(s) and then wishes to make statements about the population on the basis of the sample analysis or analyses. More often than not, sample has to be designed. Here we may only mention that the problem of designing samples should be tackled in such a fashion that the samples may yield accurate information with a minimum amount of research effort. Usually one or more forms of probability sampling, or what is often described as random sampling, are used.

To obtain data free from errors introduced by those responsible for collecting them, it is necessary to supervise closely the staff of field workers as they collect and record information. Checks may be set up to ensure that the data collecting staff perform their duty honestly and without prejudice. "As data are collected, they should be examined for completeness, comprehensibility, consistency and reliability."

The data collected must be processed and analysed. This includes steps like coding the interview replies, observations, etc.; tabulating the data; and performing several statistical computations. To the extent possible, the processing and analysing procedure should be planned in detail before actual work is started. This will prove economical in the sense that the researcher may avoid unnecessary labour such as preparing tables for which he later finds he has no use or on the other hand, re-doing some tables because he failed to include relevant data. Coding should be done carefully to avoid error in coding and for this purpose the reliability of coders needs to be checked. Similarly, the accuracy of tabulation may be checked by having a sample of the tables re-done. In case of mechanical tabulation the material (i.e., the collected data or information) must be entered on appropriate cards, which is

usually done by punching holes corresponding to a given code. The accuracy of punching is to be checked and ensured. Finally, statistical computations are needed and as such averages, percentages and various coefficients must be worked out. Probability and sampling analysis may as well be used. The appropriate statistical operations, along with the use of appropriate tests of significance should be carried out to safeguard the drawing of conclusions concerning the study.

Last of all comes the question of reporting the findings. This is the task of communicating the findings to others and the researcher must do it in an efficient manner. The layout of the report needs to be well planned so that all things relating to the research study may be well presented in simple and effective style.

Thus, the research design in case of descriptive/diagnostic studies is a comparative design throwing light on all points narrated above and must be prepared keeping in view the objective(s) of the study and the resources available. However, it must ensure the minimisation of bias and maximisation of reliability of the evidence collected. The said design can be appropriately referred to as a survey design since it takes into account all the steps involved in a survey concerning a phenomenon to be studied.

3) Research design in case of hypothesis-testing research studies

Hypothesis-testing research studies (generally known as experimental studies) are those where the researcher tests the hypotheses of causal relationships between variables. Such studies require procedures that will not only reduce bias and increase reliability, but will permit drawing inferences about causality. Usually experiments meet this requirement. Hence, when we talk of research design in such studies, we often mean the design of experiments.

Professor R.A. Fisher's name is associated with experimental designs. Beginning of such designs was made by him when he was working at Rothamsted Experimental Station (Centre for Agricultural Research in England). As such the study of experimental designs has its origin in agricultural research. Professor Fisher found that by dividing agricultural fields or plots into different blocks and then by conducting experiments in each of these blocks, whatever information is collected and inferences drawn from them, happens to be more reliable. This fact inspired him to develop certain experimental designs for testing hypotheses concerning scientific investigations. Today, the experimental designs are being used in researches relating to phenomena of several disciplines. Since experimental designs originated in the context of agricultural operations, we still use, though in a technical sense, several terms of agriculture (such as treatment, yield, plot, block etc.) in experimental designs.

KAHE

Unit II – Research Design and sampling

Research Design: Meaning- Classification- Features – Importance - Steps in research design - Selection of research problem - Sample design - Meaning- Concepts - Steps in sampling - Criteria for good sample design – Types of sampling – Probability and Non probability sampling techniques – Sample size - Determination of sample size.

SAMPLING

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. Then, the researcher may make use of Sampling.

STEPS IN SAMPLE DESIGN

1) 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 a dice etc. are examples of infinite universes.

2) 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 a 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.

3) Source List

It is also known as 'sampling frame' from which sample is to be drawn. It contains the names of all items of universe. 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.

4) 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 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.

5) 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 characteristics in the population or we may be interested in knowing some average or the other measure concerning the population. There may 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.

6) 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.

7) 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 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.

CHARACTERISTICS OF A GOOD SAMPLE DESIGN

- Sample design must result in a truly representative sample.
- Sample design must be such which results in a small sampling error.
- Sample design must be viable in the context of funds available for the research study.
- Sample design must be such so that systematic bias can be controlled in a better way.
- 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.

DETERMINATION OF SAMPLE SIZE

1) Nature of Universe

Universe may be either homogenous or heterogeneous in nature. If the items of the universe are homogenous, a small sample can serve the purpose. But if the items are heterogeneous, a large sample would be required. Technically, this can be termed as the dispersion factor.

2) Number of Classes Proposed

If many class-groups are to be formed, a large sample would be required because a small sample might not be able to give a reasonable number of items in each class-group.

3) Nature of Study

If items are to be intensively and continuously studied, the sample should be small. For a general survey the size of the sample should be large, but a small sample is considered appropriate in technical surveys.

4) Type of Sampling

Sampling technique plays an important part in determining the size of the sample. A small random sample is apt to be much superior to a larger but badly selected sample.

5) Standard of Accuracy and Acceptable Confidence Level

If the standard of accuracy or the level of precision is to be kept high, we shall require relatively larger sample. For doubling the accuracy for a fixed significance level, the sample size has to be increased fourfold.

6) Availability of Finance

In practice, size of the sample depends upon the amount of money available for the study purposes. This factor should be kept in view while determining the size of sample for large samples result in increasing the cost of sampling estimates.

7) Other Considerations

Nature of units, size of the population, size of questionnaire, availability of trained investigators, the conditions under which the sample is being conducted, the time available for completion of the study are a few other considerations to which a researcher must pay attention while selecting the size of sample.

SAMPLING TECHNIQUES

1) Probability Sampling

In probability sample, every unit in the population has equal chances for being selected as a sample unit.

2) Non–probability Sampling

In non probability sampling, units in the population has unequal or zero chances for being selected as a sample unit.

RANDOM SAMPLING PROCEDURES

The importance of randomness in sampling needs no emphasis, It is a means for securing a representative sample. How can a random sample be drawn? The layman tends to think that random

sampling means picking out units "at random", i.e., in a haphazard or hit-and-miss way. Experience shows that the human being is an extremely poor instrument for the conduct of a random selection. To ensure true randomness the method of selection must be independent of human judgement. There are basic procedures.

1) Lottery Method

This is the simplest and most familiar procedure of random sampling. If a sample of 10 students is to be drawn out of a list of 50 students in a section, take 50 equal size chips or slips of paper; number them from 1 to 50 each bearing only one number. Roll each slip. Put the rolled slips in a global container and thoroughly shuffle or mix them. Take 10 chips from the container one after another. Each time before drawing a chip, mix the chips in the container thoroughly. The units bearing the numbers of chips drawn constitute the random sample.

(i) Sampling with Replacement

After a number is selected by draw, it may be replaced and consequently it has a chance of being selected again. Such method is known as sampling with replacement. This is usually referred to as unrestricted random sampling.

(ii) Sampling without Replacement

Selected numbers is set aside, and so in subsequent draws, it does not get a chance of being selected again. This type of sampling is known as sampling without replacement. This is a form of restricted sampling.

2) Use of Table o Random Numbers

This is a less cumbersome, but equally valid procedure of sample selection. Tables of random numbers have been developed by Kendall and Smith (1939), Fisher and Yates (1963) and Tippett (1927). To select a random sample out of a given frame, one should simply start to read numbers

from a Table of Random Numbers at any randomly selected point and pick out numbers within the range of the frame. Let us suppose that random sample of 50 is to be selected from a College populations of 500 Commerce Students. We can use any table of random numbers.

3) Use of Computer

If the population is very large and if computer facilities are available, a computer may be used for drawing a random sample. The computer can be programmed to print out a series of random numbers as the researcher desires.

TYPES OF SAMPLING

A) Probability Sampling

1) Simple Random Sampling

This sampling technique gives each element an equal and independent chance of being selected. An equal chance means equal probability of selection, e.g., in a population of 300, each element theoretically has 1/300th chance of being selected. In a population of 1000, each element has 1/1000th chance of being selected. Equal probability selection method is described as Epsem sampling. An independent choice means that the draw of one element will not affect the chances of other elements being selected.

Where some elements are purposely excluded from the sample, the resulting sample is not a random one, Hence, all elements should be included in the sample frame to draw a random sample.

Merits

- All elements in the population have an equal chance of being selected
- Of all the probability sampling techniques, simple random sampling is the easiest to
- apply
- It is the simplest type of probability sampling to understand

- It does not required a prior knowledge of the true composition of the population
- The amount of sampling error associated with any sample drawn can easily be computed

Demerits

- It is often impractical, because of non-availability of population list, or of difficulty in enumerating the population. For example, it is difficult to get a current accurate list of households in a city of a list of landless rural agricultural labourers who migrate from area to area in search of employment or a list of households of a nomadic tribe
- The use of simple random sampling may be wasteful because we fail to use all of the known information about the population
- This technique does not ensure proportionate representation to various groups constituting the population
- The sampling error in this sampling is greater than that in other probability samples of the same size, because it is less precise than other methods
- The size of the sample required to ensure its representativeness is equally larger under this type of sampling than under other random sampling techniques
- A simple random design may be expensive in time and money

2) Stratified Random Sampling

This is an improved type of random sampling. In this method, the population is sub-divided into homogenous groups or strata, and from each stratum, random sample is drawn. For example university students may be divided on the basis of discipline, and each discipline group may again be divided into juniors and seniors; and the employees of a business undertaking may be divided into managers and non-managers and each of those two groups may be sub-divided into salary-grade wise strata.

a) Proportionate Stratified Sampling

This sampling involves drawing a sample from each stratum in proportion to the latter's share in the total population.

Specialization	No. of Students	Proportion of Each Stream
Production	40	0.4
Finance	20	0.2
Marketing	30	0.3
Rural Development	10	0.1
Total	100	1.0

Merits

- It enhances the representativeness of the sample by giving proper representation to all sub-groups in the population
- It gives higher statistical efficiency that the given by simple random sampling for a given sample size
- It is easy to carry out this sample method
- This method gives a self-weighing sample, the population mean can be estimated simply by calculating the sample mean

Demerits

- A prior knowledge of the composition of the population and the distribution of the population characteristics are required to adopt this method
- This method is very expensive in time and money. Of course its greater efficiency may offset the additional cost

- The identification of the strata might lead to classification errors. Some elements maybe included into the wrong strata. This may vitiate the interpretation of survey results.

b) Disproportionate Stratified Sampling

This method does not give proportionate representation to strata (group). It necessarily involves giving over representation to some strata and under representation to others. There may be several disproportionate schemes. All strata may be given equal weight, even though their shares in the total population vary. Alternatively some substrata may be given greater weight and others lesser weight. When is such disproportionate weighing preferable? Example : Drawing one per cent as sample irrespective on the numbers of members in the sample

Merits

- It is less time consuming compared with proportionate sampling, because the researcher is not necessarily concerned about the proportionate representativeness of his resulting sample as in the latter method
- It facilitates giving appropriate weighting to particular groups, which are small but more important

Demerits

- This method does not give each stratum proportionate representation. Hence, the resulting sample may be less representative
- This method requires a prior knowledge of the composition of the population, which is not always possible
- This method is also subject to classification errors. It is possible that the researcher may misclassify certain elements

- Though disproportionate sampling is a means for developing an optimal stratification scheme, its practical feasibility is doubtful because one generally does not know the relative variability in the strata nor the relative costs

3) Systematic Sampling

This method of sampling is an alternative to random sampling. It consists of taking every Kth item in the population after a random start with an item form 1 to K. For example, suppose it is desired to select a sample of 20 students, from a list of 300 students, divide the population total of 300 by 20, the quotient is 15. Select a number at random between 1 and 15, using lottery method or a table of random numbers. Suppose the selected number is 9. Then the students numbered 9, 24, 39 are selected as the sample.

As the interval between sample units is fixed, this method is also known as fixed interval method.

Merits

- It is much simpler than random sampling. It is easy to use
- It is easy to instruct to field investigators to use this method
- This method may require less time. A researcher operating on a limited time schedule will prefer this method
- This method is cheaper than simple random sampling
- It is easier to check whether every 'k'th has been included in the sample
- Sample is spread evenly over the population
- It is statistically more efficient than a simple random sample when population elements are ordered chronologically, by size, class, etc., Then systematic sampling gives a better representative sample

Demerits

- This method ignores all elements between two 'k'th elements selected. Further, except the first element, other selected elements are not chosen at random. Hence, this sampling cannot be considered to be a probably sampling in the strict sense of the term
- As each element does not have an equal chance of being selected, the resulting sample is not a random one. For studies aiming at estimation or generalizations, this disadvantage would be serious one
- This method may sometimes give a biased sample. If by chance, several 'k' th elements chosen represent a particular group, that group would be over-represented in the sample

4) Cluster Sampling

Where the population elements are scattered over a wider area and a list of population elements is not readily available, the use of simple or stratified random sampling method would be too expensive and time consuming. In such cases cluster sampling is usually adopted.

Cluster sampling means random selection of sampling units consisting of population elements. Each such sampling unit is a cluster of population elements. Then from each selected sampling unit, a sample of population elements is drawn by either simple random selection or stratified random selection.

Example: Suppose a researcher wants to select a random sample of 1000 households out of 40000 estimated households in a city for a survey. A direct sample of individual households would be difficult to select, because a list of households does not exist and would be too costly to prepare. Instead, he can select a random sample of a few blocks / wards. The number of blocks to be selected depends upon the average number of estimated households per block. Suppose the average number of households per block is 200, then 5 blocks comprise the sample. Since the number of households per

block varies, the actual sample size depends on the block which happen to be selected. Alternatively, he can draw a sample of more blocks and from each blocks a certain number of households may be selected by systematic sampling.

Merits

1. This method is much easier and more convenient to apply when large populations are studied or large geographical areas are covered. Even a ready list of population elements is not necessary. A researcher can simply draw a random sample of geographical sections and adopt single or multistage sampling depending on the vastness of the area covered by the study
2. The cost of this method is much less when compared with other sampling methods
3. This method promotes the convenience of field work at it could be done in compact places
4. Sampling under this method does not require more time
5. This method is flexible. Where it involves multistage sampling, it is possible to employ different types of sampling in successive stages

Demerits

- The cluster size may vary and this variation could increase the bias of the resulting sample. For example, if the researcher were to interview all adults in households in each selected street the number of adults would vary from house to house. There would be certain bias resulting from the large coverage of big families
- The sampling error in this method of sampling is greater. Thus, this method is statistically less efficient than other probability sampling methods
- Adjacent units of study (e.g. households) tend to have more similar characteristics than do units distantly apart. This affects the 'representativeness' of the sample and this effect is reflected in a greater sampling error.

5) Area Sampling

This is an important form of cluster sampling. In larger field surveys, clusters consisting of specific geographical areas like districts, taluks, villages or blocks in a city are randomly drawn. As the geographical areas are selected as sampling units in such cases, their sampling is called area sampling. It is not a separate method of sampling, but forms a part of cluster sampling.

In a country like India where a state (previously known as province) is divided into districts, districts into talukas and talukas into towns and villages, area sampling is done on the basis of these administrative units in multi-stages.

6) Multi-stage Sampling

In this method, sampling is carried out in two or more stages. The population is regarded as being composed of a number of first stage sampling units. Each of them is made up of a number of second stage units and so forth. That is, at each stage, a sampling unit is a cluster of the sampling units of the subsequent stage. First, a sample of the first stage sampling units is drawn, then from each of the selected first stage sampling unit, a sample of the second stage sampling units is drawn. The procedure continues down to the final sampling units or population elements. Appropriate random sampling method is adopted at each stage.

Merits

It results in concentration of fieldwork in compact small areas and consequently in a saving of time, labour and money

- It is more convenient, efficient and flexible than single-stage sampling
- It obviates the necessity of having a sampling frame covered the entire population

Demerits

The major disadvantage of the multi-stage sampling is that the procedure of estimating sampling error and cost advantage is complicated. It is difficult for a non-statistician to follow estimation procedure.

B. Non-Probability Sampling

1) Convenience Sampling

This is non-probability sampling. It means selecting sample units in a just 'hit anmiss' fashion. E.g. Interviewing people whom we happen to meet. This sampling also means selecting whatever sampling units are conveniently available e.g. a teacher may select students in his class. This method is also known as accidental sampling because the respondents whom the researcher meets accidentally are included in the sample.

Merits

- Cheapest and simplest
- It does not require a list of population
- It does not require any statistical expertise

Demerits

- Convenience sampling is highly biased, because of the researcher's subjectivity, and so it does not yield a representative sample
- It is the least reliable sampling method. There is no way of estimating the representativeness of the sample
- The findings cannot be generalized

2) Purposed or Judgement Sampling

This method means deliberate selection of sample units that conform to some pre-determined criteria. This is known as judgement sampling. This involves selection of cases which we judge as the most appropriate ones for the given study. It is based on the judgement of the researcher or some expert. It does not aim at securing a cross section of a population.

The chance that a particular case be selected for the sample depends on the subjective judgement of the researcher. For example, A researcher may deliberately choose industrial undertakings in which quality circles are believed to be functioning successfully and undertakings in which quality circles are believed to be a total failure

Merits

- It is less costly and more convenient
- It guarantees inclusion of relevant elements in the sample. Probability sampling plans cannot give such guarantee

Demerits

- This does not ensure the representativeness of the sample
- This is less efficient for generalizing when compared with random sampling
- This method requires more prior extensive information about the population one studies. Without such information, it is not possible to adjudge the suitability of the sample items to be selected
- The method does not lend itself for using inferential statistics, because, this sampling does not satisfy the underlying assumption of randomness.

3) Quota Sampling

This is a form of convenient sampling involving selection of quota groups of accessible sampling units by traits such as sex, age, social class etc., when the population is known to consist of various categories by sex, age, religion, social class, etc., in specific proportions, each investigator may be given an assignment of quota groups specified by the pre-determined traits in specific proportions. He can then select accessible persons, belonging to those quota groups in the area assigned to him.

Merits

- It is considerably less costly than probability sampling
- It takes less time
- There is no need for a list of population. Thus, quota sampling is a suitable method of sampling a population for which no suitable frame is available
- Field work can easily be organized. Strict supervision need not be required

Demerits

- It may not yield a precise representative sample, and it is impossible to estimate sampling error. The findings, therefore, are not generalizable to any significant extent
- Interviewers may tent to choose the most accessible persons; they may ignore slums or areas difficult to reach. Thus, they may fail to secure a representative sample within their quota groups
- Strict control of field work is difficult
- The quota of sampling is subject to higher degree of classification error, because the investigators are likely to base their classification of respondents' social status and economic status mostly on their impressions about them

- It is difficult for sampling on more than three variable dimensions. This is because the number of categories to be selected is a multiplication of the number of values in each variable. For instances, if we want to sample proportionate number of persons by sex, social status and age and these variables consist of two, three and three categories respectively.

4) Snowball Sampling

This is the colourful name for a technique of building up a list or a sample of a special population by using an initial set of its members as informants. For example, if a researcher wants to study the problem faced by Indians through some source like Indian Embassy. Then he can ask each one of them to supply names of other Indians known to them, and continue this procedure until he gets an exhaustive list from which he can draw a sample or make a census survey.

This sampling technique may also be used in socio-metric studies. For example, the members of a social group may be asked to name the persons with whom they have social contacts, each one of the persons so named may also be asked to do so, and so on. The researcher may thus get a constellation of associates and analyse it.

Merits

- It is very useful in studying social groups, informal group in a formal organization, and diffusion of information among professionals of various kinds
- It is useful for smaller populations for which no frames are readily available

Demerits

- The major disadvantages of snowball sampling is that it does not allow the use of probability statistical methods. Elements included are dependent on the subjective choice of the original selected respondents
- It is difficult to apply this method when the population is large

- It does not ensure the inclusion of all elements in the lists

SAMPLING ERROR

A survey aims at estimating or inferring selected population characteristics or parameters by studying either population or a sample of the population. The research results may either from the 'true values' of the parameters under study. Such differences are known as Errors and Biases. The errors of a survey may be classified into (a) Sampling Errors (b) Sampling Biases (c) Non-sampling errors and (d) Non-sampling biases.

1) Sampling Errors

The errors which arise because of studying only a part of the total population are called sampling errors. These may arise due to non-representativeness of the samples and the inadequacy of sample size. When several samples are drawn from a population, their results would not be identical. The degree of variation of sample results is measured by standard deviation and it is known as the standard error of the concerned statistic. As sample size increases the magnitude of the error decreases. Sample size and sampling error are thus negatively correlated.

2) Sampling Biases

The average of the estimates of a population parameter derived from an infinite number of samples is called the expected value of the estimator. The difference between this value and 'true value' of the parameter is the bias. Bias may arise (1) if the sampling is done by a non-random method. (2) if the sampling frame is incomplete or inaccurate and (3) if some sections of the population are not available / refuse to cooperate. Any of these factors will cause non-compensating errors which cannot be reduced by an increase in sample size. The only sure way of avoiding bias arising through the sampling method is to use a random method. Randomness is an essential part of the protection against selection bias.

3) Non-Sampling Errors

These are errors which arise from sources other than sampling. They include errors of observation, errors of measurement and errors of responses. Data are collected through the methods of observation or interviewing. The physical procedures of observation or interviewing are subject to imperfection which cause errors. Measurement errors consist of errors in processing and analysis. Errors of response include incorrect responses of the respondents, mistakes in noting their response etc.,

4) Non-Sampling Biases

These biases pose problems for scientific measurement. They affect both the population sample value and account for the difference between the population value and the true value. They consist of biases of observation and non-observation, response biases and process biases. Biases of observation are caused by obtaining and recording observations incorrectly. Non-observation biases arise from failure to obtain observations on some segments of the population due to either non-coverage or non-response. The latter may be due to refusals, non-at-homes, lost forms, etc., response biases consist of biases arising from imperfections in field observation or interviewing. Processing biases are produced during coding, tabulating and computing.

5) Total Error

In sampling theory, a popular model combines sampling and non-sampling errors and biases into the Total error. This total error is the square root of the sums of squares of variable errors and squares of bias. It is often called the root means square error. The variable error are caused only by sampling errors, and VE equals the standard error of sampling. Bias is mostly caused by measurement biases. The total error depends on the length of both the legs. The sampling error / standard error leg can be shortened by improving the sample design and by increasing the sample size. The length of

biases leg may be reduced by improving the tool of data collection, the precision of method of data collection, filed work, coding, processing and analysis. The measurement of sampling error does not pose much problem but the measurement of non-sampling errors require special procedures and it is a costly effect. Hence, the reduction of non-sampling error is a challenge to the researchers.

DATA COLLECTION

Data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions and evaluate outcomes.

METHODS OF DATA COLLECTION

The task of data collection begins after a research problem has been defined and research design/ plan chalked out. While deciding about the method of data collection to be used for the study, the researcher should keep in mind two types of data viz., primary and secondary. The primary data are those which are collected afresh and for the first time, and thus happen to be original in character. The secondary data, on the other hand, are those which have already been collected by someone else and which have already been passed through the statistical process. The researcher would have to decide which sort of data he would be using (thus collecting) for his study and accordingly he will have to select one or the other method of data collection. The methods of collecting primary and secondary data differ since primary data are to be originally collected, while in case of secondary data the nature of data collection work is merely that of compilation.

1) Primary Data

Data directly collected by the researcher, with respect to problem under study, is known as primary data. Primary data is also the first hand data collected by the researcher for the immediate purpose of the study.

2) Secondary Data

Secondary data are statistics that already exists. They have been gathered not for immediate use. This may be described as "Those data that have been compiled by some agency other than the user".

INTERVIEW METHOD

The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. This method can be used through personal interviews and, if possible, through telephone interviews. Interview is one of the popular methods of data collection. The term interview can be dissected into two terms as, 'inter' and 'view.'. The essence of interview is that one mind tries to read the other. The interviewer tries to assess the interviewed in terms of the aspects studied or issues analysed.

Personal interview method requires a person known as the interviewer asking questions generally in a face-to-face contact to the other person or persons. (At times the interviewee may also ask certain questions and the interviewer responds to these, but usually the interviewer initiates the interview and collects the information.) This sort of interview may be in the form of direct personal investigation or it may be indirect oral investigation. In the case of direct personal investigation the interviewer has to collect the information personally from the sources concerned. He has to be on the spot and has to meet people from whom data have to be collected. This method is particularly suitable for intensive investigations.

But in certain cases it may not be possible or worthwhile to contact directly the persons concerned or on account of the extensive scope of enquiry, the direct personal investigation technique may not be used. In such cases an indirect oral examination can be conducted under which the interviewer has to cross-examine other persons who are supposed to have knowledge about the

problem under investigation and the information, obtained is recorded. Most of the commissions and committees appointed by government to carry on investigations make use of this method.

Types of Interview

1) Structured Interview

The method of collecting information through personal interviews is usually carried out in a structured way. As such we call the interviews as structured interviews. Such interviews involve the use of a set of predetermined questions and of highly standardised techniques of recording. Thus, the interviewer in a structured interview follows a rigid procedure laid down, asking questions in a form and order prescribed. But in case of descriptive studies, we quite often use the technique of structured interview because of its being more economical, providing a safe basis for generalisation and requiring relatively lesser skill on the part of the interviewer.

2) Unstructured Interview

As against it, the unstructured interviews are characterized by a flexibility of approach to questioning. Unstructured interviews do not follow a system of pre-determined questions and standardised techniques of recording information. In a non-structured interview, the interviewer is allowed much greater freedom to ask, in case of need, supplementary questions or at times he may omit certain questions if the situation so requires. He may even change the sequence of questions. He has relatively greater freedom while recording the responses to include some aspects and exclude others. But this sort of flexibility results in lack of comparability of one interview with another and the analysis of unstructured responses becomes much more difficult and time-consuming than that of the structured responses obtained in case of structured interviews. Unstructured interviews also demand deep knowledge and greater skill on the part of the interviewer. Unstructured interview,

however, happens to be the central technique of collecting information in case of exploratory or formulative research studies.

3) Focused Interview

Focused interview is meant to focus attention on the given experience of the respondent and its effects. Under it the interviewer has the freedom to decide the manner and sequence in which the questions would be asked and has also the freedom to explore reasons and motives. The main task of the interviewer in case of a focussed interview is to confine the respondent to a discussion of issues with which he seeks conversance. Such interviews are used generally in the development of hypotheses and constitute a major type of unstructured interviews.

4) Clinical Interview

The clinical interview is concerned with broad underlying feelings or motivations or with the course of individual's life experience. The method of eliciting information under it is generally left to the interviewer's discretion.

5) Non-directive Interview

In case of non-directive interview, the interviewer's function is simply to encourage the respondent to talk about the given topic with a bare minimum of direct questioning. The interviewer often acts as a catalyst to a comprehensive expression of the respondents' feelings and beliefs and of the frame of reference within which such feelings and beliefs take on personal significance.

Merits of Interview

- More information and that too in greater depth can be obtained.
- Interviewer by his own skill can overcome the resistance, if any, of the respondents; the interview method can be made to yield an almost perfect sample of the general population.

- There is greater flexibility under this method as the opportunity to restructure questions is always there, specially in case of unstructured interviews.
- Observation method can as well be applied to recording verbal answers to various questions.
- Personal information can as well be obtained easily under this method.
- Samples can be controlled more effectively as there arises no difficulty of the missing returns; non-response generally remains very low.
- The interviewer can usually control which person(s) will answer the questions. This is not possible in mailed questionnaire approach. If so desired, group discussions may also be held.
- The interviewer may catch the informant off-guard and thus may secure the most spontaneous reactions than would be the case if mailed questionnaire is used.
- The language of the interview can be adopted to the ability or educational level of the person interviewed and as such misinterpretations concerning questions can be avoided.
- The interviewer can collect supplementary information about the respondent's personal characteristics and environment which is often of great value in interpreting results.

Demerits or Weakness of Interview

- It is a very expensive method, specially when large and widely spread geographical sample is taken.
- There remains the possibility of the bias of interviewer as well as that of the respondent; there also remains the headache of supervision and control of interviewers.
- Certain types of respondents such as important officials or executives or people in high income groups may not be easily approachable under this method and to that extent the data may prove inadequate.

- This method is relatively more-time-consuming, specially when the sample is large and recalls upon the respondents are necessary.
- The presence of the interviewer on the spot may over-stimulate the respondent, sometimes even to the extent that he may give imaginary information just to make the interview interesting.
- Under the interview method the organisation required for selecting, training and supervising the field-staff is more complex with formidable problems.
- Interviewing at times may also introduce systematic errors.
- Effective interview presupposes proper rapport with respondents that would facilitate free and frank responses. This is often a very difficult requirement.

Pre-requisites and Basic Tenets of Interviewing

- For successful implementation of the interview method, interviewers should be carefully selected, trained and briefed.
- They should be honest, sincere, hardworking, impartial and must possess the technical competence and necessary practical experience.
- Occasional field checks should be made to ensure that interviewers are neither cheating, nor deviating from instructions given to them for performing their job efficiently.
- In addition, some provision should also be made in advance so that appropriate action may be taken if some of the selected respondents refuse to cooperate or are not available when an interviewer calls upon them.
- In fact, interviewing is an art governed by certain scientific principles. Every effort should be made to create friendly atmosphere of trust and confidence, so that respondents may feel at ease while talking to and discussing with the interviewer.

- The interviewer must ask questions properly and intelligently and must record the responses accurately and completely. At the same time, the interviewer must answer legitimate question(s), if any, asked by the respondent and must clear any doubt that the latter has.
- The interviewers approach must be friendly, courteous, conversational and unbiased.
- The interviewer should not show surprise or disapproval of a respondent's answer but he must keep the direction of interview in his own hand, discouraging irrelevant conversation and must make all possible effort to keep the respondent on the track.

OBSERVATION METHOD

The observation method is the most commonly used method specially in studies relating to behavioural sciences. In a way we all observe things around us, but this sort of observation is not scientific observation. Observation becomes a scientific tool and the method of data collection for the researcher, when it serves a formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability. Under the observation method, the information is sought by way of investigator's own direct observation without asking from the respondent. For instance, in a study relating to consumer behaviour, the investigator instead of asking the brand of wrist watch used by the respondent, may himself look at the watch.

Advantages of Observation

- The main advantage of this method is that subjective bias is eliminated, if observation is done accurately.
- Secondly, the information obtained under this method relates to what is currently happening; it is not complicated by either the past behaviour or future intentions or attitudes.

- Thirdly, this method is independent of respondents' willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or the questionnaire method.
- This method is particularly suitable in studies which deal with subjects (i.e., respondents) who are not capable of giving verbal reports of their feelings for one reason or the other

Limitations of Observation

- Firstly, it is an expensive method.
- Secondly, the information provided by this method is very limited.
- Thirdly, sometimes unforeseen factors may interfere with the observational task.
- At times, the fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively.

Precautions before Adopting Observation Method of Data Collection

- While using this method, the researcher should keep in mind things like:
- What should be observed?
- How the observations should be recorded? Or
- How the accuracy of observation can be ensured?

Types of Observation

1) Structured Observation

In case the observation is characterised by a careful definition of the units to be observed, the style of recording the observed information, standardised conditions of observation and the selection of pertinent data of observation, then the observation is called as structured observation. Structured observation is considered appropriate in descriptive studies

2) Unstructured Observation

But when observation is to take place without these characteristics to be thought of in advance, the same is termed as unstructured observation. Whereas in an exploratory study the observational procedure is most likely to be relatively unstructured.

3) Participant Observation

We often talk about participant and non-participant types of observation in the context of studies, particularly of social sciences. This distinction depends upon the observer's sharing or not sharing the life of the group he is observing. If the observer observes by making himself, more or less, a member of the group he is observing so that he can experience what the members of the group experience, the observation is called as the participant observation.

Merits of Participant Observation

- The researcher is enabled to record the natural behaviour of the group.
- The researcher can even gather information which could not easily be obtained if he observes in a disinterested fashion.
- The researcher can even verify the truth of statements made by informants in the context of a questionnaire or a schedule.

Demerits of Participant Observation

- The observer may lose the objectivity to the extent he participates emotionally; the problem of observation-control is not solved; and it may narrow-down the researcher's range of experience.

