UTKAL UNIVERSITY
NAAC ACCREDITED A+

MASTER OF SOCIAL WORK

SEMESTER-I

MSW-1.4: Social Work Research and Statistics

CREDIT: 04

BLOCK: 1, 2, 3, 4 & 5

AUTHOR

Dr. P. K. Kar

C.D.O.E

ଦୂର ଓ ଅନ୍‌ଲାଇନ ଶିକ୍ଷା କେନ୍ଦ୍ର, ଉତ୍କଳ ବିଶ୍ୱବିଦ୍ୟାଳୟ

CENTRE FOR DISTANCE AND ONLINE EDUCATION

UTKAL UNIVERSITY

C.D.O.E.
Education For All

ABOUT THE UNIVERSITY

Founded in 1943, Utkal University is the 17th University of the country and the first of Orissa. It is the result of the efforts of Pandit Nilakantha Dash, Maharaja Krushna Chandra Gajapati, Pandit Godavarish Mishra and many others who envisioned a progressive education system for modern Odisha.

The University started functioning on 27 November 1943, at Ravenshaw College, Cuttack. It originated as an affiliating and examining body but shifted to its present campus spread over 400 acres of land at Vanivihar in Bhubaneswar, in 1962.

A number of Postgraduate Departments and other centres were established in the University campus. There are presently more than two hundred general affiliated colleges under the University. It has eleven autonomous colleges under its jurisdiction, twenty-eight constituent postgraduate departments, 2 constituent law colleges and a Directorate of Distance & Continuing Education. It boasts of a centre for Population Studies, a School of Women's Studies, an Academic Staff College, a pre-school and a high school. The University also offers a number of self-financing courses.

NAAC accredited in its 3rd cycle with A+ status in 2023. It is a member of the Indian Association of Universities and the Commonwealth Association of Universities.

C.D.O.E.
Education For All

CENTRE FOR DISTANCE & ONLINE EDUCATIONUTKAL UNIVERSITY: VANI VIHAR BHUBANESWAR:-751007

From the Director's Desk

The Centre for Distance and Online Education, originally established as the University Evening College way back in 1962 has travelled a long way in the last 52 years. **'EDUCATION FOR ALL'** is our motto. Increasingly the Open and Distance Learning institutions are aspiring to provide education for anyone, anytime and anywhere. CDOE, Utkal University has been constantly striving to rise up to the challenges of Open Distance Learning system. Nearly one lakh students have passed through the portals of this great temple of learning. We may not have numerous great tales of outstanding academic achievements but we have great tales of success in life, of recovering lost opportunities, tremendous satisfaction in life, turning points in career and those who feel that without us they would not be where they are today. There are also flashes when our students figure in best ten in their honours subjects. Our students must be free from despair and negative attitude. They must be enthusiastic, full of energy and confident of their future. To meet the needs of quality enhancement and to address the quality concerns of our stake holders over the years, we are switching over to self instructional material printed courseware. We are sure that students would go beyond the course ware provided by us. We are aware that most of you are working and have also family responsibility. Please remember that only a busy person has time for everything and a lazy person has none. We are sure, that you will be able to chalk out a well planned programme to study the courseware. By choosing to pursue a course in distance mode, you have made a commitment for self improvement and acquiring higher educational qualification. You should rise up to your commitment. Every student must go beyond the standard books and self instructional course material. You should read number of books and use ICT learning resources like the internet, television and radio programmes etc. As only limited number of classes will be held, a student should come to the personal contact programme well prepared. The PCP should be used for clarification of doubt and counseling. This can only happen if you read the course material before PCP. You can always mail your feedback on the course ware to us. It is very important that one should discuss the contents of the course materials with other fellow learners.

We wish you happy reading.

DIRECTOR

PAPER-4 SOCIAL WOK RESEARCH AND STATISTICS

Brief Content

Block No.	Block	Unit No.	Unit
1	Social Work Research	1	Meaning, definition, purpose of research,
		2	Scientific Method
		3	Variable and its types.
		4	Hypothesis
Block No.	**Block**	**Unit No.**	**Unit**
2	Research Design and Sampling	5	Research design.
		6	Formulation of research problem
		7	Sampling
		8	Measurement: Meaning, levels of measurement
Block No.	**Block**	**Unit No.**	**Unit**
3	Sources & Methods of Data Collection	9	Sources of data collection
		10	Research tools and Methods
		11	Advantages and disadvantages.
		12	Electronic resource and its usage.
Block No.	**Block**	**Unit No.**	**Unit**
4	Preparation of Research Proposal	13	Financial, time and personnel budgeting; processing; and analysis coding scheme.
		14	Tabulation and Diagrammatic representation of data
		15	Report writing and referencing; Agencies involved in social research;
		16	Ethical considerations of social work research; limitations of research.
Block No.	**Block**	**Unit No.**	**Unit**
5	Social Statistics	17	Meaning, use and its limitations in social work research
		18	Measures of central tendency: Arithmetic mean, median and mode.
		19	Dispersion: range, quartile deviation, standard deviation and co-efficient of variation
		20	Tests of significance: T test and chi-square test Computer Applications

CENTRE FOR DISTANCE AND ONLINE EDUCATION, UTKAL UNIVERSITY, BHUBANESWAR

Programme name: Master of Social Work (MSW) **Programme Code: 010326**

Course Name: Social Work Research and Statistics **Course code: MSW 1.4**

Semester: i **Credit: 4** **Block no. 1to 5** **Unit no. 1to 20**

EXPERT COMMITTEE

Director, CDOE, Utkal University

Dr. Gayatri Pattnaik, Utkal Univeristy

Dr. Jharana Mishra, Utkal Univeristy

Dr. Bilash Chandra Montry, Utkal Univeristy of Culture

Dr. Bijalaxmi Sahoo, Utkal Univeristy

COURSE WRITER

Dr. P. K. Kar , Utkal University

COURSE EDITORS

Dr. Bilash Chandra Montry, Utkal Univeristy of Culture

Dr. Bijalaxmi Sahoo, Utkal Univeristy

Dr. Renuka Kumar, CDOE, Utkal University

Published by:

Center for Distance and Online Education
Utkal University, Vanivihar
Bhubaneswar- 751007
Odisha

MSW 1.4

SOCIAL WORK RESEARCH AND STATISTICS

Content

BLOCK-01- SOCIAL WORK RESEARCH

Unit 01- Meaning, definition, purpose of research,

.**Unit 02-** Scientific Method

Unit 03- Variable and its types

Unit 04- Hypothesis

BLOCK-02- RESEARCH DESIGN AND SAMPLING

Unit 05- Research design: Concept and meaning

Unit 06- Formulation of research problem

Unit 07- Sampling

Unit 08- Measurement: Meaning, levels of measurement

BLOCK-03- SOURCES & METHODS OF DATA COLLECTION

Unit 09- Sources of data collection

Unit 10- Research tools and Methods

Unit 11- Advantages and disadvantages

Unit 12- Electronic resource and its usage.

BLOCK-04- PREPARATION OF RESEARCH PROPOSAL

Unit 13- Financial, time and personnel budgeting; processing; and analysis coding scheme

Unit 14- Tabulation and Diagrammatic representation of data

Unit 15- Report writing and referencing; Agencies involved in social research;

Unit 16- Ethical considerations of social work research; limitations of research

BLOCK-05- SOCIAL STATISTICS

Unit 17- Meaning, use and its limitations in social work research

Unit 18- Measures of central tendency: Arithmetic mean, median and mode

Unit 19- Dispersion: range, quartile deviation, standard deviation and co-efficient of variation

Unit 20- Tests of significance: T test and chi-square test, Computer Applications: Use and application of computer in Social Work Researchwith special to statistical package for social science (SPSS)

UNIT-1

UNIT-1

After learning this Unit students will be able

1.0. OBJECTIVES

-To know the meaning, definition and purpose of social work research.

-To understand the concept of social research and its relationship with social work research.

-To understand the meaning nature and characteristics of scientific method.

-To cognize the purpose and steps in research process.

- To know about the concepts and how they are operationalised

-To understand the variables and their types.

1.1. Meaning definition and purpose of Social Work Research

Any study to create new knowledge or to add to the existing store house of knowledge, may it be through observation or by any other methods is called research. Social work research is the application of research methods to solve problems that social workers confront in the practice of social work. It aims at providing information that can be taken into consideration by social workers provide to making decisions which are likely to affect their clients or schemes of things or agencies which are involved in the use of alternative intervention techniques or transformation or modification of client / programme / objectives etc.

The circumstances under which the application of the methods and techniques of social work research are warranted are the following.

1. Interest in the investigation or assessment of the nature and extent of the problems of client in terms of obtaining data regarding the actual or potential effectiveness of the client. For example, the social workers may be interested in cognizing as to what extent the intervention would be effective in solving client's problem of martial maladjustment
2. Desire to assess the effectiveness of the techniques of role play in comparison with group discussion. For example, interesting knowledge of drug abuse among school going children.
3. Desire to assess the opinion of members of community before taking decision to change the programme or purpose of the community organizer.

4. The anxiety to examine the effectiveness or group there by for mentally retarded children.
5. Desire of social work administers regarding the effectiveness of her programmes lunched

1.1.1. Definition of Social Work Research:

Social work research may be defined as "the study of the relationship of social workers with their clients, such as the individuals, groups or communities on various levels of interaction or thereby as well as their natural relationships and functioning within the organizational structure of social agencies. It may be briefly defined as systematic investigation into the problems in the field or social work.

From the theoretical point of view, social work research re-examines the special body or knowledge, concepts and theories. But from the utilitarian or practical view point, it tries to evolve a systematize theory and valid concepts, so as to understand the efficacy of different methods / interventions of social work to find out innovative interventions or alternate treatments.

Therefore, it can be maintained that social work research is concerned with the problems encountered by social workers. It also encompasses these complex questions which comfort the social work practitioners or planners or administrators of social work services whose solution remains in research and which are amenable to investigation.

Moreover, one finds hardly any significant difference between the utilization of the scientific methods and techniques as used by social research and social work research, notwithstanding the fact that when some research designs or procedures of social research are not suitable to social work research, it would appear to have been essential to develop the tools and techniques appropriate to social work research.

1.1.2. Purpose of Social Work Research:

As social work is a practice profession, the major purpose of social work research is to search for answers relating to interventions or practice effectiveness. To speak little

differently, the question which needs be answered through social work research is: what interventions or treatments really help or hinder the attainment of social work group objectives? That apart, it also helps in seeking answers for surmounting the difficulties faced by the practitioners of social work in the practice of their profession, Furthermore, it provides the knowledge for social work for building theory and for practice also. Another major purpose of social work has always been to focus on assessment of practitioner's work with individuals, groups, families, communities or appraisal of agencies or programmes which are concerned with the continued efforts of practitioners with many clients. A part from the aforesaid purpose, a special purpose of social work research is evaluation. In fact, this is the reason that designates social work research as evaluative research. The varieties of evaluative researches undertaken in social work research are based on impacts or effects, efficacy and effectiveness. The core of evaluative research under the rubric of social work research lies in evaluation of agencies and its projects and programmes.

For the social work professionals, the purpose of social work research is to know the problems faced by professional social workers in social work agencies and communities in its concern with social work functions.

The purpose of social work research may briefly be state below:

1. To attempt at understanding through the studies to establish, identity and measure the need for service.
2. To measure the services rendered as they relate to needs.
3. To examine, measure and evaluate the consequences of social work intervention.
4. To assess the efficacy of specific techniques of offering services.
5. To study the methodological aspects of social work.

1.2 SOCIAL RESEARCH and SOCIAL WORK RESEARCH

Search implies thorough investigation and the term 'research' which has been derived from the French word 'rechercher', 're' and 'chercher', means a critical examination of a topic or subject to discover new facts for increasing the sum total of human knowledge. It is a method for discovery of new knowledge which augments to the existing body of organized

facts, ideals and aspiration, "Research is considered to be the more formal, systematic, intensive process of carrying on the scientific method of analysis. It involves a more systematic structure of investigation, usually resulting in some sort of formal record of procedures and a report of results or conclusions". Research per se constitutes a method for the discovery of truth which necessitates critical thinking. "It comprises defining and redefining problems; formulating hypothesis or suggested solutions, collecting, organizing, and evaluating data; making deductions and reaching conclusions; and at least, carefully testing the conclusions to determine whether they fit the formulated hypotheses."

"It is the manipulation of things, concepts or symbols for the purpose of generalizing to extend, correct or verify knowledge, whether that knowledge aids in construction of theory or in the practice of an art." Thus, research is a "systemized effort to, gain new knowledge."

Research is characterised by : (i) a specific problem. Hi) involvement in original work, (Ui) resting upon a mental attitude of curiosity, (iv) requirement of an open mind, (v) resting upon the assumption that everything is subject to law and order, (vi) discovery of laws and generalizations (vii) study of cause and effect (vii) measurement and (ix) involvement in a conscious technique.

Social research implies discovery of some facts concealed in a social phenomenon or some laws governing it. It is mainly concerned with the cause and effect relationship of human behaviour and the discovery of new facts as well as the verification of old facts. Therefore, "we may define social research as the systematic method of discovering new facts or verifying old facts, their sequences, interrelationships, causal explanations and the natural laws which govern them." While studying human behaviour and social problems and discovering new interrelations, new knowledge, new facts and verifying old ones, social research applies the scientific method and tries to establish the causal connection between various human activities and the natural laws governing them by means of logical and systematized methods because the human behaviour may be motivated by certain rules and laws and does not appear haphazardly. Briefly stated, social research implies scientific investigation conducted in the field of social as well as behavioral sciences.

Social research has many **objectives** which may be discussed below:

1. **Manipulation of things, concepts and symbols** :

While dealing with things the scientist remains at the concrete level. He is able to purposefully handle things for experimentation. But at this level his results are at best

limited to the particular thing in a specific situation and none else. Therefore the concepts symbolizing the things and their properties are also dealt with, so as to make much sense to conduct controlled inquiries through abstract notions. Use of concepts or symbols in the process of manipulation not only reduces the content and load of the things but also provides the scientist with greater facility and effect.

2. Generalization: The sole purpose with which manipulation of things, concepts or symbols is undertaken is to arrive at statements of generality. It implies that the findings of controlled investigation should be a conclusion which "will enable us to expect that under certain class of conditions influencing a class of things, something will happen in a generalized manner, notwithstanding its degree. But in any case the absence in generality cannot characterise science. Therefore the propositions derived on the basis of observations and through manipulation of things, concepts or symbols may vary in their levels of generality, may maintain a high or low degree but should never reach the null point. Otherwise those will move beyond the framework of science. In this regard, Slesinger and Stephenson have given the example of a physician or automobile mechanic as playing the role of a researcher. Whereas the automobile mechanic endeavors to generalize about the automobiles, the physician attempts to make ailments for a given class of patients.

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

4. Extension of Knowledge: As a sequel to generalization the seemingly inconsistencies in the existing corpus of knowledge are brought into light and attempts are made to reconcile these inconsistencies. The new general proposition, established as an outcome of research also identifies gaps in the established system of knowledge. A gap in knowledge implies the inadequacy of the theory as well as the failure of a conceptual scheme to explain and account for certain aspects of a social phenomenon. The gap is bridged up in the light of the new empirical observations. Thus knowledge gets expanded. The expansion

of systematic knowledge occurs at least in a couple of ways. First in cognizing certain aspects of phenomena which were not examined in these terms prior to the advent of the new general proposition. Secondly in the light of new observation, the phenomena under investigation may be incorporated in a comparatively large class of phenomena, so as to be governed by a uniform law. As a result, the new system of knowledge not only accumulates more units under its conceptual scheme, but also appreciates greater depth of understanding and bettering of predictions.

5. Knowledge may be used for theory building or practical application : By seeking to explain the unexplained social phenomena, clarifying the doubtful one and correcting the misconceived facts relating to it, social research provides the scope to use the fruits of research in two possible ways : (a) theory building (.6) practical application. In its basic or pure form social research gathers knowledge for the sake of it, for building a theory in order to explain human behaviour in its totality, only for the satisfaction of knowing. For construction of theoretic models, the researcher organizes knowledge into propositions and then meaningfully articulated those propositions to constitute a more abstract conceptual system pertaining to a class of phenomena, influenced by a certain class of conditions.

In its practical or applied form, social research gathers information regarding the betterment of quality of life in social settings. The findings of social research are used as the means to an end, not construed just as an end in itself. From its utilitarian point of view the results of social research provide decision makers with proper guidelines for policy making, social welfare, amelioration of practical problems, mitigation or resolution of social conflict and tensions as well as rectification and removal of social evils.

It is generally believed that social research and social work research do not have much difference as the purpose of promoting the welfare of the humanity through investigation remains common to both. Whereas the social work research commences with practical problems, social research aims at producing such knowledge that can be of use in planning and executing the social work programmes .Social research also has the objective of accumulating knowledge for understanding the social life of human beings. The social work research is an applied research which has the purpose of gaining knowledge in order to control or change human behaviour. On the other hand, social research may have practical as well as theoretical concern. Social work research serves the objectives of social work. On the contrary, social research does not have any specific goal. The main objective is to enhance in the knowledge of any social science. Moreover the social work research renders helps to the social workers for dealing which social problems relating to their client which

may be afflicting either individual or the group or community. Social research may be of use to social worker as well as entire field of social work as it helps in enhancing the knowledge of dealing with and understating human behaviour.