4) Non-Participant Observation

But when the observer observes as a detached emissary without any attempt on his part to experience through participation what others feel, the observation of this type is often termed as non-

participant observation. (When the observer is observing in such a manner that his presence may be unknown to the people he is observing, such an observation is described as disguised observation.)

5) Uncontrolled Observation

If the observation takes place in the natural setting, it may be termed as uncontrolled observation, In non-controlled observation, no attempt is made to use precision instruments. The major aim of this type of observation is to get a spontaneous picture of life and persons. It has a tendency to supply naturalness and completeness of behaviour, allowing sufficient time for observing it. The main pitfall of non-controlled observation is that of subjective interpretation. There is also the danger of having the feeling that we know more about the observed phenomena than we actually do. Uncontrolled observation is resorted to in case of exploratory researches.

6) Controlled Observation

When observation takes place according to definite pre-arranged plans, involving experimental procedure, the same is then termed controlled observation. But in controlled observation, we use mechanical (or precision) instruments as aids to accuracy and standardisation. Such observation has a tendency to supply formalised data upon which generalisations can be built with some degree of assurance. Generally, controlled observation takes place in various experiments that are carried out in a laboratory or under controlled conditions

QUESTIONNAIRE

This method of data collection is quite popular, particularly in case of big enquiries. It is being adopted by private individuals, research workers, private and public organisations and even by governments. In this method a questionnaire is sent (usually by post) to the persons concerned with a request to answer the questions and return the questionnaire. A questionnaire consists of a number of questions printed or typed in a definite order on a form or set of forms. The questionnaire is mailed to

respondents who are expected to read and understand the questions and write down the reply in the space meant for the purpose in the questionnaire itself. The respondents have to answer the questions on their own.

The method of collecting data by mailing the questionnaires to respondents is most extensively employed in various economic and business surveys

Merits of Questionnaire

- There is low cost even when the universe is large and is widely spread geographically.
- It is free from the bias of the interviewer; answers are in respondents' own words.
- Respondents have adequate time to give well thought out answers.
- Respondents, who are not easily approachable, can also be reached conveniently.
- Large samples can be made use of and thus the results can be made more dependable and reliable.

Demerits of Questionnaire

- Low rate of return of the duly filled in questionnaires; bias due to no-response is often indeterminate.
- It can be used only when respondents are educated and cooperating.
- The control over questionnaire may be lost once it is sent.
- There is inbuilt inflexibility because of the difficulty of amending the approach once questionnaires have been despatched.
- There is also the possibility of ambiguous replies or omission of replies altogether to certain questions; interpretation of omissions is difficult.
- It is difficult to know whether willing respondents are truly representative.
- This method is likely to be the slowest of all.

Before using this method, it is always advisable to conduct 'pilot study' (Pilot Survey) for testing the questionnaires. In a big enquiry the significance of pilot survey is felt very much. Pilot survey is infact the replica and rehearsal of the main survey. Such a survey, being conducted by experts, brings to the light the weaknesses (if any) of the questionnaires and also of the survey techniques. From the experience gained in this way, improvement can be effected.

Main Aspects of a Questionnaire

Quite often questionnaire is considered as the heart of a survey operation. Hence it should be very carefully constructed. If it is not properly set up, then the survey is bound to fail. This fact requires us to study the main aspects of a questionnaire viz., the general form, question sequence and question formulation and wording. Researcher should note the following with regard to these three main aspects of a questionnaire:

1) 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 questions are presented with exactly the same wording and in the same order to all respondents. Resort is taken to this sort of standardisation to ensure that all respondents reply to the same set of questions. The form of the question may be either closed (i.e., of the type 'yes' or 'no') or open (i.e., inviting free response) but should be stated in advance and not constructed during questioning. Structured questionnaires may also have fixed alternative questions in which responses of the informants are limited to the stated alternatives. Thus a highly structured questionnaire is one in which all questions and answers are specified and comments in the respondent's own words are held to the minimum. When these characteristics are not present in a questionnaire, it can be termed as unstructured or non-structured questionnaire. More

specifically, we can say that in an unstructured questionnaire, the interviewer is provided with a general guide on the type of information to be obtained, but the exact question formulation is largely his own responsibility and the replies are to be taken down in the respondent's own words to the extent possible; in some situations tape recorders may be used to achieve this goal.

Structured questionnaires are simple to administer and relatively inexpensive to analyse. The provision of alternative replies, at times, helps to understand the meaning of the question clearly. But such questionnaires have limitations too. For instance, wide range of data and that too in respondent's own words cannot be obtained with structured questionnaires. They are usually considered inappropriate in investigations where the aim happens to be to probe for attitudes and reasons for certain actions or feelings. They are equally not suitable when a problem is being first explored and working hypotheses sought. In such situations, unstructured questionnaires may be used effectively. Then on the basis of the results obtained in pretest (testing before final use) operations from the use of unstructured questionnaires, one can construct a structured questionnaire for use in the main study.

2) Question Sequence

In order to make the questionnaire effective and to ensure quality to the replies received, a researcher should pay attention to the question-sequence in preparing the questionnaire. A proper sequence of questions reduces considerably the chances of individual questions being misunderstood. The question-sequence must be clear and smoothly-moving, meaning thereby that the relation of one question to another should be readily apparent to the respondent, with questions that are easiest to answer being put in the beginning. The first few questions are particularly important because they are likely to influence the attitude of the respondent and in seeking his desired cooperation. The opening

questions should be such as to arouse human interest. The following type of questions should generally be avoided as opening questions in a questionnaire:

- Questions that put too great a strain on the memory or intellect of the respondent;
- Questions of a personal character
- Questions related to personal wealth, etc.

Following the opening questions, we should have questions that are really vital to the research problem and a connecting thread should run through successive questions. Ideally, the question-sequence should conform to the respondent's way of thinking. Knowing what information is desired, the researcher can rearrange the order of the questions (this is possible in case of unstructured questionnaire) to fit the discussion in each particular case. But in a structured questionnaire the best that can be done is to determine the question-sequence with the help of a Pilot Survey which is likely to produce good rapport with most respondents. Relatively difficult questions must be relegated (tranfered) towards the end so that even if the respondent decides not to answer such questions, considerable information would have already been obtained. Thus, question-sequence should usually go from the general to the more specific and the researcher must always remember that the answer to a given question is a function not only of the question itself, but of all previous questions as well. For instance, if one question deals with the price usually paid for coffee and the next with reason for preferring that particular brand, the answer to this latter question may be couched (understood) largely in terms of price-differences.

3) Question Formulation and Wording

With regard to this aspect of questionnaire, the researcher should note that each question must be very clear for any sort of misunderstanding can do irreparable harm to a survey. Question should also be impartial in order not to give a biased picture of the true state of affairs. Questions should be

constructed with a view to their forming a logical part of a well thought out tabulation plan. In general, all questions should meet the following standards—(a) should be easily understood; (b) should be simple i.e., should convey only one thought at a time; (c) should be concrete and should conform as much as possible to the respondent's way of thinking. (For instance, instead of asking. "How many razor blades do you use annually?" The more realistic question would be to ask, "How many razor blades did you use last week?"

Multiple Choice Question and the Open-End Questions. In the former the respondent selects one of the alternative possible answers put to him, whereas in the latter he has to supply the answer in his own words. The question with only two possible answers (usually 'Yes' or 'No') can be taken as a special case of the multiple choice question, or can be named as a 'closed question.' There are some advantages and disadvantages of each possible form of question. Multiple choice or closed questions have the advantages of easy handling, simple to answer, quick and relatively inexpensive to analyse. They are most amenable to statistical analysis. Sometimes, the provision of alternative replies helps to make clear the meaning of the question. But the main drawback of fixed alternative questions is that of "putting answers in people's mouths" i.e., they may force a statement of opinion on an issue about which the respondent does not infact have any opinion. They are not appropriate when the issue under consideration happens to be a complex one and also when the interest of the researcher is in the exploration of a process. In such situations, open-ended questions which are designed to permit a free response from the respondent rather than one limited to certain stated alternatives are considered appropriate. Such questions give the respondent considerable latitude in phrasing a reply. Getting the replies in respondent's own words is, thus, the major advantage of open-ended questions. But one should not forget that, from an analytical point of view, open-ended questions are more difficult to handle, raising problems of interpretation, comparability and interviewer bias.

In practice, one rarely comes across a case when one questionnaire relies on one form of questions alone. The various forms complement each other. As such questions of different forms are included in one single questionnaire. For instance, multiple-choice questions constitute the basis of a structured questionnaire, particularly in a mail survey. But even there, various open-ended questions are generally inserted to provide a more complete picture of the respondent's feelings and attitudes.

Researcher must pay proper attention to the wordings of questions since reliable and meaningful returns depend on it to a large extent. Since words are likely to affect responses, they should be properly chosen. Simple words, which are familiar to all respondents, should be employed. Words with ambiguous meanings must be avoided. Similarly, danger words, catch-words or words with emotional connotations should be avoided. Caution must also be exercised in the use of phrases which reflect upon the prestige of the respondent. Question wording, in no case, should bias the answer. In fact, question wording and formulation is an art and can only be learnt by practice.

Essentials of Good Questionnaire

- To be successful, questionnaire should be comparatively short and simple i.e., the size of the questionnaire should be kept to the minimum
- Questions should proceed in logical sequence moving from easy to more difficult questions
- Personal and intimate questions should be left to the end
- Technical terms and vague expressions capable of different interpretations should be avoided in a questionnaire
- Questions may be dichotomous (yes or no answers), multiple choice (alternative answers listed) or open-ended. The latter type of questions are often difficult to analyse and hence should be avoided in a questionnaire to the extent possible

- There should be some control questions in the questionnaire which indicate the reliability of the respondent. For instance, a question designed to determine the consumption of particular material may be asked first in terms of financial expenditure and later in terms of weight. The control questions, thus, introduce a cross-check to see whether the information collected is correct or not
- Questions affecting the sentiments of respondents should be avoided
- Adequate space for answers should be provided in the questionnaire to help editing and tabulation
- There should always be provision for indications of uncertainty, e.g., "do not know," "no preference" and so on
- Brief directions with regard to filling up the questionnaire should invariably be given in the questionnaire itself
- Finally, the physical appearance of the questionnaire affects the cooperation the researcher receives from the recipients and as such an attractive looking questionnaire, particularly in mail surveys, is a plus point for enlisting cooperation.
- The quality of the paper, along with its colour, must be good so that it may attract the attention of recipients

Types of Questionnaire

1) Structured non disguised Questionnaire

Here, questions are structured so as to get the facts. The interviewer will ask the questions strictly as per the pre arranged order. Structured, non disguised is widely used in market research. Questions are presented with exactly the same wording and same order to all the respondents. The reason for standardizing question is, to ensure that all respondents reply the same question. The

purpose of the question is clear. The researcher wants the respondent to choose one of the five options given above. This type of questionnaire is easy to administer. The respondents have no difficulty in answering. Because it is structured, the frame of reference is obvious. In a non-disguised type, the purpose of the questionnaire is known to the respondent.

2) Structured disguised Questionnaire

This type of questionnaire is least used in Marketing research. This type of Questionnaire is used to find, peoples' attitude, when a direct undisguised question produces a bias. In this type of questionnaire what comes out is "What does the respondent know rather than what he feels". Therefore attempt in this method is to find the respondent's attitude.

3) Non-Structured and disguised Questionnaire

The main objective is to conceal the topic of enquiry by using a disguised stimulus. Though the stimulus is standardized by researcher, respondent is allowed to answer in an unstructured manner. The assumption made here is that individuals reaction is an indication of respondent's basic perception. Projective techniques are examples of Non structured disguised technique. The techniques involve the use of a vague stimulus, that an individual is asked to expand or describe or build a story, three common types under this category are (a) Word association (b) Sentence completion (c) Story telling.

4) Non structured - Non disguised Questionnaire

Here the purpose of the study is clear, but the responses to the question is open ended. Example: "How do you feel about the cyber law currently in practice and its need for further modification"? The initial part of the question is constant. After presenting the initial question, the interview becomes very unstructured as the interviewer probes more deeply. Respondents subsequent answer determines the direction the interviewer takes next. The question asked by interviewer varies

from person to person. This method is called "Depth interview". The major advantage of this method is freedom permitted to the interviewer.

PILOT STUDY

A pilot study is a research study conducted before the intended study. Pilot studies are usually executed as planned for the intended study, but on a smaller scale. Although a pilot study cannot eliminate all systematic errors or unexpected problems, it reduces the likelihood of making a Type I or Type II error. Both types of errors make the main study a waste of effort, time, and money.

Reasons to Employ a Pilot Study

There are many reasons to employ a pilot study before implementing the main study. Here are a few good reasons:

- To test the research process and/or protocol. These are often referred to as feasibility studies because the pilot study tests how possible the design is in reality. For example, are the study resources adequate, including time, finances, materials? Are there are any other logistical problems that need to be addressed?
- To identify variables of interest and decide how to operationalize each one. For instance, what are the indicators of composite variables? How will variables be measured and/or computed?
- To test an intervention strategy and identify the components that are most important to the facilitation of the intervention.
- To test methodological changes to implementation or administration of an instrument and/or train personnel on the administration of instruments.
- To develop or test the efficacy of research instruments and protocols. Are there confusing or misleading questions? Is it possible to maintain maximum objectivity and reduce observer drift?

- To estimate statistical parameters for later analyses. Certain statistical analyses require the sample size is sufficiently large and contains enough variability to detect differences between groups, given there any real differences to be detected.

SCHEDULE

This method of data collection is very much like the collection of data through questionnaire, with little difference which lies in the fact that schedules (proforma containing a set of questions) are being filled in by the enumerators who are specially appointed for the purpose. These enumerators along with schedules, go to respondents, put to them the questions from the proforma in the order the questions are listed and record the replies in the space meant for the same in the proforma.

In certain situations, schedules may be handed over to respondents and enumerators may help them in recording their answers to various questions in the said schedules. Enumerators explain the aims and objects of the investigation and also remove the difficulties which any respondent may feel in understanding the implications of a particular question or the definition or concept of difficult terms. This method requires the selection of enumerators for filling up schedules or assisting respondents to fill up schedules and as such enumerators should be very carefully selected

The enumerators should be trained to perform their job well and the nature and scope of the investigation should be explained to them thoroughly so that they may well understand the implications of different questions put in the schedule. Enumerators should be intelligent and must possess the capacity of cross-examination in order to find out the truth. Above all, they should be honest, sincere, hardworking and should have patience and perseverance. This method of data collection is very useful in extensive enquiries and can lead to fairly reliable results. It is, however, very expensive and is usually adopted in investigations conducted by governmental agencies or by some big organisations. Population census all over the world is conducted through this method.

DIFFERENCE BETWEEN QUESTIONNAIRE AND INTERVIEW SCHEDULE

1. The questionnaire is generally sent through mail to informants to be answered as specified in a covering letter, but otherwise without further assistance from the sender. The schedule is generally filled out by the research worker or the enumerator, who can interpret questions when necessary.

2. To collect data through questionnaire is relatively cheap and economical since we have to spend money only in preparing the questionnaire and in mailing the same to respondents. Here no field staff required. To collect data through schedules is relatively more expensive since considerable amount of money has to be spent in appointing enumerators and in importing training to them. Money is also spent in preparing schedules.

3. Non-response is usually high in case of questionnaire as many people do not respond and many return the questionnaire without answering all questions. Bias due to non-response often remains indeterminate. As against this, non-response is generally very low in case of schedules because these are filled by enumerators who are able to get answers to all questions. But there remains the danger of interviewer bias and cheating.

4. In case of questionnaire, it is not always clear as to who replies, but in case of schedule the identity of respondent is known.

5. The questionnaire method is likely to be very slow since many respondents do not return the questionnaire in time despite several reminders, but in case of schedules the information is collected well in time as they are filled in by enumerators.

6. Personal contact is generally not possible in case of the questionnaire method as questionnaires are sent to respondents by post who also in turn return the same by post. But in case of schedules direct personal contact is established with respondents.

7. Questionnaire method can be used only when respondents are literate and cooperative, but in case of schedules the information can be gathered even when the respondents happen to be illiterate.

8. Wider and more representative distribution of sample is possible under the questionnaire method, but in respect of schedules there usually remains the difficulty in sending enumerators over a relatively wider area.

9. Risk of collecting incomplete and wrong information is relatively more under the questionnaire method, particularly when people are unable to understand questions properly. But in case of schedules, the information collected is generally complete and accurate as enumerators can remove the difficulties, if any, faced by respondents in correctly understanding the questions. As a result, the information collected through schedules is relatively more accurate than that obtained through questionnaires.

10. The success of questionnaire method lies more on the quality of the questionnaire itself, but in the case of schedules much depends upon the honesty and competence of enumerators.

11. In order to attract the attention of respondents, the physical appearance of questionnaire must be quite attractive, but this may not be so in case of schedules as they are to be filled in by enumerators and not by respondents.

12. Along with schedules, observation method can also be used but such a thing is not possible while collecting data through questionnaires.

SECONDARY DATA

Secondary data means data that are already available i.e., they refer to the data which have already been collected and analysed by someone else. When the researcher utilises secondary data,

then he has to look into various sources from where he can obtain them. In this case he is certainly not confronted with the problems that are usually associated with the collection of original data.

Secondary data may either be published data or unpublished data. Usually published data are available in: (a) various publications of the central, state are local governments; (b) various publications of foreign governments or of international bodies and their subsidiary organisations; (c) technical and trade journals; (d) books, magazines and newspapers; (e) reports and publications of various associations connected with business and industry, banks, stock exchanges, etc.; (f) reports prepared by research scholars, universities, economists, etc. in different fields; and (g) public records and statistics, historical documents, and other sources of published information.

The sources of unpublished data are many; they may be found in diaries, letters, unpublished biographies and autobiographies and also may be available with scholars and research workers, trade associations, labour bureaus and other public/ private individuals and organisations. Researcher must be very careful in using secondary data. He must make a minute scrutiny because it is just possible that the secondary data may be unsuitable or may be inadequate in the context of the problem which the researcher wants to study. In this connection Dr. A.L. Bowley very aptly observes that it is never safe to take published statistics at their face value without knowing their meaning and limitations and it is always necessary to criticise arguments that can be based on them.

CHARACTERISTICS OF SECONDARY DATA

1) Reliability of Data

The reliability can be tested by finding out such things about the said data:

(a) Who collected the data? (b) What were the sources of data? (c) Were they colleted by using proper methods? (d) At what time were they collected? (e) Was there any bias of the compiler? (f) What level of accuracy was desired? Was it achieved?

2) Suitability of Data

The data that are suitable for one enquiry may not necessarily be found suitable in another enquiry. Hence, if the available data are found to be unsuitable, they should not be used by the researcher. In this context, the researcher must very carefully scrutinize the definition of various terms and units of collection used at the time of collecting the data from the primary source originally. Similarly, the object, scope and nature of the original enquiry must also be studied. If the researcher finds differences in these, the data will remain unsuitable for the present enquiry and should not be used

3) Adequacy of Data

If the level of accuracy achieved in data is found inadequate for the purpose of the present enquiry, they will be considered as inadequate and should not be used by the researcher. The data will also be considered inadequate, if they are related to an area which may be either narrower or wider than the area of the present enquiry

KAHE

Class: III B.COM PA **Course Name: Research Methodology**
Course Code: 17PAU501A **Unit III**
BATCH: 2017-2020

MEASURES OF CENTRAL TENDENCY

Raw data are difficult to comprehend. Classification facilitates, many a time, quick and easy understanding of diversified nature of data. A single representative value serves the purpose in a better manner.

Quantitative data in a mass exhibit certain general characteristics. They show a tendency to concentrate at certain values, usually somewhere in the centre of the distribution. Measures of this tendency are called measures of central tendency or averages. This tendency toward centralization, though not universal, has established the expression "measure of central tendency" to describe an average. The terms is imbedded in statistical language, but it is not always pertinent.

An average is a value which is typical or representative of a set of data.

A measure of central tendency gives a single representative value for a set of usually unequal values. The single value is the point of location around which the individual values of the set cluster. The measures of central tendency are hence known as 'measures of location'. They are popularly called averages.

ARITHMETIC MEAN

Arithmetic mean is the total of the values of the items divided by their number.

Direct Method

Sum 1: The expenditure of 10 families in Rupees are given below.

Family	A	B	C	D	E	F	G	H	I	J
Exp.	30	70	10	75	500	8	42	250	40	36

Calculate the Arithmetic Mean.

Family	**Expenditure (Rs.)**
A	30
B	70
C	10

Family	Expenditure (Rs.)
D	75
E	500
F	8
G	42
H	250
I	40
J	36
Total	**∑X= 1061**

AM = A + (∑fd/n)

= 150 + (192 / 12)

= 150 + 16

= Rs. 166

Short Cut Method

Sum 2: The expenditure of 10 families in Rupees are given below

Family	A	B	C	D	E	F	G	H	I	J
Exp.	30	70	10	75	500	8	42	250	40	36

Solution

Family	Expenditure (Rs.) X	d=X-A; A = 100
A	30	-70
B	70	-30
C	10	-90
D	75	-25
E	500	400
F	8	-92
G	42	-58
H	250	150
I	40	-60
J	36	-64
Total		**∑d=61**

N = 12 and ∑d = 192

AM = A + (∑d/N)

= 100 + (61/10)

= 100 + 6.10

= Rs. 106.10

Sum 3: The monthly income of 12 families in a town is given below:

S.No.	1	2	3	4	5	6	7	8	9	10	11	12
Income	280	180	96	98	104	85	80	94	100	75	600	200

Calculate the arithmetic mean by taking 150 as the assumed mean.

Solution

S.No.	**Income (Rs.) X**	**d=X-A; A=150**
1	280	130
2	180	30
3	96	-54
4	98	-52
5	104	-46
6	85	-65
7	80	-70
8	94	-56
9	100	-50
10	75	-75
11	600	450
12	200	50
Total	**-**	**192**

N = 12 and $\sum d = 192$

AM = $A + (\sum d/N)$

= 150 + (192 / 12)

= 150 + 16

= Rs. 166

Class: III B.COM PA **Course Name: Research Methodology**
Course Code: 17PAU501A **Unit III**
BATCH: 2017-2020

Step Deviation Method

Sum 4: Find the arithmetic mean by step deviation method.

Marks	20	30	40	50	50	60	70	80	90	90

Solution

Marks X	**d=X-A/C A= 50; C=10**
20	-3
30	-2
40	-1
50	0
50	0
60	1
70	2
80	3
90	4
90	4
Total	**∑d=8**

N = 10 ; ∑d = 8

AM = A + (C∑d/N)

= 50 + (10 X 8) / 10

= 50 + 8

= 58

Arithmetic Mean for Discrete Series

Sum 5: Calculate the mean number of persons per house. Given

No. of Persons per house	2	3	4	5	6	Total
No. of Houses	10	25	30	25	10	100

Solution:

No. of Persons per house X	**No. of Houses f**	**fx**
2	10	20
3	25	75
4	30	120
5	25	125

6	10	60
Total	**N=100**	**∑fx=400**

AM = ∑fx / N

= 400 / 100

= 4

Arithmetic Mean for Discrete Series – Short Cut Method

Sum 6: Calculate the arithmetic mean.

Marks	40	50	54	60	68	80	Total
No, of Students	10	18	20	39	15	8	110

Solution

Let A = 60

Marks (X)	**No. of Students (f)**	**d=X-A; A=60**	**fd**
40	10	-20	-200
50	18	-10	-180
54	20	-6	-120
60	39	0	0
68	15	8	120
80	8	20	160
Total	**N=110**	**--**	**∑fd=-220**

AM = A + (∑fd/N)

= 60 + (-220 / 110)

= 60 – 2

= 58

Arithmetic Mean for Discrete Series – Step Deviation Method

Sum 7: Calculate the arithmetic mean from the following discrete series.

Daily Wages (Rs.)	75	100	120	150	200	Total
No. of Labourers	5	12	20	14	9	60

Solution

Daily Wage (Rs.) X	No. of Labourers (f)	d=X-A/C; A=120 C=5	fd
75	5	-9	-45
100	12	-4	-48
120	20	0	0
150	14	6	84
200	9	16	144
Total	**N=60**	**--**	**∑fd=135**

AM = A + (C∑fd/N)

= 120 + (5 X 135) / 60

= 120 + 11.25

= Rs. 131.25

Continuous Series – Exclusive Class Intervals

Sum 8: Calculate Arithmetic mean for the following:

Marks	20-30	30-40	40-50	50-60	60-70	70-80
No. of Students	5	8	12	15	6	4

Solution

Marks	No. of Students (f)	Mid value (m)	fm
20-30	5	25	125
30-40	8	35	280
40-50	12	45	540
50-60	15	55	825
60-70	6	65	390
70-80	4	75	300
Total	**N=50**	**--**	**∑fm=2460**

AM = A + ∑fm/N

= 2460 / 50

= 49.20

Sum 9: From the following data, compute arithmetic mean by short cut method

Marks Obtained	0-10	10-20	20-30	30-40	40-50	50-60
No. of Students	5	10	25	30	20	10

Solution

Marks Obtained	**No.of Students (f)**	**Mid value (m)**	**d=m-A A=25**	**fd**
00-10	5	5	-20	-100
10-20	10	15	-10	-100
20-30	25	25	0	0
30-40	30	35	10	300
40-50	20	45	20	400
50-60	10	55	30	300
Total	**N=100**	**--**	**--**	**∑fd=800**

AM = A + ∑fd/N

= 25 + (800/100)

= 25 + 8

= 33

Continuous Series –Inclusive Class Intervals

Sum 10: The annual profits of 90 companies are given below. Find the arithmetic mean.

Annual Profit (Rs.)	0-19	20-39	40-59	60-79	80-99
No. of Companies	5	17	32	24	12

Solution

Annual Profit (Rs.)	**No. of Companies (f)**	**Mid value (m)**	**fm**
00-19	5	9.5	47.50
20-39	17	29.5	501.50

40-59	32	49.5	1584.00
60-79	24	69.5	1668.00
80-99	12	89.5	1074.00
Total	**N=90**	**--**	**∑fm=4875.00**

AM = ∑fm/N

= 4875 / 90

= Rs. 54.17

Continuous Series –Inclusive Class Intervals – Step Deviation Method

Sum 11: The annual profits of 90 companies are given below. Find the arithmetic mean.

Annual Profit (Rs.)	0-19	20-39	40-59	60-79	80-99
No. of Companies	5	17	32	24	12

Solution

Annual Profit	**No. of Companies (f)**	**Mid value (m)**	**d=m-A/C A=49.5; C=20**	**fd**
00-19	5	9.5	-2	-10
20-39	17	29.5	-1	-17
40-59	32	49.5	0	0
60-79	24	69.5	1	24
80-99	12	89.5	2	24
Total	**N=90**	**--**	**--**	**∑fd=21**

AM = A + (C∑fd/N)

= 49.5 + (20 X 21)/90

= 49.5 + 4.67

= Rs. 54.17

Continuous Series Less than Cumulative Frequencies

Sum 12: Calculate the mean height

Height Below (Cms)	150	155	160	165	170	175	180	185
No. of Soldiers	0	23	77	152	266	419	472	500

Solution

Height Below (Cms)	No. of Soldiers	Height (cms)	No. of Soldiers (f)	Mid value (m)	d=m-A/C A=167.5 C=5	fd
150	0	150-155	23	152.50	-3	-69
155	23	155-160	54	157.50	-2	-108
160	77	160-165	75	162.50	-1	-75
165	152	165-170	114	167.50	0	0
170	266	170-175	153	172.50	1	153
175	419	175-180	53	177.50	2	106
180	472	180-185	28	182.50	3	84
185	500					
Total	**-**	**-**	**N=500**	**-**	**-**	**∑fd=91**

Mean Height = A + (C∑fd/N)

= 167.50 + (5 X 91)/500

= 167.50 + 0.91

= 168.41 Cms.

Continuous Series – More than Cumulative Frequencies

Sum 13: Calculate the arithmetic mean from the following data:

Weight (Above)	20	25	30	35	40
No. of Boys	160	145	100	50	9

Solution

Weight Above (Kgs.)	No. of Boys	Weight (Kgs.)	No. of Boys (f)	Mid value (m)	fm
20	160	20-25	15	22.50	337.50
25	145	25-30	45	27.50	1237.50
30	100	30-35	50	32.50	1625.00
35	50	35-40	41	37.50	1537.50
40	9	40-	9	42.50	382.50
Total	**-**	**-**	**N=60**	**-**	**∑fm=5120**

AM = ∑fm/N

= 5120/160

= 32.00 Kgs.

MEDIAN

Median is the value of the middle most items when all the items are in the order of magnitude.

Arithmetic mean is calculated on the basis of magnitudes or values of all the items. But median is concerned with the position or place of the items in a series. 'Which is the middle most item' is the question.

Median divides the series into two equal parts. Half of the items will be equal to or less than the median; half of the items will be equal or more than the median.

Sum1: Find median for the following

6	9	21	5	7	-2	0	32	9

Solution

Values in Ascending Order

-2, 0, 5, 6, 7, 9, 9, 21, 32

Position of Median is = N+1/2 = 9+1/2 = 5

Median = 7 (It is the value at 5^{th} Position)

Sum2: Find Median for the following data

57	58	61	42	38	65	72	66

Solution

Values in ascending order: 38, 42, 57, 58, 61, 65, 66, 72

Position of Median is = N+1/2

= 8+1/2

= 4.5, a fraction

Value at (N/2 = 8/2) = 4^{th} Position = 58

Value at (N/2+1 = 4+1) = 5^{th} Position = 61

Median = Value at 4^{th} Position + Value at 5^{th} Position / 2

= 58 + 61 / 2

= 59.5

Discrete Series

Sum 3: Consider the following data and compute Median.

Value (X)	0	1	2	3	Total
Frequency (f)	1	2	5	3	11

Solution

Value (x)	Frequency (f)	Cumulative Frequency (cf)
0	1	1
1	2	3
2	5	8
3	3	11
Total	**11**	

The position of median is N+1/2 = 11+1/2 = 6. 6 lies between the cumulative frequencies 3 and 8. Hence, the value at 6^{th} position is the value corresponding to the cumulative frequency 8 in the table. It is 2. Hence, Median = 2.

When N+1/2 is a fraction, the two middle most items are to be identified in a similar manner and the mean of those two items is to be found.

Sum 4: The marks (out of a maximum of 10) scored by the students of a class are given below. Find the Median mark.

Mark	3	4	5	6	7	8	9	10	Total
No. of Students	1	5	6	7	10	15	10	5	59

Solution

Mark (x)	No. of Students (f)	Cumulative Frequency (cf)
3	1	1
4	5	6
5	6	12

6	7	19
7	10	29
8	15	44
9	10	54
10	5	59
Total	**59**	

N+1/2 = 59+1/2=30. When all the 59 items are in ascending order, which is in 30th position? It is included in cf=44. Median =8

Sum 5: Find the median from the following data.

Wages (Rs.)	50	75	100	150	250	Total
No. of Labourers	8	14	10	5	3	40

Solution

Wage Rs. (X)	**No. of Labourers (f)**	**Cumulative Frequency (cf)**
50	8	8
75	14	22
100	10	32
150	5	37
250	3	40
Total	**40**	

Median = N+1/2

= 40+1/2

= 20.5

Looking corresponding to cf=22, Wage (x) at 20th position= 75.

Wage (x) at 21st position 75.

Median = Wage at 20th Position + Wage at 21st Position / 2

= 75+75/2

= Rs. 75

Sum 6: Find the Median

No. of Cars sold in a Day	10	15	17	18	21	Total
No. of Days	4	16	12	5	3	40

Solution:

No. of Cars sold in a Day (x)	No. of Days (f)	Cumulative Frequency (cf)
10	4	4
15	16	20
17	12	32
18	5	37
21	3	40
Total	**40**	

N+1/2 = 40+1/2

= 20.5

X at 20th position = 15

X at 21st position = 17

Median = X at 20th Position + X at 21st Position / 2

= 15+17/2

= 16

Median – Continuous Series

Sum 7: Calculate the median height

Height (cms.)	145-150	150-155	155-160	160-165	165-170	170-175
No. of Students	2	5	10	8	4	1

Solution

Height (cms.)	No. of Students (f)	Cumulative Frequency (cf)
145-150	2	2
150-155	5	7
155-160	10	17
160-165	8	25

165-170	4	29
170-175	1	30
Total	**N=30**	

L = 155; f = 10; i = 160-155=5; cf = 7

M = L + (i (N_2-cf)/f)

= 155 + (5 (15-7))/10

= 155 + (5X8)/10

= 155 + 4

= 159 Cms.

Sum 8: Calculate the median from the following data:

Marks	10-25	25-40	40-55	55-70	70-85	85-100
Frequency	6	20	44	26	3	1

Solution

Marks	**Frequency (f)**	**Cumulative Frequency (cf)**
10-25	6	6
25-40	20	26
40-55	44	70
55-70	26	96
70-85	3	99
85-100	1	100

N/2 = 100/2 = 50

Median class interval: 40-55

L = 40; f = 44; cf = 26; i = 55 – 40 = 15

M = L + (i (N_2-cf)/f)

= 40 + (15 (50-26))/44

= 40 + (15 X 24) / 44

= 40 + 8.18

= 48.18 (or) 48

Sum 9: Calculate the median for the following:

Value	0-9	10-19	20-29	30-39	40-49	50-59	60-69
Frequency	328	720	640	598	524	378	244

Solution

Value	**Frequency (f)**	**True Class Intervals**	**Cumulative Frequency (cf)**
00-09	328	-0.5-9.5	328
10-19	720	9.5-19.5	1048
20-29	664	19.5-29.5	1712
30-39	598	29.5-39.5	2310
40-49	524	39.5-49.5	2834
50-59	378	49.5-59.5	3212
60-69	244	59.5-69.5	3456
Total	**N=3456**		

Second order limit – First upper limit = 10 – 9
= 1
Half of the difference = ½ = 0.5

0.5 has been added to each upper limit to and 0.5 has been subtracted from each lower limit to get the boundaries of the true class intervals. It is the required form for the calculation of median.

N/2 = 3456/2 = 1728. Hence, the median class interval is 29.5 – 39.5

L = 29.5; f = 598; i = 39.5-29.5=10 ; cf = 1712

M = L + (i (N_2-cf)/f)

= 29.5 + (10 (1728-1712))/598

= 29.5 + (10 X 16) / 598

= 29.5 + 0.27

= 29.77

Sum 10: Convert the following "Less than cumulative frequency" distribution into an ordinary "Frequency distribution" and then calculate the median age.

Age (Less than)	10	20	30	40	50	60	70	80
No. of Persons	4	16	40	76	96	112	120	125

Solution

Age (Less than)	**Number of Persons**	**Age**	**No. of Persons (f)**	**Cumulative Frequency (cf)**
10	4	00-10	4	
20	16	10-20	12	4
30	40	20-30	24	16
40	76	30-40	36	40
50	96	40-50	20	76
60	112	50-60	16	96
70	120	60-70	8	112
80	125	70-80	8	120
Total	**.**	**-**	**N=125**	125

N/2 = 125/2 = 62.5; Median class interval: 30-40

L = 30; f = 36; i = 40-30=10; cf=40

M = $L + (i (N_2-cf)/f)$

= 30 + (10 (62.5-40))/36

= 30 + 6.25

= 36.25

Sum 11: Calculate the Median

Annual Expenditure on Ad.	0	4	8	12	16
No. of Years	50	35	25	15	6

Class: III B.COM PA **Course Name: Research Methodology**
Course Code: 17PAU501A **Unit III**
BATCH: 2017-2020

Solution

Annual Expenditure on Ad. (more than Rs. Lakhs)	No. of Years	Annual Expenditure on Ad. (Rs. Lakhs)	No. of Years (f)	Cf
0	50	0-4	15	15
4	35	4-8	10	25
8	25	8-12	10	35
12	15	12-16	9	44
16	6	16-	6	50
Total			**N=50**	**-**

N/2 = 50/2 = 25; Median class: 4-8

L = 4; f = 10; i = 8-4=4; cf = 15

M = $L + (i (N_2-cf)/f)$

= $4 + (4 (25-15))/10$

= $4 + (4X10)/10$

= 4+4

= Rs. 8 Lakhs

Sum 12: Compute median from the following data:

Mid values	115	125	135	145	155	165	175	185	195
Frequency	6	25	48	72	116	60	38	22	3

Solution

Mid values (m)	Frequency (f)	Class Interval	Cumulative Frequency (cf)
115	6	110-120	6
125	25	120-130	31
135	48	130-140	79
145	72	140-150	151
155	116	150-160	267
165	60	160-170	327
175	38	170-180	365
185	22	180-190	387
195	3	190-200	390
Total	**N=390**		

Difference between successive mid values = 10

Half of the difference = 5

N/2 = 390/2 = 195

Median class: 150-160

L = 150; f = 116 ; i = 160-150=10; cf = 151

M = L + (i (N_2-cf)/f)

= 150 + (10 (195-151))/116

= 150 + (10X44)/116

= 150 + 3.79

= 153.79

MODE

Mode is the value which has the greatest frequency density.

Z or M_o denotes Mode.

According to Croxton and Cowden, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values". Mode is called the most typical or fashionable value of a distribution.

In individual observations and discrete series, the mode is most often available by inspection. The value which has the greatest frequency is mode. Such values have the greatest frequency density also. The difference between the greatest frequency and the next lower frequency may be nominal in a few cases. Mode is then determined on the basis of the greatest frequency density. That is, on the basis of its frequency and the neighbouring frequencies of each value. It is found out by forming a grouping table and analysis table of frequencies.

In the words of Murray R Spiegel, "The mode of a set of numbers is that value which occurs with the greatest frequency, i.e. it is the most common value. The mode may not exist, and even if it does exist it may not be unique". There is no mode when all the observations occur equal number of times. If one value occurs distinctly more times

than any other value, that value is the mode. The set which has only one mode is said to be unimodal. Sets with two modes are said to be bimodal. Sets which have more than two modes are said to be multimodal. In a few situations due to fluctuations of sampling it becomes a difficult task to identify a single value with greatest frequency in a sample even though the population is undoubtedly unimodal.

Sum 1: Determine the mode on the following:

1. 320, 395, 342, 444, 551, 395, 425, 417, 395, 401, 390, 400
2. 3, 6, 7, 5, 8, 4, 9
3. 25, 32, 24, 27, 32, 27, 25, 32, 24, 27, 25, 24
4. 0, 2, 5, 6, 9, 5, 6, 14, 6, 15, 5, 6, 5

Solution

Mode = 395	Because its frequency, 3, is higher than others. The frequency of others is 1 each. (This is an example for unimodal distribution)
No Mode	Because all the values have equal frequency (1). (This is an example for a distribution which has no mode)
No Mode	Because all the values occur equal number of times (3 times each). (This is an example for a distribution which has no mode)
Modes 5& 6	Because they occur equal number of times and they occur greater number of times than other values. (This is an example for bimodal distribution. Grouping table of frequencies also shows that 5 and 6 have greatest and equal frequency densities. Mode is said to be ill defined in this case. Hence, the answers are not given)

Discrete Series

Sum 2: Determine the Mode:

Size of Dress	18	20	22	24

No. of Sets Produced	55	120	108	45

Solution:

Mode = 20

Sum 3: Find the Modal Size

Size of Shoes	3	4	5	6	7	8	9
No. of Pairs Sold	10	25	32	38	61	47	34

Solution

Mode = 7

Sum 4: Calculate the mode from the following:

Size	10	11	12	13	14	15	16	17	18
Frequency	10	12	15	19	20	8	4	3	2

Solution

Greatest frequency is 20. Mode need not be 14 because the difference between the greatest frequency 20 and the next lower frequency 19 is very small. Further, 19 has the support of the neighbouring frequency 15 while 20 has the support of 8 only.

Grouping table and analysis table are formed as explained earlier.

Grouping Table

Size (x)	**Frequency (f) (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
10	10					
		22				
11	12			37		
			27			
12	15				**46**	
		34				
13	19					**54**
			39			
14	**20**			**47**		

Size (x)	Frequency (f) (1)	(2)	(3)	(4)	(5)	(6)
		28				
15	8				32	
			12			
16	4					15
		7				
17	3			9		
18	2					

Analysis Table

Size X	1	2	3	4	5	6	Total
10							-
11					1		1
12		1			1	1	3
13		1	1	1	1	1	5
14	1		1	1		1	4
15				1			1
16							-
17							-
18							-

Mode = 13

Continuous Series

Sum 5: Calculate the Mode.

Daily Wage in Rs.	50-60	60-70	70-80	80-90	90-100
No. of Labourers	40	62	75	100	65

Solution

L=80	The lower boundary of the modal class interval

f_1=100	The frequency of the modal class
f_0=75	The frequency of the class preceding the modal class and
f_2=65	The frequency of the class succeeding the modal class and
i=90-80=10	The size of the modal class interval

$Z = L + (iD_1/ (D_1+D_2))$

$= 80 + (10 X 25/ (25 + 35))$

$= 80 + (250/60)$

$= 80 + 4.17$

$= 84.17$

Sum 6: Find out the mode for the following data using grouping and analysis table.

CI	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40
Frequency	9	12	15	16	17	15	10	13

Solution

Grouping Table

Size (x)	**Frequency (f) (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
0-5	9					
		21				
5-10	12			36		
			27			
10-15	15				**43**	
		31				
15-20	16					**48**
			33			
20-25	**17**			**48**		
		32				
25-30	15				42	
			25			
30-35	10					38
		23				
35-40	13					

Class: III B.COM PA **Course Name: Research Methodology**
Course Code: 17PAU501A **Unit III**
BATCH: 2017-2020

Analysis Table

Size X	1	2	3	4	5	6	Total
00-05							-
05-10					1		1
10-15					1	1	2
15-20			1	1	1	1	4
20-25	1	1	1	1		1	5
25-30		1		1			2
30-35							-
35-40							-

Modal Class Interval = 20 – 25

L = 20; i = 25-20=5; D_1 = 17-16=1; D_2 = 17-15=12

Z = L + (iD_1/ (D_1+D_2))

= 20 + (5X1/(1+2))

= 20 + (5/3)

= 20 + 1.67

= 21.67

Sum 7: Calculate the mode

Interval	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16
Frequency	45	50	65	70	30	25	20	18

Solution

Grouping Table

Size (x)	Frequency (f) (1)	(2)	(3)	(4)	(5)	(6)
0-2	45					
		95				
2-4	50			**160**		
			115			
4-6	65				**185**	
		135				

Size (x)	Frequency (f) (1)	(2)	(3)	(4)	(5)	(6)
6-8	**70**					**165**
			100			
8-10	30			125		
		55				
10-12	25				75	
			45			
12-14	20					63
		38				
14-16	18					

Analysis Table

Size X	1	2	3	4	5	6	Total
0-2				1			1
2-4			1	1	1		3
4-6		1	1	1	1	1	5
6-8	1	1			1	1	4
8-10						1	1
10-12							-
12-14							-
14-16							-

The interval 4-6 does not have greatest frequency, But it has greatest frequency density consider.

L=4	The lower boundary of the modal class interval
f_1=50	The frequency of the class preceding the modal class
f_2=70	The frequency of the class succeeding the modal class
i=6-4=2	The size of the modal class interval

$Z = L + (if_2/ (f_1+f_2))$

$= 4 + (2X70/(50+70))$

$= 4 + (140/120)$

$= 4 + 1.17$

$= 5.17$

Sum 8: Calculate the mode

Marks	0-19	20-39	40-59	60-79	80-99
No. of Students	5	20	35	20	12

Solution

Marks	**No. of Students**	**Marks**
0-19	5	-0.5-19.5
20-39	20	19.5-39.5
40-59	35	39.5-59.5
60-79	20	59.5-79.5
80-99	12	79.5-99.5

Greatest Frequency = 35 ; Modal Class = 39.5 – 59.5

L = 39.5; i=59.5-39.5=20; D_1= 35-20=15; D_2= 35-20=15

$Z = L + (iD_1/ (D_1+D_2))$

$= 39.5 + (20X15/(15+15))$

$= 39.5 + (300/30)$

$= 39.5 + 10$

$= 49.5$

Sum 9: Calculate the mode.

Annual Profit (Rs.)	10-19	20-29	30-39	40-49	50-59	60-69
No. of Companies	12	15	16	17	10	9

Solution

Grouping Table

Size (x)	**Frequency (f) (1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
9.5-19.5	12					
		27				
19.5-29.5	15			**43**		

Size (x)	Frequency (f) (1)	(2)	(3)	(4)	(5)	(6)
			31			
29.5-39.5	16				**48**	
		33				
39.5-49.5	**17**					**43**
			27			
49.5-59.5	10			36		
		19				
59.5-69.5	9					

Analysis Table

Size X	1	2	3	4	5	6	Total
9.5-19.5				1			1
19.5-29.5			1	1	1		3
29.5-39.5		1	1	1	1	1	5
39.5-49.5	1	1			1	1	4
49.5-59.5						1	1
59.5-69.5							-

29.5-39.5 has the greatest frequency density. Its frequency is not the greatest. Hence, consider

L=29.5	The lower boundary of the modal class interval
f_1=15	The frequency of the class preceding the modal class
f_2=17	The frequency of the class succeeding the modal class
i=39.5-29.5-=10	The size of the modal class

Z = $L + (if_2/ (f_1+f_2))$

= 29.5 + (10X17/(15+17))

= 29.5 + (170/32)

= 29.5 + 5.31

= Rs. 34.81

Sum 10: Calculate the Mode

No. of Days Absent	No. of Students
Less than 3	10
Less than 6	25
Less than 9	38
Less than 12	48
Less than 15	51
Less than 18	52

Solution

Data are written in continuous class intervals first. Corresponding frequencies are then found.

No. of Days Absent	No. of Students	Number of Days Absent	No. of Students
Less than 3	10	0-3	10
Less than 6	25	3-6	15
Less than 9	38	6-9	13
Less than 12	48	9-12	10
Less than 15	51	12-15	3
Less than 18	52	15-18	1

Greatest frequency = 15; Modal class interval = 3-6

L=3; i=6-3=3; D_1=15-10=5; D_2=15-13=2

Z = L + (iD_1/ (D_1+D_2))

= 3 + (3X5/(5+2))

= 3 + (15/7)

= 3 + 2.14

= 5.14

Sum 11: Calculate the mode of the following frequency distribution

Wages Above (Rs.)	0	20	40	60	80	100
No. of Workers	50	45	34	16	6	0

Solution

Data are written in continuous class intervals first. Corresponding frequencies are then found;

Wages Above (Rs.)	**No. of Workers**	**Wages (Rs.)**	**No. of Workers**
0	50	00-20	5
20	45	20-40	11
40	34	40-60	18
60	16	60-80	10
80	6	80-100	6
100	0		

Greatest Frequency = 18; Modal Class Interval = 40-60

L=40; i=60-40=20; D_1=18-11=7; D_2=18-10=8

Z = L + (iD_1/ (D_1+D_2))

= 40 + (20 X 7/(7+8))

= 40 + (140/15)

= 40 + 9.33

= 49.33

Sum 12: Calculate the mode

Central Value	70	90	110	130	150
Frequency	43	78	83	125	87

Solution

Data are written in continuous class interval form.

Central Value	Frequency	Class Intervals
70	43	60-80
90	78	80-100
110	83	100-120
130	125	120-140
150	87	140-160

Greatest Frequency = 125

Modal Class Interval : 120-140

L=120; i=140-120=20; D_1=125-83=42; D_2=125-87=38

Z = L + (iD_1/ (D_1+D_2))

= 120 + (20X42/(42+38))

= 120 + (840/80)

= 120 + 10.50

= 130.50

MEASURES OF DISPERSION

In a series, all the items are not equal. There is difference or variation among the values. The degree of variation is evaluated by various measures of dispersion.

Averages are central values. They enable comparison of two or more sets of data. They are not sufficient to depict the true nature of the sets. For example, consider the following marks of two students

Student I	Student II
68	85
75	90
65	80
67	25
70	65

Both have got a total of 345 and an average of 69 each. The fact is that the second student has failed in one paper. When the averages alone are considered, the two student are equal.

Less variation is a desirable characteristic. First student has less variation. That is, he is almost equally good in all the subjects. To quote Simpson and Kafka, "An average does not tell the full story. It is hardly fully representative of a mass, unless we know the manner in which the individual items scatter around it. A further description of the series is necessary if we are to guage how representative the average is".

STANDARD DEVIATION

Standard deviation is the root mean square deviation of the values from their arithmetic mean.

SD is the abbreviation and σ (read, sigma) is the symbol. Mean square deviation of the values from their AM is Variance and is denoted by σ^2. SD is the positive square root of variance. Karl Pearson introduced the concept of standard deviation in 1893. SD is also called root mean square deviation. It is a mathematical deficiency of mean deviation to ignore negative sign. Standard deviation possesses most of the desirable properties of a good measure of dispersion. The corresponding relative measure is Coefficient of Variation. It is very popular and so extremely used as raise a doubt whether there is any other relative measure of dispersion.

Coefficient of Variation = Standard Deviation / Arithmetic Mean X 100

Sum 1: Find SD for the following: 77, 73, 75, 70, 72, 76, 75, 72, 74, 76

Solution

X	$X=X-\overline{X}$	X^2
77	3	9
73	-1	1
75	1	1
70	-4	16
72	-2	4
76	2	4
75	1	1
72	-2	4
74	0	0
76	2	4
$\sum X=740$	$\sum (X-\overline{X})=0$	$\sum X^2=44$

Arithmetic Mean, $\overline{X}$ $= \sum X/N$

$= 740/10$

$= 74$

Standard Deviation, σ $= \sqrt{\sum X^2/N}$

$= \sqrt{44/10}$

$= \sqrt{4.4}$

$= 2.10$

Sum 2: 10 Students of B.Com. class of a College have obtained the following marks in Statistics out of 100 marks. Calculate the standard deviation.

S.No.	1	2	3	4	5	6	7	8	9	10
Marks	5	10	20	25	40	42	45	48	70	80

Solution

S.No.	Marks X	X^2
1	5	25
2	10	100
3	20	400
4	25	625
5	40	1600
6	42	1764
7	45	2025
8	48	2304
9	70	4900
10	80	6400
Total	**$\sum X=385$**	**$\sum X^2=20143$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{20143/10 - (385/10)^2}$$

$$= \sqrt{20143 - (38.5)^2}$$

$$= \sqrt{20143 - 148225}$$

$$= \sqrt{532.05}$$

$$= 23.07$$

Deviations taken from Assumed Mean

Sum 3: For the data below, calculate standard deviation 40, 50, 60, 70, 80, 90, 100

Solution

X	X=X-A; A=70	d^2
40	-30	900
50	-20	400
60	-10	100
70	0	0

80	10	100
90	20	400
100	30	900
Total	**$\sum d=0$**	**$\sum d^2=2800$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{2800/7 - (0/7)^2}$$

$$= \sqrt{400 - 0^2}$$

$$= \sqrt{400}$$

$$= 20$$

Step Deviation Method

Sum 4: Given below are the marks obtained by 5 B.Com. Students

Roll No.	101	102	103	104	105
Marks	10	30	20	25	15

Calculate Standard Deviation

Solution

Roll No.	**Marks (X)**	**d=X-A/C: A=20; C=5**	**d^2**
101	10	-2	4
102	30	2	4
103	20	0	0
104	25	1	1
105	15	-1	1
Total		**$\sum d=0$**	**$\sum d^2=10$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times C$$

$$= \sqrt{10/5 - (0/5)^2} \text{ X } 5$$

$$= \sqrt{2 - 0^2} \text{ X } 5$$

$$= \sqrt{2} \text{ X } 5$$

$$= 1.4142 \text{ X } 5$$

$$= 7.07$$

Discrete Series – Deviation taken from Actual Mean

Sum 5: Calculate the Standard Deviation of the following series.

X	6	9	12	15	18
f	7	12	13	10	8

Solution

X	**f**	**fX**	**$X=X-\bar{X}$**	**X^2**	**fX^2**
6	7	42	-6	36	252
9	12	108	-3	9	108
12	13	156	0	0	0
15	10	150	3	9	90
18	8	144	6	36	288
Total	**N=50**	**$\sum fx=600$**	**-**	**-**	**$\sum fx^2=738$**

Arithmetic Mean $= \sum fx/N$
$= 600 / 50$
$= 12.00$

Standard Deviation, $\sigma = \sqrt{\sum fX^2/N}$
$= \sqrt{738/50}$
$= \sqrt{14.76}$
$= 3.84$

Discrete Series – Direct Method

Sum 6: Calculate the Standard Deviation

No. of Goals Scored in a Match	0	1	2	3	4	5

No. of Matches	1	2	4	3	0	2

Solution:

X	**f**	**fX**	**fx^2**
0	1	0	0
1	2	2	2
2	4	8	16
3	3	9	27
4	0	0	0
5	2	10	50
Total	**N=12**	**$\sum fx=29$**	**$\sum fx^2=95$**

$$\sigma = \sqrt{\frac{\Sigma fx'^2}{N} - \left(\frac{\Sigma fx'}{N}\right)^2} \times C$$

$$= \sqrt{95/12 - (29/12)^2}$$

$$= \sqrt{7.9167 - (2.4167)^2}$$

$$= \sqrt{7.9167 - 5.8404}$$

$$= \sqrt{2.0763}$$

$$= 1.44$$

Discrete Series – Deviations taken from Assumed Mean

Sum 7: Calculate Standard Deviation from the following data

X	6	9	12	15	18
f	7	12	19	10	2

X	**f**	**d=X-A; A=12**	**fd**	**fd^2**

6	7	-6	-42	252
9	12	-3	-36	108
12	19	0	0	0
15	10	3	30	90
18	2	6	12	72
Total	**N=50**	**-**	**$\sum fd=-36$**	**$\sum fd^2=522$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{522/50 - (-36/50)^2}$$

$$= \sqrt{10.44 - (0.72)^2}$$

$$= \sqrt{10.4400 - 0.5184}$$

$$= \sqrt{9.9216}$$

$$= 3.15$$

Discrete Series – Deviations taken from Assumed Mean

Sum 8: The weekly salaries of a group of employees are given in the following table. Find the mean and standard deviation of the salaries.

Salary (Rs.)	75	80	85	90	95	100
No. of Persons	3	7	18	12	6	4

Solution:

Salary (Rs.)	**No. of Persons (f)**	**d=X-A/C A=85; C=5**	**fd**	**fd^2**
75	3	-2	-6	12
80	7	-1	-7	7
85	18	0	0	0
90	12	1	12	12
95	6	2	12	24
100	4	3	12	36
Total	**N=50**	**-**	**$\sum fd=23$**	**$\sum fd^2=91$**

Arithmetic Mean $= A + (C\sum fd/N)$

$= 85 + (5X23/50)$

= 85 + 2.3

= Rs.87.30

Continuous Series – Deviation taken from Actual Mean

Sum 9: Find the Standard Deviation.

Class Interval	0-10	10-20	20-30	30-40	40-50	Total
Frequency	2	5	9	3	1	20

Solution

Class Interval	**Frequency (f)**	**Mid Value (m)**	**fm**	**m-X X=23**	**$(m-X)^2$**	**$f(m-X)^2$**
0-10	2	5	10	-18	324	648
10-20	5	15	75	-8	64	320
20-30	9	25	225	2	4	36
30-40	3	35	10	12	144	432
40-50	1	45	45	22	484	484
Total	N=20		∑fm=460	-	-	**$\sum f(m-X)^2=1920$**

Arithmetic Mean = ∑fm/N

= 460/20

= 23

SD σ = $\sqrt{\sum f(m-\bar{X})^2/N}$

= $\sqrt{1920/20}$

= $\sqrt{96}$

= 9.80

Continuous Series – Direct Method

Sum 10: The following data were obtained while observing the life span of a few neon lights of a company. Calculate SD.

Life Span (Yrs.)	4-6	6-8	8-10	10-12	12-14	Total
No. of Neon Lights	10	17	32	21	20	100

Solution

Life Span (Yrs.)	No. of Neon Lights (f)	Mid Value (m)	fm	fm^2
4-6	10	5	50	250
6-8	17	7	119	833
8-10	32	9	288	2592
10-12	21	11	231	2541
12-14	20	13	260	3380
Total	N=100		$\sum fm=948$	$\sum fm^2=9596$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{9596/100 - (948/100)^2}$$

$$= \sqrt{95.96 - (9.48)^2}$$

$$= \sqrt{95.9600 - 89.8704}$$

$$= \sqrt{6.0896}$$

$$= 2.47$$

Continuous Series – Deviation taken from Assumed Mean

Sum 11: Calculate the standard deviation of the following series.

No. of Students in 00 (Below)	2	6	10	14	18	22	26
No. of Colleges	0	7	19	42	61	72	80

Solution

No. of Students in 00	No. of Colleges	No. of Students in 00	No. of Colleges f	Mid value (m)	d=m-A A=12	fd	fd^2

(Below)							
2	0	2-6	7	4	-8	-56	448
6	7	6-10	12	8	-4	-48	192
10	19	10-14	23	12	0	0	0
14	42	14-18	19	16	4	76	304
18	61	18-22	11	20	8	88	704
22	72	22-26	8	24	12	96	1152
26	80						
Total	**-**	**-**	**N=80**	**-**	**-**	**$\sum fd=156$**	**$\sum fd^2=2800$**

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2}$$

$$= \sqrt{2800/80 - (156/80)^2}$$

$$= \sqrt{35 - (1.95)^2}$$

$$= \sqrt{35.0000\text{-}3.8025}$$

$$= \sqrt{31.1975}$$

$$= 5.59$$

Continuous Series – Step Deviation Method

Sum 12: Calcuate the standard deviation of the following frequency distribution.

Annual Profit (Rs.)	20-40	40-60	60-80	80-100	100-120	120-140	140-160
No. of Banks	10	14	25	48	33	24	16

Solution

Annual Profit (Rs.)	**No. of Banks**	**Mid value (m)**	**d=m-A A=90; C=20**	**fd**	**fd^2**
20-40	10	30	-3	-30	90
40-60	14	50	-2	-28	56
60-80	25	70	-1	-25	25
80-100	48	90	0	0	0
100-120	33	110	1	33	33
120-140	24	130	2	48	96
140-160	16	150	3	48	144

Total	N=170	-	-	$\sum fd=46$	$\sum fd^2=444$

$$\sigma = \sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times C$$

$$= \sqrt{444/170 - (46/170)^2} \text{ X } 20$$

$$= \sqrt{2.6118 - (0.2706)^2} \text{ X } 20$$

$$= \sqrt{2.6118 - 0.0732} \text{ X } 20$$

$$= \sqrt{2.5386} \text{ X } 20$$

$$= 1.5933 \text{ X } 20$$

$$= \text{Rs. } 31.87$$

CORRELATION

The term correlation refers to the relationship between the variables. Simple correlation refers to the relationship between two variables. There may be fluctuation or co-variation between the values of the variables. The direction of change and the closeness of the relationship are found.

DEFINITION

Correlation Analysis attempts to determine the degree of relationship between variables- Ya-Kun-Chou.

. Correlation is an analysis of the covariation between two or more variables.- A.M.Tuttle.

Types of Correlation

1) Positive and Negative Correlation

It depends upon the direction of change of the variables. If the two variables tend to move together in the same direction (ie) an increase in the value of one variable is accompanied by an increase in the value of the other, (or) a decrease in the value of one

variable is accompanied by a decrease in the value of other, then the correlation is called positive or direct correlation. Price and supply, height and weight, yield and rainfall, are some examples of positive correlation.

If the two variables tend to move together in opposite directions so that increase (or) decrease in the value of one variable is accompanied by a decrease or increase in the value of the other variable, then the correlation is called negative (or) inverse correlation. Price and demand, yield of crop and price, are examples of negative correlation.

2) Linear and Non-linear Correlation

If the ratio of change between the two variables is a constant then there will be linear correlation between them.

Consider the following

X	2	4	6	8	10	12
Y	3	6	9	12	15	18

Here the ratio of change between the two variables is the same. If we plot these points on a graph we get a straight line.

If the amount of change in one variable does not bear a constant ratio of the amount of change in the other. Then the relation is called Curvi-linear (or) non-linear correlation. The graph will be a curve.

3) Simple and Multiple Correlation

When we study only two variables, the relationship is simple correlation. For example, quantity of money and price level, demand and price. But in a multiple correlation we study more than two variables simultaneously. The relationship of price, demand and supply of a commodity are an example for multiple correlation.

4) Partial and Total Correlation

The study of two variables excluding some other variable is called Partial correlation. For example, we study price and demand eliminating supply side. In total correlation all facts are taken into account.

KARL PEARSON'S COEFFICIENT OF CORRELATION

This is also called product moment correlation coefficient. This is denoted by r. This is covariance between the two variables divided by the product of their standard deviations. This can be calculated by using any one of the formulae. Choice of formula depends on the nature of the data.

Sum 1: The following table gives aptitude test scores and productivity indices of 8 randomly selected workers:

Aptitude Scores	57	58	59	59	60	61	62	64
Productivity Index	67	68	65	68	72	72	69	71

Calculate the correlation coefficient between Aptitude score and productivity index.

Solution

X	**Y**	**X=X-X X=60**	**Y=Y-Y Y=69**	**XY**	**X^2**	**Y^2**
57	67	-3	-2	6	9	4
58	68	-2	-1	2	4	1
59	65	-1	-4	4	1	16
59	68	-1	-1	1	1	1
60	72	0	3	0	0	9
61	72	1	3	3	1	9
62	69	2	0	0	4	0
64	71	4	2	8	16	4
$\sum x=480$	$\sum y=552$	$\sum x=0$	$\sum y=0$	$\sum xy=24$	$\sum x^2=36$	$\sum y^2=44$

Mean of X = $\sum X/N$ = 480/8 = 60

Mean of Y = $\sum Y /N$ = 552/8 = 69

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}$$

$= 24 / \sqrt{36 \text{ X } 44}$

$= 0.6030$

Sum 2: Compute the coefficient of Correlation between X –Advertisement Expenditure and Y-Sales

X	10	12	18	8	13	20	22	15	5	17
Y	88	90	94	86	87	92	96	94	88	85

Solution

X	**Y**	**XY**	**X^2**	**Y^2**
10	88	880	100	7744
12	90	1080	144	8100
18	94	1692	324	8836
8	86	688	64	7396
13	87	1131	169	7569
20	92	1840	400	8464
22	96	2112	484	9216
15	94	1410	225	8836
5	88	440	25	7744
17	85	1445	289	7225
$\Sigma X=140$	**$\Sigma Y=900$**	**$\Sigma XY=12718$**	**$\Sigma x^2=2224$**	**$\Sigma y^2=81130$**

$$r = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{\sqrt{[\, n\Sigma x^2 - (\Sigma x)^2 \,][\, n\Sigma y^2 - (\Sigma y)^2 \,]}}$$

$= 10 \text{ X } 12718 - 140 \text{ X } 900 / \sqrt{10 \text{ X } 2224 - (140)^2} \sqrt{10 \text{ X } 81130 - (900)^2}$

$= 1180 / \sqrt{2640 \text{ X } 1300}$

$= 0.6370$

Sum 3: Calculate the coefficient of correlation between Expenditure on Advertisement in Rs. '000 (X) and Sales in Rs. Lakhs (Y) after allowing a time lag of two months.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
X	40	45	47	50	53	60	57	51	48	45
Y	75	69	65	64	70	71	75	83	90	92

Solution: As a time lag of two months is to be allowed, the following pairs of values are available

X	**Y**	**XY**	**X^2**	**Y^2**
40	65	2600	1600	4225
45	64	2880	2025	4096
47	70	3290	2209	4900
50	71	3550	2500	5041
53	75	3975	2809	5625
60	83	4980	3600	6889
57	90	5130	3249	8100
51	92	4692	2601	8464
$\sum X=403$	**$\sum Y=610$**	**$\sum XY=31097$**	**$\sum x^2=20593$**	**$\sum y^2=47340$**

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[\,n\sum x^2 - (\sum x)^2\,]\,[\,n\sum y^2 - (\sum y)^2\,]}}$$

$$= 8 \text{ X } 31097 - 403 \text{ X } 610 \,/\, \sqrt{8 \text{ X } 20593 - (403)^2}\sqrt{8 \text{ X } 47340 - (610)^2}$$

$$= 2946 \,/\, \sqrt{2335 \text{ X } 6620}$$

$$= 0.7493$$

Sum 4: From the following data, compute the coefficient of correlation between X and Y.