1.3 Scientific Method:

1.3.1. Nature

Scientific method is "an objective, logical and systematic method of analysis of phenomena, devised to permit the accumulation of reliable knowledge". . It is characterized by an intellectual attitude and is not confined to a field of specific subject matter, but rather to- a procedure or mode of investigation. As Karl Pearson emphasizes, "the scientific method is one and the same in all branches and the method of all logically trained minds the unity of all sciences consists alone in their method, not in their materials". Since facts may be related to any field, the facts themselves do not make the science, but by the method they are dealing with. Thus, science is free from any particular subject matter. It is not wrapped with any particular body of facts.4 It deals with physical as well as psychical process by taking into account the entire knowable universe for its subject. It deals with man as well as nature, bringing into its fold everything to which its method can be extended. Hence a science is not made by the nature of things with which it is concerned but by the method through which the things are dealt with. As a systematic step by step procedure following the logical processes of reasoning,' scientific method aims at gaining knowledge of the universe. It never belongs to a particular body of knowledge. George Lundberg considers scientific method as one consisting of systematic observation, classification, and interpretation of data. The main difference between day-to-day generalization and the scientific method lies in the degree of formality, rigorousness, verifiability and the general validity of the latter. According to Wolfe, "any mode of investigation by which science has been built up and is being developed is entitled to be called scientific method." Scientific method is marked, says Karl Pearson, "by three features, viz., careful and accurate classification of facts and observation of their correlation and sequence; discovery of Scientific Laws with the aid of creative imagination; and self-criticism."

1.3.2. Characteristics of Scientific Method

According to Cohen and Nagel, the first characteristic feature of scientific method is its nature of 'non-imposition', implying that the method aims only at discovering facts as they actually are and not as they 'ought to be'. Secondly, scientific method deals with some specific issues. The scientific investigation is completed when the "felt-problems" are solved. Thirdly, it is believed that scientific method follows a path of systematic doubt and is ever prepared to discard any theory when the established facts so demand. Fourthly, the scientific enquiry follows a circular path as it not only proceeds from weighing evidence, appraisal of facts to experience but also from experience to facts. Being more particular about the consistency of the method followed rather than the result likely to be obtained, scientific method claims rationality. Its widespread desire for truth suspends all value judgments. Fifthly, scientific enquiry seeks 'verification and proof which is made possible through the combinations of relevant observation and logical verification of the phenomenon. David Easton has also laid down certain assumptions of scientific method like : 00 Regularities, (ii) Verification, (Hi) Techniques (iv) Quantification, (v) Values, (vi) Systematisation, {vii) Pure Science, and (viii) Integration. Wilkinson and Bhandarkar have identified certain 'articles of faith' upon which the scientific method is based. These are : (i) Reliance on empirical evidence, (ii) Use of relevant concepts, (Hi) Commitment to objectivity, (iv) Ethical neutrality, (v) Generality (vi) Predictions based on probabilities, and (vi) Public methodology affording testing of conclusions through replication.

Out of the above discussion, we may deduce the following salient principles of scientific method : (i) Regularities (ii) Empiricism (Hi) Use of concepts (iv) Verifiability (v) Objectivity (vi) Ethical neutrality (vii) Generality (viii) Predictability (ix)Relativism (x) Skepticism (xi) Quantification (xii) Systematization and (xiii) Public methodology.

1. Regularities:

   It is believed in the scientific method that the phenomena occur in the universe in a regular and patterned manner. It is the task of science to ascertain these patterns in the natural world. A scientific study must be made public by making known to others as to how the conclusions are reached. Different individuals can investigate independents and are most likely to arrive at the same conclusion. There is nothing secret or personal about it because science is a collective, cooperative endeavor geared to the discovery of facts and unless the methodology of scientific enquiry is made public, it would not enable the fellow scientists or critics to replicate the initial enquiry

for verification. Repeated replications strengthen the conclusions and lend added evidence to it. At the same time, it also ensures that mistakes, if any, in the initial enquiry are not only repeated, but also removed in the process. Therefore Dewey says that it is "a method of knowing which is self-corrective in operation, that learns from failures as from successes." Modern science, in contrast to ancient science, has developed through the exposure of its methods and conclusions to critical scrutiny. Criticism has always been the very life-blood of science, says Karl Pearson.

2. Empiricism :

Empiricism implies that ascientific investigation must be conducted empirically. In other words, our views about some or the other aspect of society must be based on clear and definite factual evidence. Truth is established on the basis of evidence. Conclusion is admitted when it is based on evidence. Such evidence must be produced by observing the relevant social reality with the help of human senses, such as sight, hearing, taste, smell and touch. Relevant data are gathered by observation and experimentation. Nothing is left td speculation. Social reality may, at present, be observed directly with the help of human senses or indirectly with the support of some instruments so as to aid and extend the ability to observe. The validity and reliability of data are thoroughly checked and carefully analyzed by employing appropriate methods. On the basis of the findings of analysis, conclusion is reached.

3. Use of Concepts:

Concepts are the building blocks of theory. A fact is a logical construct of concepts. A concept abstracted from the sense perceptions and should not be confused with the phenomenon itself. Since the ordinary language fails to convey adequately the implications of scientific terms, science evolves its own language. These linguistic apparatus of science are manipulated along with symbols with a view to contribute immensely to the established body of systematic knowledge. The scientist constantly depends upon relevant concepts for moving gradually from concrete sense data to the higher levels of abstraction.

4. Verifiability :

Verifiability presupposes that the phenomena must be capable of being observed and measured. Scientific method presupposes that knowledge in order to be valid, should consist of propositions amenable to empiricism. All evidence must be based on observation. Science, being empirical, claims that knowledge must be referred to concrete human experiences so as to make verification possible. Lundberg believes that, "if the verification of deduction involves condition of observation which is impracticable or impossible of attainment the theory is metaphysical rather than scientific." In order to bring greater exactitude, verification must also be accompanied by measurement.

5. Objectivity :

By objectivity it is meant that the scientific investigation must not be influenced by the subjective biases of the investigator. Rather the phenomena are observed in its true form. The man of science is committed to the belief that to go nearer to the goal of truth, he must above all things, believe that the phenomenon world is a reality, independent of beliefs, hopes or fear or fears of any individual, all of which we find out not by intuition and speculation but by actual observation.[8]

According to Lundberg, "the first requisite of all sound knowledge is the determination and ability to get at naked facts and not to be influenced by mere appearances or by prevalent notions or by one's wishes' Objectivity is the hall mark of scientific method. Green visualizes objectivity as "the willingness and ability to examine evidence dispassionately." The main criterion of objectivity is that the conclusion should not vary from person to person, all person should reach the same conclusion. The scientific man must above all things, have a detached view as the phenomena in which he himself is involved as on observer.

In the opinion of J. Galtung, objectivity is the composite of 'intra-subjectivity', and 'inter subjectivity.' Testing of intra-subjectivity pre-supposes that repeated observation of phenomenon by the same observer will produce the constant data. On the other hand the

test of inter-subjectivity presupposes that repeated observation of a constant phenomenon by different observers will always provide them with constant data.

As the very purpose of science is to find out the naked truth, objectivity is fundamental to all sciences and essential for verification. In the words of Lundberg, "it permits repetition of observation under practically identical conditions. This facilitates the verification of observation by many observers."[1] Although objectivity apparently appears to be very easy, in real terms, it is very difficult to be achieved. Personal views, concepts and beliefs of the investigator do influence his study. Hence, scientific man must "above all things....strive at self-elimination in his judgment and provide an argument which is as true for each individual mind as his own."

6.Ethical Neutrality:

Scientific method demands that the investigator maintains an ethically neutral attitude in his pursuit of knowledge. Science never passes normal judgment on facts by designating them as good and bad. In his professional capacity, the man of science is not supposed to take sides on issues jof moral or ethical nature. Scientific method reserves science on normative questions. As Schrodinger says, "Science never imposes anything, science states. Science aims at nothing but making true and adequate statements about its objects."

7.Generality :

Principles evolved through scientific method are universal. The conclusions drawn through scientific investigation apply to all cases and all circumstances. The conclusions are not affected by the factors of time and space. In the words of Maclver, "Such a law is simply another name for a carefully described and uniformly recurring sequence of conditions." The scientist is constantly and necessarily obliged to discover "under the surface level of diversity the thread of uniformity."13 The primary aim of science is to trace order in nature. To this end, science seeks to ascertain the common characteristics of types of objects and general laws or condition of events,"14 Scientific principles hold true irrespective of the temporal and spatial order. "Science is not interested in individual objects or individual groups of objects as such" However, various branches of science do not attain the same level of generalization. The degree of maturity of science is directly proportional to its generalizing potential,

8. Predictability:

Science can make prediction by its logical reasoning and inferences establishing the cause and effect relationship among different phenomena. The foundation of science is based upon a faith in causality that the past and future belong to the same continuum. Based on the "law of uniformity of nature" stating that the nature will behave similarly under similar conditions, science believes that predictions about phenomena must rest on the bed-rock of the trend repeatedly observed. It also believes that probably the same trend would manifest itself in some concrete effects. Predictability depends upon two essential conditions, such as the fixity of cause and effect relationship and the stability of causative factors. Prediction in the domain of science is grounded in the established knowledge concerning order among facts.

However, the scientific expectation may not always be accurate. Science can only make prediction about the state of things on the basis of the law of causation and law of uniformity of nature with certain degree of accuracy. "Scientific knowledge is a body of statements of varying degrees of certainty, some most unsure, some nearly sure, some absolutely uncertain."

9. Relativism;

Relativism implies that the results obtained through scientific method are never considered as absolute truths. Propositions found valid in the light of scientific method under certain circumstances may be questioned on the face of new evidence. Results of scientific investigation are only tentative and never considered as permanent. They have got relative credibility as a proposition is considered valid so long as it is not refuted in the future. Relativism as a principle of scientific method further holds that no notions are sacred to the scientist, no propositions are privileged to the researcher or no truths are absolute to an investigator.

10. Skepticism:

Skepticism is that principle of science which holds that the scientist must possess the capacity to view skeptically the validity of prevailing social theories. He must not be swayed away by the popularity of a given proposition despite its general acceptance. The scientist is free to be skeptical of any explanation on the ground that they not only lack in authenticity but also sufficiency of evidence. "

11. Quantification:

All observations in the domain of science must be quantified for precision. All sets of verified generalizations which form the basis for scientific inquiry are to be accepted in mathematical language.

12. Systematization:

While dealing with the empirical truths and analysing the true nature of these empirical objects, science follows a systematic and formal method. Such a rigorous method of analysis and generalisation enables the votaries of science to re-examine the results in different occasions. In the realm of science a couple of methods is widely prevalent- inductive and deductive. In inductive method, the particular truths are gathered gradually and continually in empirical situation till the most general truth is established. On the contrary, in deductive methods the truths of the propositions are not questioned, conclusions are drawn from those self-evident propositions. Thus induction proceeds from particular to general and the reverse process is evinced in the deductive method for the discovery of the truth that lies concealed within a set of statements. For the extreme deductionists, a set of self evident propositions stand at the head of the system and from these other theorems are to be derived by the process of reasoning. On the other hand the extreme inductionist or empiricist's view of the matter is that science must construct its axioms from sense data moving from particulars to arrive at the most general axioms in a gradual and continual manner. The true method of science is induction for deduction with a view to reconstructions because it borrows the elements of' formal truth' and 'material truth' from deduction and induction s respectively and thereafter applies logical reasoning for establishing its own truth. Larabee holds in this regard, "If extreme rationalist (deductionist) is like a spider spinning out theories from within, the extreme empiricist (inductionist) is to be compared to an ant which piles useless heaps of facts. Better than either the spider or the ant is the

bee, which_ selectively gathers pollen and transforms it into honey....

13. Public Methodology : The method used in scientific inquiry is always made public because science is a public institution and a collective, co-operative endeavor aiming at the discovery of facts. Science not only learns from successes, but also from failures as it is a method of knowing which is self-corrective in operation." Public methodology enables the fellow scientists to replicate the initial inquiry so as to lend added credence and support to conclusions. The basic requirement of science being frequentative verification of conclusions, repetitive searches are to be conducted which may lend confirmation to the prevalent corpus of science, help intend necessary modifications in it or even reject it.

1.3.3 Problems of the Social Sciences

Although the social sciences are firmly inclined to adopt the scientific view point like the physical sciences, those cannot claim the same degree of success in an analogous manner. In the past century, the knowledge of social behaviour accumulated by the social sciences is much broader and deeper than ever before. A generation ago, the American sociologists were caught in the controversy about how scientific are the Social Sciences. A majority of sociologists, notwithstanding the difficulties of direct observation of opinions and sentiments and their conclusive verification, failed to exclude the subjective phenomena. George Lundberg went to the extent of claiming that there were no essential differences between social behaviour and physical phenomena and one could be as readily objectified as the other. He also believed that sociological research methodology should be analogous to that of physics, embracing experimentation and exact measurement. P.W. Bridgemen, an eminent physicist expressed somewhat similar view by proposing that a scientifically minded man should ignore those elements of man's interior life which are not subject to observation, the oldest method of science.

The social scientists, in their endeavor for developing the methods of social research, have borrowed extensively from the physical sciences, but they have not ruled out necessary alterations to cope with their subject matter. Simultaneously, keeping in touch with changing times, attempts are on to improve these methods and standardize the tool and techniques of research to attain the accuracy at par with the physical sciences. Nevertheless, it remains a difficult task to demonstrate that they have come closer to the physical sciences in terms of their methods[21] due to the reasons that whereas the physical or natural

1.4. PURPOSE AND STEPS IN RESEARCH PROCESS

Research process involves a series of steps progressing from the formulation of research problem through survey of literature, formulation of hypothesis, preparation of research design, determination of sample design, collection of data, execution of the project, analysis and interpretation of data and testing of hypothesis to presentation of the conclusions. Such a series of actions, concerning a lot of shifts from one kind of activity to the other do not, however, reflect a singular concentration on any particular step at any point of time. Rather, they are required to be effectively carried out in the desired sequence at different points on specific operations, with the anticipation at each step the requirement of the subsequent steps in the process of research.

Although certain steps are designated as the major ones, those are only the groupings of classes of operations out of hundreds, each contributing to social research in its own way. Similarly the sequential and consecutive steps never imply that the process of research corresponds to their neat execution as each of these steps are involved in a set of inter-related operations and each one of these operations will determine the quality of research. Because of their effect on the quality of research, a small operational lapse anywhere may have adverse effect on the standard of investigation.

It is visualized that the research process travels a circular path because the findings of research studies may most often generate fresh ideas and problems which need further investigation. Furthermore, the aforesaid major steps may overlap in time sequence and their activities may be functionally interrelated, involving both backward as well as forward linkages. Whereas backward linkage implies substantial impact of the subsequent steps on the preceding steps in the process of research, the forward linkage indicates the tremendous impact of the preceding steps on the structure of the following steps. Practically, the major steps involving their corresponding activities may overlap continuously instead of adhering to a strictly prescribed sequence, because the researcher should always bear it in mind that the different steps involved the process of research are neither mutually exclusive nor separate and distinct. In such a situation, it is but quite natural that the researcher will have to constantly expect at each step in the process of research the exigency of the subsequent steps. Generally the following order is maintained as a useful procedural guideline in the sequence of the major steps of social research:

1. Formulation of Research Problem.

2. Review of Literature.
3. Formulation of Hypotheses.
4. Working out the Research Design.
5. Defining the Universe of Study.
6. Determining Sample Design.
7. Administering the Tools of Data Collection.
8. Analysis of Data.
9. Testing of Hypotheses.
10. Generalisation and Interpretation.
11. Reporting the Research.

In real terms research begins with a problem which needs solution. Such a perception on the part of the researcher, first of all, falls within the general area of interest indicating to locate the problem either in the quest of some intellectual pursuit or for some practical concern, such as finding out a practical solution to a problem, evaluation of a programme in the light of new facts, gathering of relevant facts for social planning or even for policy making. While selecting a problem for research, the Social Scientists are likely to be influenced by their own personal values as well as the prevalent social conditions. As scientists differ with regard to their values and societies differ in respect of their preference of different spheres, the choice of topics in social research vary widely.

As the general topic fails to provide the ability to examine the relevance of data, adopt the methods or to organize them, the need for formulation of a specific problem is always felt. This makes the goal of the researcher clear-cut. It not only guides the researcher in exploration, but also progressively sharpens the focus of questions by narrowing down the coverage like a pinpoint. For example, if the general topic is compared with the base of a pyramid, the specific topic may resemble the apex of it.

In any case, formulation of a problem, arising out of theoretical situation or practical concern, is not an easy task, as it appears to be. In real terms it is a uphill task, so much so that even a scientist of the stature of Charles Darwin has gone to the length of saying that "Looking back, I think it was more difficult to see what the problems were than to solve them." As a problem involves some difficulty experienced by the investigator, the formulation of the problem should make its various components explicit in such a manner that it will justify the saying that "a problem well put is half solved" R.K. Merton has identified three important questions as the three principal components involved in the process of formulation

of a problem in the field of research in soft sciences : (i) What one want's to know ?, (ii) Why one wants to seek answers to those particular questions? and (Hi) What may be the possible answers to the originating questions ? These three questions correspond to the components of the originating questions, the rationale and the specifying questions respectively.

There are at least three types of originating questions: (t) originating questions calling for discovering a particular body of social facts, (ii) originating questions directing attention to the research for uniformities between classes of variables, and (Hi) questions addressing to a variety of institutional spheres. As regards the rationale component in the progressive formulation of a problem, the statement of reasons pertaining to the tenability of a question is made. It also seeks to justify the answers' contribution to theoretical or practical concerns. The fundamental requirement of a rationale is to widen the base of scientifically consequential question and to eschew the scientifically trivial ones. R.K.. Merton holds the view that "the rationale states the case for question in the court of scientific opinion."

The theoretical rationale attempts to justify the contribution, likely to be made by the answers to questions, in terms of the enlargement of the scope of the prevalent ideas or concepts or theory. It may also throw light on the observed inconsistencies in existing ideas and examine the nature of inconsistencies in terms of its spuriousness or reality. On the other hand, the practical rationale acts as a pointer to justify as to how the answers to the research questions will bring about certain desired practical values. However, a question raised for the practical concern may also have its bearing on the theoretical system.