	X	**Y**
Sum of squares of deviations from the arithmetic mean	8250	724
Sum of products of deviations of X an Y from respective means	2350	
No. of pairs of observations	10	

Solution

$$r = \frac{\Sigma XY}{\sqrt{\Sigma X^2 . \Sigma Y^2}}$$

$= 2350 / \sqrt{8250} \ \sqrt{724}$

$= 0.9615$

Sum 5: Compute the coefficient of Correlation between X –Advertisement Expenditure and Y-Sales

X	10	12	18	8	13	20	22	15	5	17
Y	88	90	94	86	87	92	96	94	88	85

Solution

X	**Y**	**u=X-a/c A=15;C=1**	**v=Y-b b=90;d=1**	**uv**	**U^2**	**V^2**
10	88	-5	-2	10	25	4
12	90	-3	0	0	9	0
18	94	3	4	12	9	16
8	86	-7	-4	28	49	16
13	87	-2	-3	6	4	9
20	92	5	2	10	25	4
22	96	7	6	42	49	36
15	94	0	4	0	0	16
5	88	-10	-2	20	100	4
17	85	2	-5	-10	4	25
Total	**-**	**$\sum u$=-10**	**$\sum v$=0**	**$\sum uv$=118**	**$\sum u^2$=274**	**$\sum v^2$=130**

$$r = \frac{n\Sigma uv - (\Sigma u)(\Sigma v)}{\sqrt{\left[n\Sigma u^2 - (\Sigma u)^2\right].\left[(n\Sigma v^2) - (\Sigma v)^2\right]}}$$

$= 10 \text{ X } 118 - (-10)\ (0) / \sqrt{10 \text{ X } 274 - (-10)^2} \sqrt{10 \text{ X } 130 - (0)^2}$

$= 1180 / \sqrt{2640} \ \sqrt{1300}$

$= 0.6370$

Sum 6: From the following data find the percentage of variation in Y that is explained by the variation in X; N=11; $\sum X=117$; $\sum Y=260$; $\sum X^2=1313$; $\sum Y^2=6580$; $\sum XY=2827$

Solution

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$= 11 \times 2827 - (117)(260) \sqrt{11 \times 1313 - (117)^2} \sqrt{11 \times 6580 - (260)^2}$$

$$= 677 / \sqrt{754} \sqrt{4780}$$

$$= 0.3566$$

Sum 7: Find Karl Pearson's Coefficient of Correlation from the marks secured by 10 students in Accountancy and Statistics

Accountancy	45	70	65	30	90	40	50	75	85	60
Statistics	35	90	70	40	95	40	60	80	80	50

Solution

X	**Y**	**X=X-X X=61**	**Y=Y-Y Y=64**	**XY**	**X^2**	**Y^2**
45	35	-16	-29	464	256	841
70	90	9	26	234	81	676
65	70	4	6	24	16	36
30	40	-31	-24	744	961	576
90	95	29	31	899	841	961
40	40	-21	-24	504	441	576
50	60	-11	-4	44	121	16
75	80	14	16	224	196	256

85	80	24	16	384	576	256
60	50	-1	-14	14	1	196
∑x=610	**∑y=640**	**∑x=0**	**∑y=0**	**∑xy=3535**	**$\sum x^2$=3490**	**$\sum y^2$=4390**

Mean of X = ∑X/N = 610/10 =61
Mean of Y = ∑Y/N = 640/10=64

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$$

= 3535 / $\sqrt{3490}\sqrt{4390}$

= 0.9031

Sum 8: Calculate Karl Pearson's Coefficient of Correlation from the following data

Roll Nos.	**1**	**2**	**3**	**4**	**5**
Marks in Accountancy	48	35	17	23	47
Marks in Statistics	45	20	40	25	45

Solution

X	**Y**	**X=X-X X=34**	**Y=Y-Y Y=35**	**XY**	**X^2**	**Y^2**
48	45	14	10	140	196	100
35	20	1	-15	-15	1	225
17	40	-17	5	-85	289	25
23	25	-11	-10	110	121	100
47	45	13	10	130	169	100
∑x=170	**∑y=175**	**∑x=0**	**∑y=0**	**∑xy=280**	**$\sum x^2$=776**	**$\sum y^2$=550**

Mean of X = ∑X/N = 170/5 = 34
Mean of Y = ∑Y/N = 175/5 = 35

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$$

= 280 $\sqrt{776 \text{ X } 550}$

= 280 / 653.299

= 0.429

Sum 9: The following table gives indices of industrial production of registered unemployed (in hundred thousand). Calculate the value of the coefficient so obtained.

Year	1991	1992	1993	1994	1995	1996	1997	1998
Index of Production	100	102	104	107	105	112	103	99
Number Unemployed	15	12	13	11	12	12	19	26

Solution

X	**Y**	**X=X-X X=34**	**Y=Y-Y Y=35**	**XY**	**X^2**	**Y^2**
100	15	-4	0	0	16	0
102	12	-2	-3	6	4	9
104	13	0	-2	0	0	4
107	11	3	-4	-12	9	16
105	12	1	-3	-3	1	9
112	12	8	-3	-24	64	9
103	19	-1	4	-4	1	16
99	26	-5	11	-55	25	121
$\sum x=832$	**$\sum y=120$**	**$\sum x=0$**	**$\sum y=0$**	**$\sum xy=92$**	**$\sum x^2=120$**	**$\sum y^2=184$**

Mean of X = $\sum X/N$ = 832/8 = 104

Mean of Y = $\sum Y/N$ = 120/8 = 15

$$r = \frac{\sum XY}{\sqrt{\sum X^2 . \sum Y^2}}$$

$= -91 \sqrt{120 \text{ X } 184}$

$= 0.619$

SPEARMAN'S RANK CORRELATION

The Karl Pearson's method is based on the assumption that the population being studied is normally distributed. When it is known that the population is not normal or when the shape of the distribution is not known, there is need for a measure of correlation that involves no assumption about the parameter of the population.

It is possible to avoid making any assumption about the populations being studied by ranking the observations according to size and basing the calculations on the ranks rather than upon the original observations. It does not matter which way the items are ranked, item number one may be the largest or it may be the smallest. Using ranks rather than actual observations gives the coefficient of the rank correlation.

This method of finding out co-variability or the lack of it between two variables was developed by the British Psychologist Charles Edward Spearman in 1904. This measure is especially useful when quantitative measures for certain factors (such as in the evaluation of leadership ability or the judgement of female beauty) cannot be fixed, but the individual in the group can be arranged in order thereby obtaining for each individual a number indicating his (her) rank in the group.

$$\rho = 1-\left[\frac{6\sum d^2}{N\,(N^2-1)}\right]$$

Where there is no tie. d-difference between X and Y ranks.

$$\rho = 1-\left[\frac{6\,(\sum d^2 + (m(m^2-1)/12)}{N\,(N^2-1)}\right]$$

When one value occurs m times

$$\rho = 1-\left[\frac{6\,(\sum d^2 + (m(m^2-1)/12) + ((m(m^2-1)/12)}{N\,(N^2-1)}\right]$$

When more than one value is repeated

It is calculated when ranks are given or when rank correlation coefficient is required. Rank correlation coefficient also lies between -1 and +1.

Sum 1: Rankings of 10 trainees at the beginning (x) and at the end (y) of a certain course are given below:

	A	B	C	D	E	F	G	H	I	J

X	1	6	3	9	5	2	7	10	8	4
Y	6	8	3	7	2	1	5	9	4	10

Calculate Spearman's rank correlation coefficient.

Solution

X	**Y**	**d**	**D^2**
1	6	-5	25
6	8	-2	4
3	3	0	0
9	7	2	4
5	2	3	9
2	1	1	1
7	5	2	4
10	9	1	1
8	4	4	16
4	10	-6	36
Total	**-**	**$\sum d=0$**	**$\sum d^2=100$**

$$\rho = 1-\left(\frac{6\sum d^2}{N(N^2-1)}\right)$$

$$\rho = 1-\left(\frac{6X100}{10X99}\right)$$

$= 1 – 0.6061$

$= 0.3939$

Sum 2: From the data given below, calculate the rank correlation coefficient.

X	21	36	42	37	25
Y	47	40	37	42	43

Solution

X	**Y**	**X**	**Y**	**d**	**D^2**
21	47	5	1	4	16
36	40	3	4	-1	1
42	37	1	5	-4	16
37	42	2	3	-1	1

25	43	4	2	2	4
Total	**-**	**-**		**$\sum d=0$**	**$\sum d^2=38$**

$$\rho = 1-\left(\frac{6\sum d^2}{N(N^2-1)}\right)$$

$$\rho = 1-\left(\frac{6X38}{5(5^2-1)}\right)$$

$$\rho = 1-\left(\frac{6X38}{5 X 24}\right)$$

$$= 1 - 1.9$$

$$= -0.9$$

Sum 3: Find the rank correlation coefficient for the percentage of marks secured by a group of 8 students in Economics and Statistics.

Marks in Economics	50	60	65	70	75	40	70	80
Marks in Statistics	80	71	60	75	90	82	70	50

Solution

X	**Y**	**X**	**Y**	**d**	**D^2**
50	80	7	3	4	16
60	71	6	5	1	1
65	60	5	7	-2	4
70	75	3.5	4	-0.5	0.25
75	90	2	1	1	1
40	82	8	2	6	36
70	70	3.5	6	-2.5	6.25
80	50	1	8	-7	49
Total	**-**	**-**	**-**	**$\sum d=0$**	**$\sum d^2=113.50$**

$$\rho = 1-\left[\frac{6\ (\sum d^2 + (m(m^2-1)/12)}{N\ (N^2-1)}\right]$$

$$\rho = 1-\left[\frac{6\ (113.5 + 0.5)}{8\ (8^2-1)}\right]$$

$$\rho = 1-\left[\frac{6\ X\ 114}{8\ X\ 63}\right]$$

$= 1-1.3571$

$= -0.3571$

Sum 4: Marks obtained by 8 students in Accountancy and Statistics are given below. Compute rank correlation.

X	15	20	28	12	40	60	20	80
Y	40	30	50	30	20	10	30	60

Solution:

$$\rho = 1-\left[\frac{6\ (\sum d^2 + (m(m^2-1)/12) + ((m(m^2-1)/12)}{N\ (N^2-1)}\right]$$

$$\rho = 1-\left[\frac{6\ (81.5+0.5+2)}{8\ (8^2-1)}\right]$$

$$\rho = 1-\left[\frac{6\ X\ 84}{8\ X\ 63}\right]$$

$= 0$

REGRESSION

After having established the fact that two variables are closely related we may be interested in estimating (predicting) the value of one variable given the value of another.

For example, if we know that advertising and sales are correlated we find out expected amount of sales for a given advertising expenditure or the required amount of expenditure for attaining a given amount of sales. Similarly, if we know that the yield of rice and rainfall are closely related we may find out the amount of rain required to achieve a certain production figure. Regression analysis reveals average relationship between two variables and this makes possible estimation or prediction.

The dictionary meaning of the term 'regression' is the act of returning or going back. The term 'regression' was first used by Sir Francis Galton (1822-1911) in 1877 while studying the relationship between the height of fathers and sons. This term was introduced by him in the paper 'Regression towards fathers and sons revealed a very interesting relationship, i.e., tall fathers tend to have tall sons and short fathers short sons, but the average height of the sons of a group of tall fathers is less than that of the fathers and the average fathers. The line describing the tendency to regress or going back was called by Galton s 'Regression Line'. The term is still used to describe that line drawn for a group of points to represent the trend present, but it no longer necessarily carries the original implication of "stepping back" that Galton intended. These days there is a growing tendency to the modern writers to use the term estimating line instead of regression line because the expression estimating line is more clarificatory in character.

DEFINITIONS

Regression is the measure of the average relationship between two or more variables in terms of the original units of the data.

The term 'regression analysis' refers to the methods by which estimates are made of the values of a variable from a knowledge of the values of one or more other variables and to the measurement of the errors involved in this estimation process – Morris Hamburg.

One of the most frequently used techniques in economics and business research to find a relation between two or more variables that are related causally, is regression analysis – Taro Yamane

Regression analysis attempts to establish the 'nature of the relationship' between variables – that is, to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting" – Ya Lun Chou

It is clear from the above definitions that regression analysis is a statistical device with the help of which we are in a position to estimate (or predict) the unknown values of one variable from known values of another variable. The variable which is used to predict the variable of interest is called the independent variable or exploratory variable and the variable we are trying to predict is called the dependent variable or "explained variable. The independent variable is denoted by X and the dependent variable by Y. The analysis used is called the simple regression analysis – simple because there is only one predictor or independent variable, and linear because of the assumed linear relationship between the dependent and the independent variables. The term "linear" means that an equation of a straight line of the form Y=a+bx, where a and b are constants, is used to describe the average relationship that exists between the two variables.

It should be noted that the term 'dependent' and 'independent' refer to the mathematical or functional meaning of dependence – they do not imply that there is necessarily any cause and effect relationship between the variables. What is meant is simply that estimates of values of the dependent variable Y may be obtained for given values of the independent variable X from a mathematical function involving X and Y. In that sense, the values of Y are dependent upon the values of X. The X variable may or may not be causing change in the Y variable. For example, while estimating sales of a product from figures on advertising expenditures, sale is generally taken as the dependent variable. However, there may or may not be causal connection between these two factors in the sense that changes in advertising expenditures cause change in sales. In fact, in

certain cases, the cause-effect relation may be just opposite of what appears to be the obvious one.

Uses of Regression Analysis

Regression analysis is a branch of statistical theory that is widely used in almost all the scientific disciplines. In economics it is the basic technique for measuring or estimating the relationship among economic variables that constitute the essence of economic theory and economic life. For example, if we know that two variables, Price (X) and Demand (Y), are closely related we can find out the most probable value of X for a given value of Y or the most probable value of Y for a given value of X. Similarly, if we closely related, we can find out the expected price for a certain price for a certain amount of tax levy. Thus, we find that the study of regression is of considerable help to the economists and businessmen. The uses of regression are not confined to economics and business field only. Its applications are extended to almost all the natural, physical and social sciences. The regression analysis attempts to accomplish the following:

1. Regression analysis provides estimates of values of the dependent variable from values of the independent variable. The device used to accomplish this estimation procedure is the regression line. The regression line describes the average relationship existing between X and Y variables. i.e., it displays mean values of X for given values of Y. The equation of this line, known as the regression equation, provides estimates of the dependent variable when values of the independent variable are inserted into the equation.

2. A second goal of regression analysis is to obtain a measure of the error involved in using the regression line as a basis for estimation. For this purpose the standard error of estimate is calculated. This is a measure of the scatter or spread of the observed values of Y around the corresponding values estimated from the regression line. If the line fits the data closely, that is, if there is little scatter of the observations around the regression line, good estimates can be made of the Y variable. On the other hand, if

there is a great deal of scatter or the observations around the fitted regression line, the line will not produce accurate estimates of the dependent variable

3. With the help of regression coefficients we can calculate the correlation coefficient. The square of correlation coefficient (r) called coefficient of determination, measures the degree of association of correlation that exists between the two variables. It assesses the proportion of variance in the dependent variable that has been accounted for by the regression equation. In general, the greater the value of r^2 the better is the fit and the more useful the regression equations as a predictive device.

DIFFERENCE BETWEEN CORRELATION AND REGRESSION

Correlation	Regression
Correlation is the relationship between variables. It is expressed numerically	Regression means going back. The average relation between the variables is given as an equation
Between two variables, none is identified as independent variable	One of the variables is independent variable and the other is dependent variable in any particular context
Correlation does not mean causation. One variable need not be the cause and the other, effect	Independent variable may be the 'the cause' and dependent variable, 'the effect'
There is spurious or nonsense correlation	There is no such possibility. Regression is considered only when the variables are related
Correlation coefficient is independent of change of origin and scale	Regression coefficients are independent of change or origin but are affected by change of scale
Correlation coefficient is a number -1 and +1	The two regression coefficients have the same sign, + or -. One of them can be greater than 1 numerically. But they can not be greater than 1 numerically simultaneously.
Correlation coefficient is not in any unit of measurement	Each regression coefficient is in the unit of measurement of the dependent variable
Correlation coefficient indicates the direction of co-variation and the closeness of the linear relation between two variables	Regression equations give the value of the dependent variable corresponding to any value of the independent variable

Class: III B.COM PA **Course Name: Research Methodology**
Course Code: 17PAU501A **Unit III**
BATCH: 2017-2020

Correlation	Regression
The significance of the sample correlation coefficient can be tested. The limits between which the population correlation coefficient is expected to lie can be found	Target can be reached. The value of the independent variable can be chosen so as to get the target value of the dependent variable. For example, a specific amount can be spent on advertisement to get the targeted revenue.

Sum 1: From the following data, obtain the two regression equations:

X	6	2	10	4	8
Y	9	11	5	8	7

Solution

X	Y	XY	X^2	Y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\sum X=30$	**$\sum Y=40$**	**$\sum XY=214$**	**$\sum X^2=220$**	**$\sum Y^2=340$**

Let the regression equation of Y on X be Y=A+BX

The normal equations are $\sum Y = NA + B\sum X$

$\sum XY = A\sum X + B\sum X^2$

By substituting the value from the table,

$5A + 30B = 40$ Say (1)

$30A + 220B = 214$ Say (2)

(1) X 6 $30A + 180B = 240$ Say (3)

(2) – (3) $40B = -26$

$B = -26 / 40$

$= -0.6500$

From (1), $5A-30X0.6500 = 40$

A = 40 + 19.5 / 5

= 11.90

The regression equation of Y on X is

Y = 11.90 – 0.6500X

Let the regression equation of X on Y be X = A + BY

The normal equations are $\sum X = NA + B\sum Y$

$\sum XY = A\sum Y + B\sum Y^2$

By substituting the values from the table,

5A + 40B = 30 Say (4)

40A + 340 B = 214 Say (5)

(4) X 8 40A = 320 B = 240 Say (6)

(5) – (6) 20B = - 26

B = -26/20

= -1.300

From (4), 5A+40 X (-1.30) = 30

A = 30 + 52 /5

= 16.40

The regression equation of X on Y is X = 16.40 – 1.300

Sum 2: You are given the following data:

	X	**Y**
Arithmetic Mean	36	85
Standard Deviation	11	8
Correlation coefficient between X and Y	0.66	

(a) Find the two regression equations

(b) Estimate the value of X when Y=75

Solution

bxy = rσx/σy = 0.66 x 11 / 8 = 0.9075

byx = rσy/σx = 0.66 X 8 / 11 = 0.4800

a) Regression equation of Y on X

$Y - \bar{Y} = byx\ (X-\bar{X})$

$Y - 85 = 0.4800\ (X-36)$

$= 0.4800\ X - 17.28$

$Y = 67.72 + 0.4800\ X$

Regression equation of X on Y

$X - \bar{X} = bxy\ (Y-\bar{Y})$

$X - 36 = 0.9075\ (Y-85)$

$= 0.9075 - 77.14$

$X = 0.9075\ Y - 41.14$

b) When Y=75, $X = 0.9075\ X\ 75 - 41.14$

$= 26.92$

Sum 3: From the following information on values of two variables X and Y find the two regression lines and the correlation coefficient.

N=10; $\sum X=20$; $\sum Y=40$; $\sum X^2=240$; $\sum Y^2=410$; $\sum XY=200$

Solution

$\bar{X} = \sum X/N = 20/10 = 2.00$

$\bar{Y} = \sum Y/N = 40/10 = 4.00$

$$bxy = \frac{N\sum XY-(\sum X)(\sum Y)}{N\sum Y^2-(\sum Y)^2} \text{ as } \sum X\neq 0 \text{ and } \sum Y\neq 0$$

$$= \frac{10\ X\ 200 - 20\ X\ 40}{10\ X\ 410 - (40)^2}$$

$$= \frac{2000 - 800}{4100\text{-}1600}$$

$$= \frac{1200}{2500} = 0.4800$$

$$byx = \frac{N\sum XY-(\sum X)(\sum Y)}{N\sum X^2-(\sum X)^2} \text{ as } \sum X\neq 0 \text{ and } \sum Y\neq 0$$

$$= \frac{1200}{10X240-(20)^2}$$

$$= \frac{1200}{2000} = 0.6000$$

Regression equation of Y on X

$Y - \bar{Y} = byx\ (X-\bar{X})$

$Y-4 = 0.6000\ (X-2)$

$= 0.6000\ X - 1.20$

$Y = 2.80 + 0.6000X$

Regression equation of X on Y

$X - \bar{X} = bxy\ (Y-\bar{Y})$

$X-2 = 0.4800\ (Y-4)$

$= 0.4800\ Y - 1.92$

$X = 0.08 + 0.4800Y$

Sum 4: Calculate the two regression equations from the following data:

X	10	12	13	12	16	15
Y	40	38	43	45	37	43

Also estimate Y when X=20.

Solution

X	**Y**	**XY**	**X^2**	**Y^2**
10	40	400	100	1600
12	38	458	144	1444
13	43	559	169	1849

12	45	540	144	2025
16	37	592	256	1369
15	43	645	225	1849
$\sum X=78$	**$\sum Y=246$**	**$\sum XY=3192$**	**$\sum X^2=1038$**	**$\sum Y^2=10136$**

$\bar{X} = \sum X/N = 78/6 = 13.00$

$\bar{Y} = \sum Y/N = 246/6 = 41.00$

$$bxy = \frac{N\sum XY-(\sum X)(\sum Y)}{N\sum Y^2-(\sum Y)^2}$$

$$= \frac{6 \text{ X } 3192 - 78 \text{ X } 246}{6 \text{ X } 10316 - (246)^2}$$

$$= \frac{19152 - 19188}{60816-60516}$$

$$= \frac{-36}{300} = 0.1200$$

$$byx = \frac{N\sum XY-(\sum X)(\sum Y)}{N\sum X^2-(\sum X)^2}$$

$$= \frac{-36}{6X1038-(78)^2}$$

$$= \frac{-36}{6228-6084} = 0.6000$$

$$= \frac{-36}{} = 0.2500$$

144

Regression equation of Y on X

$Y - \bar{Y} = byx (X-\bar{X})$

$Y-41 = -0.2500 (X-13)$

$= -0.2500 X + 3.25$

$Y = 44.25 + 0.25X$

When X = 20, Y=44.25-0.25 X 20 = 39.25

Regression equation of X on Y

$X - \bar{X} = bxy (Y-\bar{Y})$

$X-13 = -0.1200 (Y-41)$

$= -0.1200 Y + 4.92$

$X = 17.92 - 0.12Y$

Sum 5: From the data given below, find two regression equations

Marks in Mathematics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Solution

X	**Y**	**$X=X-\bar{X}$ $\bar{X}=32$**	**$Y=Y-\bar{Y}$ $\bar{Y}=38$**	**XY**	**X^2**	**Y^2**
25	43	-7	5	-35	49	25
28	46	-4	8	-32	16	64
35	49	3	11	33	9	121
32	41	0	3	0	0	9
31	36	-1	-2	2	1	4
36	32	4	-6	-24	16	36
29	31	-3	-7	21	9	49
38	30	6	-8	-48	36	64
34	33	2	-5	-10	4	25
32	39	0	1	0	0	1
$\sum X=320$	**$\sum Y=380$**	**$\sum X=0$**	**$\sum Y=0$**	**$\sum XY=-93$**	**$\sum X^2=140$**	**$\sum Y^2=398$**

Solution

$\overline{X} = \sum X/N = 320/10 = 32.00$

$\overline{Y} = \sum Y/N = 380/10 = 38.00$

$bxy = \sum xy/\sum y^2 = -93/398 = -0.2337$

$byx = \sum xy/\sum x^2 = -93/140 = -0.6643$

a) Regression equation of Y on X

$Y - \overline{Y} = byx\ (X-\overline{X})$

$Y-38 = -0.6643\ X\ (X-32)$

$= -0.6643\ X + 21.26$

$= 59.26 - 0.6643X$

Regression equation of X on Y

$X - \overline{X} = bxy\ (Y-\overline{Y})$

$X-32 = -0.2337\ (Y-38)$

$= -0.2337Y + 8.88$

$= 40.88 - 0.2337Y$

KAHE

UNIT IV

Hypothesis - Meaning - Sources -Types - Formulation - Data Analysis - Z test (mean, diff. of mean, diff. of proportion) - t-test (mean) - Paired t-test - Chi square test - Introduction to theoretical concept of ANOVA - Factor Analysis and Discriminant Analysis.

HYPOTHESIS

Hypothesis is usually considered as the principal instrument in research. Its main function is to suggest new experiments and observations. In fact, many experiments are carried out with the deliberate object of testing hypotheses. Decision-makers often face situations wherein they are interested in testing hypotheses on the basis of available information and then take decisions on the basis of such testing. In social science, where direct knowledge of population parameter(s) is rare, hypothesis testing is the often used strategy for deciding whether a sample data offer such support for a hypothesis that generalization can be made. Thus hypothesis testing enables us to make probability statements about population parameter(s). The hypothesis may not be proved absolutely, but in practice it is accepted if it has withstood a critical testing. Before we explain how hypotheses are tested through different tests meant for the purpose, it will be appropriate to explain clearly the meaning of a hypothesis and the related concepts for better understanding of the hypothesis testing techniques.

What is Hypothesis?

Ordinarily, when one talks about hypothesis, one simply means a mere assumption or some supposition to be proved or disproved. But for a researcher hypothesis is a formal question that he intends to resolve. Thus a hypothesis may be defined as a proposition or a set of proposition set forth as an explanation for the occurrence of some specified group of phenomena either asserted merely as a provisional conjecture to guide some investigation or accepted as highly probable in the light of established facts. Quite often a research hypothesis is a predictive statement, capable of being tested by scientific methods, that relates an independent variable to some dependent variable. For example, consider statements like the following ones:

"Students who receive counselling will show a greater increase in creativity than students not receiving counselling" Or

"the automobile A is performing as well as automobile B."

These are hypotheses capable of being objectively verified and tested. Thus, we may conclude that a hypothesis states what we are looking for and it is a proposition which can be put to a test to determine its validity.

CHARACTERISTICS OF HYPOTHESIS

1. Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.
2. Hypothesis should be capable of being tested. In a swamp of untestable hypotheses, many a time the research programmes have bogged down. Some prior study may be done by researcher in order to make hypothesis a testable one. A hypothesis "is testable if other

deductions can be made from it which, in turn, can be confirmed or disproved by observation."

3. Hypothesis should state relationship between variables, if it happens to be a relational hypothesis.
4. Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and he should develop such hypotheses.
5. Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.
6. Hypothesis should be consistent with most known facts i.e., it must be consistent with a substantial body of established facts. In other words, it should be one which judges accept as being the most likely.
7. Hypothesis should be amenable to testing within a reasonable time. One should not use even an excellent hypothesis, if the same cannot be tested in reasonable time for one cannot spend a life-time collecting data to test it.
8. Hypothesis must explain the facts that gave rise to the need for explanation. This means that by using the hypothesis plus other known and accepted generalizations, one should be able to deduce the original problem condition. Thus hypothesis must actually explain what it claims to explain; it should have empirical reference.

FORMULATION OF HYPOTHESIS

1. A hypothesis may originate in different ways. A cultural environment may give rise to it. In India, for example, religion and custom dominate the way of life. This has had it reaction on economic values and individual initiative in various walks of life. Such a situation could give rise to any number of hypothesis; sociological, cultural, political and economic
2. A second source of hypothesis is folk wisdom or current popular beliefs and practices suggesting both the problems and the hypothesis
3. Analogies are often a spring of valuable hypothesis. Students of sociology in the course of their studies would have come across analogies wherein a society is compared to a biological organism, the natural law to the social law, thermo-dynamics to social dynamics etc.,
4. The history of science provides an eloquent testimony to the fact that personal and idiosyncratic experiences of the scientist contributes a great deal to the type and form of questions he may ask as also to the kinds of tentative answers to these questions that he can provide
5. Hypotheses may also rest on the findings of other studies. The researcher on the basis of the findings of other studies may hypothesise that similar relationship between specified variables will hold good in the present study, too. This is a very common way of researchers who design their study with a view to replicating another study conducted in a different concrete context
6. Another source of hypothesis formulation in cases which are expectations to accepted theory
7. Personal experience and individual reaction may give rise to hypotheses

8. A hypotheses may turn from a body of theory which, by way of logical deduction, may lead to the production that if certain conditions are present, certain results will follow. Theory is indeed an extremely fertile seed-bed of hypotheses.

PROCEDURE FOR HYPOTHESIS TESTING

To test a hypothesis means to tell (on the basis of the data the researcher has collected) whether or not the hypothesis seems to be valid. In hypothesis testing the main question is: whether to accept the null hypothesis or not to accept the null hypothesis? Procedure for hypothesis testing refers to all those steps that we undertake for making a choice between the two actions i.e., rejection and acceptance of a null hypothesis. The various steps involved in hypothesis testing are stated below:

1) Making a Formal Statement

The step consists in making a formal statement of the null hypothesis (H0) and also of the alternative Hypothesis (Ha). This means that hypothesis should be clearly stated, considering the nature of the research problem. For instance, Mr. Mohan of the Civil Engineering Department wants to test the load bearing capacity of an old bridge which must be more than 10 tons, in that case he can state his hypotheses as under:

Null Hypothesis H_0 : μ =10 tons

Alternative Hypothesis H_a: μ> 10 tons

Take another example. The average score in an aptitude test administered at the national level is 80. To evaluate a state's education system, the average score of 100 of the state's students selected on random basis was 75. The state wants to know if there is a significant

difference between the local scores and the national scores. In such a situation the hypotheses may be stated as under:

Null Hypothesis $H_0 : \mu = 80$

Alternative Hypothesis $H_a: \mu \neq 80$

The formulation of hypotheses is an important step which must be accomplished with due care in accordance with the object and nature of the problem under consideration. It also indicates whether we should use a one-tailed test or a two-tailed test. If Ha is of the type greater than (or of the type lesser than), we use a one-tailed test, but when Ha is of the type "whether greater or smaller" then we use a two-tailed test.

2) Selecting a Significance Level

The hypotheses are tested on a pre-determined level of significance and as such the same should be specified. Generally, in practice, either 5% level or 1% level is adopted for the purpose. The factors that affect the level of significance are: (a) the magnitude of the difference between sample means; (b) the size of the samples; (c) the variability of measurements within samples; and (d) whether the hypothesis is directional or non-directional (A directional hypothesis is one which predicts the direction of the difference between, say, means). In brief, the level of significance must be adequate in the context of the purpose and nature of enquiry.

3) Deciding the Distribution to Use

After deciding the level of significance, the next step in hypothesis testing is to determine the appropriate sampling distribution. The choice generally remains between normal distribution

and the t-distribution. The rules for selecting the correct distribution are similar to those which we have stated earlier in the context of estimation.

4) Selecting a Random Sample and Computing an Appropriate Value

Another step is to select a random sample(s) and compute an appropriate value from the sample data concerning the test statistic utilizing the relevant distribution. In other words, draw a sample to furnish empirical data.

5) Calculation of the Probability

One has then to calculate the probability that the sample result would diverge as widely as it has from expectations, if the null hypothesis were in fact true.

6) Comparing the Probability

Yet another step consists in comparing the probability thus calculated with the specified value for α□, the significance level. If the calculated probability is equal to or smaller than the α value in case of one-tailed test (and α □/2 in case of two-tailed test), then reject the null hypothesis (i.e., accept the alternative hypothesis), but if the calculated probability is greater, then accept the null hypothesis. In case we reject H_0, we run a risk of (at most the level of significance) committing an error of Type I, but if we accept H_0, then we run some risk (the size of which cannot be specified as long as the H_0 happens to be vague rather than specific) of committing an error of Type II.

PARAMETRIC TEST

't' Test

When the size of sample is small (less than 30). In particular, it will no longer be possible for us to assume (a) that the random sampling distribution of a statistic is approximately normal

and (b) that values given by the sample data are sufficiently close to the population values and can be used in their place for the calculation of the standard error of the estimate.

The removal of these assumptions makes it necessary to use entirely new techniques to deal with the problems of small samples. The division between the theories of large and small samples is, therefore, a very real one, though it is not always easy to draw a precise line of demarcation. It should be noted that as a rule, the methods and the theory of small samples are applicable to large samples, though the reverse is not true.

While dealing with small samples our main interest is not to estimate the population values as in true in large samples; rather our interest lies in testing a given hypothesis, i.e., in ascertaining whether observed values could have arisen by sampling fluctuations from some value given in advance. For example, if a sample of 15 gives a correlation coefficient of +0.4, we shall be interested not so much in the value of the correlation in the parent population, but more generally whether this value could have been arisen from an uncorrelated population. i.e. whether it is significant of correlation in the parent population.

It should be noted that the investigator who works with very small samples must know that his estimates will vary widely from sample to sample. Moreover, he must be satisfied with relatively wide confidence intervals. Precision of statement is less, of course, the wider the intervals employed. Each inference drawn from large sample results in far more precise in the limits it sets up than is an inference based on a much smaller sample.

Student's 't' Distribution

Theoretical work on t-distribution was done by W.S. Gosset (1876-1937) in the early 1900. Gosset was employed by the Guinness and Son, a Dulbin bravery, Ireland, which did not permit employees to publish research findings under their own names. So Gosset adopted the pen name "Student" and published his findings under their name. Thereafter, the t-distribution is commonly called Student's t-distribution or simply student's distribution.

The t-distribution is used when sample size is 30 or less and the population standard deviation is unknown.

The 't' statistic is defined as

$t = X-\mu/S \times \sqrt{n}$

Where $S = \sqrt{\sum(X-X)^2/ n-1}$

Sum 1: The manufacturer of a certain make of electric bulbs claims that his bulbs have a mean life of 25 months with a standard deviation of 5 months. A random sample of 6 such bulbs gave the following values

Life of Months	24	26	30	20	20	18

Can you regard the producer's claim to be valid at 1% level of significance? (Given that the table values of the appropriate test statistics at the said level are 4.032, 3.707 and 3.499 for 5, 6 and 7 degrees of freedom respectively)

Solution

Let us taken the hypothesis that there is no significant difference in the mean life of bulbs in the sample and that of the population, Applying t-test:

X	**$(X-\bar{X})$ X**	**X^2**
24	+1	1
26	+3	9
30	+7	49
20	-3	9
20	-3	9
18	-5	25
$\sum X=138$		$\sum X^2=102$

$t = X-\mu/SD \times \sqrt{n}$

Average of X= $\sum X/n$ = 138/6 = 23

$$SD = \sqrt{\sum X^2/n-1} = \sqrt{102/5} = \sqrt{20.4} = 4.517$$

$$= \frac{(23-25)}{4.517}\sqrt{6} = \frac{2 \text{ X } 2.449}{4.517} = 1.084$$

$\nu = n-1 = 6-1 = 5$

For $\nu=5$ $t_{0.01}=4.032$

The calculated value of t is less than the table value. The hypothesis is accepted. Hence, the producer's claim is not valid at 1% level of significance.

Sum 2: A random sample of size 16 has 53 as mean. The sum of the squares of the deviation taken from mean is 135. Can this sample be regarded as taken from the population having 56 as mean? Obtain 95% and 99% confidence limits of the mean of the population. (For v=15, $t_{0.05}$=2.13 for v=15, $t_{0.01}$=2.95)

Solution

Let us take the hypothesis that there is no significant difference between the sample mean and hypothetical mean, Applying t-test.

$$t = \frac{\bar{X}-\mu}{SD} X \sqrt{n}$$

$$SD = \sqrt{\sum X^2/n-1} = \sqrt{135/15} = 3$$

$$= (53-56) / 3 \quad X\sqrt{16}$$

$$= 3 \text{ X } 4/ 3$$

$$= 4$$

v=16-1=15. For v=16, $t_{0.05}$=2.13.

The calculated value is more than the table value. The hypothesis is rejected. Hence, the sample has not come from a population having 56 as mean.

Sum: 3 The life time of electric bulbs for a random sample of 10 from a large consignment gave the following data:

Item	1	2	3	4	5	6	7	8	9	10
Life in '000 hours	4.2	4.6	3.9	4.1	5.2	3.8	3.9	4.3	4.4	5.6

Can we accept the hypothesis that the average life time of bulbs is 4000 hours.

Solution

Let us take the hypothesis that there is no significant difference in the sample mean and the hypothetical population mean

X	**$(X-\bar{X})$**	**$(X-\bar{X})^2$**
4.2	-0.2	0.04
4.6	+0.2	0.04
3.9	-0.5	0.25
4.1	-0.3	0.09
5.2	+0.8	0.64
3.8	-0.6	0.36
3.9	-0.5	0.25
4.3	-0.1	0.01
4.4	0.0	0.00
5.6	+1.2	1.44
$\sum X=44$		$\sum(X-X)^2=3.12$

$t= \bar{X}-\mu/S \; X \;\; n$

Where $S=\sqrt{\sum(X-X)^2/ n-1}$

$$t = \frac{4.4-4}{0.589}\sqrt{10} = \frac{0.4 \times 3.162}{0.589} = 0.2148$$

$\nu = n-1 = 10-1 = 9$

For $\nu=9$ $t_{0.05}=2.262$

The calculated value of t is less than the table value. The hypothesis is accepted. The average life time of the bulbs could be 4000 hours.

Sum 4: Two types of drugs were used on 5 and 7 patients for reducing their weight.

Drug A was imported and drug B indigenous. The decrease in the weight after using the drugs for six months was as follows:

Drug A	10	12	13	11	14		
Drug B	8	9	12	14	15	10	9

Is there a significant difference in the efficacy of the two drugs? If not, which drug should you buy (For $\nu=10$, $t_{0.05}=2.228$)

Solution

Let us take the hypothesis that there is no significant difference in the efficacy of the two drugs.

X_1	$(X_1 - \bar{X}_1)$	$(X_1 - \bar{X}_1)^2$	X_2	$(X_2 - \bar{X}_2)$	$(X_2 - \bar{X}_2)^2$
10	-2	4	8	-3	9
12	0	0	9	-2	4
13	+1	1	12	+1	1
11	-1	1	14	+3	9

14	+2	4	15	+4	16
			10	-1	1
			9	-2	4
$\sum X_1=60$		$\sum (X_1 - X_1)^2=10$	$\sum X_2=77$		$\sum (X_2 - X_2)^2=44$

$$t = \frac{\bar{X}_1 + \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$S = \sqrt{\frac{\sum(X_1-\bar{X}_1)^2 + \sum(X_2-\bar{X}_2)^2}{n_1+n_2-2}}$$

$$S = \sqrt{\frac{10 + 44}{5+7-2}} = \sqrt{\frac{54}{10}} = 2.324$$

$$= \frac{12-11}{2.324} \quad \frac{5 \times 7}{5+7} = \frac{1.708}{2.324} = 0.735$$

$\nu = n_1+n_2-2 = 5+ 7 -2 = 10$

$\nu = 10$, $t_{0.05}=2.228$

The calculated value of t is less than the table value, the hypothesis is accepted. Hence, there is no significance in the efficacy of two drugs. Since drug B is indigenous and there is no

difference in the efficacy of imported and indigenous drug, we should buy indigenous drug, i.e., B.

Sum 5 : For a random sample of 10 persons, fed on diet A, the increased weight in pounds in a certain period were:

10	6	16	17	13	12	8	14	15	9

For another random sample of 12 persons, fed on diet B, the increase in the same period were:

7	13	22	15	12	14	18	8	21	23	10	17

Test whether the diets A and diet B differ significantly as regards their effect on increase in weight, Given the following.

Degrees of Freedom	19	20	21	22	23
Value at t at 5% level	2.09	2.09	2.08	2.07	2.07

Solution: Let us take the null hypothesis that A and B do not differ significantly weight regard to their effect on increase in weight.

X_1	$(X_1 - \bar{X}_1)$	$(X_1 - \bar{X}_1)^2$	X_2	$(X_2 - \bar{X}_2)$	$(X_2 - \bar{X}_2)^2$
10	-2	4	7	-8	64
6	-6	36	13	-2	4
16	+4	16	22	+7	49
17	+5	25	15	0	0
13	+1	1	12	-3	9
12	0	0	14	-1	1
8	-4	16	18	+3	9
14	+2	4	8	-7	49
15	+3	9	21	+6	36

9	-3	9	23	+8	64
			10	-5	25
			17	+2	4
$\sum X_1=120$		$\sum (X_1 - X_1)^2=120$	$\sum X_2=180$		$\sum (X_2 - X_2)^2=314$

Mean increase in weight of 10 persons fed on diet A

Mean increase in weight of 10 persons fed on diet B

$$t = \frac{\bar{X}_1 + \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1+n_2}}$$

$$S = \sqrt{\frac{\sum(X_1-X_1)^2 + \sum(X_2-X_2)^2}{n_1+n_2-2}}$$

$$SD = \frac{120 + 314}{10+12-2} = \frac{434}{20} = 4.66$$

$X_1=12$, $X_2=15$, $n_1=10$, $n_2 = 12$, $S=4.66$. Substituting the values in the above formula:

$$t = \frac{12 - 15}{4.66} \times \frac{10 \times 12}{10 + 12} = \frac{3}{4.66} \times 2.34 = 1.51$$

For v=20, the table value of t at 5 per cent level is 2.09. The calculated value is less than the table value and hence the experiment provides no evidence against the hypothesis. We, therefore, conclude that diets A and B do not differ significantly as regards their effect on increase in weight is concerned.

Sum 6: In a test given to two groups of students, the marks obtained are as follows:

I Group	18	20	36	50	49	36	34	49	41
II Group	29	28	26	35	30	44	46		

Examine the significance of difference between the arithmetic mean of the marks secured by the students of the above two groups. (The value of t at 5% level of significance for v=14 is 2.14)

Solution

Let us take the hypothesis that there is no significant difference in the arithmetic mean of the marks secured by the students of the two groups.

Group I X_1	$(X_1 - \overline{X}_1)$	$(X_1 - \overline{X}_1)^2$	**Group II** X_2	$(X_2 - \overline{X}_2)$	$(X_2 - \overline{X}_2)^2$
18	-19	361	29	-5	25
20	-17	289	28	-6	36
36	-1	1	26	-8	64
50	+13	169	35	+1	1
49	+12	144	30	-4	16
36	-1	1	44	+10	100
34	-3	9	46	+12	144
49	+12	144			
41	+4	16			
$\sum X_1=333$		$\sum (X_1 -$	$\sum X_2=238$		$\sum (X_2 - X_2)^2=386$

		$X_1)^2$=1134			

$$t = \frac{\bar{X}_1 + \bar{X}_2}{S} \times \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$S = \sqrt{\frac{\sum(X_1 - X_1)^2 + \sum(X_2 - X_2)^2}{n_1 + n_2 - 2}}$$

$$SD = \frac{1134 + 386}{9 + 7 - 2} = 10.42$$

$$t = \frac{37 - 34}{10.42} = \frac{9 \times 7}{9 + 7} = \frac{3}{10.42} \times 1.984 = 0.571$$

$\nu = n_1 + n_2 - 2 = 9 + 7 - 2 = 14$; For $\nu = 14$, $t_{0.05} = 2.14$

The calculated value of t is less than the table value and hence the hypothesis hold true. We, therefore, conclude that the mean marks of the students of the two groups do not differ significantly.

F-Test

The F-test is named in honor of the great Statistician R.A. Fisher. The object of F –test is to find out whether the two independent estimates of population variance differ significantly, or whether the two samples may be regarded as drawn from the normal populations having the same variance. For carrying out the test of significance, we calculate the ratio F. F is defined as

$F=S_1^2/S_2^2$, Where $S_1^2 = \sum(X_1-\overline{X_1})^2/n_1-1$ and

$$S_2^2 = \sum(X_2-\overline{X_2})^2/n_2-1$$

It should be noted that S_1^2 is always the larger estimate of variance, i.e., $S_1^2>S_2^2$

$$F = \frac{\text{Larger estimate of Variance}}{\text{Smaller estimate of Variance}}$$

The calculated value of F is compared with the table value for v1 and v2 at 5% or 1% level of significance. If calculated value of F is greater than the table value then the F ratio is considered significant and the null hypothesis is rejected. On the other hand, if the calculated value of F is less than the table value the null hypothesis is accepted and it is inferred that both the samples have come from the population having same variance.

Since F Test is based on the ratio of two variances, it is also known as the Variance Ratio Test. The ratio of two variances follows a distribution called the F distribution named after the famous statistician R.A. Fisher.

Sum 1: Two random samples were drawn from two normal populations and their values are :

A	66	67	75	76	82	84	88	90	92		
B	64	66	74	78	82	85	87	92	93	95	97

Test whether the two populations have the same variance at the 5% level of significance. (F=3.36) at 5% level of significance level v1=10 and v2=8.

Solution

Let us take the hypothesis that the two populations have the same variance

A X_1	**(X_1-X_1); $\bar{X}1$**	**X_1^2**	**B X_2**	**$(X_2-\bar{X}_2)$; X_2**	**c X_2^2**
66	-14	196	64	-19	361
67	-13	169	66	-17	289
75	-5	25	74	-9	81
76	-4	16	78	-5	25
82	2	4	82	-1	1
84	4	16	85	2	4
88	8	64	87	4	16
90	10	100	92	9	81
92	12	144	93	10	100
			95	12	144
			97	14	196
$\sum X_1=720$	**$\sum X_1=0$**	**$\sum X_1^2=734$**	**$\sum X_2=913$**	**$\sum X_2=0$**	**$\sum X_2^2=1298$**

Average of $X_1 = \sum X_1/n1 = 720 / 9 = 80$;

Average of $X_2 = \sum X_2/n2 = 913/ 11 = 83$

$S_1^2 = \sum(X_1-X_1)^2/n_1-1$

$= 734 / 9-1 = 91.75$

$S_2^2 = \sum(X_2-X_2)^2/n_2-1$

$= 1298 / 11-1 = 129.80$

$F = S_1^2/S_2^2$

$= 129.8 / 191.75$

$= 1.415$

The calculated value of F is less than the table value. The hypothesis is accepted. Hence, it may be calculated that the two populations have the same variance.

Sum 2: In a sample of 8 observations, the sum of squared deviations of items from the mean was 84.4. In another sample of 10 observations, the value was found to be 102.60. Test whether the difference is significant at 5% level.

You are given that at 5% level, critical value of F for v1=7 and v2=9 degrees of freedom is 3.29 and for v1=8 and v2=10 degrees of freedom, its value is 3.07.

Solution

Let us take hypothesis that the difference in the variance of the two samples is not significant. We are given

$S_1^2 = \sum(X_1-X_1)^2/n_1-1 = 84.4/7 = 12.06$

$S_2^2 = \sum(X_2-X_2)^2/n_2-1 = 102.3/9 = 11.40$

$F = S_1^2/S_2^2 = 12.06/11.40 = 1.06$

The calculated value of F is less than the table value. Hence, we accept the hypothesis and conclude that the difference in the variance of two samples is not significant at 5% level.

Sum 3: Two samples are drawn from two normal populations. From the following data test whether the two samples have the same variance at 5% level.

Sample 1	60	65	71	74	76	82	85	87		
Sample 2	61	66	67	85	78	63	85	86	88	91

Solution: Let us take the hypothesis that the two populations have the same variance.

$F = S_1^2/S_2^2$

Sample 1 X_1	**$(X_1-\bar{X}_1)$; X1**	**X_1^2**	**Sample 2 X_2**	**$(X_2-\bar{X}_2)$; X_2**	**X_2^2**
60	-15	225	61	-16	256
65	-10	100	66	-11	121
71	-4	16	67	-10	100

74	-1	1	85	8	64
76	1	1	78	1	1
82	7	49	63	-14	196
85	10	100	85	8	64
87	12	144	86	9	81
			88	11	121
			91	14	196
$\sum X_1=600$	**$\sum X_1=0$**	**$\sum X_1^2=636$**	**$\sum X_2=770$**	**$\sum X_2=0$**	**$\sum X_2^2=1200$**

Average of X_1 = 600/8 = 75

Average of X_2 = 770/10 = 77

$S_1^2 = \sum(X_1-X_1)^2/n_1-1$ = 636 / 8-1 = 90.857

$S_2^2 = \sum(X_2-X_2)^2/n_2-1$ = 1200/10-1 = 133.33

$F = S_1^2/S_2^2$

= 133.33/90.857

= 1.467

For v1=9 and v2=7, $F_{0.05}$ = 3.68. The calculated value of F is less than the table value. The hypothesis holds good and hence we conclude that the two populations have the same variance.

Sum 4: The following data present the yields in Quintals of common ten subdivisions of equal area of two agricultural plots.

Plot 1	6.2	5.7	6.5	6.0	6.3	5.8	5.7	6.0	6.0	5.8
Plot 2	5.6	5.9	5.6	5.7	5.8	5.7	6.0	5.5	5.7	5.5

Test whether two samples taken from two random populations have the same variance. (5% point of F for v1=9 and v2=9 is 3.18)

Solution: Let us take the null hypothesis that the samples come from populations having the same variance

Plot 1 X_1	**$(X_1-\bar{X}_1)$; X1**	**X_1^2**	**Plot 2 X_2**	**$(X_2-\bar{X}_2)$; X_2**	**c X_2^2**
6.2	0.2	0.04	5.6	-0.1	0.01
5.7	-0.3	0.09	5.9	0.2	0.04
6.5	0.5	0.25	5.6	-0.1	0.01
6.0	0	0	5.7	0	0
6.3	0.3	0.09	5.8	0.1	0.01
5.8	-0.2	0.04	5.7	0	0
5.7	-0.3	0.09	6.0	0.3	0.09
6.0	0	0	5.5	-0.2	0.04
6.0	0	0	5.7	0	0
5.8	-0.2	0.04	5.5	-0.2	0.04
$\sum X_1=60$	**$\sum X_1=0$**	**$\sum X_1^2=0.64$**	**$\sum X_2=57$**	**$\sum X_2=0$**	**$\sum X_2^2=0.24$**

$F = S_1^2/S_2^2$

$S_1^2 = \sum(X_1-X_1)^2/n_1-1 = 0.64 / 9 = 0.071$

$S_2^2 = \sum(X_2-X_2)^2/n_2-1 = 0.24 / 9 = 0.027$

$F = S_1^2/S_2^2 = 0.071/0.027 = 2.63$

The value of F for 9 and 6 at 5% level of significance is 3.18. The calculated value is less than the table value. The hypothesis holds true. Hence, the samples come from population having the same variance.

CHI-SQUARE

The chi-square test is an important test amongst the several tests of significance developed by statisticians. Chi-square, symbolically written as χ^2 (Pronounced as Ki-square), is a statistical measure used in the context of sampling analysis for comparing a variance to a theoretical variance. As a non-parametric test, it "can be used to determine if categorical data shows dependency or the two classifications are independent. It can also be used to make comparisons between theoretical populations and actual data when categories are used." Thus, the chi-square test is applicable in large number of problems. The test is, in fact, a technique through the use of which it is possible for all researchers to (i) test the goodness of fit; (ii) test the significance of association between two attributes, and (iii) test the homogeneity or the significance of population variance.

Chi-square is an important non-parametric test and as such no rigid assumptions are necessary in respect of the type of population. We require only the degrees of freedom (implicitly of course the size of the sample) for using this test. As a non-parametric test, chi-square can be used (i) as a test of goodness of fit and (ii) as a test of independence.

As a test of goodness of fit, χ^2 test enables us to see how well does the assumed theoretical distribution (such as Binomial distribution, Poisson distribution or Normal distribution) fit to the observed data. When some theoretical distribution is fitted to the given data, we are always interested in knowing as to how well this distribution fits with the observed data. The chi-square test can give answer to this. If the calculated value of χ^2 is less than the table value at a certain level of significance, the fit is considered to be a good one which means that the divergence between the observed and expected frequencies is attributable to fluctuations of sampling. But if the calculated value of χ^2 is greater than its table value, the fit is not considered to be a good one.

As a test of independence, χ^2 test enables us to explain whether or not two attributes are associated. For instance, we may be interested in knowing whether a new medicine is effective in controlling fever or not, χ^2 test will helps us in deciding this issue. In such a situation, we proceed with the null hypothesis that the two attributes (viz., new medicine and control of fever) are independent which means that new medicine is not effective in controlling fever. On this

basis we first calculate the expected frequencies and then work out the value of χ^2 . If the calculated value of χ^2 is less than the table value at a certain level of significance for given degrees of freedom, we conclude that null hypothesis stands which means that the two attributes are independent or not associated (i.e., the new medicine is not effective in controlling the fever). But if the calculated value of χ^2 is greater than its table value, our inference then would be that null hypothesis does not hold good which means the two attributes are associated and the association is not because of some chance factor but it exists in reality (i.e., the new medicine is effective in controlling the fever and as such may be prescribed). It may, however, be stated here that χ^2 is not a measure of the degree of relationship or the form of relationship between two attributes, but is simply a technique of judging the significance of such association or relationship between two attributes.

In order that we may apply the chi-square test either as a test of goodness of fit or as a test to judge the significance of association between attributes, it is necessary that the observed as well as theoretical or expected frequencies must be grouped in the same way and the theoretical distribution must be adjusted to give the same total frequency as we find in case of observed distribution. χ^2 is then calculated as follows:

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

If two distributions (observed and theoretical) are exactly alike, $\chi 2 = 0$; but generally due to 2 sampling errors, $\chi 2$ is not equal to zero and as such we must know the sampling distribution of $\chi 2$ so that we may find the probability of an observed $\chi 2$ being given by a random sample from the hypothetical universe. Instead of working out the probabilities, we can use ready table which gives probabilities for given values of $\chi 2$. Whether or not a calculated value of $\chi 2$ is significant can be ascertained by looking at the tabulated values of $\chi 2$ for given degrees of freedom at a certain level of significance. If the calculated value of $\chi 2$ is equal to or exceeds the table value, the difference between the observed and expected frequencies is taken as significant,

but if the table value is more than the calculated value of $\chi 2$, then the difference is considered as insignificant i.e., considered to have arisen as a result of chance and as such can be ignored.

As already stated, degrees of freedom play an important part in using the chi-square distribution and the test based on it, one must correctly determine the degrees of freedom. If there are 10 frequency classes and there is one independent constraint, then there are (10 – 1) = 9 degrees of freedom. Thus, if 'n' is the number of groups and one constraint is placed by making the totals of observed and expected frequencies equal, the d.f. would be equal to (n – 1). In the case of a contingency table (i.e., a table with 2 columns and 2 rows or a table with two columns and more than two rows or a table with two rows but more than two columns or a table with more than two rows and more than two columns), the d.f. is worked out as follows: d.f. = (c – 1) (r – 1), where 'c' means the number of columns and 'r' means the number of rows.

CONDITIONS FOR THE APPLICATION OF $\chi 2$ TEST

1. Observations recorded and used are collected on a random basis.
2. All the items in the sample must be independent.
3. No group should contain very few items, say less than 10. In case where the frequencies are less than 10, regrouping is done by combining the frequencies of adjoining groups so that the new frequencies become greater than 10. Some statisticians take this number as 5, but 10 is regarded as better by most of the statisticians.
4. The overall number of items must also be reasonably large. It should normally be at least 50, howsoever small the number of groups may be.
5. The constraints must be linear. Constraints which involve linear equations in the cell frequencies of a contingency table (i.e., equations containing no squares or higher powers of the frequencies) are known are know as linear constraints.

Sum 1: A die is thrown 132 times with following results:

Number turned up	1	2	3	4	5	6
Frequency	16	20	25	14	29	28

Is the die unbiased?

Solution

Let us take the hypothesis that the die is unbiased. If that is so, the probability of obtaining any one of the six numbers is 1/6 and as such the expected frequency of any one number coming upward is 132 ×1/6 = 22. Now we can write the observed frequencies along with expected frequencies and work out the value of $\chi 2$ as follows:

No. Turned UP	Observed Frequency	Expected Frequency	(Oi – Ei)	(Oi – Ei)	(Oi – Ei) /Ei
1	16	22	–6	36	36/22
2	20	22	–2	4	4/22
3	25	22	3	9	9/22
4	14	22	–8	64	64/22
5	29	22	7	49	49/22
6	28	22	6	36	36/22

$\sum (O-E)^2/E = 9$

Hence, the calculated $\chi^2=9$

DF = (n-1) (6-1)=5

The table value of $\chi 2$ for 5 degrees of freedom at 5 per cent level of significance is 11.071. Comparing calculated and table values of $\chi 2$, we find that calculated value is less than the table value and as such could have arisen due to fluctuations of sampling. The result, thus, supports the hypothesis and it can be concluded that the die is unbiased.

Sum 2: Find the value of □2 for the following information:

Class	*A*	*B*	*C*	*D*	*E*

Observed frequency	8	29	44	15	4
Theoretical (or expected) frequency	7	24	38	24	7

Solution : Since some of the frequencies less than 10, we shall first re-group the given data as follows and then will work out the value of $\chi 2$

Class	**Observed Frequency**	**Expected Frequency**	**O-E**	**$(O-E)^2/E$**
A and *B*	(8 + 29) = 37	(7 + 24) = 31	6	36/31
C	44	38	6	36/38
D and *E*	(15 + 4) = 19	(24 + 7) = 31	–12	144/31

$\sum (O-E)^2/E = 6.76$ (Approximate)

Sum 3: Genetic theory states that children having one parent of blood type A and the other of blood type B will always be of one of three types, A, AB, B and that the proportion of three types will on an average be as 1 : 2 : 1. A report states that out of 300 children having one A parent and B parent, 30 per cent were found to be types A, 45 per cent per cent type AB and remainder type B. Test the hypothesis by $\chi 2$ test.

Solution: The observed frequencies of type A, AB and B is given in the question are 90, 135 and 75 respectively.

The expected frequencies of type A, AB and B (as per the genetic theory) should have been 75, 150 and 75 respectively.

We now calculate the value of $\chi 2$ as follows:

Type	**Observed Frequency**	**Expected Frequency**	**O-E**	**$(O-E)^2$**	**$(O-E)^2/E$**
A	90	75	15	225	225/75 = 3
AB	135	150	–15	225	225/150 = 1.5

B	75	75	0	0	0/75 = 0

$\sum (O-E)^2/E = 3 + 1.5 = 4.5$

Table value of $\chi 2$ for 2 d.f. at 5 per cent level of significance is 5.991.

The calculated value of $\chi 2$ is 4.5 which is less than the table value and hence can be ascribed to have taken place because of chance. This supports the theoretical hypothesis of the genetic theory that on an average type A, AB and B stand in the proportion of 1 : 2 : 1.

Sum 4: The table given below shows the data obtained during outbreak of smallpox:

Particulars	Attacked	Not attacked	Total
Vaccinated	31	469	500
Not vaccinated	185	1315	1500
Total	**216**	**1784**	**2000**

Test the effectiveness of vaccination in preventing the attack from smallpox. Test your result with the help of $\chi 2$ at 5 per cent level of significance.

Solution:

Let us take the hypothesis that vaccination is not effective in preventing the attack from smallpox i.e., vaccination and attack are independent.

Group	Observed Frequency	Expected Frequency	O-E	$(O-E)^2$	$(O-E)^2/E$
AB	31	54	–23	529	529/54 = 9.796
Ab	469	446	+23	529	529/44 = 1.186
aB	158	162	+23	529	529/162 = 3.265
ab	1315	1338	–23	529	529/1338 = 0.395

$\sum (O-E)^2/E = 14.462$

Degrees of freedom in this case = $(r - 1)(c - 1) = (2 - 1)(2 - 1) = 1$.

The table value of χ^2 for 1 degree of freedom at 5 per cent level of significance is 3.841. The calculated value of χ^2 is much higher than this table value and hence the result of the experiment does not support the hypothesis. We can, thus, conclude that vaccination is effective in preventing the attack from smallpox.

Sum 5: Two research workers classified some people in income groups on the basis of sampling studies. Their results are as follows:

Investigators	**Income groups**			**Total**
	Poor	**Middle**	**Rich**	
A	160	30	10	200
B	140	120	40	300
Total	**300**	**150**	**50**	**500**

Show that the sampling technique of at least one research worker is defective.

Solution

Let us take the hypothesis that the sampling techniques adopted by research workers are similar (i.e., there is no difference between the techniques adopted by research workers). This being so, the expectation of A investigator classifying the people in

OF	EF	(O-E)	$(O-E)^2/E$
160	120	40	1600/120 = 13.33
30	60	–30	900/60 = 15.00
10	20	–10	100/20 = 5.00
140	180	–40	1600/180 = 8.88
120	90	30	900/90 = 10.00

40	30	10	100/30 = 3.33

$\sum (O-E)^2/E = 55.54$

Degrees of freedom = $(c - 1)(r - 1)$ = = (3 – 1) (2 – 1) = 2.

The table value of $\chi 2$ for two degrees of freedom at 5 per cent level of significance is 5.991. The calculated value of $\chi 2$ is much higher than this table value which means that the calculated value cannot be said to have arisen just because of chance. It is significant. Hence, the hypothesis does not hold good. This means that the sampling techniques adopted by two investigators differ and are not similar. Naturally, then the technique of one must be superior than that of the other.

Sum 6: In an anti a malarial campaign in a certain area, quinine was administered to 812 persons out of a total population of 3248. The number of fever cases is shown below:

Treatment	**Fever**	**No Fever**	**Total**
Quinine	20	792	812
No Quinine	220	2216	2436
Total	**240**	**3008**	**3248**

Discuss the usefulness of Quinine in checking malaria.

Solution Let us take the hypothesis that quinine is not effective in checking malaria.

Observed Frequency	**Expected Frequency**	**$(O-E)^2$**	**$(O-E)^2/E$**
20	60	1600	26.667
220	180	1600	8.889
792	752	1600	2.128
2216	2256	1600	0.709
			$(\sum((O-E)^2/E)=38.393$

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

= 38.393

v= (r-1)(c-1) = (2-1) (2-1)=1

v=1, $\chi^2_{0.05}$ = 3.84

The calculated value of χ^2 is greater than the table value. The hypothesis is rejected. Hence, quinine is useful in checking malaria.

Sum 7: Based on information on 1000 randomly selected fields about the tenancy status of the cultivation of these fields and use of fertilizers, collected in an agro-economic survey, the following classification was noted:

Treatment	Owned	Rented	Total
Using Fertilizer	416	184	600
Not using Fertilizer	64	336	400
Total	**480**	**520**	**1000**

Would you conclude that owner cultivators are more inclined towards the use of fertilizer at 5% level? Carry out chi-square test as per testing procedure.

Solution

Let us take the hypothesis that ownership of fields and the use of fertilizers are independent attributes.

Observed Frequency	Expected Frequency	$(O-E)^2$	$(O-E)^2/E$
416	288	16384	56.889
64	192	16384	85.333
184	312	16384	52.513
336	208	16384	78.769

			(∑((O-E)^2/E)=273.504

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

= 273.504

v= (r-1)(c-1) = (2-1) (2-1)=1

v=1, $\chi^2_{0.05}$ = 3.84

The calculated value of χ^2 is much more than the table value. The hypothesis is rejected. Hence, it can be concluded that owner's cultivators are more inclined towards the use of fertilizers.

Sum 8: In an experiment on immunization of cattle from tuberculosis, the following results were obtained.

Particulars	**Affected**	**Not Affected**
Inoculated	12	26
Not Inoculated	16	6

Calculate χ^2 and discuss the effect of vaccine in controlling suspectability to tuberculosis. (5% value of χ^2 for one degree of freedom =3.84)

Solution:

Let us take the hypothesis that the vaccine is not effective in controlling susceptibility to tuberculosis.