The component of specifying questions in the process of formulating a research problem aims at transforming the originating questions into a series of observations in a particular concrete situation, necessitating collection of empirical data, so as to seek possible answers to the originating questions in terms that satisfy the rationale fruitfully. Review of Related Literature

Since an "effective research is based on past knowledge, an investigator must always take advantage of the knowledge which have been preserved or accumulated earlier. It not only helps the researcher in avoiding duplication and formulating useful hypothesis, but also provides him with the evidence that he is familiar with what is already known and what is still unknown and untested in the field.

Review of related literature implies the analysis of the summary of the writings of recognized authorities and the earlier researches in the particular area. According to J.W. Best "Practically all human knowledge can be found in books and libraries. Unlike other

animals...man builds upon the accumulated and recorded knowledge of the past. In the words of C.V. Good "the keys to the vast store house of published literature may open doors to sources of significant problems and explanatory hypothesis and provide helpful orientation for definition of the problem, background for selection of procedure and comparative data for interpretation of results."

The essentiality of the review of literature remains in the fact that it provides the researcher with an indication of the direction, updates information relating to researcher's own problem, avoids the replication of the study of findings, provides the scope for analogy and formulation of hypothesis.

The main objectives of review of related literature are : {i) to provide theories, ideas, explanations or hypotheses, which are likely to be helpful in the formulation of research problem, (ii) to avoid overlapping studies (Hi) to be a fertile source for formulating hypothesis, (iv) to suggest the methods of data collection, procedures, to explore the sources of data and statistical techniques appropriate to the solution of the problem, (u) to gather comparative data and findings of earlier researches which are of use in the interpretation of data and analysis of results, (vi) to enable the investigator to gain expertise in his area of interest, and (vii) to keep the researcher abreast of the most recent development in his area of activity.

1.5 : CONCEPTS AND OPERATIONALISATION OF CONCEPTS:

Concepts are the building blocks of theory. A fact is a logical construct of concepts. A concept is abstracted from the sense perceptions and should not be confused with the phenomenon itself. Since the ordinary language fails to convey adequately the implications of scientific terms, science evolves its own language. These linguistic apparatus of science are manipulated along with symbols with a view to contribute immensely to the established body of systematic knowledge. The scientist constantly depends upon relevant concepts for moving gradually from concrete sense data to the higher levels of abstraction. Thus concepts are mental images and we assign terms to these images so that we can communicate effectively.

As regards operationalisation of concept, the researcher is concerned with spelling out exactly and precisely how the concept will be measured. He will have to describe the operations that will be undertaken in measuring the concept. The process of developing operational definition is not necessarily linear, or step by step in an ordered manner. When attempting to operationalize, the researcher may have to modify his conceptual definition or when

attempting to select a data collection method, his operationalisation may be affected. This is a circular process.

1.6 VARIABLES AND ITS TYPES

A variable is anything that is subject to change. It is a concept which can take on different qualitative or quantitative values. The concepts like income, height, weight are the examples of variables. When we simply say lower income or higher income, large size or small size, we treat the property "income" or "size" or an individual or landed property respectively qualitatively. But when we refer the income of an individual as Rs. 15,000 per month or the size of the farm as 2 acres, we treat the "income" or "size" quantitatively. The phenomena which are capable of taking on quantitatively different values are called continuous variables. A continuous variable can assume any numerical value within a specific range. On the other hand the non-continuous variables are expressed in integer values. The individual values concerning such variables fall on the scale only with different gaps. For example, age is a continuous variable, contrary to the number of off-springs which is non-continuous variable.

A distinction is also made between dependent variable and independent variable In experimental research the independent variable is called manipulated variable and the dependent variable, which is the measure of some dimension of behavior is called a natural variable. The dependent variables is so called because it depends upon or is a consequence of the other variable. For example, if weight is depend upon age, the weight is a dependent variable and age is an independent variable stated otherwise, independent variables are construed to be the presumed causes and dependent variables are the presumed effects. In a study the researcher manipulates or mea the independent variable so as to determine its effect on other variables, cause the dependent variables. On the other hand the dependent variable is measure 1 ^ study how it has responded to a cause or treatment.

There may be one or more independent variables which are related to the purpose of study, but are likely affect the dependent variable. Such variables are cat extraneous variable. In a good research design, the researcher always tries to minimize the effect of extraneous variable(s) by exercising 'control'. Unless the extraneonous independent variables are controlled in an experiment, the effect can not be attributed entirely to the independent variable for the purpose of study.

If the researcher fails to restrain the experimental conditions, due to lack *a* 'control' over the extraneous variables, a confounded relationship exists between the dependent and independent variables. Such a relationship indicates the influence of* extraneous variables.

For the purpose of convenience an experimental-organismic classification is made where the experimental variables are those variables which are supposed to be active and can be manipulated by the researcher or others. On the contrary, (lie organismic variables are not usually subject to manipulation or exposure. These , variables, otherwise known as attributes, are qualities already possessed by human beings, for example parents' education, preference for political party, religious affiliation etc. However, some organismic variables like anxiety are also subject to manipulation by the experimenter, in spite of its variation from the experimental manipulation.

The variables used in multivariate analysis are also classified into different categories, such as *(i)* explanatory variable and criterion variable, *(ii)* observable variables I and latent variables, *(iii)* Discrete and pseudo variables etc.

Depending upon the external criterion or internal criterion, the explanatory criterion dichotomy is made. If X is considered to be the cause of Y, then the former is the explanatory variable and Y the latter is described as criterion variable. In other words, X is the independent or causal variable and Y is the dependent or resultant variable. In economics the relationship between X and Y, the two variables, is stated thus.

$$Y = f(X)$$

Where Y is output (Dependent variable), X is the aggregate input (Independent variable)

We gather the following type of information in the General theory of Keynesian equilibrium model :

$$Y - f(X) \qquad f$$

Where Y is the dependent variable and is represented by *(i)* Unemployment a society's resources and *(ii)* reduction of its national product. X is represented by Propensity to consume *(ii)* the marginal efficiency of capital, and the rate of interest on the supply of money and demand for liquidity.

As regards the case of other social sciences in general and sociology in particular, we study the social aspect of behaviour oriented towards other individuals, groups and

institutions and as such it is really difficult to specify variables and relations because of a large number of non-economic variables to produce simple coherent models. While specifying any particular variables, the social scientists incorporate human beings, environment and stimulus under independent variables and the response i.e. change in the human behaviour brought about by stimuli as dependent variables.

When explanatory variables are directly observable in some situations, the same are described as observable variables. On the other hand, there may be some unobservable variables which influence the criterion variable as latent variables.

The discrete variables are expressed in integer values only. On the other hand a continuous variable can assume any numerical value, any real value even in decimal points. The dummy variable or pseudo variable is useful in algebraic manipulations in multivariate analysis. Therefore, it has got a technical connotation.

1.7. HYPOTHESIS: SOURCES, FORMULATION, ATTRIBUTES AND TYPES

In order to make the problem explicit and in order to focus attention in its solution, it is essential to start with certain known theories. Research, in real terms, depends upon a continuous interplay of theory and facts, upon a continuous stimulation of facts by theory and theory by facts. Theory is initiated by facts and facts lead to the rejection or reformulation of existing theory. Facts may also redefine or clarify the theory. Hampel has compared a scientific theory to a network in which the terms and concepts are represented by knots and definitions and hypothesis by threads connecting the knots. From certain observational data we derive an interpretative string to some points in the theoretical framework. Then we proceed through definitions and hypothesis to other points from which another interpretative string permits to the plane of observation. Theory thus gives meaning to empirically observed facts and puts them systematically. Theory is also built upon facts and various facts put in a theoretically framework may be analysed and interpreted in a logical manner. Grounded on old facts and with the help of theoretical framework, new facts are discovered. In the process, certain deductions are formulated which are called hypotheses. Thus "after internalizing the problem, after turning back on experience for possible solutions, after observing relevant phenomena, the scientist may formulate a hypothesis."

"A Hypothesis is a conjectural statement, a tentative proposition about relation between two or more phenomena or variables". It is a tentative generalisation, the validity of which remains to be tested. At its initial stage, a hypothesis may be an imagined idea or a hunch or a mere guess. It is in the form of a declarative sentence and always indicates relation of one or more variable(s) with other variable(s) in a general or specific way. It is mostly based on accumulated knowledge. A hypothesis is made to examine the correct explanation of a phenomena through investigation, to observe facts on the basis of collected data. If on the basis of verification, the hypothesis is found to be valid, a theory is obtained. Thus, hypothesis a theory entertained in order to study the facts and find out the validity of the theory. The etymological -meaning of hypothesis, therefore, is a theory which is not full reasoned, derived out of the combination of two words Iiypo[1] and 'thesis' meaning 'less than' and 'reasoned theory of rational view point' respectively. Accordingly Mill defines hypothesis as "any supposition which we make (either without actual evidence or an evidence avowedly insufficient) in order to endeavor to deduce conclusions in accordance with facts which are known to be real, under the idea that if the conclusions to which the hypothesis leads are known truths, the hypothesis itself either must be or at least likely to be, true". Likewise, Goode and Hatt define it as "a proposition which can be put to test to determine validity". P.V. Young says that a hypothesis "is provisional central idea which becomes the basis for fruitful investigation, known as working theory" Coffey defines hypothesis as "an attempt at explanation : a provisional supposition made in order to explain scientifically some facts or phenomena". Hypothesis is not a theory, rather hypotheses are linked and related to theory which is more elaborate in nature as compared to hypothesis. Therefore . William H. George, while distinguishing between theory and hypothesis, described theory as 'elaborate hypothesis'. Hypothesis is not a claim of truth, but a claim for truth and hence serves as a bridge in the process of investigation which begins with a problem and ends with resolution of the problem. In the words of Cohen and Nagel "a hypothesis directs our search for the order."

1.7.1 Criteria for formulation of Hypothesis

There exist two criteria for formulation of a good hypothesis.' First, it is a statement about the relations between variables. Secondly it carries clear implications for testing the stated relations. Thus, these couple of criteria imply that the hypotheses comprise two or more variables which are measurable or potentially measurable and that they specify the

way in which they are related. A statement which fails to meet these criteria is no scientific hypothesis in the true sense of the term. However, there are legitimate hypotheses, formulated in factor analytic studies. The following examples may be cited in order to justify how the couple of criteria apply to hypotheses :

1. More intelligent persons will be less hostile than those of lower level of intelligence.
2. Group study contributes to higher grade achievement[3] In the first hypothesis, we visualise a relation stated between

one variable, 'intelligence', and another variable 'hostility/ Furthermore, measurement of these variables is also easily conceivable. In the second example, a relation has also been stated between the variables 'group study' and 'grade achievement.' There exists the possibility of the measurement of the variables are thus there is implication for testing the hypotheses. Thus both the criteria are satisfied.

1.7.2Attributes of a Valid or Usable Hypothesis

1. the most essential attribute of a valid hypothesis is that it should be capable of empirical verification, so that it has to be ultimately confirmed or refuted. Otherwise it will remain a proposition only. Therefore it should be formulated in such a way that it is possible to deduce certain inferences which in turn can be tested by observation in the field. It should not be a mere -moral judgment. As the basis of objectivity, the most essential condition of scientific method, empirical test, concerning the verification of facts and figures enables generalisations which do not differ from person to person. The concepts incorporated in the hypothesis should be explicitly defined and must have unambiguous empirical correspondence.

2. Secondly, the hypothesis must be conceptually clear, definite and certain. It should not be vague or ambiguous. It should be properly expressed. The concepts should not only be formally defined in a clear-cut manner, but also operationally. If a hypothesis is loaded with un-defined or ill-defined concepts, it moves beyond empirical test because, understandably, there is no standard basis for cognizing what observable facts would constitute its test. Hypotheses stated in vague terms do not lead anywhere. Therefore, while formulating the hypothesis, the researcher should take care to incorporate such concepts which are not only commonly accepted, but also communicable so that it would ensure continuity in research.

3. Thirdly, hypothesis must be specific and predictions indicated should be spelled out. A general hypothesis has limited scope in the sense that it may only serve as an indicator of an area of investigation rather than serving the hypothesis. A hypothesis of grandiose scope is simply not amenable to test. Narrower hypothesis involves a degree of humility and specific hypothesis is of any real use. A hypothesis must provide answer to the problem which initiated enquiry.

4. Fourthly, the possibility of actually testing the hypothesis can be approved. A hypothesis should be formulated in such a way that its conceptual content can be easily translated to understand the observable reality. If the hypothesis is not the closest to things observable, it would not be possible to test their accord with empirical facts. The concepts involved in the hypothesis should be such that the possibility of generating operational definitions can be ensured and deductions can be made, According to Cohen and Nagel, "hypothesis must be formulated in such a manner that deductions can be made from it and consequently, a decision can be reached as to whether it does or does not explain the facts considered."

5. Fifthly, the hypothesis should be related to a body of theory and should possess theoretical relevance. It must provide theoretical rationale by seeking answer to question as to what will be the theoretical gains of testing the hypothesis. If the hypothesis is derived from a theory, research will enable to confirm support, correct or refute the theory. Science being the constant interplay of theory and fact, gains immensely from such testing. If the hypotheses are selected at random and in piece meal, they cannot be studied in relation to broader theoretical framework. In the words of Goode and Hatt, "When research is systematically based upon a body of existing theory, a genuine contribution is more likely to result." In other words, to be worth doing a hypothesis must not only be carefully stated, but it should possess theoretical relevance."

6. Finally, the hypothesis should be related to available techniques. The hypothesis, in order to be workable, should be capable of being tested and measured to existing methods and techniques of scientific nature. According to Goode and Hatt, "the theories who do not know what techniques are available to test hypothesis is a "poor way to formulate usable question."[6]. On the contrary, if a new or original theory is in the process of evolution, it would make the work of the investigator easier for propounding a new theory. In this regard, Goode and Hatt have correctly stated, "In many serious sociological discussions research frontiers are

continuously challenged by the assertion that various problems ought to be investigated even though the investigations are presently impossible."[7] Knowledge of the available techniques at the time of formulations of hypothesis is merely a sensible requirement which applies to any problem in its earlier stages in order to judge its researchability. But "this is not be taken on absolute injunction against the formulation of hypothesis which at present are too complex to be handled by the contemporary technique."[8]

1.7.3. Sources of Hypothesis: Science develops along with the development of the human society. The development of science gives birth to hypothesis which is the result of the constant transformation of socio -cultural milieu, in socio-economic relations, human behaviour, values and attitudes.

Locating the origin of hypothesis, Goode and Hatt have identified the following sources :

2. General Culture in which a Science Develops.

A cultural pattern influences the thinking process of the people and the hypothesis may be formulated to test one or more of these ideas. Cultural values serve to direct research interests. The function of culture has been responsible for developing today's science to a great dimension. In the words of Goode and Hatt, "to say that the hypotheses are the product of the cultural values does not make them scientifically less important than others, but it does at least indicate that attention has been called to them by the culture itself. For example in the Western society race is thought to be an important determinant of human behaviour. Such a proposition can be used to formulate a hypothesis. We may also cite metaphysical bias and metaphysical ideas of Indian culture to have been responsible for the formulation of certain types of hypotheses. It implies that cultural elements of common cultural pattern may form a source of the formulation of hypotheses.

2. Scientific Theory.

A major source of hypothesis is theory. A theory binds a large body of facts by positing a consistent\ and lawful relationship among a set of general concepts representing those facts. Further generalizations are formed on the basis of the knowledge of theory. Corollaries are drawn from the theories. These generalizations or corollaries constitute a part of hypothesis. Since theories deal with abstractions which cannot be directly observed and can only remain in the thought process, a scientific hypothesis which is concerned with observable facts and observable relationship between facts can only be used for the purpose of selecting some of the facts as concrete

instances of the concepts and for making a tentative statement about the existence of a relation among the selected facts with the purpose of subjecting the relation to an empirical test.[11] A hypothesis emerges as a deduction from theory. Hence, hypotheses become "working instruments of theory"[11] every worthwhile theory' provides for the formulation of additional hypothesis. "The hypothesis is the backbone of all scientific theory construction; without it, confirmation or rejection of theories would be impossible." The hypotheses when tested are "either proved or disproved and in turn constitute further tests of the original theory." Thus the hypothetical type of verbal proposition forms the link between the empirical propositions or facts, and the theories. The validity of a theory can be examined only, by means of scientific predictions or experimental hypothesis.

3. Analogies, Observation of a similarity between two phenomena may be a source of formation of a hypothesis aimed at testing similarity in any other respect. Julian Huxley has pointed out that "casual observation in nature or in the framework of another science may be a fertile source of hypothesis."[15]. The success of a system in one discipline can be used in other discipline also. The theory of ecology is based on the observation of certain plants in certain geographical conditions. As such, it remains in the domain of Botany. On the basis of that the hypothesis of human ecology could be conceived. Hypothesis of social physics is also based on analogy. "When the hypothesis was born out by social observation, the same term was taken into sociology. It has become an important idea in sociological theory"[16] Although analogy is not always considered, at the time of formulation of hypothesis; it is generally satisfactory when it has some structural analogies to other well established theories. For the systematic simplicity of our knowledge, the analogy of a hypothesis becomes inversely helpful. Formulation of an analogous hypothesis is construed as an achievement because by doing so its interpretation is made easy.

4. Consequences of Personal, Idiosyncratic Experience as the Sources of Hypothesis. Not only culture, scientific theory and analogies provide the sources of hypothesis, but also the way in which the individual reacts to each of these is also a factor in the statement of hypotheses. Certain facts are present, but every one of us is not able to observe them and formulate a hypothesis.