Observed Frequency	**Expected Frequency**	**(O-E)^2**	**(O-E)^2/E**
12.5	17.7	27.04	1.528
15.5	10.3	27.04	2.625

25.5	20.3	27.04	1.332
6.5	11.7	27.04	2.311
			$(\sum((O-E)^2/E)=7.796$

$$\chi^2 = \Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

= 7.796

v= (r-1)(c-1) = (2-1) (2-1)=1

v=1, $\chi^2_{0.05}$ = 3.84

Since the calculated value of χ^2 is greater than the table value, the hypothesis is not true. We, therefore, conclude that vaccine is effective in controlling susceptibility to tuberculosis.

Sum 9: From the data given below about the treatment of 250 patients suffering from a disease, state whether the new treatment is superior to the conventional treatment.

Treatment	**Favourable**	**Unfavourable**	**Total**
New	140	30	170
Conventional	60	20	80
Total	**200**	**50**	**250**

Solution

Observed Frequency	**Expected Frequency**	**$(O-E)^2$**	**$(O-E)^2/E$**
140	136	16	0.118
60	64	16	0.250
30	34	16	0.471
20	16	16	1.000

			$(\sum((O-E)^2/E)=1.839$

$$\chi^2 = \Sigma \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

= 1.839

v= (r-1)(c-1) = (2-1) (2-1)=1

v=1, $\chi^2_{0.05}$ = 3.84

The calculated value of χ^2 is less than the table value. The hypothesis is accepted. Hence, there is no significant difference between the new and conventional treatment.

Sum 10: 1000 studentsat College level are graded according to IQ and their economic conditions. Use Chi-square test to find out whether there is any association between economic conditions and the level of IQ.

Economic Condition	**Intelligent Quotient**			**Total**
	High	**Medium**	**Low**	
Rich	160	300	140	600
Poor	140	100	160	400
Total	**300**	**400**	**300**	**1000**

Solution

Let us take the hypothesis that there is no association between economic conditions and the level of IQ. On the basis of this hypothesis the expected frequencies corresponding to (a) and (b) are:

Observed Frequency	**Expected Frequency**	**$(O-E)^2$**	**$(O-E)^2/E$**
160	180	400	2.222
140	120	400	3.333

300	240	3600	15.000
100	160	3600	22.500
140	180	1600	8.889
160	120	1600	13.333
			(∑((O-E)^2/E)=65.277

$$\chi^2 = \Sigma \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

= 65.277

v= (r-1)(c-1) = (2-1) (3-1)=2

v=2, $\chi^2_{0.05}$ = 5.99

The calculated value of χ^2 is much greater than the table value. The hypothesis is rejected. Hence, there is association between economic condition and the level of IQ.

IMPORTANT NONPARAMETRIC

Tests of hypotheses with 'order statistics' or 'nonparametric statistics' or 'distribution-free' statistics are known as nonparametric or distribution-free tests. The following distribution-free tests are important and generally used

1. Sign Tests

The sign test is one of the easiest parametric tests. Its name comes from the fact that it is based on

the direction of the plus or minus signs of observations in a sample and not on their numerical magnitudes. The sign test may be one of the following two types:

(a) One sample sign test;

(b) Two sample sign test.

(a) One sample sign test

The one sample sign test is a very simple non-parametric test applicable when we sample a continuous symmetrical population in which case the probability of getting a sample value less than mean is 1/2 and the probability of getting a sample value greater than mean is also 1/2. To test the null hypothesis $\mu\ \mu = H0$ against an appropriate alternative on the basis of a random sample of size '*n*', we replace the value of each and every item of the sample with a plus (+) sign if it is greater than $\mu H0$, and with a minus (–) sign if it is less than $\mu H0$. But if the value happens to be equal to $\mu H0$, then we simply discard it. After doing this, we test the null hypothesis that these + and – signs are values of a random variable, having a binomial distribution with $p = 1/2$*. For performing one sample sign test when the sample is small, we can use tables of binomial probabilities, but when sample happens to be large, we use normal approximation to binomial distribution.

(b) Two sample sign test (or the sign test for paired data)

The sign test has important applications in problems where we deal with paired data. In such problems, each pair of values can be replaced with a plus (+) sign if the first value of the first sample (say *X)* is greater than the first value of the second sample (say *Y*) and we take minus (–) sign if the first value of *X* is less than the first value of *Y*. In case the two values are equal, the concerning pair is discarded. (In case the two samples are not of equal size, then some of the values of the larger sample left over after the random pairing will have to be discarded.) The testing technique remains the same as started in case of one sample sign test.

2. Fisher-Irwin Test

Fisher-Irwin test is a distribution-free test used in testing a hypothesis concerning no difference among two sets of data. It is employed to determine whether one can reasonably assume, for example, that two supposedly different treatments are in fact different in terms of the results they produce. Suppose the management of a business unit has designed a new training programme which is now ready and as such it wishes to test its performance against that of the old training programme.

3. McNemer Test

McNemer test is one of the important nonparametric tests often used when the data happen to be nominal and relate to two related samples. As such this test is specially useful with before-after measurement of the same subjects. The experiment is designed for the use of this test in such a way that the subjects initially are divided into equal groups as to their favourable and unfavourable views about, say, any system. After some treatment, the same number of subjects are asked to express their views about the given system whether they favour it or do not favour it

4. Wilcoxon Matched-pairs Test (or Signed Rank Test)

In various research situations in the context of two-related samples (i.e., case of matched paires such as a study where husband and wife are matched or when we compare the output of two similar machines or where some subjects are studied in context of before-after experiment) when we can determine both direction and magnitude of difference between matched values, we can use an important non-parametric test viz., Wilcoxon matched-paires test. While applying this test, we first find the differences (di) between each pair of values and assign rank to the differences from the smallest to the largest without regard to sign. The actual signs of each difference are then put to corresponding ranks and the test statistic T is calculated which happens to be the smaller of the two sums viz., the sum of the negative ranks and the sum of the positive ranks.

5. Rank Sum Tests

Rank sum tests are a whole family of test, but we shall describe only two such tests commonly used viz., the U test and the H test. U test is popularly known as Wilcoxon-Mann-Whitney test, whereas

H test is also known as Kruskal-Wallis test.

a) **Wilcoxon-Mann-Whitney test (or U-test)**

This is a very popular test amongst the rank sum tests. This test is used to determine whether two independent samples have been drawn from the same population. It uses more information than the sign test or the Fisher-Irwin test. This test applies under very general conditions and requires only that the populations sampled are continuous. However, in practice even the

violation of this assumption does not affect the results very much. To perform this test, we first of all rank the data jointly, taking them as belonging to a single sample in either an increasing or decreasing order of magnitude. We usually adopt low to high ranking process which means we assign rank 1 to an item with lowest value, rank 2 to the next higher item and so on. In case there are ties, then we would assign each of the tied observation the mean of the ranks which they jointly occupy. For example, if sixth, seventh and eighth values are identical, we would assign each the rank $(6 + 7 + 8)/3 = 7$. After this we find the sum of the ranks assigned to the values of the first sample (and call it R1) and also the sum of the ranks assigned to the values of the second sample (and call it R2). Then we work out the test statistic i.e., U.

b) **The Kruskal-Wallis test (or H test)**

This test is conducted in a way similar to the U test described above. This test is used to test the null hypothesis that 'k' independent random samples come from identical universes against the alternative hypothesis that the means of these universes are not equal. This test is analogous to the one-way analysis of variance, but unlike the latter it does not require the assumption that the samples come from approximately normal populations or the universes having the same standard deviation.

MULTIVARIATE ANALYSIS

Techniques which simultaneously analyse more than two variables on a sample of observations can be categorized as multivariate techniques. We may as well use the term 'multivariate analysis' which is a a collection of methods for analyzing data in which a number of observations are available for each object.

Growth of Multivariate Techniques

Of late, multivariate techniques have emerged as a powerful tool to analyse data represented in terms of many variables. The main reason being that a series of univariate analysis carried out separately for each variable may, at times, lead to incorrect interpretation of the result. This is so because univariate analysis does not consider the correlation or inter-dependence among the variables.

Characteristics and Applications

Multivariate techniques are largely empirical and deal with the reality; they possess the ability to analyse complex data. Accordingly in most of the applied and behavioural researches, we generally resort to multivariate analysis techniques for realistic results. Besides being a tool for analyzing the data, multivariate techniques also help in various types of decision-making.

A simplified way. In other words, multivariate techniques transform a mass of observations into a smaller number of composite scores in such a way that they may reflect as much information as possible contained in the raw data obtained concerning a research study. Thus, the main contribution of these techniques is in arranging a large amount of complex information involved in the real data into a simplified visible form. Mathematically, multivariate techniques consist in "forming a linear composite vector in a vector subspace, which can be represented in terms of projection of a vector onto certain specified subspaces. For better appreciation and understanding of multivariate techniques, one must be familiar with fundamental concepts of linear algebra, vector spaces, orthogonal and oblique projections and univariate analysis.

Even then before applying multivariate techniques for meaningful results, one must consider the nature and structure of the data and the real aim of the analysis. We should also not forget that multivariate techniques do involve several complex mathematical computations and as such can be utilized largely with the availability of computer facility.

FACTOR ANALYSIS

There are several methods of factor analysis, but they do not necessarily give same results. As such factor analysis is not a single unique method but a set of techniques. Important methods of factor analysis are:

(i) the centroid method;

(ii) the principal components method;

(ii) the maximum likelihood method.

(i) Factor: A factor is an underlying dimension that account for several observed variables. There can be one or more factors, depending upon the nature of the study and the number of variables involved in it.

(ii) Factor-loadings: Factor-loadings are those values which explain how closely the variables are related to each one of the factors discovered. They are also known as factor-variable correlations. In fact, factor-loadings work as key to understanding what the factors mean. It is the absolute size (rather than the signs, plus or minus) of the loadings that is important in the interpretation of a factor.

(iii) Communality (h2): Communality, symbolized as h2, shows how much of each variable is accounted for by the underlying factor taken together. A high value of communality means that not much of the variable is left over after whatever the factors represent is taken into consideration.

CLUSTER ANALYSIS

Cluster analysis consists of methods of classifying variables into clusters. Technically, a cluster consists of variables that correlate highly with one another and have comparatively low correlations with variables in other clusters. The basic objective of cluster analysis is to determine how many mutually and exhaustive groups or clusters, based on the similarities of profiles among entities, really exist in the population and then to state the composition of such groups. Various groups to be determined in cluster analysis are not predefined as happens to be the case in discriminant analysis.

Steps: In general, cluster analysis contains the following steps to be performed:

(i) First of all, if some variables have a negative sum of correlations in the correlation matrix,one must reflect variables so as to obtain a maximum sum of positive correlations for the matrix as a whole.

(ii) The second step consists in finding out the highest correlation in the correlation matrix and the two variables involved (i.e., having the highest correlation in the matrix) form the nucleus of the first cluster.

(iii) Then one looks for those variables that correlate highly with the said two variables and includes them in the cluster. This is how the first cluster is formed.

(iv) To obtain the nucleus of the second cluster, we find two variables that correlate highly but have low correlations with members of the first cluster. Variables that correlate highly with the

said two variables are then found. Such variables along the said two variables thus constitute the second cluster.

(v) One proceeds on similar lines to search for a third cluster and so on.

DISCRIMINANT ANALYSIS

Regression based statistical technique used in determining which particular classification or group (such as 'ill' or 'healthy') an item of data or an object (such as a patient) belongs to on the basis of its characteristics or essential features. It differs from group building techniques such as cluster analysis in that the classifications or groups to choose from must be known in advance.

Discriminant function analysis is a statistical analysis to predict a categorical dependent variable (called a grouping variable) by one or more continuous or binary independent variables (called predictor variables). The original dichotomous discriminant analysis was developed by Sir Ronald Fisher in 1936. It is different from an ANOVA or MANOVA, which is used to predict one (ANOVA) or multiple (MANOVA) continuous dependent variables by one or more independent categorical variables. Discriminant function analysis is useful in determining whether a set of variables is effective in predicting category membership.

Discriminant analysis is used when groups are known a priori (unlike in cluster analysis). Each case must have a score on one or more quantitative predictor measures, and a score on a group measure.[3] In simple terms, discriminant function analysis is classification - the act of distributing things into groups, classes or categories of the same type.

Moreover, it is a useful follow-up procedure to a MANOVA instead of doing a series of one-way ANOVAs, for ascertaining how the groups differ on the composite of dependent variables. In this case, a significant F test allows classification based on a linear combination of predictor variables. Terminology can get confusing here, as in MANOVA, the dependent variables are the predictor variables, and the independent variables are the grouping variables**.**

REGRESSION

After having established the fact that two variables are closely related we may be interested in estimating (predicting) the value of one variable given the value of another. For

example, if we know that advertising and sales are correlated we find out expected amount of sales for a given advertising expenditure or the required amount of expenditure for attaining a given amount of sales. Similarly, if we know that the yield of rice and rainfall are closely related we may find out the amount of rain required to achieve a certain production figure. Regression analysis reveals average relationship between two variables and this makes possible estimation or prediction.

The dictionary meaning of the term 'regression' is the act of returning or going back. The term 'regression' was first used by Sir Francis Galton (1822-1911) in 1877 while studying the relationship between the height of fathers and sons. This term was introduced by him in the paper 'Regression towards fathers and sons revealed a very interesting relationship, i.e., tall fathers tend to have tall sons and short fathers short sons, but the average height of the sons of a group of tall fathers is less than that of the fathers and the average fathers. The line describing the tendency to regress or going back was called by Galton s 'Regression Line'. The term is still used to describe that line drawn for a group of points to represent the trend present, but it no longer necessarily carries the original implication of "stepping back" that Galton intended. These days there is a growing tendency to the modern writers to use the term estimating line instead of regression line because the expression estimating line is more clarificatory in character.

DEFINITIONS

Regression is the measure of the average relationship between two or more variables in terms of the original units of the data.

The term 'regression analysis' refers to the methods by which estimates are made of the values of a variable from a knowledge of the values of one or more other variables and to the measurement of the errors involved in this estimation process – Morris Hamburg.

One of the most frequently used techniques in economics and business research to find a relation between two or more variables that are related causally, is regression analysis – Taro Yamane

Regression analysis attempts to establish the 'nature of the relationship' between variables – that is, to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting" – Ya Lun Chou

It is clear from the above definitions that regression analysis is a statistical device with the help of which we are in a position to estimate (or predict) the unknown values of one variable from known values of another variable. The variable which is used to predict the variable of interest is called the independent variable or exploratory variable and the variable we are trying to predict is called the dependent variable or "explained variable. The independent variable is denoted by X and the dependent variable by Y. The analysis used is called the simple regression analysis – simple because there is only one predictor or independent variable, and linear because of the assumed linear relationship between the dependent and the independent variables. The term "linear" means that an equation of a straight line of the form Y=a+bx, where a and b are constants, is used to describe the average relationship that exists between the two variables.

It should be noted that the term 'dependent' and 'independent' refer to the mathematical or functional meaning of dependence – they do not imply that there is necessarily any cause and effect relationship between the variables. What is meant is simply that estimates of values of the dependent variable Y may be obtained for given values of the independent variable X from a mathematical function involving X and Y. In that sense, the values of Y are dependent upon the values of X. The X variable may or may not be causing change in the Y variable. For example, while estimating sales of a product from figures on advertising expenditures, sale is generally taken as the dependent variable. However, there may or may not be causal connection between these two factors in the sense that changes in advertising expenditures cause change in sales. In fact, in certain cases, the cause-effect relation may be just opposite of what appears to be the obvious one.

Uses of Regression Analysis

Regression analysis is a branch of statistical theory that is widely used in almost all the scientific disciplines. In economics it is the basic technique for measuring or estimating the relationship among economic variables that constitute the essence of economic theory and

economic life. For example, if we know that two variables, Price (X) and Demand (Y), are closely related we can find out the most probable value of X for a given value of Y or the most probable value of Y for a given value of X. Similarly, if we closely related, we can find out the expected price for a certain price for a certain amount of tax levy. Thus, we find that the study of regression is of considerable help to the economists and businessmen. The uses of regression are not confined to economics and business field only. Its applications are extended to almost all the natural, physical and social sciences. The regression analysis attempts to accomplish the following:

1. Regression analysis provides estimates of values of the dependent variable from values of the independent variable. The device used to accomplish this estimation procedure is the regression line. The regression line describes the average relationship existing between X and Y variables. i.e., it displays mean values of X for given values of Y. The equation of this line, known as the regression equation, provides estimates of the dependent variable when values of the independent variable are inserted into the equation.
2. A second goal of regression analysis is to obtain a measure of the error involved in using the regression line as a basis for estimation. For this purpose the standard error of estimate is calculated. This is a measure of the scatter or spread of the observed values of Y around the corresponding values estimated from the regression line. If the line fits the data closely, that is, if there is little scatter of the observations around the regression line, good estimates can be made of the Y variable. On the other hand, if there is a great deal of scatter or the observations around the fitted regression line, the line will not produce accurate estimates of the dependent variable
3. With the help of regression coefficients we can calculate the correlation coefficient. The square of correlation coefficient (r) called coefficient of determination, measures the degree of association of correlation that exists between the two variables. It assesses the proportion of variance in the dependent variable that has been accounted for by the regression equation. In general, the greater the value of r^2 the better is the fit and the more useful the regression equations as a predictive device.

DIFFERENCE BETWEEN CORRELATION AND REGRESSION

Correlation	Regression
Correlation is the relationship between variables. It is expressed numerically	Regression means going back. The average relation between the variables is given as an equation
Between two variables, none is identified as independent variable	One of the variables is independent variable and the other is dependent variable in any particular context
Correlation does not mean causation. One variable need not be the cause and the other, effect	Independent variable may be the 'the cause' and dependent variable, 'the effect'
There is spurious or nonsense correlation	There is no such possibility. Regression is considered only when the variables are related
Correlation coefficient is independent of change of origin and scale	Regression coefficients are independent of change or origin but are affected by change of scale
Correlation coefficient is a number -1 and +1	The two regression coefficients have the same sign, + or -. One of them can be greater than 1 numerically. But they can not be greater than 1 numerically simultaneously.
Correlation coefficient is not in any unit of measurement	Each regression coefficient is in the unit of measurement of the dependent variable
Correlation coefficient indicates the direction of co-variation and the closeness of the linear relation between two variables	Regression equations give the value of the dependent variable corresponding to any value of the independent variable

Correlation	Regression
The significance of the sample correlation coefficient can be tested. The limits between which the population correlation coefficient is expected to lie can be found	Target can be reached. The value of the independent variable can be chosen so as to get the target value of the dependent variable. For example, a specific amount can be spent on advertisement to get the targeted revenue.

Sum 1: From the following data, obtain the two regression equations:

X	6	2	10	4	8
Y	9	11	5	8	7

Solution

X	Y	XY	X^2	Y^2
6	9	54	36	81
2	11	22	4	121
10	5	50	100	25
4	8	32	16	64
8	7	56	64	49
$\sum X=30$	**$\sum Y=40$**	**$\sum XY=214$**	**$\sum X^2=220$**	**$\sum Y^2=340$**

Let the regression equation of Y on X be Y=A+BX

The normal equations are $\sum Y = NA + B\sum X$

$\sum XY = A\sum X + B\sum X^2$

By substituting the value from the table,

$5A + 30B = 40$ Say (1)

$30A + 220B = 214$ Say (2)

(1) X 6 $30A + 180B = 240$ Say (3)

(2) – (3) $40B = -26$

$B = -26 / 40$

= -0.6500

From (1), 5A-30X0.6500 = 40

A = 40 + 19.5 / 5

= 11.90

The regression equation of Y on X is

Y = 11.90 – 0.6500X

Let the regression equation of X on Y be X = A + BY

The normal equations are $\sum X = NA + B\sum Y$

$\sum XY = A\sum Y + B\sum Y^2$

By substituting the values from the table,

5A + 40B = 30 Say (4)

40A + 340 B = 214 Say (5)

(4) X 8 40A = 320 B = 240 Say (6)

(5) – (6) 20B = - 26

B = -26/20

= -1.300

From (4), 5A+40 X (-1.30) = 30

A = 30 + 52 /5

= 16.40

The regression equation of X on Y is X = 16.40 – 1.300

Sum 2: You are given the following data:

	X	**Y**
Arithmetic Mean	36	85
Standard Deviation	11	8
Correlation coefficient between X and Y	0.66	

(a) Find the two regression equations

(b) Estimate the value of X when Y=75

Solution

bxy = $r\sigma x/\sigma y$ = 0.66 x 11 / 8 = 0.9075

byx = $r\sigma y/\sigma x$ = 0.66 X 8 / 11 = 0.4800

a) Regression equation of Y on X

$Y - \bar{Y} = byx\ (X-\bar{X})$

Y – 85 = 0.4800 (X-36)

= 0.4800 X – 17.28

Y = 67.72 + 0.4800 X

Regression equation of X on Y

$X - \bar{X} = bxy\ (Y-\bar{Y})$

X – 36 = 0.9075 (Y-85)

= 0.9075 – 77.14

X = 0.9075 Y – 41.14

b) When Y=75, X = 0.9075 X 75 – 41.14

= 26.92

Sum 3: From the following information on values of two variables X and Y find the two regression lines and the correlation coefficient.

N=10; $\sum X=20$; $\sum Y=40$; $\sum X^2=240$; $\sum Y^2=410$; $\sum XY=200$

Solution

$\bar{X} = \sum X/N = 20/10 = 2.00$

$\bar{Y} = \sum Y/N = 40/10 = 4.00$

$$bxy = \frac{N\sum XY-(\sum X)(\sum Y)}{N\sum Y^2-(\sum Y)^2} \text{ as } \sum X\neq 0 \text{ and } \sum Y\neq 0$$

10 X 200 – 20 X 40

$$= \frac{}{10 \times 410 - (40)^2}$$

$$= \frac{2000 - 800}{4100\text{-}1600}$$

$$= \frac{1200}{2500} = 0.4800$$

$$byx = \frac{N\sum XY-(\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2} \text{ as } \sum X \neq 0 \text{ and } \sum Y \neq 0$$

$$= \frac{1200}{10X240-(20)^2}$$

$$= \frac{1200}{2000} = 0.6000$$

Regression equation of Y on X

$Y - \bar{Y} = byx (X-\bar{X})$

$Y-4 = 0.6000 (X-2)$

$= 0.6000 X - 1.20$

$Y = 2.80 + 0.6000X$

Regression equation of X on Y

X – X = bxy (Y-Y)

X-2 = 0.4800 (Y-4)

= 0.4800 Y – 1.92

X = 0.08 + 0.4800Y

Sum 4: Calculate the two regression equations from the following data:

X	10	12	13	12	16	15
Y	40	38	43	45	37	43

Also estimate Y when X=20.

Solution

X	**Y**	**XY**	**X^2**	**Y^2**
10	40	400	100	1600
12	38	458	144	1444
13	43	559	169	1849
12	45	540	144	2025
16	37	592	256	1369
15	43	645	225	1849
$\sum X=78$	**$\sum Y=246$**	**$\sum XY=3192$**	**$\sum X^2=1038$**	**$\sum Y^2=10136$**

$\bar{X} = \sum X/N = 78/6 = 13.00$

$\bar{Y} = \sum Y/N = 246/6 = 41.00$

$$bxy = \frac{N\sum XY-(\sum X)(\sum Y)}{\text{----------}}$$

$$= \frac{N\sum Y^2 - (\sum Y)^2}{6 \text{ X } 3192 - 78 \text{ X } 246}$$

$$= \frac{6 \text{ X } 10316 - (246)^2}{19152 - 19188}$$

$$= \frac{60816 - 60516}{-36}$$

$$= \frac{-36}{300} = 0.1200$$

$$byx = \frac{N\sum XY - (\sum X)(\sum Y)}{N\sum X^2 - (\sum X)^2}$$

$$= \frac{-36}{6X1038 - (78)^2}$$

$$= \frac{-36}{6228 - 6084} = 0.6000$$

$$= \frac{-36}{144} = 0.2500$$

Regression equation of Y on X

$Y - \bar{Y} = byx\ (X - \bar{X})$

Y-41 = -0.2500 (X-13)

= -0.2500 X + 3.25

Y = 44.25 + 0.25X

When X = 20, Y=44.25-0.25 X 20 = 39.25

Regression equation of X on Y

$X - \bar{X} = bxy\ (Y-\bar{Y})$

X-13 = -0.1200 (Y-41)

= -0.1200 Y + 4.92

X = 17.92 - 0.12Y

Sum 5: From the data given below, find two regression equations

Marks in Mathematics	25	28	35	32	31	36	29	38	34	32
Marks in Statistics	43	46	49	41	36	32	31	30	33	39

Solution

X	**Y**	**$X=X-\bar{X}$** **$\bar{X}=32$**	**$Y=Y-\bar{Y}$** **$\bar{Y}=38$**	**XY**	**X^2**	**Y^2**
25	43	-7	5	-35	49	25
28	46	-4	8	-32	16	64
35	49	3	11	33	9	121
32	41	0	3	0	0	9
31	36	-1	-2	2	1	4
36	32	4	-6	-24	16	36

29	31	-3	-7	21	9	49
38	30	6	-8	-48	36	64
34	33	2	-5	-10	4	25
32	39	0	1	0	0	1
$\sum X=320$	**$\sum Y=380$**	**$\sum X=0$**	**$\sum Y=0$**	**$\sum XY=-93$**	**$\sum X^2=140$**	**$\sum Y^2=398$**

Solution

$\overline{X} = \sum X/N = 320/10 = 32.00$

$\overline{Y} = \sum Y/N = 380/10 = 38.00$

$bxy = \sum xy/\sum y^2 = -93/398 = -0.2337$

$byx = \sum xy/\sum x^2 = -93/140 = -0.6643$

a) Regression equation of Y on X

$Y - \overline{Y} = byx\ (X-\overline{X})$

$Y-38 = -0.6643\ X\ (X-32)$

$= -0.6643\ X + 21.26$

$= 59.26 - 0.6643X$

Regression equation of X on Y

$X - \overline{X} = bxy\ (Y-\overline{Y})$

$X-32 = -0.2337\ (Y-38) = -0.2337Y + 8.88 = 40.88 - 0.2337Y$

KAHE

UNIT V

Interpretation - Meaning - Techniques of Interpretation - Report writing - Significance - Report Writing - Steps in Report Writing - Layout of report - Types of Reports - Oral Presentation - Executive Summary - Mechanics of Writing Research Report - Precautions for Writing Report - Norms for using Tables - Charts - and Diagrams - Appendix - Norms for using Index and Bibliography.

INTERPRETATION

After collecting and analyzing the data, the researcher has to accomplish the task of drawing inferences followed by report writing. This has to be done very carefully, otherwise misleading conclusions may be drawn and the whole purpose of doing research may get vitiated. It is only through interpretation that the researcher can expose relations and processes that underlie his findings. In case of hypotheses testing studies, if hypotheses are tested and upheld several times, the researcher may arrive at generalizations. But in case the researcher had no hypothesis to start with, he would try to explain his findings on the basis of some theory. This may at times result in new questions, leading to further researches. All this analytical information and consequential inference(s) may well be communicated, preferably through research report, to the consumers of research results who may be either an individual or a group of individuals or some public/private organisation.

MEANING OF INTERPRETATION

Interpretation refers to the task of drawing inferences from the collected facts after an analytical and/or experimental study. In fact, it is a search for broader meaning of research findings. The task of interpretation has two major aspects viz., (i) the effort to establish continuity in research through linking the results of a given study with those of another and (ii) the establishment of some explanatory concepts. "In one sense, interpretation is concerned with relationships within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to include the results of other research, theory and hypotheses.". Thus, interpretation is the device through which the factors that seem to explain what has been observed by researcher in the course of the study can be better understood and it also provides a theoretical conception which can serve as a guide for further researches.

Why Interpretation?

Interpretation is essential for the simple reason that the usefulness and utility of research findings lie in proper interpretation. It is being considered a basic component of research process because of the following reasons:

1. It is through interpretation that the researcher can well understand the abstract principle that works beneath his findings. Through this he can link up his findings with those of other studies, having the same abstract principle, and thereby can predict about the concrete world of events. Fresh inquiries can test these predictions later on. This way the continuity in research can be maintained.
2. Interpretation leads to the establishment of explanatory concepts that can serve as a guide for future research studies; it opens new avenues of intellectual adventure and stimulates the quest for more knowledge.

3. Researcher can better appreciate only through interpretation why his findings are what they are and can make others to understand the real significance of his research findings.
4. The interpretation of the findings of exploratory research study often results into hypotheses for experimental research and as such interpretation is involved in the transition from exploratory to experimental research. Since an exploratory study does not have a hypothesis to start with, the findings of such a study have to be interpreted on a post-factum basis in which case the interpretation is technically described as 'post factum' interpretation.

TECHNIQUES OF INTERPRETATION

- InSSSterpretation is not an easy job and it requires a great skill on the part of the investigator. The investigator gets the required expertise to apply the techniques. The techniques of interpretation are given below:

1) Relationship between Variables

The basic object of every analytical research is to find out the relationship between any two variables. There may be three types of relationship

- Symmetrical Relationship
- Reciprocal Relationship
- Asymmetrical Relationship

The interpretation of data can be made with the help of these relationships

2) Percentages

Percentages are used in making comparison between two or more series of data. They are also used to describe the relationships.

3) Averages

There are three forms of averages such as arithmetic mean, median, mode. Though there are other measures of central tendency, the above three measures are commonly used. Instead of using long statistical tables, the use of average makes the interpretation very simple.

4) Dispersion

Dispersion refers to the amount or the magnitude of the spread. Measures of dispersion include range, inter quartile range, average deviation and standard deviation. These measures help to interpret the data more scientifically

PRECAUTIONS IN INTERPRETATION

1. At the outset, researcher must invariably satisfy himself that (a) the data are appropriate, trustworthy and adequate for drawing inferences; (b) the data reflect good homogeneity; and that (c) proper analysis has been done through statistical methods
2. The researcher must remain cautious about the errors that can possibly arise in the process of interpreting results. Errors can arise due to false generalization and/or due to wrong interpretation of statistical measures, such as the application of findings beyond the range of observations, identification of correlation with causation and the like. Another major pitfall is the tendency to affirm that definite relationships exist on the basis of confirmation of particular hypotheses. In fact, the positive test results accepting the hypothesis must be interpreted as "being in accord" with the hypothesis, rather than as "confirming the validity of the hypothesis". The researcher must remain vigilant about all such things so that false generalization may not take place. He should be well equipped with and must know the correct use of statistical measures for drawing inferences concerning his study.
3. He must always keep in view that the task of interpretation is very much intertwined with analysis and cannot be distinctly separated. As such he must take the task of interpretation as

a special aspect of analysis and accordingly must take all those precautions that one usually observes while going through the process of analysis viz., precautions concerning the reliability of data, computational checks, validation and comparison of results.

4. He must never lose sight of the fact that his task is not only to make sensitive observations of relevant occurrences, but also to identify and disengage the factors that are initially hidden to the eye. This will enable him to do his job of interpretation on proper lines. Broad generalisation should be avoided as most research is not amenable to it because the coverage may be restricted to a particular time, a particular area and particular conditions. Such restrictions, if any, must invariably be specified and the results must be framed within their limits.
5. The researcher must remember that "ideally in the course of a research study, there should be constant interaction between initial hypothesis, empirical observation and theoretical conceptions. It is exactly in this area of interaction between theoretical orientation and empirical observation that opportunities for originality and creativity lie." He must pay special attention to this aspect while engaged in the task of interpretation.

REPORT WRITING

Research report is considered a major component of the research study for the research task remains incomplete till the report has been presented and/or written. As a matter of fact even the most brilliant hypothesis, highly well designed and conducted research study, and the most striking generalizations and findings are of little value unless they are effectively communicated to others. The purpose of research is not well served unless the findings are made known to others. Research results must invariably enter the general store of knowledge. All this explains the significance of writing research report. There are people who do not consider writing of

report as an integral part of the research process. But the general opinion is in favour of treating the presentation of research results or the writing of report as part and parcel of the research project. Writing of report is the last step in a research study and requires a set of skills somewhat different from those called for in respect of the earlier stages of research. This task should be accomplished by the researcher with utmost care; he may seek the assistance and guidance of experts for the purpose.

STEPS IN WRITING REPORTS

Research reports are the product of slow, painstaking (careful / meticulous), accurate inductive (logical / reasonable) work

1) Logical Analysis of the Subject Matter

It is the first step which is primarily concerned with the development of a subject. There are two ways in which to develop a subject – (a) logically and (b) Chronologically (arranging data as per time of occurrence). The logical development is made on the basis of mental connections and associations between the one thing and another by means of analysis. Logical treatment often consists in developing the material from the simple possible to the most complex structures. Chronological development is based on a connection or sequence in time or occurrence. The directions for doing or making something usually follow the chronological order.

2) Preparation of the Final Outline

It is the next step in writing the research report. Outlines (hints) are the framework upon which long written works are constructed. They are in aid (help) to the logical organization of the material and a reminder of the points to be stressed in the report.

3) Preparation of the Rough Draft

This follows the logical analysis of the subject and the preparation of the final outline. Such a step is of utmost importance for the researcher now sits to write down what he has done in the context of his research study. He will write down the procedure adopted by him in collecting the material for his study along with various limitations faced by him, the technique of analysis for adopted by him, the broad findings and generalizations and the various suggestions he wants to offer regarding the problem concerned

4) Rewriting and Polishing of the Rough Draft

This step happens to be most difficult part of all formal writing. Usually this step requires more time than the writing of the rough report. The careful revision makes the difference between a mediocre (ordinary) and a good piece of writing. While rewriting and polishing, one should check the report for weaknesses in logical development or presentation. The researcher should also 'see whether or not the material, as it is presented, has unity and cohesion (organization); does the report stand upright and firm and exhibit a definite pattern, like a marble arch? Or does it resemble an old wall of moldering cement and loose bricks. In addition, the researcher should give due attention to the fact that in his rough draft he has been consistent (reliably / steady) or not. He should check the mechanics of writing – grammar, spelling and usage

5) Preparation of the Final Bibliography

Next in order comes the task of the preparation of the final bibliography. The bibliography, which is generally appended (add on) to the research report, is a list of books in some way pertinent to the research which has been done. It should contain all those works which the researcher has consulted. The bibliography should be arranged alphabetically and may be

divided into two parts; the first part may contain the names of books and pamphlets, and the second part may contain the name of magazines and newspaper articles. Generally, this pattern of bibliography is considered convenient and satisfactory from the point of view of reader, thought it is not the only way of presenting bibliography

6) Writing the Final Draft

This consists the last step. The final draft should be written in a concise (brief) and objective style and in simple language, avoiding vague expressions such as "it seems". While writing the final draft, the researcher must avoid abstract (theoretical) terminology and technical jargon. Illustrations and examples based on common experiences must be incorporated in the final draft as they happen to most effective in communicating the research findings to others. A research report should not be dull (boring), but must enthuse (motivate) people and maintain interest and must show originality. It must be remembered that every report should be an attempt to solve some intellectual problem and must contribute to the solution of a problem and must add to the knowledge of both the researcher and the reader

LAYOUT OF THE RESEARCH REPORT

The layout of the report means as to what the research report shouldcontain. A comprehensive layout of the research report should comprise

(A) Preliminary pages

(B)The main text; and

(C) The end matter.

(A) Preliminary Pages

In its preliminary pages the report should carry a title and date, followed by acknowledgements inthe form of 'Preface' or 'Foreword'. Then there should be a table of

contents followed by list oftables and illustrations so that the decision-maker or anybody interested in reading the report caneasily locate the required information in the report.

(B) Main Text

The main text provides the complete outline of the research report along with all details. Title of theresearch study is repeated at the top of the first page of the main text and then follows the otherdetails on pages numbered consecutively, beginning with the second page. Each main section of thereport should begin on a new page. The main text of the report should have the following sections

(i) Introduction

(ii) Statement of findings and recommendations

(iii) The results

(iv) The implications drawn from the results

(v) The summary.

i) Introduction

The purpose of introduction is to introduce the research project to the readers. Itshould contain a clear statement of the objectives of research i.e., enough background should begiven to make clear to the reader why the problem was considered worth investigating. A briefsummary of other relevant research may also be stated so that the present study can be seen in thatcontext. The hypotheses of study, if any, and the definitions of the major concepts employed in thestudy should be explicitly stated in the introduction of the report.The methodology adopted in conducting the study must be fully explained. The scientific readerwould like to know in detail about such thing: How was the study carried out? What was its basicdesign? If the study was an

experimental one, then what were the experimental manipulations? If thedata were collected by means of questionnaires or interviews, then exactly what questions were

asked (The questionnaire or interview schedule is usually given in an appendix)? If measurementswere based on observation, then what instructions were given to the observers? Regarding thesample used in the study the reader should be told: Who were the subjects? How many were there?How were they selected? All these questions are crucial for estimating the probable limits ofgeneralizability of the findings. The statistical analysis adopted must also be clearly stated. In additionto all this, the scope of the study should be stated and the boundary lines be demarcated. The variouslimitations, under which the research project was completed, must also be narrated.

(ii) Statement of findings and recommendations

After introduction, the research report mustcontain a statement of findings and recommendations in non-technical language so that it can beeasily understood by all concerned. If the findings happen to be extensive, at this point they should beput in the summarised form.

(iii) Results

A detailed presentation of the findings of the study, with supporting data in the form of tables and charts together with a validation of results, is the next step in writing the main text of thereport. This generally comprises the main body of the report, extending over several chapters. Theresult section of the report should contain statistical summaries and reductions of the data rather thanthe raw data. All the results should be presented in logical sequence and splitted into readily identifiablesections. All relevant results must find a place in the report. But how one is to decide about what isrelevant is the basic question. Quite often guidance comes primarily from the research problem andfrom the hypotheses, if any, with which the study was concerned. But

ultimately the researcher mustrely on his own judgement in deciding the outline of his report. "Nevertheless, it is still necessary thathe states clearly the problem with which he was concerned, the procedure by which he worked onthe problem, the conclusions at which he arrived, and the bases for his conclusions.

(iv) Implications of the results

Toward the end of the main text, the researcher should again putdown the results of his research clearly and precisely. He should, state the implications that flowfrom the results of the study, for the general reader is interested in the implications for understandingthe human behaviour.

v) Summary

It has become customary to conclude the research report with a very brief summary,resting in brief the research problem, the methodology, the major findings and the major conclusionsdrawn from the research results.

(C) End Matter

At the end of the report, appendices should be enlisted in respect of all technical data such asquestionnaires, sample information, mathematical derivations and the like ones. Bibliography of sourcesconsulted should also be given. Index (an alphabetical listing of names, places and topics along withthe numbers of the pages in a book or report on which they are mentioned or discussed) shouldinvariably be given at the end of the report. The value of index lies in the fact that it works as a guideto the reader for the contents in the report.

MECHANICS OF WRITING A RESEARCH REPORT

There are very definite and set rules which should be followed in the actual preparation of theresearch report or paper. Once the techniques are finally decided, they should be scrupulously

adheredto, and no deviation permitted. The criteria of format should be decided as soon as the materials forthe research paper have been assembled. The following points deserve mention so far as the mechanicsof writing a report are concerned:

1. Size and physical design: The manuscript should be written on unruled paper 8 1 $2^2 \times 11^2$ insize. If it is to be written by hand, then black or blue-black ink should be used. A margin of at leastone and one-half inches should be allowed at the left hand and of at least half an inch at the right handof the paper. There should also be one-inch margins, top and bottom. The paper should be neat andlegible. If the manuscript is to be typed, then all typing should be double-spaced on one side of thepage only except for the insertion of the long quotations.

2. Procedure: Various steps in writing the report should be strictly

3. Layout: Keeping in view the objective and nature of the problem, the layout of the report shouldbe thought of and decided and accordingly adopted (The layout of the research report and varioustypes of reports have been described in this chapter earlier which should be taken as a guide forreport-writing in case of a particular problem).

4. Treatment of quotations: Quotations should be placed in quotation marks and double spaced,forming an immediate part of the text. But if a quotation is of a considerable length (more than fouror five type written lines) then it should be single-spaced and indented at least half an inch to the rightof the normal text margin.

5. The footnotes: Regarding footnotes one should keep in view the followings:

(a) The footnotes serve two purposes viz., the identification of materials used in quotations in the report and the notice of materials not immediately necessary to the body of the researchtext but still of supplemental value. In other words, footnotes are meant for cross references,citation of authorities and sources, acknowledgement and elucidation or

explanation of apoint of view. It should always be kept in view that footnote is not an end nor a means ofthe display of scholarship. The modern tendency is to make the minimum use of footnotesfor scholarship does not need to be displayed.

(b) Footnotes are placed at the bottom of the page on which the reference or quotation which they identify or supplement ends. Footnotes are customarily separated from the textualmaterial by a space of half an inch and a line about one and a half inches long.

(c) Footnotes should be numbered consecutively, usually beginning with 1 in each chapterseparately. The number should be put slightly above the line, say at the end of a quotation.At the foot of the page, again, the footnote number should be indented and typed a littleabove the line. Thus, consecutive numbers must be used to correlate the reference in thetext with its corresponding note at the bottom of the page, except in case of statisticaltables and other numerical material, where symbols such as the asterisk (*) or the like onemay be used to prevent confusion.

(d) Footnotes are always typed in single space though they are divided from one another by double space.

6. Documentation style: Regarding documentation, the first footnote reference to any given workshould be complete in its documentation, giving all the essential facts about the edition used. Suchdocumentary footnotes follow a general sequence. The common order may be described as under:

(i) Regarding the single-volume reference

1. Author’s name in normal order (and not beginning with the last name as in a bibliography) followed by a comma;

2. Title of work, underlined to indicate italics;

3. Place and date of publication;

4. Pagination references (The page number).

(iii) Regarding works arranged alphabetically

(iv) Regarding periodicals reference

1. Name of the author in normal order;

2. Title of article, in quotation marks;

3. Name of periodical, underlined to indicate italics;

4. Volume number;

5. Date of issuance;

6. Pagination.

(v) Regarding anthologies and collections reference

Quotations from anthologies or collections of literary works must be acknowledged notonly by author, but also by the name of the collector.

(vi) Regarding second-hand quotations reference

In such cases the documentation should be handled as follows:

1. Original author and title;

2. "quoted or cited in,";

3. Second author and work.

7. Punctuation and abbreviations in footnotes

The first item after the number in the footnote is the author's name, given in the normal signature order. This is followed by a comma. After thecomma, the title of the book is given: the article (such as "A", "An", "The" etc.) is omitted and onlythe first word and proper nouns and adjectives are capitalized. The title is followed by a comma.Information concerning the

edition is given next. This entry is followed by a comma. The place ofpublication is then stated; it may be mentioned in an abbreviated form, if the place happens to be afamous one such as Lond. for London, N.Y. for New York, N.D. for New Delhi and so on. Thisentry is followed by a comma. Then the name of the publisher is mentioned and this entry is closedby a comma. It is followed by the date of publication if the date is given on the title page. If the dateappears in the copyright notice on the reverse side of the title page or elsewhere in the volume, thecomma should be omitted and the date enclosed in square brackets [c 1978], [1978]. The entry isfollowed by a comma. Then follow the volume and page references and are separated by a commaif both are given. A period closes the complete documentary reference. But one should rememberthat the documentation regarding acknowledgements from magazine articles and periodical literaturefollow a different form as stated earlier while explaining the entries in the bibliography.

8. Use of statistics, charts and graphs

A judicious use of statistics in research reports is oftenconsidered a virtue for it contributes a great deal towards the clarification and simplification of thematerial and research results. One may well remember that a good picture is often worth more thana thousand words. Statistics are usually presented in the form of tables, charts, bars and line-graphsand pictograms. Such presentation should be self explanatory and complete in itself. It should besuitable and appropriate looking to the problem at hand. Finally, statistical presentation should be neatand attractive.

9. The final draft

Revising and rewriting the rough draft of the report should be done with greatcare before writing the final draft. For the purpose, the researcher should put to himself questionslike:

Are the sentences written in the report clear? Are they grammatically correct? Do they saywhat is meant'? Do the various points incorporated in the report fit together logically? "Having atleast one colleague read the report just before the final revision is extremely helpful. Sentences thatseem crystal-clear to the writer may prove quite confusing to other people; a connection that hadseemed self evident may strike others as a non-sequitur. A friendly critic, by pointing out passagesthat seem unclear or illogical, and perhaps suggesting ways of remedying the difficulties, can be aninvaluable aid in achieving the goal of adequate communication.

10. Bibliography

Bibliography should be prepared and appended to the research report

11. Preparation of the index

At the end of the report, an index should invariably be given, thevalue of which lies in the fact that it acts as a good guide, to the reader. Index may be prepared bothas subject index and as author index. The former gives the names of the subject-topics or conceptsalong with the number of pages on which they have appeared or discussed in the report, whereas thelatter gives the similar information regarding the names of authors. The index should always bearranged alphabetically. Some people prefer to prepare only one index common for names of authors,subject-topics, concepts and the like ones.

PRECAUTIONS FOR WRITING RESEARCH REPORTS

Research report is a channel of communicating the research findings to the readers of the report. A good research report is one which does this task efficiently and effectively. As such it must be prepared keeping the following precautions in view:

1. While determining the length of the report (since research reports vary greatly in length), one should keep in view the fact that it should be long enough to cover the subject but short enough to maintain interest. In fact, report-writing should not be a means to learning more and more about less and less.
2. A research report should not, if this can be avoided, be dull; it should be such as to sustain reader's interest.
3. Abstract terminology and technical jargon should be avoided in a research report. The report should be able to convey the matter as simply as possible. This, in other words, means that report should be written in an objective style in simple language, avoiding expressions such as "it seems," "there may be" and the like.
4. Readers are often interested in acquiring a quick knowledge of the main findings and as such the report must provide a ready availability of the findings. For this purpose, charts, graphs and the statistical tables may be used for the various results in the main report in addition to the summary of important findings.
5. The layout of the report should be well thought out and must be appropriate and in accordance with the objective of the research problem.
6. The reports should be free from grammatical mistakes and must be prepared strictly in accordance with the techniques of composition of report-writing such as the use of quotations, footnotes, documentation, proper punctuation and use of abbreviations in footnotes and the like.
7. The report must present the logical analysis of the subject matter. It must reflect a structure wherein the different pieces of analysis relating to the research problem fit well.

8. A research report should show originality and should necessarily be an attempt to solve some intellectual problem. It must contribute to the solution of a problem and must add to the store of knowledge.
9. Towards the end, the report must also state the policy implications relating to the problem under consideration. It is usually considered desirable if the report makes a forecast of the probable future of the subject concerned and indicates the kinds of research still needs to be done in that particular field.
10. Appendices should be enlisted in respect of all the technical data in the report.
11. Bibliography of sources consulted is a must for a good report and must necessarily be given.
12. Index is also considered an essential part of a good report and as such must be prepared and appended at the end.
13. Report must be attractive in appearance, neat and clean, whether typed or printed.
14. Calculated confidence limits must be mentioned and the various constraints experienced in conducting the research study may also be stated in the report.
15. Objective of the study, the nature of the problem, the methods employed and the analysis techniques adopted must all be clearly stated in the beginning of the report in the form of introduction.

TYPES OF REPORTS

1. Research reports vary greatly in length and type. In each individual case, both the length and the form are largely dictated (determined / ordered) by the problems at hand
2. For instance, business firms prefer reports in the letter form, just one or two pages in length.

3. Banks, Insurance organizations and financial institutions are generally fond of (having a liking for) the short balance-sheet type of tabulation for their annual reports to their customers and shareholders
4. Mathematicians prefer to write the results of their investigations in the form of algebraic (numerical) notations (Symbols + / -)
5. Chemists (Scientist trained in Chemistry) report their results in symbols and formulae (method)
6. Students of literature usually write long reports presenting the critical analysis of some writer or period or the like with a liberal use of quotations from the works of the author under discussion
7. In the field of education and psychology, the favorite form in the report on the results of experimentation accompanied by the detailed statistical tabulations
8. Clinical psychologists and social pathologists (diagnosis of disease) frequently find it necessary to make use of the case-history (all the relevant information previous gatherer) form
9. News items in the daily papers are also forms of report writing. They represent firsthand on-the-scene accounts of the events described or compilation (collection) of interviews with persons who were on the scene (area). In such report the first paragraph usually contains the important information in detail and the succeeding paragraphs contain material which is progressively less and less important
10. Book – reviews which analyze the content of the book and report on the author's intentions, his success or failure in achieving his aims, his language, his style, scholarship (learning /

research / study), bias or his point of view, such reviews also happen to be a kind of short report

11. The reports prepared by governmental bureaus (agency) special commissions, and similar other organizations are generally very comprehensive (full / complete) reports on the issues involved. Such reports are usually considered as important research products
12. Similarly, Ph.D.theses and dissertation are also a form of report-writing, usually completed by students in academic institutions
13. The above narration throws light on the fact that the results of a research investigation can be presented in a number of ways viz., a technical report, popular report,
14. Which method of presentation to be used in a particular study depends on the circumstances under which the study arose (take place) and the nature of the results
15. A technical report is used whenever a full written report of the study is required whether for record-keeping or for public dissimilation
16. A popular report is used if the research results have policy implications.

A) TECHNICAL REPORT

In technical report the main emphasis is on (i) the methods employed (ii) assumptions made in the course of the study (iii) the detailed presentation of the findings including their limitations and supporting data

1) Summary of Results

A brief review of the main findings just in two or three pages

2) Nature of the Study

Description of the general objectives of study, formulation of the problem in operational terms, the working hypothesis, the type of analysis and data required

3) Methods Employed

Specific methods used in the study and their limitations. For instance, in sampling studies we should give details of sample design viz. sample size, sample selection etc.,

4) Data

Discussion of data collected, their sources, characteristics and limitations. If secondary data are used, their suitability of the problem at hand be fully assessed.

5) Analysis of Data and Presentation of Findings

The analysis of data and presentation of the findings of the study with supporting data in the form of tables and charts be fully narrated (explained). This, in fact, happens to be the main body of the report usually extending over several chapters

6) Conclusions

A detailed summary of the findings and the policy implications drawn from the results be explained

7) Bibliography

Bibliography of various sources consulted be prepared and attached

8) Technical Appendices

Appendices be given for all technical matters relating to questionnaire, mathematical derivations, elaboration on particular technique of analysis and the like ones

9) Index

Index must be prepared and be given invariably in the report at the end. Even in technical report, simple presentation and ready availability of the findings remain an important consideration and as such the liberal use of charts and diagrams is considered desirable

B) POPULAR REPORT

The popular report is one which emphasis on simplicity and attractiveness. The simplification should be sought (required) through clear writing, minimization of technical, particularly mathematical, details and liberal use of charts and diagrams. Attractive layout along with large print, many subheadings, even an occasional cartoon now and then is another characteristic feature of the popular report

1) Findings and their Implications

Emphasis (importance) in the report is given on the findings of most practical interest and on the implication of these findings

2) Recommendation for Action

Recommendations for action on the basis of the findings of the study is made in this section of the report

3) Objectives of the Study

A general review of how the problem arise is presented along with the specific objectives of the project under study

4) Methods Employed

A brief and non-technical description of the methods and techniques used, including a short review of the data on which the study is based, is given in this part of the report

5) Results

This section constitutes the main body of the report wherein the results of the study are presented in clear and non-technical terms with the liberal use of all sorts of illustrations such as charts, diagrams and the like ones

6) Technical Appendices

More detailed information on methods used, forms etc, is presented in the form of appendices. But the appendices are often not detailed if the report is entirely meant for general public

ORAL PRESENTATION

1. At times oral presentation of the results of the study is considered effective, particularly in cases where policy recommendations are indicated by project results.
2. The merit of this approach lies in the fact that it provides an opportunity for give-and-take decisions which generally lead to a better understanding of the findings and their implications.
3. But the main demerit of this sort of presentation is the lack of any permanent record concerning the research details and it may be just possible that the findings may fade (weaken) away from people's memory even before an action is taken.
4. In order to overcome this difficulty, a written report may be circulated before the oral presentation and referred to frequently during the discussion.
5. Oral presentation is effective when supplemented by various visual devices. Use of slides, wall charts and blackboards is quite helpful in contributing to clarity and in reducing the boredom, if any. Distributing a board outline, with a few important tables and charts concerning the research results, makes the listeners attentive who have a ready outline on which to focus their thinking. This very often happens in academic institutions where the researcher discusses his research findings and policy implications with others either in a seminar or in a group discussion.

6. Thus, research results can be reported in more than one ways, but the usual practice adopted, in academic institutions particularly, is that of writing the Technical Report and then preparing several research papers to be discussed at various forums in one form or the other. But in practical field and with problems having policy implications, the technique followed is that of writing a popular report. Researches done on governmental account or on behalf of some major public or private organisations are usually presented in the form of technical reports.

Executive Summary

An executive summary, or management summary, is a short document or section of a document, produced for business purposes, that summarizes a longer report or proposal or a group of related reports in such a way that readers can rapidly become acquainted with a large body of material without having to read it all. It usually contains a brief statement of the problem or proposal covered in the major document(s), background information, concise analysis and main conclusions. It is intended as an aid to decision-making by managers and has been described as the most important part of a business plan.

An executive summary differs from an abstract in that an abstract will usually be shorter and is typically intended as an overview or orientation rather than being a condensed version of the full document. Abstracts are extensively used in academic research where the concept of the executive summary is not in common usage. "An abstract is a brief summarizing statement... read by parties who are trying to decide whether or not to read the main document", while "an executive summary, unlike an abstract, is a document in miniature that may be read in place of the longer document"

Structure

- Be approximately 5-10% of the length of the main report
- Be written in language appropriate for the target audience
- Consist of short, concise paragraphs
- Begin with a summary
- Be written in the same order as the main report
- Only include material present in the main report
- Make recommendations
- Provide a justification
- Have a conclusion
- Be readable separately from the main report

WRITING AN ABSTARCT

Abstracts have always served the function of "selling" your work. But now, instead of merely convincing the reader to keep reading the rest of the attached paper, an abstract must convince the reader to leave the comfort of an office and go hunt down a copy of the article from a library (or worse, obtain one after a long wait through inter-library loan). In a business context, an "executive summary" is often the only piece of a report read by the people who matter; and it should be similar in content if not tone to a journal paper abstract.

Parts of an Abstract

Motivation

Why do we care about the problem and the results? If the problem isn't obviously "interesting" it might be better to put motivation first; but if your work is incremental progress on a problem that is widely recognized as important, then it is probably better to put the problem statement first to indicate which piece of the larger problem you are breaking off to work on.

This section should include the importance of your work, the difficulty of the area, and the impact it might have if successful.

Problem statement

What problem are you trying to solve? What is the scope of your work (a generalized approach, or for a specific situation)? Be careful not to use too much jargon. In some cases it is appropriate to put the problem statement before the motivation, but usually this only works if most readers already understand why the problem is important.

Approach

How did you go about solving or making progress on the problem? Did you use simulation, analytic models, prototype construction, or analysis of field data for an actual product? What was the extent of your work (did you look at one application program or a hundred programs in twenty different programming languages?) What important variables did you control, ignore, or measure?

Results

What's the answer? Specifically, most good computer architecture papers conclude that something is so many percent faster, cheaper, smaller, or otherwise better than something else. Put the result there, in numbers. Avoid vague, hand-waving results such as "very", "small", or "significant." If you must be vague, you are only given license to do so when you can talk about orders-of-magnitude improvement. There is a tension here in that you should not provide numbers that can be easily misinterpreted, but on the other hand you don't have room for all the caveats.

Conclusions

What are the implications of your answer? Is it going to change the world (unlikely), be a significant "win", be a nice hack, or simply serve as a road sign indicating that this path is a waste of time (all of the previous results are useful). Are your results general, potentially generalizable, or specific to a particular case?

Other Considerations

- An abstract must be a fully self-contained, capsule description of the paper. It can't assume (or attempt to provoke) the reader into flipping through looking for an explanation of what is meant by some vague statement. It must make sense all by itself. Some points to consider include:
- Meet the word count limitation. If your abstract runs too long, either it will be rejected or someone will take a chainsaw to it to get it down to size. Your purposes will be better served by doing the difficult task of cutting yourself, rather than leaving it to someone else who might be more interested in meeting size restrictions than in representing your efforts in the best possible manner. An abstract word limit of 150 to 200 words is common.
- Any major restrictions or limitations on the results should be stated, if only by using "weasel-words" such as "might", "could", "may", and "seem".
- Some publications request "keywords". These have two purposes. They are used to facilitate keyword index searches, which are greatly reduced in importance now that on-line abstract text searching is commonly used. However, they are also used to assign papers to review committees or editors, which can be extremely important to your fate.

So make sure that the keywords you pick make assigning your paper to a review category obvious (for example, if there is a list of conference topics, use your chosen topic area as one of the keyword tuples).

Norms for using tables,charts and Diagrams

1. Ensure that display items are self-explanatory: Some readers (and certainly reviewers and journal editors) turn their attention to the tables and figures before they read the entire text, so these display items should be self-contained.

2. Refer, but don't repeat: Use the text to draw the reader's attention to the significance and key points of the table/figure, but don't repeat details.

3. Be consistent: Ensure consistency between values or details in a table (e.g., abbreviations, group names, treatment names) and those in the text.

4. Give clear, informative titles: Table and figure titles should not be vague9,19 but should concisely describe the purpose or contents of the table/figure and should ideally draw the reader's attention to what you want him/her to notice (e.g., Advantages and disadvantages of using sleep therapy with patients suffering from schizophrenia). Also ensure that column heads, axis labels, figure labels, etc., are clearly and appropriately labelled.

5. Adhere to journal guidelines: Check what your target journal has to say about issues like the number of tables and figures, the style of numbering, titles, image resolution, file formats, etc., and follow these instructions carefully.

Guidelines for tables

1. Combine repetitive tables: Tables and figures that present repetitive information will impair communication rather than enhance it. Examine the titles of all your tables and figures and check if they talk about the same or similar things. If they do, rethink the presentation and combine or delete the tables/graphs.

2. Divide the data: When presenting large amounts of information, divide the data into clear and appropriate categories and present them in columns titled accurately and descriptively.

3. Watch the extent of data in your tables: If the data you have to present is extensive and would make the tables too cluttered or long, consider making the tables a part of the Appendix or supplemental material.

4. De-clutter your table: Ensure that there is sufficient spacing between columns and rows7 and that the layout does not make the table look too messy or crowded.

Guidelines for figures

1. Ensure image clarity: Make sure that all the parts of the figure are clear.Use standard font; check that labels are legible against the figure background; and ensure that images are sharp

2. Use legends to explain the key message:Figure legends are pivotal to the effectiveness of a figure. Use them to draw attention to the central message as well as to explain abbreviations and symbols.

3. Label all important parts: Label the key sections and parts of schematic diagrams and photographs, and all axes, curves, and data sets in graphs and data plots.

4. Give specifics: Include scale bars in images and maps; specify units wherever quantities are listed; include legends in maps and schematics; and specify latitudes and longitudes on maps.

1. The title clearly describes what the table is about.

2. The column heads are descriptive and clearly indicate the nature of the data presented.

3. The data is divided into categories for clarity.

4. It is self-contained and can be understood quite well even without reference to the entire paper.

5. Superscript letters and notes are used to offer additional, clarifying information.

6. Sufficient spacing is present between columns and rows; the layout is clean; and the font is legible.

APPENDIX: NORMS FOR USING INDEX AND BIBLIOGRAPHY

CONVENTIONS RELATING TO PREPARATION OF RESEARCH REPORTS.