Referring to Fleming's discovery of penicillin, Backrach has maintained that such discovery is possible only when the scientist is prepared to be impressed by the 'unusual'. An unusual event struck Fleming when he noted that the dish containing bacteria had a green mould and

the bacteria were dead. Usually he would have washed the dish and have attempted once again to culture the bacteria. But normally, he was moved to bring the live bacteria in close contact with the green mould, resulting in the discovery of penicillin. The example of Sir Issac Newton, the discoverer of the theory of Gravitation, is another glaring example of this type of 'personal experience'. Although prior to Newton's observation, several persons had witnessed the falling of the apple, he was the right man to formulate the theory of gravitation on the basis of this phenomenon. Thus emergence of a hypothesis is a creative manner. To quote McGuigan, "to formulate a useful and valuable hypothesis, a scientist needs first sufficient experience in that area, and second the quality of the genius." In the field of social sciences, an illustration of individual perspective may be visualized in Veblen's work. Thorstein Veblen's own community background was replete with negative experiences concerning the functioning of economy and he was a 'marginal man', capable of looking at the capitalist system objectively. Thus, he could be able to attack the fundamental concepts and postulates of classical economics and in real terms Veblen could experience differently to bear upon the economic world, resulting in the making of a penetrating analysis of our society. Such an excellent contribution of Veblen has, no doubt, influenced social science since those days.

2.7.4. Types of Hypothesis

Hypotheses may be of various kinds. It may be crude or refined. A crude hypothesis is at the lower level of abstraction, indicating only the kind of data to be collected, not leading to higher theoretical research. On the contrary, the refined hypothesis appears to be more significant in research. It may be in the form of describing something in a given instance, that a particular object, situation or event has certain characteristics. It may be in the form of counting the frequencies or of association among the variables. It may be in the form of causal relationship that a particular characteristic or occurrence is one of the causes determining the other.

On the basis of levels of abstraction, Goode and Hatt have distinguished three broad types of hypotheses. First, there are the simple level of hypotheses indicating merely the uniformity in social behaviour. They are the most exact and the least abstract, as they state the existence of presence of empirical uniformities. Often it is said that such hypotheses do not involve much verification or do not require testing at all and they merely add up facts. But it is not correct to say so. Even empirical researches describing certain facts need testing of hypotheses and testing may result in providing with an altogether different profile. Secondly, there are complex ideal hypotheses at a higher level of abstraction. These are more

complex and aim at testing the existence of logically derived relationships between empirical uniformities. They are in the form of generalisation, and therefore are also a little abstract. But empirical relationships are important in their context. Such hypotheses are useful in developing tools of analysis and in providing constructs for further hypothesizing. Thirdly, there are hypotheses which are very complex and quite abstract. They are concerned with the interrelations of multiple analytic variables. They lead to the formulation of a relationship between changes in one property and changes in another.

The above kinds of hypotheses may be explained in an example. On the basis of empirical data we may show statistical regularity by wealth, religion region, size of community culture, tradition, health etc. First, we may formulate hypotheses in a simple manner on the basis of statistical regularity. Secondly, in order to formulate a complex ideal hypothesis we may combine all the factors together. As regards the formulation of the third category of hypothesis, more abstraction is brought in. Only one of the factors can be studied at a time, such as relationship between religion and fertility or wealth and fertility, and all other variables may be controlled. Obviously, it is a very abstract way of handling the problem, because people may be affected £y a multiplicity of variables. Yet, we are interested in studying the cause and effect relationship of one factor at one time. Hence, this level of hypothesizing is not only more abstract, simultaneously it is more sophisticated and provides scope for further research.

1.7.5 Attributes of Hypothesis

In any scientific investigation, the role of hypothesis is indispensable as it always guides and gives direction to scientific research. Research remains unfocused without a hypothesis. Without it, the scientist is not in a position to decide as to what to observe and how to observe. He may at best beat around the bush. In the words of Northrop, "The function of hypothesis is to direct our search for order among facts, the suggestions formulated in any hypothesis may be solution to the problem, whether they are, is the task of the enquiry". Several near consequences are provided in the process of deductive development of hypothesis. In the process of conducting experiments for confirming the hypothesis, scores of new facts develop and expand the horizon of knowledge of the scientist. Since hypothesis is concerned with explaining facts, the rejection of hypothesis is not futile. Rather, it is worthwhile in the sense that it can be of great service in pointing out the way to true hypothesis. Even a false hypothesis is capable of showing the direction of inquiry. Realizing the indispensability of hypothesis in a scientific investigation, Cohen and Nagel observe,

'Hypotheses are required at every stage of an inquiry. It must not be forgotten that what are called general principles or laws can be applied to a present, still unterminated inquiry only with some risk. For they may not, in fact, be applicable. The general laws of any science function as hypothesis, which guide the inquiry in all its phases". Thus, there is little doubt that the importance of hypothesis in the field of scientific research is tremendous. At least five reasons may be attributed to justify hypothesis as a significant device for scientific research.

First, it is an operating tool of theory. It can be deduced from other hypotheses and theories. If it is correctly drawn and scientifically formulated, it enables the researcher to proceed on correct line of study. Due to this progress, the investigator becomes capable of drawing proper conclusions. In the words of Goode and Hatt, "without hypothesis the research is unfocused, a random empirical wandering. The results cannot be studied as facts with clear meaning. Hypothesis is a necessary link between theory and investigation which leads to discovery and addition to knowledge.

Secondly, the hypothesis acts as a pointer to enquiry. Scientific research has to proceed in certain definite lines and through hypothesis the researcher becomes capable of knowing specifically what he has to find out by determining the direction provided by the hypothesis. Hypotheses acts like a pole star or a compass to a sailor with the help of which he is able to head in the proper direction.

Thirdly, the hypothesis enables us to select relevant and pertinent facts and makes our task easier. Once, the direction and points are identified the researcher is in a position to eliminate the irrelevant facts and concentrate only on the relevant facts. Highlighting the role of hypothesis in providing pertinent facts, P.V. Young has stated, "The use of hypothesis prevents a blind research and indiscriminate gathering of masses of data which may later prove irrelevant to the problem under study". For example, if the researcher is interested in examining the relationship between broken home and juvenile delinquency, he can easily proceed in the proper direction and collect pertinent information succeeded only when he has succeed in formulating a useful hypothesis.

Fourthly, the hypothesis provides guidance by way of providing the direction, pointing to enquiry, enabling to select pertinent facts and helping to draw specific conclusions.. It saves the researcher from the botheration of 'trial and error* which causes loss of money, energy and time.

Finally, the hypothesis plays a significant role in facilitating advancement of knowledge beyond one's value and opinions. In real terms, the science is incomplete without hypotheses.

1.8. Key Words: Social Research, Social Work Research, Scientific Method, Concepts, Variables, Hypothesis.

1.9. References:

1. Whitley, Frederick, The Elements of Research
2. Goode and Hatt, Methods in social research
3. Young.P.V., Social Survey and Social Research
4. Kalton & Moser, Survey Method in Social Investigation
5. Lundeberg, Social Research

UNIT-II

UNIT-II

2.0 Objectives:

-To learn about research design and its types.

-To understand how the research problem is formulated.

-To know the concept of sampling, its definition, principles, types and procedures, population and universe.

-To cognize the meaning of measurement, levels of measurement.

-To know the meaning of validity and reliability and their types.

2.1. REASERACH DESIGN:

Once the research problem is formulated, a specific topic is assigned and the hypothesis is formulated, the next stage is to work out a research design. Preparing research design is an important stage in the process of conducting a research. Kerlinger defines a research design as “the plan, strategy of investigation purporting to answer research questions and control variance." The term 'plan' implies the overall scheme or programme of the research embracing on outline of what the researcher intends to do, ranging from the stage of formulation of hypotheses and their working implications to the final stage of data analysis. The term ‘structure’ intends to define the research study in a more specific way as the outline. The term 'strategy' is used in a more specific way than 'plan* and involves the methods and techniques- for collection of data and their analyses so as to achieve the precise research objectives. Miller defines "designed research" as "the planned sequence of the entire process involved in conducting a research study". According to P.V. Young, "Research design is the logical and systematic planning and directing of a piece of research." Selltiz and others define research design as "a catalogue of the various phases and facts relating to the formulation of a research effort. It is an arrangement of the essential conditions for collection and analysis

of data in a form that aims to combine relevance to research purpose with economy with a procedure."

In the words of Ackoff "Design is the" process of making decisions before the situation arises in which the decision has to be carried out. It is a process of deliberate anticipation directed towards bringing an unexpected situation under control." E.A. Schuman says that "A

research design represents a compromise dictated by many practical considerations that go into social research. He says further, "A research design is not a highly specific plan to be followed without deviation, but rather a series of guide posts to keep one headed in the right direction."

According to Jahoda, Deutsch and Cook, "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."

Thus, it becomes amply clear from the above definitions that research design is nothing but a scheme of work to be undertaken by a researcher at various stages, facilitating the research work in a systematic manner and conducting the various operations methodologically. Research design acts as a guide to achieve the goal of the researcher step by step in a calculative and cautious manner within a prescribed time limit and "specified cost. If the study is/not completed within the time limit, it will not only increase the cost but also cause a series of other problems associated with research, affecting the quality of research. Therefore, "the challenge of a research design is to translate the general scientific model into a practical research operation. Research design will refer to the entire process of planning and carrying out a research study". It involves the arrangement of conditions and observations in such a way that alternative answers to the questions taken up in the research are ruled out, containing a built in system of checks against all the factors that might affect the validity of the research outcome.

. According to P.V. Young, a research design should be able to provide answer the following queries

(i) What is the study about and what type of data are required?

ii) What is the purpose of the study? What is its scope?

(iii) What are the sources of the needed data- ?

(iv)What should be the place or area of study?

(v) What time, approximately, is required for the study?

(vi) What should be the amount of material or number of cases for the study?

(vii) What type of sampling should be used? (viii) What method of data collection would be appropriate? (ix) How will the data be analyzed? (x) What should be the approximate expenditure? (xi) What would be the methodology of study? (xii) What should be the specific nature of study? Keeping in view the aforesaid design decisions, the researcher may split the overall practical research design into the following phases.

(a) The sampling design, dealing with the method of selecting items to be observed for the given study; (6) the observational design, specifying the conditions under which the observations are to be made;

(c) the statistical design, taking into account the quantitative and statistical aspects of the design which concerns with the questions of how many items are to be observed and how the information and' data gathered are to be analyzed.

(d) The operational design, relating to the use of specific technique for the operation of the model already designed. It deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

2.2. Basic Purposes of Research Design

From what has been stated above, we can derive two basic -purposes: (a) To provide answers to the research questions, (6) to control variance. Indeed, these research purposes are achieved by the researcher himself, not by the research design.

As regards the first purpose, a research is designed to enable the researcher to arrive at an objective, accurate, valid and economic solution of the given problem to the maximum possible extent. Since scientific research begins with a provisional supposition in the form of a hypothesis, the major purpose of the design is to provide the research with a valid test of the hypothesis on the basis of empirical evidence obtained by the researcher by utilizing the least amount of money, manpower and time and maximum possibility of its being approved by other investigators engaged in the given area of investigation. By providing a sort of blue print for the variation of hypothesis, presuming the relation between two or more variables on the basis of empirical facts and by directing the process of observation in terms of determining the facts relevant to the research problem, how and where to look for them and how many observations to make, the research design becomes indispensable on the part of any researcher in scientific investigation. Furthermore, it also indicates whether or not the variables of the research are to be manipulated or selected, what specific values of the manipulated or selected variables are to be utilised in the scientific investigation, how a conceptual variable can be converted into observable facts. Research design also makes

specification of the method to be adopted for manipulation of the independent variable and for measurement of the dependent variable along with suggesting the ways by which the collected data for research are to be analysed and determining the level of statistical analysis appropriate to the research situation. "The design of an experiment and its analysis are interrelated. Indeed it is often said that one should not do an experiment without knowing how it is to be analysed." This statement of Riecken and Boruch not only applies to experimental design, but also holds good for all kinds of designs of research.

The second purpose of research is to control the effects of the potentially relevant independent variables on the behaviour of the research subjects. It merely facilitates the process of obtaining answer to relevant questions in the research study and
enables the investigator to exercise control over experimental, extraneous and error variances pertaining to the particular research problem being studied. The validity of the research findings would be affected if these variables are not controlled. In a real world, any observed event of behaviour is influenced by a multiplicity of facts and events. The behaviour being "a real World event involving overt or covert responses by one or more actors to a task or situation" and task being "any impending sequence of acts guided by a goal" both behaviour and task involve a complexity of events. Each of these may be used as an independent variable. Of course, consideration of a variable as an independent one depends upon the researcher's interest or the nature of the research problem. For example, job satisfaction, educational achievement, individual production, restriction of birth rate and similar other effects are explainable on the basis of the influence of a number of related or unrelated facts and events. But it is not possible to incorporate each one of these variables within the same research undertaking. On the contrary, a researcher must keep himself restricted only to a limited number which are used as the more explicitly relevant variables in a given research. If they happen to be active variables, their values are deliberately changed and thereby they are manipulated in order to be controlled.

2.3. Characteristic Features of a Good Research Design

Designing a research, particularly in the field of social sciences is very complex as the selection of a method or methods of logic and planning of the design was not always guarantee sound results. As a blue print, the research design may at best be only tentative and useful to the extent of providing the researcher with a series of guide posts to keep him headed ir[1] the right direction. Although every design has its own strengths and weaknesses and

simultaneously the possibility of a single perfect research design is difficult, a good research design is often believed to possess characteristic features like flexibility, appropriateness, efficiency, economically sound and so on. A design which minimizes bias and maximizes the reliability of data is construed as a good design.

Similarly the design giving the smallest experimental error is considered to the best design and the design yielding maximal information covering various aspects of a problem is construed as the most efficient design because it is appropriate to the research problem. Hence, consideration of a design as good depends too much upon the objective of the research problem and also the nature of the problem under investigation. A single design can never serve the purpose of all types of research problems because what appears to be suitable in one case may be lacking in one respect or the other in the context of some other research problems. A good research design should always fulfill the following four conditions; objectivity, reliability, validity and generalisability of the findings.

(a) **Objectivity**. The findings are said to be objective when they pertain to the method of data collection and the scoring of the responses. The objectivity in respect of the procedure may be judged by the degree of agreement between the final scores assigned to various persons by more than one independent observer.

The more the agreement among the observers, the more objective are the observation, recording and evaluation of the responses. Therefore, a good research design should permit fairly objective measuring instruments in which every observer visualizing a performance comes to the same conclusion.

(b) **Reliability**. The question of reliability of knowledge is usually raised when the presence of a problem arouses in the knower a demand, not only for something more than mere conjecture, but for something for which it shall be useful in a given situation and perhaps in other similar situations.[10] Reliable knowledge means any claim that is substantiated as trustworthy for a given purpose.

(c) **Validity.** Validity implies self-consistency or absence of self-contradiction. It is identified with formal truth or self-consistency. A valid reasoning conforms to the rules of correct reasoning. It is that type of reasoning where conclusions automatically follow from the premises legitimately.

(d) **Generalisability**. The degree of generalisability is known in terms of the replicability and reproducibility of the findings in spite of different measures and settings respectively.

Elements of Research Design

1. **Selection of Research Problem.**
2. **Selection of Units of Analysis.**
3. **Choice of Variable**
4. **Identification of Relationships.**
5. **The Nature of Causal Relationship.**
8. **Operationlisation of Concepts and Definitions.**
7. **Formulation of Hypotheses.**

1. **Selection of Research Problem.** As regards the selection of topic for research, anything that is social and empirical is a relevant problem for social research. The factors which affect the decisions on selection of topic in social sciences are : (i) the structure and state of a discipline (ii) social problems (iii) other determinants like the availability of grants for particular themes, the popularity and prestige of the particular area 'of research, public interest and motivation of the researcher etc., and (iv) practical considerations.

2. **Selection of Units of Analysis**. Determination of the units of analysis is a key factor in social research. In general, the purpose of the study dictates the selection of the appropriate unit of analysis. The objects or events or entities under investigation are referred to as units of analysis in social sciences.

3. **Choice of Variable.** Since a social scientist is primarily interested in studying the relationship among some characteristics or properties of the observed units which are subject to variation over cases, over time or over both cases and time, it is necessary for a researcher to decide which variables should be the focus of research. Explanatory variables are known as the variables under focus. They are of two types dependent and independent. The former one is the variable the researcher is interested in explaining and predicting. Dependent variable is the presumed effect. The

independent variable is the presumed cause,

The extraneous variables are those which are not the direct focus of research. Those are of two types : controlled and uncontrolled. The controlled variables are held constant or prevented from varying during the course of observation. A Apart from the above classification of the variables, a typology of quantitative and qualitative variables is also made. Whereas a quantitative variable implies values or categories consisting of numbers, qualitative variables represent certain qualities, attributes or discrete categories.

4. **Identification of Relationship**. In real terms, very many social researchers directly aim at developing and testing relationships, apart from gaining familiarity of a phenomenon or description of communities or groups or exploration of a situation or event. However, on the whole, research findings largely depend on particular anticipated relationships. Therefore, identification of the anticipated relationship and the guiding theoretical premises assume greater importance.

5. **The Nature of Causal Relationship.** Causal relationships constitute the heart of scientific understanding. These are very much required for purposes of explanation and prediction. In order to establish causality, the social scientists take help of three types of evidence: association, direction and non-spuriousness.

Statistical association, such as a pattern of change in one variable is related to the other variable, indicates that former is the cause. Causal relationships are determined in terms of strong and weak associations.

Another criterion required for establishing a causal connection between events is that the direction of influence should be from cause to effect. In other words', cause must precede its effect.

The third criterion needed to establish a causal relationship between events is non-spuriousness which implies that in order to inter a causal relationship from an observed correlation there should be sufficient reason to believe that no hidden factors have contributed to a spurious relationship. Ideally, the researcher must show that the connection between the variables is held constant.

6. **Operationalisation of Concepts.** Since concepts serve a number of important functions, clarity and precision in the usage of concepts are to be achieved by definitions which must contain the distinctive characteristics or qualities of the phenomenon under investigation.

Concepts, in order to be operationally existent, should be established through operational definitions which are instrumental in specifying the contextual meaning of concepts and providing the framework of their application. Briefly stated, the operational definitions serve as a link between the conceptual theoretical level and the observational empirical level.

7. **Formulation of Hypothesis**. In order to state the research questions in a precise manner so as to give clear indication of what is to be observed and what kind of information will be gathered, the research questions must be stated in the form of hypotheses. Hypotheses are tentative generalisations which are expected but based on unconfirmed relationship between two or more variables.

2.4. Types of Research Design

On the basis of research purpose, research studies may broadly be categorized into four types:

(i) Exploratory or Formulative Design

(ii) Descriptive Design

(iii) Diagnostic Design

(iv) Experimental Design.

The nature of each of these designs is explained below.

(i) Exploratory or Formulative Design

The main purpose of exploratory study is to gather information which will help in future for formulation of a precise research problem. On the basis of the collected facts the researcher may be able to formulate sound hypotheses for further research. It may also enable the researcher to get himself acquainted with the phenomena which he expects to investigate at a later stage. The aim of an exploratory or formulative study may be clarification of concepts, establishing priorities for future research and collection of data about the actual conditions which affect an intended research.

Requirement of Exploratory Design

The essentials for exploratory or formulative design are:

(a) Review of pertinent literature

(b) Experience Survey

(c) Analysis of Insight Stimulating cases.

(a) Review of Pertinent Literature. While proceeding in the path of research the researcher has to take help from the work already done by his predecessors. By doing so, he will not only save himself from the problem of trial and error but also minimize the expenditure of his energy. Apart from reviewing available literature pertaining to the problem under investigation, the researcher may also take into account the literature pertinent to analogous problems.

(6) Experience Survey. Because of the complicated nature of social problems, the researcher is not in a position to collect all the required materials about a particular problem from one place. At times the researcher has to contact the persons who have earned enough of experience to understand and analyse the social reactions. The researcher should take advantage of their experience in a very intelligent manner.

Taking good advantage of the experience of the persons involves the following steps :

(i) Selection of respondents

(ii) Questioning of respondents

(a) Selection of Respondents: Formulation of a correct exploratory design requires that the investigator should make proper selection of the respondents. For this purpose he should select only those respondents who are dependable and who have actual knowledge regarding the problem under investigation. The selection of the respondents may be made either directly or indirectly. In direct selection the investigator chooses those persons who are well known for their knowledge in the problem area. In case of indirect selection the investigator chooses those persons who are indirectly concerned with the problem. Hence, the selection of the respondents should not be confined to a particular group, rather it should be many-sided.

ii) Questioning of the Respondents: Proper questioning of the respondents ensures relevant information. Therefore while framing the questions, due attention should be given on clarity of concepts. For this purpose, the investigator should consult the books and the relevant portions of the bibliographical schemes adequately.

(c) Analysis of Insight Stimulating Cases: Analysis of insight stimulating cases includes all those events, incidents and phenomena that stimulate the researcher. Such cases invoke in the investigator the thinking regarding the formulation of the hypotheses. In this regard, the attitude of the investigator, intensity of the case study and integrative power of investigators appear to be very important.

As regards the attitude of the investigator, receptivity and sensitivity are needed. These qualities enable the investigator to take stock of various developments occurring in his field of study and make steady progress.

Intensive case study involves studying the subject matter in all its dimensions and verifications, in the background of history.

In this regard, the groups, the community and groups of individuals may be treated as the units of study.

Integrative power of the investigator is considered important because, on that basis he is able to collect even the minutest possible information regarding the subject matter. What appears significant, in this regard, is his attention on new observations rather than on experimentation.

(ii) Descriptive Research Design

The purpose of descriptive type of design is to describe some event, situation, people, group or community or some phenomena. Fundamentally, it is a fact finding exercise which focuses on relatively far dimensions of a well defined entity, aiming at precise and systematic measurement of some dimensions of a phenomenon. Usually, a descriptive design involves detailed numerical descriptions, such as distribution of the population of a community by age, sex, caste or education. The researcher may also take recourse to descriptive design for estimating the proportion of people in a particular geographical locality in respect of their specific views or attitudes.

However, the procedure followed in descriptive design is broadly analogous, notwithstanding the differences evinced in their field, formulation of hypotheses, objectives, for treatment of the problem and in matters of field expansion.

(iii) Diagnostic Research Design

Being concerned with the express characteristics and existing social problems, the diagnostic research design endeavors to find out relationship between express causes and also suggests ways and means for the solution. Thus, the diagnostic studies are concerned with discovering and testing whether certain variables are associated. Such studies may also aim at determining the frequency with which something occurs or the ways in which a phenomenon is associated with some other factors.

Diagnostic studies are mostly motivated by hypotheses. A primary description of a problem serves the basis so as to relate the hypotheses with the source of the problem and only those data which form and corroborate the hypotheses are collected. As regards the objectives of diagnostic research design, it is based on such knowledge which can also be

motivated or put into practice in the solution of the problem. Therefore, it is obvious that the diagnostic design is concerned with both the case as well as the treatment.

Diagnostic studies seek immediate to timely solution of the causal elements. The researcher, before going through other references, endeavors to remove and solve the factors and the causes responsible for giving rise to the problem.

The research design of diagnostic studies demands strict adherence to objectivity for elimination of any chances of personal bias or prejudice. Utmost care is taken while taking decisions regarding the variables, nature of observation to be made in the field, the type of evidence to be collected and tools of data collection. Simultaneously the research economy should not be lost sight of. Any faulty decision in these regard will result in wastage of time, energy and money.

Usually the first step in such designing is accurate formulation of research problem wherein research objectives are precisely stated and principal areas of investigation are properly linked. Otherwise the investigator will find it difficult to ensure the collection of required data in a systematic manner. Simultaneously, the clarification of concepts and the operational definition of the terms should also be ensured so as to make them amenable to measurement.

At the next stage certain decisions regarding collection of data are taken. In this regard, the researcher should always bear in mind the advantages and disadvantages of the method to be employed and at the same time the nature of research problem, type of data needed, degree of desired accuracy etc. should be considered. That apart, while collecting data, effort must be made to maintain objectivity to the maximum possible extent.

In order to surmount the financial constraints, paucity of time, a representative sample of the research universe should be drawn so as to gather relevant information. A wide range of sampling techniques are prevalent which must be made use of, appropriately by the researchers.

At the stage of analysis of data, the researcher must take proper care in placing each item in the appropriate category, tabulating of data, applying statistical computations and so on.

Sufficient care must be taken to avoid potential errors due to faculty procedures of analysis of data. Advance decisions regarding the mode of tabulation, whether manual or by machine, accuracy of tabulating procedures, statistical application etc. will be of immense help in this regard.

(iv) Experimental Design

The concept of experimental design in sociological research refers to systematic study of human relations by making the observations under conditions of control. In the words of Jahoda and Cook, 'an experiment maybe considered as a way of organizing the collection of evidence so as to permit one to make inference about the tenability of a hypothesis. According to Chapin, "experiment is simply observation under controlled conditions. When observation alone fails to disclose the factors that operate in a given problem, it is necessary for the scientist to resort to experiment."

In real terms, experimentation is resorted to when it is not possible to solve the problem through observation and general knowledge. The core of the experimental method lies in drawing inferences by observation of human relations under controlled conditions. Since a number of factors are in operation in every complex social situation, the social scientist, while seeking to describe the single causal relation of factor A to factor B, must attempt to create an artificial situation wherein all other factors, such as C, D, E etc., are controlled. Such a state is achieve4 by selecting two groups which are equal in all significant receipts and choosing either of the groups as 'experimental group, and the other as the 'control group', and thereafter exposing the 'experimental group' to the assumed causal variable, while keeping the 'control' group under control. After a specific time period, the two groups are compared in terms of the 'assumed effect'. The assumed causal variable and the assumed effect are otherwise called the independent variable and dependent variable respectively. Required evidence for testing causal relations among variables, already stated in the form of a hypothesis, is generated by the above method of experiment.

Demonstration of causal relationship among variables in experimental design involves three clear cut operations; such as demonstrating co variation, eliminating spurious relationships and establishing the time order of occurrence. We have already discussed the first two operations elsewhere in this chapter. Here we will discuss the third operation which is concerned with establishing the time order of occurrence. This necessitates that the researcher should demonstrate that one phenomenon occurs first or gets transformed before the other phenomenon with the premise that the phenomenon which is yet to occur cannot be the determinant of the present or past phenomena.

Experimental design enables the researcher to draw causal inferences. It also smoothens the observation of independent variable causing assumed effect.

The three components of experimental design are: comparison, manipulation, and control.

Through comparison, the correlation between variables is known. It also enables us to demonstrate the association between two or variables.

Through manipulation the researcher establishes the time order of events. The major evidence which become essential to determine the sequence of events is that-a change occurs only after the activation of the independent variable. In other words the independent variable precedes the dependent variable.

- **Types of Experimental Design**

There are numerous ways in which experiments can be done in the field of social sciences. In their work "Experimental and Quasi-Experimental Designs of Research on Teaching", Donald T. Campbell and Julian C. Stanley have mentioned more than a hundred ways of conducting experiments which maybe designated as experimental design. But from the analytical point of view seven broad categories may be mentioned;

(i) After only design.
(ii) Before-after design.
(iii) Before-after with control group design.
(iv) Four group-six study design.
(v) After only with control group design.
(vi) Ex-post facto design.
(vii) Factorial design.

(i) After only design. Among all categories of experimental designs, after only design appears to be the simplest. This consists in measuring the dependent variable only after the experimental subjects have been exposed to the experimental variable. This design is considered more appropriate as an exploratory study than a real experiment.

(ii) Before-after design. As the name suggests, in this design measurement of the dependent variable is taken before as well as after exposure of the subject to the experimental variable, and the difference between the two measurements is taken to be the effect of the experimental variable. For example if the measured value of the dependent variable before exposure of the subject to the experimental variable is noted as 'A' and its measured value after exposure of the subject to experimental variable is noted as ⁽B' then the effect of the experimental variable is taken to be (B—-A).

(iii) Before-after with control group design. In this design the research has a control group against which the results of the experimental group are compared. The control group and experimental groups are selected in such a way that both the groups are similar and interchangeable. The control group is measured before as well as after without being exposed to the experimental variable. Hence, there may hardly be any difference between before and after measurements. But if there is any difference between before and after measurement, it represents the result of uncontrolled variables.

On the other hand, the experimental variable is introduced in the experimental group. The difference between before and after measurements in respect of the experimental group is construed as the result of experimental variable as well as the uncontrolled variables. To know the exact effect of the experimental variable, the researcher deducts the difference between the two measurements of the controlled group from the difference of the two measurements of the experimental group. The following notation explains this:

	Control Group	**Experiment Group**
Whether before-measurement is taken ?	Yes **(A)**	Yes **(C)**
Whether experimental variable is introduced ?	**No**	Yes
Whether after-measurement is taken ?	Yes **(B)**	Yes **(D)**
Effect of experimental variable	**(D-O—(B—A)**	

(iv) Four group-six study design. In this type of design two experimental groups and two control groups are taken. Measurements are made in six cases, *i.e.* before-measurement, and after-measurement in respect of experimental group-I, after-measurement in experimental group-II, before and after measurements in respect of control group-I; and only after measurement in control group-II.

Before measurements in all the four identical groups will be almost the same. If the before-measurements have no effect on the variable being studied, the two

experimental groups should provide the same after-measurements and, similarly, the two control groups should also give the same after measurements. However, the results of in the two experimental groups are most likely to be different from the results of the two control groups, if the experimental variable exerts any influence.

(v) After only with control group design. This is also known as two group-two study design, which is a modification of the four group-six study design. Here, the researcher does not study the experimental variable under different conditions. Hence, the effect of experimental variable is determined simply by finding out the differences between the after-measurements in respect of experimental and control groups. It so happens because if before-measurements of the experimental group-II and control group-II are taken, those are likely to be the same due to the identical characteristics of the groups. On this presumption, the researcher may very well ignore them.

We can express the design by the following notation:

	Experimental Group	**Control Group**
Whether before measurement is taken ?	**No**	No
Whether experimental variable is introduced ?	Yes	Yes
Whether after-measurement is taken ?	Yes **(A)**	Yes **(B)**
Effect of experimental variable	**(A—B)**	

(vi) Ex-post facto design. In Ex-post facto design the experimental and control groups are selected after the introduction of the experimental variable. Thus, it can be called as a variation of the after-only design. The main advantage of this design is that the test subjects are not influenced towards the subject by their knowledge of being tested. It also enables the researcher to introduce the experimental variable according to his own will and to control his observations.

(vii) Factorial design. All categories of experimental designs discussed above are designed to test experimental variable at one level only. But, on the other hand, the factorial designs enable the experimenter the testing of two or more variables simultaneously.

2.5 FORMULATION OF RESEARCH PROBLEM

In real terms research begins with a problem which needs solution. Such a perception on the part of the researcher, first of all, falls within the general area of interest indicating to locate the problem either in the quest of some intellectual pursuit or for some practical concern, such as finding out a practical solution to a problem, evaluation of a programme in the light of new facts, gathering of relevant facts for social planning or even for policy making, While selecting a problem for research, the social scientists are likely to be influenced by their own personal values as well as the prevalent social conditions. As scientists differ with regard to their values and societies differ in respect of their preference of different spheres, the choice of topics in social research vary widely.

In any case, formulation of a research problem, arising out of a theoretical situation or practical concern, is not an easy task, as it appears to be. Actually it an uphill task, so much so that even a scientist of the stature of Charles Darwin has gone to the length of saying that "looking back, I think it was more difficult to see what the problems were than to solve them". R.K. Merton has identified three important questions as the three principal components involved in the process of formulation of a problem in the field of research in soft sciences: (i) What one wants to know? (ii)Why one wants to seek answers to those particular questions? (iii)What may be the possible answers to the originating questions? These tree questions correspond to the components of the originating questions, the rationale and the specifying questions respectively.

2.6. SAMPLING: DEFINITION

Sampling is the process of obtaining information regarding the entire research population or aggregate or totality by examining only a part of it. It is a picture in miniature to represent the larger whole, to enable the researcher to make a judgment or inference about the whole with the presumption that the sample data will provide valid and reliable conclusions.

Sampling is not only typical to the field of science, rather in our day-to-day life, we take to such method. For example, a cook presses a few pods of boiled rice in the cooking pot to be able to know that it is ready to be served. Similarly, while examining the quality of food grain one need not examine each and every grain contained in the

bag. Rather an examination of a handful of grains gives the impression regarding the entire bag.

We frequently practice some crude versions of sampling, because in practice, census enquiry is not always possible as it involves a great deal of time, money and energy and therefore, the census method may be practically beyond the capacity of an ordinary researcher. Probably, Government is the sole institution which can completely enumerate all the items. Even then government takes ten years of time for population census. Moreover, many a time it is well nigh impossible to touch every item in the population and seldom it is possible to obtain sufficient accuracy by studying only a part of the total population. If the universe is small, it needs to be emphasized that it is of no use to take recourse to a sample survey.

2.6.1. Population and Universe

Thus a sample is a portion selected from the population or universe. However, the terms 'population' and 'universe' have been used in social research in a very specific sense. Population refers to the total number of cases with a given characteristic or characteristics, or all the members of a given set or class. Broadly speaking, 'universe' constitutes all the individuals, things, events, documents or observations either on a single or many individuals etc. belonging to designated category characterizing specific attribute which a particular study should principally cover. For example, if an investigation is intended to determine the average per capita income of the inhabitants in a particular urban centre, the universe will comprise all the earning people in that urban centre.

A 'universe' or 'population' consists of sub-population, *e.g.* the girl students in a college making the sub-population of the population which consists of all the college students. A sub-population or stratum of the entire universe may be defined by one or more specification, which make the division of the population into mutually exclusive sections or strata.

A single unit or member of the population is called population element. According to J. Simond "Sample is a collection of observation for which one has data to work with." Almost any set of observations for which one has data makes a sample.

A 'population' or 'universe' may be divided into two types :

(i)) Finite universe. *(ii)* Infinite Universe

A universe containing a finite member of observations or items is known as finite universe *e.g.* the students in a college, workers in a factory, the population of a city etc.

On the contrary, a population having an infinite number of observations or with the number of observations so as to appear practically infinite, is called as an infinite universe *e.g.* the floating population of a pilgrimage, number of people in a festival or fair etc. However, infinite populations do not create a problem in sampling method but rather they are considered better than the finite universe.

Another type of classification of the universe is the dichotomy of—

(i) Existent universe, (it) Hypothetical universe.

An existent universe comprises concrete objects such as number of buses in a bus-stand, number of scooters in a particular city etc.

On the other hand, if the universe does not consist of concrete objects, rather it comprises imaginary objects, it is called hypothetical universe. The population of the throws of coin or dice in infinite number of times is hypothetical universe.

The researcher makes a decision concerning sampling unit before he actually selects sample. A sample is a finite sub-set of the population, selected from it with the objective of investigation properties. A sample unit may be a geographical one or it may be a social unit or even it may be an individual *e.g. .a* state, district or a village. Even a construct unit such as house, flat etc. may be called a sampling unit in a geographical arena.

A sample unit may also be a social unit like family, club, school etc. or may be an individual.

The number of units in the sample is known as sample size. In other words, sample size refers to a number of items selected from the universe to constitute a sample. Decision regarding the size of sample is a major problem before a researcher. Therefore, the sample size has to be more than a small unit and at the same time less than large. And optimum sample is one which fulfils the requirements of efficiency, representativeness, reliability and flexibility.

2.6.2. Utility of Sampling:

Before we go to discuss the types of sampling design involving implementing designs, it is worthwhile to understand the advantages and disadvantages of sampling in a general way. As regards its advantages, Prof. R.A Fisher has summed up the advantages of

sampling over census in just four words, such as speed, economy, adaptability and scientific approach.

If the sampling plan is properly designed and is carefully executed, it is most likely yield to better results, even sometimes better than those obtained through census method.