Organization of the Research Report Most scientific research reports, irrespective of the field, parallel the method of scientific reasoning. That is: the problem is defined, a hypothesis is created, experiments are devised to test the hypothesis, experiments are conducted, and conclusions are drawn. The exact format of scientific reports is often discipline dependent with variations in order and content. The student is encouraged to adopt the format that is most appropriate to the discipline of the research. Many journals offer a formatting template to aid the author. One example of such a framework is as follows:

- Title
- Abstract
- Introduction
- Experimental Details or Theoretical Analysis
- Results
- Discussion
- Conclusions and Summary
- References

KAHE

S.No	Questions	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6	Answer
1	A study to gain familiarity with a phenomenon to achieve new insights is :	Exploratory research	Descriptive Research	Diagnostic Research	Hypothesis testing Research			Exploratory research
2	Study to determine the frequency with which something occurs uis called	Diagnostic Research	Descriptive Research	Exploratory research	Hypothesis testing Research			Diagnostic Research
3	Study to portray accurately characteristics of a particular individual ,	Exploratory research	Descriptive Research	Diagnostic Research	Hypothesis testing Research			Descriptive Research
4	In social science and business research we often use:	Ex post facto research	Hypothesis testing Research	Diagnostic Research	Exploratory research			Ex post facto research
5	If the researcher has no control over the variables it is termed as :	Exploratory research	Ex post facto research	Diagnostic Research	Hypothesis testing Research			Ex post facto research
6	Critical evaluation made by the researcher with the facts and information already	Exploratory research	Analytical Research	Diagnostic Research	Hypothesis testing Research			Analytical Research
7	Gathering knowledge for knowledge sake is termed as :	Exploratory research	Pure Research	Diagnostic Research	Hypothesis testing Research			Pure Research
8	Research aiming to find an solutions for an immediate problem is called:	Applied Research	Exploratory research	Diagnostic Research	other research			Applied Research
9	Research to find reason , why people think or do certain things is an example	Qualitative research	Quantitative Research	Applied Research	Fundamental research			Qualitative research
10	Data based research coming up with conclusions which are capable of being verified	Empirical Research	Pure Research	Conceptual Research	Basic Research			Empirical Research
11	To develop new concepts or to reinterpret existing ones, philosophers and thinkers	Empirical Research	Conceptual Research	Pure Research	Basic Research			Conceptual Research

12	A researcher is free to pick up a problem in	Conclusion-oriented Research	Decision-oriented Research	Historical Research	Diagnostic Research			Conclusion-oriented Research
13	A researcher will not be free to embark upon research according to his own	Historical Research	Decision-oriented Research	Conclusion-oriented Research	Diagnostic Research			Decision-oriented Research
14	Decision-oriented Research is always for the need of the	Researcher	Society	Decision maker	Others			Decision maker
15	Research carried out on over several time periods are called	One-time Research	Longitudinal Research	Field setting Research	Clinical Research			Longitudinal Research
16	Research studies with substantial structure and specific hypothesis to be	Formalized Research	Empirical Research	Historical Research	One-time Research			Formalized Research
17	Characteristics of research that allows research results to be verified by replicating	Logical	Empirical	Historical	Replicable			Replicable
18	The methods or techniques used by researchers in performing research	Research Methods	Research techniques	Research Methodology	Research process			Research Methods
19	The way to systematically solve the research problem is called:	Research Methods	Research techniques	Research process	Research Methodology			Research Methodology
20	Conceptual Literature comprises of :	Concepts and Theories	Earlier studies	Scientific enquiry	Specific Details			Concepts and Theories
21	Empirical literature comprises of :	Concepts and Theories	Earlier studies	Scientific enquiry	Specific Details			Earlier studies
22	if we want to identify the reasons for human behavior we can best apply	Qualitative research	Quantitative research	Fundamental research	Analytical research.			Qualitative research
23	Attitude or opinion research is	qualitative research	quantitative research	conceptual research	fundamental research.			qualitative research

24	To determine the frequency with which something occurs or associated with	descriptive research	formulative research	diagnostic research	hypothesis research			diagnostic research
25	Research means _____	find out new things	find out old things	find out methods	find out techniques			find out new things
26	. Depth interview defined as	Medge Jhon	Dr. Fay. B. Karpf	P.V. Young	Morton.			P.V. Young
27	Primary data is _____ source of collection	Second	Indirect	non direct	Direct			Direct
28	Over Rapporting _____	to collect the data	for good interaction	should be avoided	should be used			should be avoided
29	Sample size should be possible with the help of	Society	Available resources	Person	hypothesis			Available resources
30	Good research is	Progressive	Systematic	Informative	non informative			Systematic
31	Failure to set out clearly the objectives of research are bound to lead to	Partly true	partly false	false	true			true
32	Which one if the following is not the essentials of sampling?	Representative	Adequacy	Homogeneity	Reliability			Homogeneity
33	The main objective of research design is	Budgeting on time	Budgeting on money	Number of respondents to be selected	prepare a structure			prepare a structure
34	The concept which can take on different quantitative values is called as _____	variable	information	sources	data			variable
35	The pre determined plots or the blocks and different treatment are used as	units	experimental units	tratments	control groups			experimental units

36	Which one of the following is not the source of research problem?	Specialization	Observation	Consultations	Resource development			Resource development
37	The formal, systematic and intensive process of carrying on a scientific	Research Design	Research	Interpretation	Research analysis			Research
38	The research, which has the purpose of improving a product or a process testing	Statistical research	Applied research	Domestic research	Biological research			Applied research
39	The research that describes what was in the past is	Schedule	Research	Historical research	Ancient research			Historical research
40	According to the nature of the sample chosen for the study what varies	Data Collection	Sampling	Methodology	Research Problem			Data Collection
41	The device which would retain the actual wording of the respondents is	Recording	)Writing	Copying	Editing			Recording
42	It refers to the relationship between facts or to the of them in some meaningful	Science	Fact	Theory	Economics			Theory
43	What is regarded as an empirically verifiable observation?	Science	Research	Theory	Social Sciences			Research
44	It is a tentative statement about something validity of which is usually unknown	Null hypothesis	Hypothesis	Research hypothesis	Statistical hypothesis			Hypothesis
45	Formulative research studies emphasise on	Bibliographical survey	Discovery of insights	Sample representatives	Specific prediction.			Discovery of insights
46	Research depends upon	random method	systematic method	horizontal method	vertical method			systematic method
47	Which of the following method of sampling belongs to the category of	Judgement sampling	Quota sampling	Convenience sampling	Stratified sampling			Convenience sampling

48	Basic principle of an experimental design is	Duplication	Replication	Simplification	Multiplication			Replication
49	The content of a structured interview is	Statement based on experience	Controlled observation	Predetermined questions	Feelings and beliefs			Predetermined questions
50	Social microscope is	Statistical data	Interview	Observation	Case study method			Case study method
51	In an experimental research a group is exposed to usual conditions is termes	experimental group	cntrol group	confounded relationship	non experimental group			cntrol group
52	A Standard test must have the following quality	) Objectivity	Length	Interest	Nonverbal content			Objectivity
53	---------- constitutes the blue print collection measurement and analysis	Applied thinking	Operational research	Subjective assessment	Research design			Research design
54	Phenomena, which can take on qualitatively different values even in decimal	Non-continuous	Continuous variable	Independent variables	dependent variables			Continuous variable
55	_____ research is based on the measurement of quantity or amount	Quantitiative	Qualititative	experimental	non experimental			Quantitiative
56	_____ research is concerned with qualitative phenomena	Quantitative	Qualitative	non experimental	experimental			Qualitative
57	_____ research is related to some abstract idea or theory	Conceptual	Empirical	Conceptual & Empirical	applied			Conceptual
58	_____ research relies on experience or observation alone	Conceptual	Empirical	Field setting	Historical			Empirical
59	_____ research, which utilizes historical sources like documents,	Field setting	Conceptual	Empirical	Historical			Historical

60	_____ scale is simply a system of assigning number	Ordinal	Interval	Nominal	ratio			Nominal

S.NO	Questions	Option 1	Option 2	Option 3	Option 4	Option 5
1	Deliberate sampling is also known as :	Purposive sampling	Probability Sampling	Random Sampling	Judgment sampling	
2	When population elements are selected for inclusion in the sample based on the ase of access:	Convenience sampling	Judgment sampling	Random Sampling	Probability Sampling	
3	Researcher's Judgment used for selecting items, which he considers as representative of the population, is	Judgment Sampling	Convenience sampling	Random Sampling	Systematic Sampling	
4	Simple random sampling is also called:	Chance Sampling	Convenience sampling	Random Sampling	Systematic Sampling	
5	To select a sample as every 10^{th} house on one side of a street is an example of:	Non Probability sampling	Systematic Sampling	Convenience sampling	Random Sampling	
6	To draw a sample from non-homogeneous group, the sampling used is:	Stratified Sampling	Deliberate Sampling	Convenience sampling	Random Sampling	
7	Quota Sampling is an important form of :	Probability Sampling	Non-Probability Sampling	Convenience Sampling	Systematic Sampling	
8	Grouping the population and selecting groups for inclusion in the sampling is called:	Cluster Sampling	Area Sampling	Stratified Sampling	Systematic Sampling	
9	Sampling that helps the researcher to carry out research where there is no specific list of population concerned	Cluster Sampling	Systematic Sampling	Stratified Sampling	Area Sampling	

10	Sampling determined according to mathematical decisions on the basis o information yielded as Survey	Sequential Sampling	Multi-Stage Sampling	Area sampling	Quota Sampling	
11	Data Collected by filling up the Schedules by the enumerators on the basis of replies given by respondents.	Questionnaire	Schedule	Interview	observation	
12	Method of selecting items to be observed for the given study is called:	Sampling design	Statistical Design	Operational Design	Observational Design	
13	Research Design that relates to conditions under which observations are to be made is called:	Sampling design	Statistical Design	Operational Design	Observational Design	
14	A concept which can take on different quantitative values is called a	Variable	Research	Research Design	data	
15	_____ data are those which are collected afresh and for the first time	Primary	Secondary	case study method	warranty cards	
16	_____ data are to be originally collected	Secondary	Primary	warranty cards	case study method	
17	_____ data's are those which have already been collected by someone else.	Primary	Secondary	Primary & Secondary	case study method	
18	_____ Method is most commonly used method specially ;in studies relating to behavioural	Interview	Questionnaires	Schedules	Observation	
19	_____ Most of collecting data involves presentation of oral verbal stimuli	Questionnaires	Interview	Observation	Schedule	

20	The method of collecting inform through personal interview is usually carried out in a _____ way	Structured	unstructured	formal	informal	
21	_____ Interview is meant to focus attention on the given experience of the respondent and its	Focused	Clinical	Structured	Directive	
22	_____ Interview is concerned with broad underlying feelings orMotivations	Unstructured	Sttructured	Clinical	Non-directive	
23	The main Number of sources of data is	2	3	4	1	
24	Number of methods of collection of primary data is	2	3	4	5	
25	Number of questions in a questionnaire should be	5	10	maximum	minimum	
26	Sources of secondary data are	Published sources	Unpublished sources	Neither Published sources nor Unpublished	both Published sources and Unpublished sources	
27	compared with primary data , secondary data are	more reliable	less reliable	equally reliable	none of these	
28	In Quantitative classification data are classified on the basis of	attributes	time	location	magnitudes	
29	A ------- source is one that itself collects the data.	Primary	Secondary	Published	un published	

30	The data which is compiled from the records of others is called-----data	Primary	Secondary	un published	Published	
31	What type of data will be original in character	unpublised	source data	primary data	secondary data	
32	. What type of data are those which have already been by someone else	secondary data	primary data	source data	unpublised	
33	Primary data can be collected through ,	direct method	indirect method	other mathods	direct and indirec method	
34	Which method of data collection is used in studies relating to behavioral sciences	mailed questionnaire	through post	observation method	indirect collection	
35	Which type of data collection is most commonly used method	mailed questionnaire	through post	observation method	indirect collection	
36	If observations takes place in the natural setting, it may be termed as	uncontrolled observation	controlled observation	personal observation	controlled observation	
37	If observation takes place according to definite prearranged plans, it is called	uncontrolled observation	controlled observation	personal observation	uncontrolled observation	
38	_____ is asking questions face to face	indirect method	mailed questionnaire	through post	personal interview.	
39	_____ is meant to focus attention on the given experience of the respondent and its effects.	facing interview	sending post	facing indirect interview	mailing through net.	

40	In _____ interview, the interviewers function is to simply encourage the respondent to talk	Direct	in direct	through post	through mail	
41	More information and depth can be obtained in _____ method	through mail	questionnaire	indirect interview	interview	
42	Which method of data collection is very popular	questionnaire method	pilot study	mailed questionnaire	through post	
43	Which method of data collection is used especially in case of big enquiries	questionnaire method	pilot study	mailed questionnaire	through post	
44	Which method of data collection is very much like the collection of data through questionnaires	schedules	inquires	finding	posting method	
45	Journals, books, magazines etc.. are useful sources of collecting	primary data	secondary data	case study method	warranty cards	
46	Case study method is a very popular method for ,	quantitative method	qualitative method	non qualitative	non quantitative	
47	Which method involves careful and complete observation of a unit	pilot study	questionnaire study	case study	schedule	
48	The collected raw data to detect errors and are called ,	editing	coding	classification	tabulation	
49	When editing is done to assure that the data are	accurate	informal	formal	additional	

50	Pilot study should be taken for,	pre- testing the questionnaire	post testing the questionnaire	pre- testing the hypothesis	post testing the hypothesis	
51	Questionnaire should be contain	simple and easy	complex	not understandable	maximum	
52	t-test is an important	parametric test	. non-parametric test	normal distribution test	. statistical test	
53	Random sampling is	Chance sampling	Non probability sampling	Complex sampling	Deliberate sampling.	
54	Probability sampling is	accidental sampling	quata sampling	snow ball sampling	systematic sampling.	
55	Non – probability sampling is	systematic sampling.	Random sampling.	Cluster sampling.	Quota sampling.	
56	Which is not a method of non-probability sampling.	accidental sampling	quata sampling.	Purposive sampling	Random sampling.	
57	Which is not a method of probability sampling.	Systematic sampling.	cluster sampling.	Area sampling.	Purposive sampling.	
58	What is to be Quota sampling is a method of	Probability sampling	Unrestricted sampling	Non-probability sampling	cluster sample	
59	Random sampling conducted to test a questionnaire?	Experiment	Research	Pilot survey	Interview.	
60	The aim of schedule is	To collect the data in an objective manner	Reliability	Subjective	consistent	

Option 6	Answer
	Purposive sampling
	Convenience sampling
	Judgment Sampling
	Chance Sampling
	Systematic Sampling
	Stratified Sampling
	Non-Probability Sampling
	Cluster Sampling
	Area Sampling

	Sequential Sampling
	Schedule
	Sampling design
	Observational Design
	Variable
	Primary
	Primary
	Secondary
	Observation
	Interview

	Structured
	Focused
	Sttructured
	2
	4
	minimum
	Neither Published sources nor Unpublished sources
	less reliable
	magnitudes
	Secondary

	Published
	primary data
	secondary data
	direct and indirec method
	observation method
	observation method
	uncontrolled observation
	controlled observation
	personal interview.
	facing interview

	in direct
	interview
	questionnaire method
	questionnaire method
	schedules
	secondary data
	qualitative method
	case study
	editing
	accurate

	pre- testing the questionnaire
	simple and easy
	parametric test
	Chance sampling
	systematic sampling.
	Quota sampling.
	Random sampling.
	Purposive sampling.
	Non-probability sampling
	Experiment
	To collect the data in an objective manner

	Questions	Option 1	Option 2	Option 3	Option 4	Option 5
S.NO	A predictive statement that relates a dependant variable with an independent variable is:	Null Hypothesis	Alternate Hypothesis	Research hypothesis	non hypothesis	
1	In an Experimental Hypothesis testing research when a group is exposed to usual conditions it is termed as	Experimental Group	Control Group	Experimental units	Control units	
2	In an Experimental Hypothesis testing research when a group is exposed to Special conditions it is termed as	Experimental Group	Control Group	Control units	Experimental units	
3	Survey of people who have had practical experience with the problem to be studied is called:	Feedback Survey	Experience Survey	Census	interview	
4	Assigning numbers of basket ball players in order to identify them is an example of:	Nominal Scale	Ordinal Scale	Interval scale	Ratio Scale	
5	Rank orders represent ordinal scales and are frequently used in research relating to:	Quantitative phenomena	Qualitative phenomena	experimental	discrete	
6	A students rank in his graduation class involves the sue of an	Nominal Scale	Ordinal Scale	Interval scale	Ratio Scale	
7	A Fahrenheit scale is an example of an:	Nominal Scale	Ordinal Scale	Interval scale	Ratio Scale	
8	The extent to which a measuring instrument provides adequate coverage of the topic under study is:	Content Validity	Criterion Validity	Concurrent Validity	Construct Validity	
9	The usefulness of a test in predicting some future performance is termed as:	Content Validity	Predictive Validity	Concurrent Validity	Construct Validity	

10	Procedures of assigning numbers to various degrees of opinion , attitude and other concept is called:	Scaling	Measuring	Responding	Reliability	
11	Categorical Scales are also known as:	Nominal Scale	Ordinal Scale	Interval scale	Ratio Scale	
12	Comparative Scales are also known as:	Nominal Scale	Ordinal Scale	Interval scale	Ranking Scale	
13	Scales that used to measure only one attribute of the respondent or object are called	Unidimensional Scale	Multidimensional Scale	Ordinal Scales	None of the above	
14	Approach in which the scale is developed on ad hoc basis is called:	Arbitrary Approach	Consensus Approach	Item analysis approach	Others	
15	The method of evaluation of items to be included in the instrument by panel of judges during scale construction is	Arbitrary Approach	Consensus Approach	Item analysis approach	Others	
16	The scale constructed through the Item analysis approach are called:	Arbitrary scales	Differential Scales	Summated scales	Cumulative Scales	
17	The scale constructed through the Consensus scale approach are called:	Arbitrary scales	Differential Scales	Summated scales	Cumulative Scales	
18	Likert scale is an example of :	Arbitrary scales	Differential Scales	Summated scales	Cumulative Scales	
19	.Individual observations are called ------	raw data	grouped data	ungrouped data	master data	
20	Which one is Geographical classification?	1990-91	North	Male	442	

21	In discrete frequency distribution values are given---------	Class interval	grouped data	ungrouped data	raw data	
22	In continuous frequency distribution values are given---------	Class interval	grouped data	raw data	ungrouped data	
23	Classification according to class-intervals would yield	raw data	discrete data	qualitative data	grouped data	
24	In Qualitative classification data are classified on the basis of	attributes	time	location	class intervals	
25	In chronological classification data are classified on the basis of	class intervals	attributes	time	location	
26	In Geographical classification data are classified on the basis of	area	attributes	time	location	
27	Phenomena which can take on quantitatively different values even in decimal points are called :	Discrete variables	Extraneous Variables	Continuous Variables	Independent Variables	
28	Variables that can be expressed only in integer values are called	Discrete variables	Extraneous Variables	Continuous Variables	Independent Variables	
29	Age is an example of :	Discrete variables	Extraneous Variables	Independent Variables	Continuous Variables	
30	Number of Children is an example of :	Discrete variables	Independent Variables	Continuous Variables	Extraneous Variables	
31	Variable that is antecedent to the dependant variable is called:	Discrete variables	Extraneous Variables	Continuous Variables	Independent Variables	

32	Behavioral changes that occur as a result of environmental manipulations are examples of :	Discrete variables	Extraneous Variables	Dependant Variables	Independent Variables	
33	Independent variables that are not related to the study but may affect the dependant variable are termed as :	Independent Variables	Discrete Variables	Continuous Variables	Extraneous Variables	
34	Whatever effect is noticed on dependant variable as a result of extraneous variable is technically	normal error	Experimental Error	Statistical Error	standard error	
35	Clinical interview is concerned with the course of individuals	Experience	Motivation	Life experience	work experience	
36	_____ method consists in contacting respondents on telephone itself	Telephone interview	Personal interview	Structured interview	Unstructured interview	
37	_____ interviews is concerned with the course of individuals life experience	unstructured	structured	clinical	non directive	
38	_____ method is consisting in contacting respondents in telephone itself	telephone interview	personnel interview	structured interview	unstructured interview	
39	____ method is plays an important part in individual surveys	telephone interview	personal interview	Observation	schedule	
40	telephone interview method consists in contacting respondent on _____ itself	direct	indirect	face to face	telephone	
41	_____ method of data collection is quite popular particularly in case of enquires	questionnaire	schedule	interview	observation	
42	_____ method of data collection is very much like the collection of data through	questionnaire	schedules	interview	observation	

43	warranty cards are usually _____ sized cards	normal	a.4 size cards	postal	store card	
44	____ is a definite plan for obtaining a sample for a given population	respondents	selected items	sample design	technique	
45	schedules method of data collection is very much like the collection of data through	questionnaires	schedules	interview	observation	
46	research has its special significance for solving --------- problems	situational	economical	operational	social	
47	Research has its special significance in solving various operational &_____problemsin business industry	planning	scientific	structure	Business	
48	Research has its special significance in solving various operational & Planningproblems of _____	Business	Industry	finance	business and industry	
49	_____ scale places events in order	ordinal	ratio	interval	Nominal	
50	In interval scale, the _____ are adjusted in terms	Sequence	Interval	Numeric	ratio	
51	. In _____ scale, the intervals are adjusted in terms	Ordinal	Nominal	Interval	Ratio	
52	_____ scale have an absolute or true zero of measurement	Ordinal	Nominal	interval	Interval	
53	Ratio scale have an absolute or two zero of _____	Measurement	Values	scales	sources	

54	. _____ scale represents the actual amount of variables	Ratio	Ordinal	Nominal	Interval	
55	Ratio scale represents the actual amount of _____	Numbers	Ratio	Values	Variable	
56	Geometric & harmonic means calculated in _____ scale	Nominal	Ordinal	Interval	Ratio	
57	_____ & _____ means calculated in ratio scale	Geometric	harmonic	mean	median	
58	_____ design is needed to facilitate the smooth sailing of the various research operation	Research	Structure	Numeric	experimental	
59	Prediction or a hypothesized relationship is to be tested by _____ method	Research	Experiment	Structured	Scientific	

Option 6	Answer
	Research hypothesis
	Control Group
	Experimental Group
	Experience Survey
	Nominal Scale
	Qualitative phenomena
	Ordinal Scale
	Interval scale
	Content Validity
	Predictive Validity

	Scaling
	Ratio Scale
	Ranking Scale
	Unidimensional Scale
	Arbitrary Approach
	Consensus Approach
	Summated scales
	Differential Scales
	Summated scales
	ungrouped data
	North

	grouped data
	Class interval
	grouped data
	attributes
	time
	area
	Continuous Variables
	Discrete variables
	Continuous Variables
	Discrete variables
	Independent Variables

	Dependant Variables
	Extraneous Variables
	Experimental Error
	Life experience
	Telephone interview
	clinical
	telephone interview
	telephone interview
	telephone
	questionnaire
	schedules

	postal
	sample design
	questionnaires
	operational
	planning
	business and industry
	ordinal
	Interval
	Interval
	Interval
	Measurement

	Ratio
	Variable
	Ratio
	Geometric and hormonic
	Research
	Scientific

S.NO	Questions	Option 1	Option 2	Option 3	Option 4	Option 5
1	The word ---------- is used to indicate various statistical measures like mean,	standard deviation, correlation etc, in the universe.	parameter	statistic	hypothesis	
2	The term STATISTIC refers to the statistical measures relating to the	population	sample	universe	sample unit	
3	Type two error is called as_____	alpha error	standard error	beta error	statistic error	
4	The hypothesis under test is .	Simple hypothesis___	alternative hypothesis	null hypothesis	complex hypothesis	
5	Level of significance is the probability of	Type I error	Type II error	Not committing error	committing error	
6	A test based on a test statistic is classified as	Randomized test	non-randomized test	sequential test	Bayes test	
7	Degrees of freedom is related to .	No. of observations in a set	hypothesis under test	no. of independent observations in a set]	No.of samples under test	
8	A critical function provides the basis for ----------.	accepting H_0	rejecting H_0	no decision about H_0	no decision about Ha	
9	Student's t-test is applicable in case of ----------.	small samples	for sample of size between 5 and 30	large samples	optimum samples	
10	If the sample size is ------- 30 then those samples may be regarded as large samples.	less than	greater than	less than or greater than	smaller than	

11	The random sampling distribution of statistics is approximately,	normal	binomial	poission	quantitative	
12	Reject H_0 when it is true is known as	Type I error	Type II error	Correct decision	wrong decision	
13	Reject H_0 when it is false is known as	Type I error	Type II error	wrong decision	Type II error	
14	Accept H_0 when it is true is known as,	Type I error	wrong decision	Correct decision	Type II error	
15	Accept H_0 when it is false is known as,	Type I error	Type II error	wrong decision	Correct decision	
16	If $\|Z\| < Z_a$ the null hypothesis is,	rejected	accepted	zero	not accepted	
17	Any statistical measure computed from population data is known as,	parameter	sample	statistic	event	
18	Any statistical measure computed from population data is known as,	event	sample	statistic	parameter	
19	A part of the population selected for study is called a	parameter	statistic	sample	event	
20	The standard deviation of the sampling distribution is known as,	standard error	proportion	hypothesis	normal error	
21	Testing of hypothesis was initiated by	Neyman	spearman	poison	kothari. C. R	

22	The value of 5 % level of significance is	2.58	1.96	1,64	2.33	
23	The value of 1 % level of significance is,	2.58	1.96	1,64	2.33	
24	type one error is denoted as	alpha error	beta error	normal error	standard error	
25	The vertical scale of a bar diagram possesses data of	Percentage	Frequency	Nominal scale	Discrete data.	
26	"Frequency Polygon" is formed in connecting	Variables	Quantitative data	Numerical numbers	Mid point	
27	The second part of the Bibliography contains	Author name	Publications	Double authors	Magazines.	
28	Who used the term ANOVA for the first time?	RA fisher	Sne décor	Cohen	Kothari	
29	----------refers to the process of assigning numerals or symbols to answers of response	Coding	Editing	Classification	tabulation	
30	The analysis, which studies the joint variation of two or more variables	Casual	Correlation	Multiple regression	Canonical.	
31	. ---------- is the most commonly or frequently occurring value in a series in Research	Median	Mean	Mode	standard deviation	
32	.The deviation, which is used mostly in research studies and is regarded as a satisfaction	Mean deviation	Standard deviation	Skewness	Mode	

33	The test used to judge the significance of difference between the means of two samples	Karl Pearson's test	T- test	F-test	Chi square-test	
34	Who used the term ANOVA for the I time?	RA Fisher	Sne décor	Kothari	Cohen	
35	.Data are converted into symbols at what stage?	Graphing	Coding	Tabulation	Figural depiction.	
36	What is to be conducted to test a questionnaire?	Pilot survey	Experiment	Pretest	Interview	
37	The research study, which is based on describing the characteristic of a particular individual or group	Experience survey	Descriptive	Diagnostic	Exploratory	
38	Under which sampling design does every item of universe has an equal chance of inclusion	Non-probability	Probability	Random	None	
39	The whole data put in concise, precise and logical order called	Objects	Sampling	Tabulation	Arrangement	
40	It is an approach where scale is developed in Adhoc basis	Arbitrary approach	consensus approach	interval scale	ratio scale	
41	I t is used to measure concepts	consensus approach	arbitrary approach	interval scale	ratio scale	
42	_____ is used to find out the relationship between the items	Arbitrary approach	consensus approach	ratio scale	factor scale	
43	_____ method are used to find out the individual items	consensus approach	arbitrary approach	item analysis approach	factor approach	

44	_____ is used to assigning number to symbols	nominal scale	ordinal scale	interval scale	ratio scale	
45	_____ places events in order	nominal scale	ordinal scale	interval scale	ratio scale	
46	Normally rank orders represent in _____	ordinal scale	nominal scale	interval scale	ratio scale	
47	_____ scale intervals are adjusted in terms of some rule	ordinal scale	nominal scale	interval scale	ratio scale	
48	_____ is an absolute or true zero measurement	ratio scale	nominal scale	interval scale	ordinal scale	
49	_____ represents the actual amount of variables	ordinal scale	nominal scale	interval scale	ratio scale	
50	It is used to deduct errors and omissions	coding	editing	classification	tabulation	
51	____ consists in the review of the reporting forms by the investigator	field editing	central editing	classification	coding	
52	It refers to the process of assigning numerals or other symbols to answers	coding	editing	classification	tabulation	
53	___ can be done by hand or by mechanical or electronical devices	coding	editing	classification	tabulation	
54	Multiplication and division is possible in _____	ordinal scale	nominal scale	ratio scale	factor scale	

55	_____ or a hypothesized relationship is to be tested by scientific method	Orders	Values	Prediction	Thoughts	
56	_____ is an underlying dimension that account of factor analysis	factor	sign test	multivariate techniques	correlation	
57	_____ is not a method of factor analsis	centroid method	principal components method	maximum liklihood method	minimum liklihood method	
58	the term path analysis is first introduced by	sewall wright	kothari	gilbreth	poission	
59	_____ analysis consists methods of classifying variables into	Cluster	factor	structure	MDS	
60	_____ Test is based on plus or minus signs	MDS	Cluster	sign test	MDS	

Option 6	Answer
	parameter
	universe
	beta error
	null hypothesis
	Type I error
	Randomized test
	hypothesis under test
	rejecting H_0
	small samples
	greater than

	normal
	Type I error
	Correct decision
	Correct decision
	Type II error
	accepted
	statistic
	statistic
	sample
	standard error
	Neyman

	1.96
	2.58
	alpha error
	Discrete data.
	Mid point
	Magazines.
	RA fisher
	Coding
	Multiple regression
	Mean
	Standard deviation

	T- test
	RA Fisher
	Coding
	Pretest
	Diagnostic
	Probability
	Tabulation
	Arbitrary approach
	arbitrary approach
	factor scale
	item analysis approach

	nominal scale
	ordinal scale
	ordinal scale
	interval scale
	ratio scale
	ratio scale
	editing
	field editing
	coding
	tabulation
	ratio scale

	Prediction
	factor
	minimum liklihood method
	sewall wright
	Cluster
	sign test

S.NO	Questions	Option 1	Option 2	Option 3	Option 4	Option 5	Option 6
1	A simple table represent	only one factor or variable	always two factor or variable	two or more number of factors or	multiple variable		
2	A complex table represent:	only one factor variable	always two factor variable	two or more factor or variables.	multiple variable		
3	The headings of the rows given in the first column of a table are called	stub	captions	titles	prefatory notes.		
4	The column headings of a table are known as	sub-title	stubs	reference notes	captions		
5	A frequency distribution can be	discrete	continuous	compound	discrete and continuous		
6	Frequency of a variable is always,	in percentage	a fraction	an integer	a point		
7	The data given as, 5,12,16,24,35,44 will be called as	continuous series	a discrete series	an individual series	time series		
8	Charts and graphs are the presentation of numerical facts by means of:	points and lines	area and other geometrical forms	symbols	points lines symbols		
9	Graphs and charts facilitate:	comparison of values	to know the trend	to know relationship	to know the objective		

10	The purpose served by diagrams and chart is,	simple presentation of data	to avoid tabulation	to avoid textual form	to avoid chart		
11	Choice of a particular chart depends on	the purpose of study	the nature of data	the type of audience	the source of data		
12	Which of the following is one-dimensional diagram	Bar- diagram	Pie-diagram	cylinder	a graph		
13	Which of the following is not a two-dimensional diagram	square diagram	multiple bar diagram	rectangular diagram	pie-chart.		
14	Non-dimensional diagrams are also known as,	cubes	spheres	pictograms	charts		
15	An alternative chart to pie-chart is,	step bar diagram	Rctangular chart	Sphere	charts		
16	Pie-chart represents the components of a factory by,	percentages	angles	sectors	circles		
17	Histogram is suitable for the data presented as	continuous grouped frequency	discrete grouped frequency	individual series	multiple series		
18	Which one is considered a major component of the research study	interpretation	research report	finding	draft		
19	Research task remains incomplete till the _____ has been presented.	report	objective	finding	suggestions		

20	The purpose of the research is not well served unless the _____ are made known to others	interpretation	research report	finding	draft		
21	What is the last step in a research study	writing report	writing finding	writing limitations	writing suggestions		
22	What report are the product of slow,painstaking,accurate inductive work	interpretation	research report	finding	draft		
23	What is the first step in report writing	logical analysis of the subject	objective	interpretation	findings		
24	Which is the finial step in report writing.	writing report	writing finding	writing draft	writing suggestions		
25	The two ways in which to develop a subject in logical analysis of subject matter are	logically	chronologicall y	simple	order		
26	What are the frame work upon which long written works are constructed	outline	draft	finding	interpretation.		
27	What is usually appended to the research work	editing	coding	bibliography	tabulation		
28	Which one follows the logical analysis of the subject and the preparation of the final out line	rough draft	bibliography	main draft	content page		
29	Which page carry title, acknowledgements, preface or forward, table of contents, list of contests, illustrations etc…	introduction part	preliminary pages	finding pages	objective pages		

30	The _____ provides the complete outline of the research report along with all details	preliminary text	finding text	main text	conclusion pages.		
31	Which one is the detailed presentation of the findings of the study	rough draft	bibliography	result	main draft		
32	.Which contains appendices in respect of all technical data such as questionnaires, sample information, mathematical derivations etc…	end matter	finding text	main draft	bibliography		
33	In a technical report _____ must be invariably given at the end of the report	interpretation	index	finding	result.		
34	The _____ is one which gives emphasis on simplicity and attractiveness	article report	research report	popular report	technical report		
35	Which one of the results of the study is considered effective especially in thecases where policy recommendations are indicated by project results	oral presentations	writing presentations	verbal presentations	writing limitations		
36	Oral presentations is effective when supplemented by various	logical devices	visual devices	coding devices	ending devices		
37	Which should be avoided in a research report	abstract terminology	technical jargon	abstract terminology and technical	footnotes		
38	_____ should slow originality and should necessarily be on attempt tosolve some intellectual problem	interpretation	research report	finding	draft		

39	Which one is a fundamental component of research process	interpretation	research report	draft	finding		
40	_____ refers to the task of drawing inferences to collected facts	interpretation	research report	finding	draft		
41	objectives, nature of problem, methods employed and the analysis technique adopted must be clearly stated at the _____ of the report	specifying	interpreting	beginning	findings		
42	Usefulness and utility of research finding lie in proper	interpretation	research report	finding	analysis		
43	_____ requires great skill and dexterity on the part of researcher	analysis	interpretation	research report	finding		
44	The researcher must remain caution about the _____ that can possibly arise in the process of interpreting results	analysis	error	findings	conclusions		
45	Which one should be considered while interpreting a given data	validity	reliability	practicality	accessability		
46	the main text provides the complete	analysis	text	outline	structure		
47	_____ is the last step of research report	writing main draft	writing final draft	writing suggesstions	writing introduction		
48	_____ comsidered as major component of research stydy at final	research report	research hypothesis	research design	research structure		

49	which one is not the steps in writing research report	logical analysis of subject matter	preparation of final outline	preparation of rough draft	formulating hypothesis		
50	_____ is appended in the to the research report	bibiliography	hypothesis	design	sources		
51	_____ report shuold give siplicity and attractiveness	popular report	technical jargon	technical report	short report		
52	which one is not the type of research report	technical report	popular report	short report	hypothesis report		
53	the bibiliography should be placed in	end of the research report	beginning og the research report	middle of the research report	side of the research report		
54	surveys are conducted in _____ studies	descriptive	experimental	quantitative	qualitative		
55	which of the following is not the essential element of report writing	research methods	research methodology	reference notes	conclusion		
56	testing hypothesis is a	inferential statistics	descriptive statistics	data preparation	data analysis		
57	which of the following is non probability sampling	snowball	random	cluster	stratified		
58	uniting various qualitative methods with quantitative methods can be called as	coalesce	triangulation	bipartitle	impassive		
59	_____ Test is also known as h Test	Kruskal Wallis test	sign test	MDS	Factor analysis		

Answer
multiple variable
two or more factor or
stub
captions
discrete and continuous
an integer
an individual series
points lines symbols
comparison of values

simple presentation of data
the nature of data
Bar-diagram
pie-chart.
pictograms
Rctangular chart
angles
continuous grouped frequency
research report
report

finding
writing report
research report
logical analysis of the subject
writing draft
logical and chronological
outline
bibliography
rough draft
preliminary pages

main text
result
end matter
index
popular report
oral presentations
visual devices
abstract terminology and
research report

interpretation
interpretation
beginning
interpretation
interpretation
error
reliability
outline
writing final draft
research report

formulating hypothesis
bibiliography
popular report
short report
end of the research report
descriptive
research methodology
inferential statistics
snowball
triangulation
Kruskal Wallis test