(a) Greater Speed: Speed emphasizes use of less time for collection of data as an universe in miniature is to be inspected and examined. Less time is consumed not only in the conduct of sampling inquiry, but also in the processing, editing, and analysis of data. As such, adoption of sampling method becomes absolutely necessary where the results are urgently needed.

(b) Reduced Cost: The sampling method involves reduction of cost of inquiry and as such this is much more economical than census method. Because of lack of finances the developing countries may not be able to afford to a complete census. The sample method will be of immense help in conducting socio-economic surveys in those countries.

(c) Administrative Convenience: When small samples are used, the researcher finds it easy to give more attention to each return and to check their accuracy. Since a complete census requires a huge administrative setup involving large number of personnel, trained investigators and moreover proper co-ordination between various operating agencies, the organization and administration conducting a sample survey is in a relatively more convenient position. Contrary to the census method the sample survey requires less number of personnel or trained investigators and smaller field of inquiry.

(d)Reliability: A carefully designed and scientifically executed sample survey always yields more reliable results than those obtained on the basis of a complete census survey due to the following reasons:

(i) It is always possible to determine the sampling errors and thereby ascertain the degree of reliability of the results through the use of different devices. *{ii)* In a sample survey the non-sampling errors can be effectively controlled and minimized by:

(a) Engaging sufficiently qualified, skilled and trained personnel.

(b) Providing proper training to the field investigators.

(c) Adequate accuracy.

(d) Utilizing refined equipments and developed statistical techniques and processing and analyzing the relatively limited data.

(e) By ensuring the follow up work in case of non-response and incomplete response.

On the other hand, in case of a complete census, the non-sampling errors caused due to measuring and recording of observations, inaccuracy and incompleteness of information, location of unity, incomplete response, non-responses, training of investigators, interpretation of questions, bias of investigator etc. appear to be more serious.

(ii) The effective reduction of non-sampling errors in a sample survey more than compensates errors in the estimates due to sampling procedure. Therefore, the results, it provides become more accurate and reliable ones.

(iii) Sampling procedure has also got an edge over the census method in the sense that the efficiency of sampling procedure lies in the accuracy of result at par with the complete census by the use of some type of sampling check procedures.

(e) Wider Scope: Although it generally appears that a complete census can only provide detailed information due to its coverage of all the units in the universe, nevertheless, sampling procedure results in considerable saving of time, money, labour power and simultaneously provides obtaining more detailed and exhaustive information of the limited number of units selected in the sample. Moreover, in many a case, a complete enumeration of the universe is well nigh impracticable. In such cases where the investigation requires adequately trained personnel and more sophisticated equipments for data collection, processing and analyzing the data, it is rather inconceivable. In those cases sampling procedure is readily adaptable.

(f) Suitable for infinite or hypothetical population:

Where the universe appears to be too large or infinite, the sampling procedure is the only means available to the investigator for estimating the parameters of a population. For example, the floating population in a fair can only be estimated by sampling method. In the analogous manner, in case of a hypothetical universe the sampling procedure is the only scientific technique or estimation of the parameters of the population. For example, in the problem of tossing a coin where the process may continue indefinitely a sampling procedure is more readily adaptable than census.

(g) Destructive testing: If in the course of inspection the units are affected adversely or destroyed, or in other words if the testing unit is destructive, the researcher is left with no other means than taking recourse to sampling procedure. In such cases, inspecting a representative sample can only be resorted to, in order to avoid complete census which will destroy all the items.

2.6.3. Principles of Sampling:

Sampling demands the exercise of great care, otherwise the results may be misleading.

If sample survey is not properly designed and carefully executed according to the principles, the results will lack in reliability and are more to be misleading. Therefore, F. Stephen says, "Samples are like medicines, they can be harmful, when they are taken carelessly without knowledge of their effects"

(1) Every good sample should have proper levels with instructions about its uses. Unless sampling design is made perfect, it might lead to serious complication in the final result, whereas the omission of few units in case of a census survey may hardly matter. Incomplete response or non-response from even one or two units in a small sample might have tremendous impact on the findings.

(2) A careful sample survey necessitates the involvement of qualified skilled and more experienced personnel, proper supervision and relatively more sophisticated tools and statistical techniques for the planning and execution of the sampling skill and for the collection, processing* and analysis of data. If the survey lacks in either of these or all, the findings of the sample survey may not be relied upon.

(3) Although, it is believed that a sample survey saves time, money and manpower, it may so happen that sometimes the sample survey might need more time, relatively more expenditure and manpower than a census survey. This happens so particularly when the sample size constitutes a large chunk of the research universe and when complicated weighted system is used.

(4) The researcher cannot resort to sampling procedure if he wants to get information about the entire universe *.i.e.* about each and every unit of the population. Moreover, as sampling procedure is well nigh impossible when the population is too heterogeneous.

2.6.4. Types and procedures of Sampling:

Different types of sample designs are based on a couple of factors such as the representation basis and elements selection technique. As regards the representation basis, the classification may be between—probability and non-probability sampling. Probability sampling is based on the finite universe and conducive to random selection. On the other hand the non-probability sampling is non-random sampling, because it is related to infinite universe. As regards the other typology based on the element selection techniques, the sample may either be restricted or unrestricted.

In case of unrestricted sample, sample element is drawn individually from the population at large, whereas in case of restricted sample and other forms of sampling are covered. Different sample designs discussed above are shown in the chart given below.

Representation Basis-------- > **Element Selection Technique** 4	**Probability** Sampling	**Non-probability Sampling**
Un-restricted Sampling	Simple Random Sampling	Haphazard Sampling or Convenience Sampling
Restricted Sampling	Complex Random sampling *(i)* Cluster Sampling *(ii)* Systematic Sampling *(Hi)* Stratified Sampling etc.	Purposive Sampling *(i)* Quota Sampling *(ii)* Judgment Sampling

In the execution of sample survey paramount importance is given to the chance on the appropriate sample design. The selection is generally made keeping in view the objective and scope of inquiry and the type of universe to be sampled.

From the practical point of view the sample designs are basically of two types:

- **Probability sampling.**
- **Non-probability sampling.**

Probability Sampling :

It provides a scientific technique of drawing samples from population in accordance with certain laws or chance in which each unit in the universe has some definite pre-assigned probability of being selected in the sample. That is why it is called chance or random sampling. For example, in a lottery method, the individual units are picked up from the entire group not by deliberate attempt but by some mechanical processes. Since the method of selection is based on blind chance, the results obtained from it can be assured in terms of probability or the errors of estimation or the significance of the result obtained from random sampling can be measured. Due to such reasons, the superiority of random sampling design over the deliberate sampling design is established. The law of statistical regularity is ensured by random sampling, implying that if the sample chosen is a random one on an average, it will be executed in same composition and characteristics of the population. Due to such reasons, random sampling is adjudged as the most efficient technique of selecting representative sample.

Random sampling may be of two types :

- Sampling without replacement.
- Sampling with replacement.

In random sampling without replacement, sampling from the finite population implies that the probability of selecting a specified unit of population at any given draw is equal to the probability of its being selected at the first draw.

In random sampling with replacement, the unit selected in any draw is replaced back before making the next draw. Thus simple random sample with replacement always amounts to sampling from an infinite population even though the population is finite.

Selection of Simple Random Sampling :

Proper care is taken to select simple random sampling. A random sample should be selected in such a manner that it will ensure the representativeness of population. This can be done by any of these following methods :-

(a) Lottery Method.

(6) Use of table of random numbers.

(c) Grid system.

(d) Alphabetical list.

Lottery Method is considered as the simplest method of drawing random sample. In such a case, the names of all the units of population or their number are written

on a slip or a card which should be as homogeneous as possible in shape, size, colour etc. to avoid human bias. Then they are mixed thoroughly in a container and thereafter lottery is drawn either blind folded or by rotating in a drum or box or in any similar devices. If the population is small then the slips or cards are put in a bag and thoroughly shuffled and then the required number or slips as units are drawn one by one. After each drawl the slips are thoroughly shuffled. The sampling units relating to numbers on the selected slips or cards will constitute random sample. An example may be cited in this regard. If we suppose that a random sample of 20 individuals is to drawn from a population of 200 individuals, we will have to assign the number from 1 to 200, one number to each individual of the population and arrange 200 individual slips bearing numbers 1 to 200. Thereafter all these slips are put in a container and shuffled thoroughly. Finally, 20 slips are drawn out one after another.

When the population to be sampled is reasonably large, we may use another lottery method in which the slips or cards are placed in a metal cylinder, thrown into large rotating drum which is operated mechanically. The rotation of drum mixes or randomizes the cards and finally a sample size in a desired size 'n' is drawn out from the container. Thus lottery method provides a sample which is quite independent of the properties of the population. Therefore, it is considered the best and commonly used method of selecting random sample. As regards simple random sampling with replacement, each slip or card drawn is replaced back in the container before making the next draw. But in a sampling without replacement the slips or cards once drawn are not kept again in the container. Therefore, a thorough mixing is necessary before drawing the next card or slip.

Use of Random Number Tables:

A relatively easy method in drawing random sample can be made by the use of table of random numbers. Since the lottery method has already been described as a cumbersome task and quite time consuming process; particularly when the population is sufficiently large and since the slips or cards used in the processes cannot be made exactly similar as some bias is likely to occur, the statisticians have designed random number tables, which have been constructed in such a manner that each of the digits 0, 1, 2, 9 appears with approximately the same frequency and independently of each other. Various statisticians like Tippet, Fisher and Yates, Kendal and Babington Smiths *etc.* have prepared tables of random numbers which can be of use for drawing random sample. The use of random number table involves following steps.

(i) 'N' units in the population numbering from 1 to N must be identified.

ii) Any page of the random number table must be selected at random and the numbers row wise or column wise or even diagonally be picked up at random.

(iii) The population units corresponding to the number selected in the above procedure constitute the random sample.

Although generally, Tippet's random number tables are used for the purpose, other random number tables also serve all practical purposes.

Tippet's table of random numbers:

L.H.C. Tippet constructed a list of 10,400 four digit numbers written at random at every page. They have been constructed out of 4,16,000 digits taken from census reports by combining them in fours. A list is given below—

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

For example, if one is interested in taking 20 units from the population of 5000 units bearing numbers from 3001 to 8000, One will have to select such figures from the above random numbers which are not less than 3001 nor greater than 8000.

Fisher and Yates tables **of random numbers** :

These comprise 15,000 digits arranged in twos, which have been obtained by drawing numbers at random from the 10th to 19th digits of A.S. Thomson's 20 figure logarithimic tables.

Kendall and Babington Smiths Table :

These random tables consist of one lakh digits, which are grouped into 25,000 sets of four digits random numbers.

(c) Grid system :

It is used for selecting a sample of area. According to this method a map of the entire area is prepared. Then a screen with squares is placed upon the map. Some of the squares are selected at random. Then the screen is placed upon the map and the areas falling within the selected squares are taken as samples.

(d) Alphabetical list (Selecting from **sequential list)**

Under this plan the names are first arranged serially according to some particular order which may be alphabetical, geographical or simply serial. Then out of the list every 10th or any other number, as the case may be, is taken up. If every tenth unit is to be selected, the selection may begin from 7th and 17th, 27th, 37th units etc. may be selected.

Merits of Simple Random Sampling:

1. Simple random sampling, being a probability sampling, bias caused due to personal judgment or discretion of the investigator is eliminated. At the same time the sample selected becomes more representative of the universe in comparison with judgment sampling.
2. The efficiency of the estimates is as curtained due to random character of the sample and due to estimation of standard errors of the sampling distribution. According to the principle of statistic regularity and principle of inertia of large number, large sample will be more representative of the universe and can provide better results.
3. Simple random sample enables us to obtain the most reliable and maximum information at the least cost-It also enables the researcher to save his time and manpower because it is highly developed.

Demerits of Simple Random Sampling:

1. Although it is easy to draw random samples from finite population with the aid of random numbers tables, this is only possible when a complete list is maintained and items are readily numbered. Hence simple random sampling requires an up-dated population *i.e.* a complete up dated list of the universe. But in real terms it is practically impossible to maintain such a list. Therefore, it restricts the user to go for this sampling design.
2. If the area covered by field survey is large, it is expected that units selected in random sample are scattered widely and therefore it may consume more time and money, involve higher cost for collection of the required information.
3. If the sample size is small it may not represent the universe in miniature and thus it may fail to reflect the true characteristics of the universe. Although it is told that sampling involves less time and engage's' less man power and money, in reality the numbering of units in the entire universe and

preparation of cards and slips becomes quite time consuming and expensive particularly when the universe is large. Moreover, in the social sciences it may not be used so effectively. Simple random sampling usually requires large sample as compared to stratified sampling for gaining greater degree of accuracy.

4. There are instances wherein simple random sampling gives results which is highly I probabilistic *i.e.* probability is very small.

Complex Random Sampling:

Probability sampling under restricted sampling technique may result in complex random sampling. Such sample may be called mix-sampling design because in real terms many of such designs may represent a combination of both probability and non-probability sampling-procedure in selecting sample. The different types of popular complex random sampling are discussed below:-

(i) Systematic Sampling:

It is the most practical way of drawing samples in selecting every *i* th item from a complete list. An element of randomness is initiated into this type of sampling by the use of random numbers and by picking up the unit with which we start. For example, if a 5% sample is required, the first item would be chosen randomly from the first 20 and thereafter every 20th item would automatically be picked-up and included in the sample. Hence, in systematic sample the investigator chooses the first unit at random and thereafter the remaining units are selected at fixed intervals and included in the sample. Of course, strictly speaking a systematic sampling is not a random sample. Nevertheless, it is considered reasonable to constitute systematic sample as random sample.

As regards the merits of systematic, sampling, it can be taken as an improvement over a simple random sampling on the ground that it is spread more evenly in the entire population.

Secondly, the systematic sampling becomes an easy task on the part of researcher and involves less cost.

Thirdly, it can be used conveniently, even in case of large population.

Demerits :

The systematic sample is also not free from its own demerits.

It may prove to be an inefficient method in sampling in case there is hidden periodicity in the population.

Secondly, if all elements of the population are ordered in such a way that they appear the representative of total population, systematic sample is considered analogous to random sampling. But in practices the systematic sampling is used on the basis of availability of lists of population.

Stratified Random Sampling :

When the universe does not constitute a homogeneous group, stratified random sampling technique is resorted to obtain a representative sample. Hence, the technique of stratified random sampling is used to obtain more representative sample if the population is heterogeneous with respect to the characteristics under study. Stratification implies the division of universe into different layers. Therefore, stratified random sampling involves the following step :

The given universe has to be stratified into number of sub-groups or sub-population, known as strata in such a manner

(i) The units within each stratum are as homogeneous as possible.

(ii) There should be marked differences between various strata. *iiii)* The different strata or sub-groups should not be overlapping. The criterion on the basis of which the entire universe is divided into various sub-groups of strata is known as stratifying factor which may be geographical, sociological, economic characteristics of the given universe, such as geographical area, economic status include occupation, level of education, sex etc.

Stratification will be effective only when it fulfills the three characteristics, such as knowing entity of the units in the subgroup, marked differences between various strata and non-overlapping strata, when the distribution is highly skewed, stratification becomes very effective and valuable. For example in a stratified sampling the population size is 'N' and there are

'K^J relatively homogeneous strata of sizes $N, N_1, N_2 \ldots\ldots N_K$

respectively such that $N = \sum_{i=1}^{K} Ni$

2. Simple random samples, without replacement, are to be drawn from each of 'K' strata. Let, simple random sampling without replacement of size *ni* be drawn from the *i* th strata (i =

k

1, 2.....k) such that ∑ ni= n, where n is the total sample size

i= 1

k

from the population size N. The sample of $n = \sum ni$, i varying

i= 1

from 1 to k units, is known as stratified random sample without replacement and technique of drawing such a sample is known as stratified random sampling.

The stratified random sampling has the following basic problems :-

(1) The researcher may be confused while determining different strata or sub-groups.

(2) He may further be confused while determining the sizes of the samples to be drawn from different strata.

In stratified sampling, the allocation of sample size n_i (i = 1, 2, ...k) i.e. the number of units to be selected from the *i* th stratum, is done either by proportional allocation, optimum allocation or disproportionate allocation.

In the proportional allocation, the items are selected from each stratum in the same proportion as they exist in the population.

The ratio of units selected from the stratum to the population size remains the same in all the strata. This principle is mathematically stated as :

$$\frac{n1}{N1} = \frac{n_2}{N_2} = \ldots \ldots \frac{n_K}{Nk} \quad \text{——}$$

In the optimum allocation, the number of units to be drawn from the various strata is determined by the principle of optimization so that : *(a) Variance* of sample estimate of the population mean is minimum. In other words, its precision is maximum for fixed total sample size *n* (6) *Variance* of the estimate is minimum for a fixed cost estimate, (c) for fixed desired precision, the cost of the sampling design is minimum.

r

In disproportionate allocation, an equal number of units are taken from each stratum without having regard to the representation of the stratum in the universe. Hence, the proportion may vary from stratum to stratum. In other words, in a disproportionate stratified sample, the number of units selected from each stratum is independent of its size.

Merits of Stratified Random Sampling :

1. Stratified random sampling, if properly constituted and executed, overcomes the drawbacks of random sampling or purposive sampling.

At the same time, it enjoys the benefits of these sampling methods, because, it not only divides the given universe into different homogeneous strata keeping in view the purposive strata characteristics but also thereafter using the technique of random sampling in drawing samples from each homogeneous sub- groups. Thus a stratified random sampling is capable of giving adequate representation in respect of each sub-group of the population and rules out a possibility of complete omission of any important group of the population.

(a) Since the stratified random sampling provides more representative sample of the universe and thereby results in less variability in comparison with other sampling designs it is considered more efficient than other methods of sampling.

(c) The stratified random sampling is also convenient from the administrative point of view because of the population divisions of universe into relatively homogeneous strata. It also involves low cost and less time in terms of collection of data and supervision of the field work.

(d) The stratified random sampling is also considered an efficient method for obtaining the results of non-precision for each other strata.

(e) Lastly, the stratified random sampling is quite effective in tackling the problems which differ quite significantly in different segment of population by considering each segment as different stratum and by approaching them independently during sampling.

Demerits of Stratified Random Sampling :

Since the success of stratified random sampling is dependent upon effective stratification of the universe into different homogeneous sub-groups, and adequacy of representation in respect of each of the strata, the results will be biased when the stratification is faulty or adequacy is not maintained.

The following are the demerits of stratified **random sampling** :-

(a) The researcher may find it difficult to stratify the universe into homogeneous strata.

(b) Appropriate size of sample which ensured the uses of simple random sampling is not so easy to be determined from each of the stratum.

(c) The error caused due to wrong stratification cannot be compensated even by taking large samples. Therefore, faulty stratification will yield biased results.

(d) In case of dis-proportional stratified sampling if the weight assigned to different strata are faulty the result in sample will not only fail to be representative of the universe, it might also yield biased results.

CLUSTER SAMPLING:

As simple random sampling and stratified random sampling-cause heavy expenses due to the coverage of large and sparsely dispersed population and since the elements chosen in sample may lack uniformity, the total area of interest which happens to be large one, it may be conveniently divided into a number of smaller non-overlapping areas and thereafter a number of these small areas, usually called cluster, maybe chosen with the ultimate sample consisting of all the units in these small areas of cluster. These clusters may be household, city wards, or various social units. However, simple random sampling methods are used to make the sampling of clusters from the universe ; then from these selected clusters the constituent elements are drawn on the basis of random sampling, *e.g.* if a social scientist desires to conduct a sample study of the problems of the aged in villages of a district, he may proceed as follows :-

First, the researcher enlists all the villages within the district and selects a sample through the procedure ensuring random method. For each of the villages included in the sample, he makes a list of aged persons under its jurisdiction, above a particular age. For each of the villages he selects a sample of desired size on the basis of simple random sampling. In this way the researcher sets probability of random sampling of elements which are consistent either to a lesser or greater extent. Thus the researcher becomes capable of avoiding enormous expenses which would have been incurred in course of simple random sampling procedure. When the researchers researcher resorts to sub sampling within the clusters it is called *Multi-stage sampling* because the sampling technique is carried out in various stages, progressively from more inclusive to less inclusive sampling units until he obtains his desired sample. Thus it is a further development of the principle of cluster sampling. It is more flexible than any other methods of sampling.

Although the cluster sampling reduces cost by confining survey to selected clusters, it certainly is less precise than random sampling. Relatively speaking, the marginal error is much greater in cluster sampling. Nevertheless, it is only used because of the economic advantage and due to the fact that estimates based on cluster sample happen to be more reliable per unit cost.

Area Sampling : While using the cluster sampling the researcher must bear the following points in the mind.

(i) The cluster should be as small as possible. *iii)* The cluster should be consistent with cost and limitations of the study must be based **on** some geographic. The clusters sub divisions.

(iv) Each cluster should contain approximately the same number of sampling units. Therefore, the cluster sampling is not found suitable if the sample comprises the private residential houses, business or industry complex, apartment etc., depicting wide variations in the number of social units.

Haphazard or Convenience Sampling :

As the name suggests, the convenience sampling is not based on any system, rather it is chosen according to the convenience of the researcher because if otherwise rigid sampling procedures are followed, the researcher may find it difficult to contact all the chosen units due to lack of money, man power and time at hand. Nevertheless, the purpose of research is not lost sight of.

Quota Sampling :

Quota sampling may be viewed as a special form of stratified sampling wherein the investigator is told in advance the sample size, so that he is aware of the number of sampling units to be examined o$ enumerated from the stratum, assigned to him. The quota of the units to be examined by the researchers from the st ratum assigned to them is fixed for each investigator. While fixing up sampling quotas certain specified characteristics are to be taken into account, such as income group, occupational category, political or religious affiliation, sex etc. However, the investigator has his own choice to select the particular units for investigations. They are merely assigned the quotas with clear-cut instructions to interview a specified number of units from each stratum. Although the investigator has got the choice for selecting randomly sample units, nevertheless, instead of making a random selection, the researcher usually applies judgment and discretion in choosing the sample. Simultaneously, he also tries to get the desired information as quickly as possible. At the same time, in , case of non-response from some of the selected sample units due to certain reasons like non-avail ability of the informant, in spite of repeated calls by the researcher or the inability or refusal of the respondent to give the required information, the investigator chooses some new units himself to maintain his quota. Thus he is most likely to incorporate some purposive units to obtain the required data.

Merits:

The quota sampling enjoys the advantages of both a stratified-cum-purposive sampling and judgment sampling. Thus it provides the best use of stratification without incurring heavy expenses involved in resorting to any probabilistic method of sampling. That apart, it also saves time and money because of the closeness of sample units to be investigated.

Secondly, quota sampling is likely to yield quite dependable results, provided it is carefully executed by the skilled and experienced investigator.

Demerits:

Quota sampling also suffers from the limitations of judgment or purposive sampling.

(i) It is likely to biased due to the personal beliefs or prejudices of researcher while selecting the units and inspecting them.

(ii) It may involve the biases on account of the replacement or substitution of sample units for those who have no responses.

(III) Estimation of sampling error is not possible because it is not based on random samples. Notwithstanding the shortcomings, quota sampling is adopted in market survey, political survey or in surveying the opinion poll, particularly when the researcher finds it difficult or, at times, impossible to make identification of different strata in the universe.

Judgement Sampling :

In judgement sampling the investigator's judgement is taken into account for selecting items, considered to be representative of the universe. For instance, a judgement sample of college students might be taken to obtain their reaction to a new method of teaching. The judgement sample is taken into consideration to obtain data quite frequently on qualitative research work so as to develop hypothesis rather than generalizing the research universe.

2.7 MEASUREMENT

Measurement is defined as assignment of numerals to rule. The measurement of variables forms the core of the research process. The main problem of the study of social phenomena is that the basic variables can not sufficiently be precise. But with the development of social research, researchers in various social fields have been evolving more accurate measurements oft. the qualities or by assigning numerical symbols of qualities doubt that quantification facilitates the establishment of s acknowledged that variety of forms of measurement are in e considered measurement as the process of assigning numb some other writers like Weyl, Stevens and Coombs held i any empirical procedure that involves the assignment of s; according to rules. Numerals constitute only one such type.

Although it is easy to assign numeral in respect of properties of some objects it is relatively difficult in respect of others. For example, for measuring the physical weight, biological age or annual income we can easily assign numbers but simultaneously it is relatively difficult to measure such things as social conformity, marital happiness or level of intelligence. However, in any case, the purpose of measurement is to make the events or subjects being analysed amenable to mathematical description and analysis. What is measured is not the object, person, state or event itself, but some characteristics of it. Hence

while counting the objects; we do not actually count measuring themselves, but only their charactertic of being present. For example, we never measure individuals, only take into account their characteristics age, weight, height etc.

2.7.1. Levels of Measurement

Nominal Scale: 'Nominal' is defined in the dictionary is 'of, pertaining to, or consisting in a name or names". As such a nominal scale is one that consists of two or more named categories into which object or individual or responses are classified. As the most elementary methods of measurement it classifies persons, objects or events into a number of mutually exclusive categories in terms of their presence or absence or on the basis of possession or non-possession of certain property or even in terms of applicability or inapplicability. If simply describes the differences between by assigning" them to categories. Therefore, the fundamental requirement of such scale is the ability to distinguish two or more categories relevant to the attributes being considered and specify criteria for placing individuals, etc. in one or the other category In simple terms, nominal measurement is 'have' vrs, 'have not' type of measurement which assigns mutually exclusive labels to identify objects by specifying relationship between the categories in terms of their difference from each other rather than implying that they represent the more or less of the characteristic. Thus the population of a district ma he classified according to religion into Hindus, Muslims and Christians or according to sex in to male and female and each social category be identified either in the form of numerals or in the form of letters. But no empirical relationship exists among the numbered categories that corresponds to the mathematical relationship between numbers assigned. The counting of numbers in each category is the only possible arithmetic operation when a nominal scale is employed. Therefore such statistics as are appropriate to counting, the mode, the coefficient of contingency and the chi - square may be utilized.

Use or nominal scales is made where the focus of the study is on t he pattern of relationship among several characteristic of a person, where the focus is on uncovering a relationship between two characteristics rather than on he mathematical form of relationship or specification of some degree of precision.

Ordinal Scale: An ordinal scale defines the relative position of objects or individuals with respect to a characteristic, with no implication as to the distance between

positions, although it places event in order. It is the lowest level of the ordered scale. This type of measurement necessitates that the objects of a set can be ranked in terms of an operationally defined characteristic. The basic requirement of this scale is that the] researcher should be able to determine whether, each object or individual being measured, is in possession of the attribute of the same amount or less than another individual. In order to determine the order of position, one must be able to place each individual or object at a single point with respect to the attribute in question. This type of measurement is possible in so far as the ordinal transitivity postulate is justified. We can describe such a procedure in three ways. First, measurement in respect of any number of objects of any kind is made cordially merely by extension to a, b, c….n, if a>b>c...>n on a particular characteristic. As regards the second way, the researcher can describe it by ranking the objects on the combined criteria of composite merit ratings scores. Third, description can also be made by using criteria like "less than", "proceeds", "is superior to" or "is above" etc. In this way the researcher need not always use the criteria "greater than" only.

As the relative position of two or more objects on some characteristics are indicated in the ordinal data, a significant amount of market research, particularly consumer-oriented research relies on them. It is quite useful in obtaining preference measurement in respect of several brands of flavors of package designs. The ordinal scales simply indicate rank order and do not provide data on how much more or less of the characteristic various objects possess. As regards the statistics applicable to this type of measurement, the researcher can use median, percentiles and rank order correlations, in addition to those applicable to nominal scales.

Interval Scales: In the internal scale not only the position of objects or persons are arranged in terms of equal to, less or greater but also the un its or intervals are equals in other words, the interval or equal measurement, a part from having conjunction with the characteristics of the first two types of scales, *i.e.* nominal and ordinal also involves equal distances in the quality being measured. For example, the distance between the positions labelled 2 and 3 on the scale is equal to the distance between the positions and 4 etc. The equal interval further ensures that the intervals can be amendable mathematical relation in terms of addition and subtraction.

The primary drawback of the interval scale is the lack of a true zero and its capacity to measure the complete absence of a trait. Moreover, the researcher adds subtracts merely intervals or distances rather than quantities or amounts. As the interval scales incorporate the concept of equality, they provide more powerful measurements than the first two scales. Therefore, more powerful statistical measures like mean as the appropriate measure of central tendency, standard deviation as the measure of dispersion and product movement correlation technique can be used with interval scales. That apart 't' test and 'F' test can also be employed for statistical significance.

Ratio Scale: This scale contains all the characteristics of interval scale. In addition to that it also contains an absolute zero, because ratios are meaningless unless there is an absolute zero point. The empirical operations necessary to establish a ratio scale are not only the methods for determining equivalence, non-equivalence rank order and the equality of intervals, but the equality of ratios.

With a ratio measurement, often called fundamental measurement, the comparison of the absolute magnitude of the numbers becomes possible. Measures of weight, time intervals, length, area, angles etc. all conform to ratio scales. For example, a person weighing 88 kg. is said to be twice as heavy as one weighing 44 kg. and a person weighing 132 kg. is three times as heavy. Generally all statistical techniques are usable with ratio scales. Manipulation with real numbers can also be carried out with ratio scale values. If one's data conform to the criteria for a ratio scale, all the relations between numbers in the conventional system of mathematics such as multiplication and division etc. obtain between the correspondingly numbered positions on the scale.

The physical scientists have the advantage to describe variables in ratio scale for maintaining precision in description, but for most of the subject matter of the behavioural sciences, the social scientists are far from having devised procedures that confirm to the requirements of a ratio scale, although Stevens maintains that attempts have been made to construct such scales in connection with the judgement of psychological attributes such as loudness and pitch.

2.8 : Reliability and Validity: Meanings and Types: Reliability of a measure implies its capacity to yield the same results in repeated applications to the same

sample of respondents or events. A measuring instrument is said to be reliable if it is capable of providing consistent results. The reliability of a measuring instrument can be defined as the extent to which it measures consistently what is seeks to measure. Although a reliable measuring instrument contributes to validity, a reliable instrument need not be a valid instrument. We may take an example in this regard. An instrument meant for measuring weight consistently overweighs objects by seven kgs. is no doubt a reliable scale, but this does not ensure a valid measure of weight. On the contrary a valid instrument is always reliable. Hence it naturally appears that reliability is not as valuable as validity although in comparative terms it is easier for the purpose of testing.

A reliable measurement is characterised by a small error element and not prone to random fluctuations from time to time, unlike an unreliable measuring instrument which resembles an elastic ruler, providing a different reading each time and making the observed scores quite unreliable.

Since different sources of error are in operation in different ways, different methods ascertain reliability are followed by either singly or in a combined manner, depending on the probable errors and the cost of each method in the situational context.

The Coefficient of Stability: As a very important aspect of reliability, stability is concerned with securing consistent results with repeated measurements in respect of the same individual and with the same instrument. The degree of stability is determined on the basis of fluctuation in the quality being measured or variation in transient personal or situational factors by comparing the results of repeated observations. The various causes of inconsistency may be attributed either to the actual changes in the quality being measured or changes in extraneous conditions. The inconsistencies may also arise out of the impacts of repeated administrations themselves. The behavioural scientists are interested in ensuring the typical position of the person in respect of a particular quality and its variation in a specific period of time. In order to ascertain the stability of results of a measuring instrument, a coefficient of correlation or some other index of agreement between the scores of two administrations is calculated. The degree to which the measure unfolds the differences among persons on a particular quality is indicated by the stability co-efficient. This procedure, called test-retest, is suitable for one form of a measure, such as questionnaire or interview or any one version of a test. The measure is applied a

couple of times to the same individuals at two different points of time and the scores so obtained are correlated. The higher the coefficient of stability, higher is the reliability of the measure. The basic assumption underlying this method is that the attitude, memory and practice from the first administration will not affect significantly the scores obtained on the second administration. However, complications may arise out of the memory or reduced anxiety, interest and motivation of the individuals on account of their familiarities with the instrument administered during the retest. Moreover due to the increasing interest and stimulation, the proceeding measurement is likely to effect actual change in the quality being measured. Another probable cause of a genuine change in the quality being measured may be attributed to the individual's acquiring of more information or his shift in attitude during the second administration.

Coefficient of equivalence: This aspect of reliability considers how much error may get introduced by different researchers or different samples being studied. While determining the equivalence, the following considerations are made: *(i)* the extent to which different investigators administer the instrument to the same individuals almost at the same time to yield the consistent results, and *(ii)* the extent to which the administration of different instruments to the same individuals almost at the same time provide consistent result. Whereas the justification with the former condition is that it is capable of providing comparable results, the rationale behind the latter condition is that the estimates of the equivalence of the positions of individuals on different instruments relates to differences caused by the items sampling. The correlation between the scores on the two forms of instruments, purporting to measure the same quality or behaviour, indicates the extent to which they are measuring the same quality insistently.

Coefficient of Homogeneity: Equivalence of results based on different samples items can also be estimated on the basis of internal analysis of responses to test items. This technique for assessing reliability is adopted when only one form of the Measure is available. The measure is taken through the determination of interval consistency. A high degree of inter correlation among items that comprise the measure indicates that a test or questionnaires etc. is internally consistent. The spilt half method of estimating reliability involves the administration of a single form of test only to group of individuals only once and thereafter its scores are divided in two halves

scores of the two halves, which may be construed as alternate forms of the same measuring instrument are correlated and the coefficient, known as the coefficient of inter consistency, is taken as a measure of reliability. Thereafter the co-efficient of equivalence for the entire test or the corrected split half reliability determined on the basis of Spearman-Brown formula. A popular technique for spliting a measure into two equivalent' halves is to assign the odd-numbered items to one sub-measure and the even numbered ' items to another sub-measure.

VALIDITY

Validity: Validity tells us whether the scale measures what it is supposed taw measure. As the most critical criterion of measurement, validity implies that the measure should be free from consistent or systematic error so that it is able to measure what; is purports to measure. Selltiz, Wrightsman and Cook have defined the validity of aa instrument as the extent to which differences in scores on it reflect true differences among individuals on the characteristics which the researcher intends to measure rather than constant or random errors. The social scientists have not yet been able to create any completely valid instrument to measure the complex human behavior and performances. Nevertheless, attempts have been made to reduce the external influences other than the characteristics being measured so as to enhance the validity of a measuring instrument. Due to the behavioural scientist's inability to determine validity with direct confirming knowledge, barring- *a* few cases for determining pragmatic and practical-validity, he seeks other relevant evidence that confirms the answers he has found with his measuring tool. The nature of the research problem and the judgment of the researcher often determine the relevant evidence. Basically, the researcher considers:-three types of validity; content validity, construct validity and criterion related validity.

Content Validity: The extent to which a measuring tool provides adequate coverage, of the topic under study speaks of its content validity. But when the measure contains, items which measure situations or subject matter beyond the domain of interest, the content validity suffers. For example, a questionnaire constructed to measure the I.Q would lack content validity if it covers items which measure the job satisfaction. Moreover, when the items in a measure do not adequately represent the relevant domain the content validity also suffers. For example, a test of a teacher's proficiency in science teaching would lack content validity if it measures his ability

in all other subjects. On the other hand, if the instrument contains a representative sample of the universe, the content validity would be considered good. Hence, for demonstrating the content validity of a test, proper definition of the universe of items must be defined in such a manner that the measure created for the purpose must have a representative sample of these items. In the absence of any numerical way for the expression of its goodness, it can be determined on the basis of intuition and judgment of a panel of persons who shall embark upon determining how well the measuring tool meets the standards. For example, a researcher interested in studying the teaching proficiency might incorporate a large number of questions for a questionnaire measure of teaching proficiency and thereafter have a group of judges who would judge the extent to which the items in the questionnaire represent the relevant universe.

Construct Validity. Construct implies the abstract variables which cannot be directly observed, such as motivation, morale, intelligence etc. The concept of construct validity which involves propositions about the relationships of construct to other variables construct or observable behavior was coined by L.J.Cronbach and P.E. Meehl. For determining the construct validity the investigator associates a set of other proposition with the result received from using the measurement. A network of highly complex deductive system linking several variables are involved in these propositions. Even a small set of variables may be linked by independent propositions.

Since the values of abstract variables are not subject to direct observation and can only be inferred from other observable variables, the latter are made operational definitions of these constructs by the investigator. The appropriateness of the operational definitions of the construct constitutes the construct validity of a measure. However the establishment of the appropriateness involves a lengthy process necessitating a series of empirical investigations. At the outset, the investigator embarks upon developing operational definition or measure of the construct and thereafter establishes hypothesis relating to scores on this measure with one or more variables which are subject to direct observation. He may also state the relationship between this particulars construct under investigation with other constructs. The set of tentative generalisations explaining ' how the variables are interlinked is described as a nomological network. Then the investigator undertakes one or more empirical studies to establish the relationship between the operational

definitions of the construct being validated with the other variables in nexus in a hypothesised manner. If the data support the predictions inherent in the nomological network, we may conclude that there is some construct validity to the extent that it does. Otherwise the investigator must either modify the operational definition of the construct or abandon the construct entirely, depending upon the extent to which the empirical data do not support the hypothesis of the nomological network.

Criteria-related validity: This form of validity relates to the researcher's ability to predict some outcome or estimate the existence of some current condition, reflecting the success of measures employed for some estimating purpose. The criterion may be either behaviour or an attitude and must possess the qualities like relevance, freedom form bias, reliability and availability. The typology of criterion related validity is made on the basis of the measurement of the predictor and criterion variables, Broadly speaking, there may be three different types of criterion related validity : (i) Predictive validity, i*i)* Concurrent validity, and *(iii)* Post-dictive validity.

(i) Predictive validity: It refers to the usefulness of a test in predicting some future performance and for that purpose the predictor variable is measured at one point of time and the criterion at some future point. For example, the test score of the Engineering students at the entry level may be taken into account to infer the probable performance of the students at the end.

(ii) Concurrent validity: As the name suggests, the predictor and criterion variables are measured concurrently or at the same point of time to determine this type of criterion- related validity. It distinguishes between persons who are different on the basis of their present status or position. For example, the present salary structure may be taken into account to infer employee’s satisfaction with salary.

(iii) Post-dictive validity: Unlike the case of predictive validity in post-dictive validity the criterion variable is measured first at one point of time and thereafter the predictor is measured at a later point. For example, in postmortem corpse is medically examined to determine the cause of death.

2.9. Key Words:

Research Design, Research Problem, Sampling, Levels of Measurement, Validity, Reliability.

2.10. References:

1. Whilty, Frederick, The Elements of Research
2. Goode and Hatt, Methods in Social Research
3. Young.P.V., Social Survey and Social Research
4. Kalton & Moser, Survey Method in Social Investigation
5. Lundeberg, Social Research

UNIT-III

UNIT-III

3.0 OBJECTIVES

-To know the sources of data.

-To learn about different methods and tools of data collection and their advantages and disadvantages.

-To learn about electronic resource and its usage.

3.1 SOURCES AND METHODS OF DATA COLLECTION

The first step in the statistical approach to a problem is the collection of numerical facts of data. These data are the raw material for the final statistical conclusions and the quality of data will greatly affect the conclusions. Thus, utmost importance must be given to this process and every possible precaution should be taken while assembling the facts. Various forms of data and the necessary steps to be followed for collection of data are discussed in this chapter. There are two types of statistical data such as primary data and secondary data. The primary data are those which the investigator originates for the purpose of specific enquiry in hand. Thus, if it is desired to study the response of the students to the new method of teaching and if the relevant facts are collected by the investigator or his research worker, such data would be termed as primary data. On the contrary, secondary data are those which are not originated by the investigator himself, but he obtains these from the records of someone else. Thus, if one desires to analyse the weather condition of different regions, one can get the required infurmation from the records of Meteorology department. These data will be called secondary data. It may be noted that the data which are viewed primary for one person may be considered secondary for another. In the above example, weather data will be primary to the meteorology department, but for a person who makes use of these data for further research, these data are secondary. In the Indian context, following are the few sources that supply the secondary data related to various social and economic aspects : (i) Statistical Abstract of India issued annually by the Union Government, («) Statistical Abstract of various States issued annually by the State Economic and Statistical Organisations, iiii) Survekshana issued quarterly by National Sample Survey Organization of India, (iv) Monthly Index of Whole sale Prices issued monthly by Labour Bureau, Government of India and (v) R.B.I. Monthly Bulletins issued by Reserve Bank of India.

The Advantages of Secondary Data are : (a) The researcher, by using secondary data, is capable of reducing considerably the cost and time involved in an investigation. At the same

time, he can utilise these saved resources for the improvement of quality of the analysis and interpretation of the findings ; (6) Most often the secondary data is more reliable as it is collected by specifically trained personnel for the purpose of research ; and (c) The researcher is in a position to increase the scope of inquiry in terms of area and time involved.

Demerits of Secondary data : This is not to say that, the secondary data are free from all demerits. In real terms, most often the exact definitions of terms and units used in a secondary data are unknown. For example, if the data relate to the wages of workers, one may not be able to know whether those include bonus and over-time allowance, etc. or not. Secondly, if one obtains data from two sources, it may be difficult for comparison in terms of definitions, units and time period covered. For example, if the data related to production of paddy, taken from two different sources, in one source it may be recorded in tons while in the other the value of output may be recorded in rupees. Thirdly, the definition of terms may not be uniform over time in a time series wherein data are taken over different points of time. Fourthly, data collected from primary sources usually show greater details, compared to secondary data wherein some information may be missing or some categories are often pooled together.

Fifthly, while copying the figures from the primary sources, the investigator or his representative is likely to commit mistakes.

3.2 METHODS AND TOOLS OF OBSERVATION

3.2.1 OBSERVATION

Observation method has occupied an important place in descriptive sociological research. It is the most significant and common technique of data collection. Analysis of questionnaire responses is concerned with what people think and do as revealed by what they put on paper. The responses in interview is revealed by what people express in conversation with the interviewer. Observation seeks to ascertain what people think and do by watching them in action as they express themselves in various situations and activities.

Observation is the process in which one or more persons observe what is occurring in some real life situation and they classify and record pertinent happenings according to some planned schemes. It is used to evaluate the overt behaviour of individuals in controlled or uncontrolled situation. It is a method of research which deals with the external behaviour of persons in appropriate situations.

According to P.V. Young, "Observation is a systematic and deliberate study through eye, of spontaneous occurrances at the time they occur. The purpose of observation is to perceive

the nature and extent of significant interrelated elements within complex social phenomena, culture patterns or human conduct".

From this definition it is clearly understood that observation is a systematic viewing with the help of the eye. Its objective is to discover important mutual relations between spontaneously occurring events and explore the crucial facts of an event or a situation. So it is clearly, visible that observation is not simply a random perceiving, but a close look at crucial facts. It is a planned, purposive, systematic and deliberate effort to focus on the significant facts of a situation.

According to Oxford Concise Dictionary, "Observation means accurate watching, knowing of phenomena as they occur in nature with regard to cause and effect or mutual relations".

This definition focuses on two important points. Firstly, in observation the observer wants to explore the cause-effect relationships between facts of a phenomenon. Secondly v facts are watched accurately, carefully and recorded by the observer.

3.2.2 CHARACTERISTICS OF OBSERVATION

Observation method is a scientific technique used for collecting data. It is different from the casual observation of men in the street. Every casual observation is not a research observation. More specifically, scientific observation must comply with the following characteristics.

1. Observation is a systematic method

Observation is not haphazard or unplanned. The length of the observation periods, the interval between them, the number of observations, the area or situation of observation and various techniques used for observation are carefully planned. Often there are systematic managements for controlling the situation if special factors are to be studied, for example study of honest behaviour, sportsman spirit, leadership qualities etc.

2. Observation is specific

It is not just looking around for general aspects of human behaviour. Rather it is directed at those specific aspects of total situation which are assumed to be significant from the stand point of the purpose of the study. The layman may frequently overlook what is crucial while observing an event or phenomenon, but the scientific observer should look for some definite things which suit his purpose of study so as to economise his time, money and effort for observation.

3. Observation is objective

Observation should be objective and free from bias as far as possible. It should generally be guided by a hypothesis. The observer must maintain ethical neutrality. He must consider hypothesis as something to be tested. But at the sametime he must maintain a flexible attitude, so that he can deviate from his original plan when such deviation appears inevitable.

4. Observation is quantitative

Although many important phenomena can not be quantified, it becomes almost an imperative to use some means for quantifying Observations in order to increase their precision and to facilitate their analysis. Even the quality should be converted into quantity, because qualitative data is subjective and quantitative one is objective and can further be interpreted in objective manner.

5. Observation is an affair of eyes:

P.V. Young remarks that observation is a systematic and deliberate study through eye. An observer gathers the data which he has seen in his own eyes. Collecting information through eyes is probably the most trustworthy technique of data collection in social research.

6. Definite aim:

Observation must have some definite aims and objectives. It should be clearly defined before the beginning of the actual observation process. Without the proper aims and objectives observation will be unsystematic and expensive.

7. The record of observation is made immediately

During the observation period it is very difficult on the part of the observer to remember each and every element of observation. He may forget much important information. If we rely on memory the factor of forgetting will enter and affect the data of observation. Therefore the observer should record all important information as soon as the observation is completed.

8. Observation is verifiable:

Observation result can be checked and verified. Observation must be verified with usual criteria of reliability, validity and usability. It may be possible to check the findings of the observation by comparing the results of different observers by repeating the study.

3.2.3. ORGANISATION OF OBSERVATION METHOD

In organizing an observation one has to go step by step in a systematic manner. The following is an adequate sequence of steps, which can be followed for the organisation of observation.

(1) Determination of the method of study:

If the research is to be conducted with the help of field-observation, the first thing to be decided is whether the phenomena can be studied by observation or not and whether the respondent will allow it to be observed by an outsider or not.

When this has been solved, a plan of action for observation has to be decided as is required for the success of the study. Generally it should focus on the topic of the study and hypothesis. The observer should also decide what kind of information would be beneficial for the test of his hypothesis or for making proper generalizations on the topic.

Once the observer has decided the above mentioned procedure, he has to choose about the proper method of study ; whether he is going to use participant or non-participant observation ; controlled or non-controlled observation. A great care has to be taken for choosing the methods of study because a wrong choice of method will make the entire research fruitless. Usually the observer should select the method of study in accordance with the nature and objectives of research problem.

Planning for observation

Observation as a research technique must always be directed by a specific purpose. It requires the expert knowledge of the observer to plan it systematically much before the start of the actual process. So observation is neither hapazard nor unplanned. The planning of an observation requires so many things. The factors one must consider in detail while planning to employ observation method are :

(i) Determination of nature and limitation of observation :

Definition of specific activities or units of behaviour to be observed should be determined by the researcher much before the observation. This helps him to decide what should be observed and what may be left out.

(ii) Determination of time, place and subjects of the study :

Before the observation, the observer has to take right decision regarding the time, place and subjects of the observation. The

observation may be of a short duration or may continue for a longer duration. The observation may take place in a natural surrounding or on a non-controlled basis or it may take place on a laboratory type experimental basis. If the observer has to go for a non-controlled observation, he has to find out a proper place where he can study the proper incident and if he wants to conduct a controlled laboratory experiment, he has to go for the

necessary arrangements. For any type of observation, the observer has to determine the right type of people and the right place for his observation.

(iii) Determination of investigator

Observation is a technical job. It can be either carried on by the researcher himself or requires a team of trained persons. If the observation has to be carried on by a team of field workers the researcher has to select the right type of field workers and he has to arrange proper training for them.

(iv) Training of the observer

It is necessary for anyone who conducts an observational study to undergo extensive training in order to ensure validity and reliability in his observations. Training in _ observation is particularly important in social sciences, because the observer is . frequently faced with determining which factors are significant out of the multiple causations occurring simultaneously. The observer should be carefully selected and in the course of training, a check should be made of their suitability for the observation of the particular phenomenon.

(v) Tools for observation

Various tools have been devised to help the observer in making more objective and reliable observations and to systematize the collection of data.

According to Kulbir Singh, tools for observation includes the following :

(a) Check Lists and Schedules

These instruments make a list of items which are significant for the observation. After each item a space is provided for the observer for brief description about that particular item. These tools enable the observers to record different information quickly and rapidly and ensure that they donot miss any relevant information. It also helps for the uniform classification of data. Some check-lists are designed in a such a manner that the researcher can arrive at a score that enables him to make comparisons with other data.

Schedules are also commonly used for this purpose. The researcher has to take great care for the drafting of these schedules. Before the actual operation one should test these schedules and make certain changes if necessary.

(b) Time sampling

This technique helps the observer to record the frequency of observable forms of occurrences during the number of definite time intervals that are systematically spaced. For

example, if a Professor desires to know the type of activity of a student, he can record observable forms of behaviour that his student exhibits during a specified five minute interval in Sociology class each day for a fortnight rather than recording everything the student does. The Professor may only tabulate the occurrences or non-occurrence of one objectively defined form of behaviour such as "the frequency of classroom participation".

Time sampling is an adequate technique because it permits observable instances of behaviour to be quantified directly. By taking into account the number of observations, a score can be obtained, which shows the number of times the subject exhibited a particular form of behaviour. These obtained scores lend themselves rapidly to statistical treatment.

(c) Behavioural diaries

Behavioural diaries are the informal methods used to collect data. When a respondent is participating in a significant incident the observer may note down a factual statement about what the subject said or did, record various information from this incident or describe various situations in which it occurred. After collecting a series of information over a period of time, the observer may gain considerable insight regarding the life of the respondent.

Some sort of daily diaries must be filled by the observer at the end of everyday observation. It should contain some of the important headings and sub-headings. It helps for the recording of different informations in an easier way and helps for the classification of data.

Anecdotes : According to Lokesh Raul, the anecdote has been the most widely used method for describing naturalistic behaviour. It is a word description of a behaviour episode. There is no set pattern for anecdotal writing, and various styles have been used for it.

Anecdote descriptions have served as useful tools in behavioural research. Such descriptions have been extensively used in examining classroom practices ; in analysing community influences on development ; in assessing change in interpersonal behaviour of hyper aggressive children undergoing residential treatment.

(d) Mechanical Instrument:

When several researches describe the incident, their reports often vary because of their personal bias, selective perception, emotional involvement or capricious memories. But mechanical instruments are unaffected by such factors. These can obtain various information of an event in an accurate way.

Motion picture and sound recording, for example, preserve the details of an incident in a reproducible form which can also be checked by other research workers. Films can be used to analyse audience reactions, to make slow-motion analysis of complex activities that could not have been studied under normal conditions and to serve many purposes. Many mechanical instruments not only give a reliable account of what happened, but also report it in a quantified form.

But although mechanical devices provides more refined and reliable data than human observers, they are subject to certain limitations, these can only be employed more easily in carefully controlled laboratory experiments. The use of these instruments sometimes alter the behaviour and activities of the respondents. It is also very expensive and time consuming.

Data Collection:

After arranging all necessary equipments required for research the researcher collects data through observation to test the hypothesis. However one should remember that the data collected should be thorough and no significant data should be missed.

According to Good (1966) planning for observation includes the following factors :

(i) An appropriate group of subjects to observe.

(ii) Selection and arrangement of any special condition for the group.

(iii) Length of each observation period, interval between periods and number of periods.

(iv) Physical position of the observer and its possible effect on the subjects.

(v) Definition of specific activities or units of behaviour to be observed.

(vi) Entry of frequencies or tallies in the record, as a total for the entire observation period or by sub-division of time within the observation period.

(vii) Scope of observation, whether for an individual or for a group.

(viii) Form of recording, including consideration of mechanical techniques and such factors as number, time, distance and spatial relationships.

(ix) Training of the observer in terms of expertise. (x) Interpreting observation.

Execution of Observation

A good observation plan may not lead to success unless it is followed up with skill and resourcefulness. An expert execution demands :

(i) Proper arrangement of special conditions for observing.

(ii) Assuming the proper physical position for observing.

(Hi) Focusing attention on specific activities or units of behaviour under observation.

(iv) Observing discretely, the length and number of periods and intervals decided upon.

(v) Handling well the recording instruments to be used. (vi) Utilising the training received in terms of expertise.

Recording and interpreting observation:

There are two common procedures of recording available for observation. They are :

(i) Simultaneous Recording : When the observers carry on recording their observation simultaneously with the occurrence of the phenomena observed, it is called as simultaneous recording.

(ii) Post-Recording : At times the observer undertakes to record his observations not simultaneously with the actual observation process, but immediately after he has observed a unit of time, when the details are fresh in his mind. Using any method out of these two depends on the nature of group, the type of activities or behaviour to be observed and the skill of the observer.

The use of various devices, such as checklists, rating scales, score cards etc. help in the proper analysis and interpretation of data collected through observation. All the data should be classified and tabulated to facilitate in drawing general conclusions.

(Hi) Generalisation : When the observer has interpreted his data, it becomes easier for him to draw various generalizations. While generalizing, the observer should follow the law of logic and statistical method.

3.2.4. TYPES OF OBSERVATION :

The observer needs to know beforehand the type of observations he is to make ; whether he has to simply note the occurrence of certain events or to make a judgement of their intensity duration as well as the apparent effect. Observation may be classified into different types according to the method used and the type of control exercised. The following are the chief types of observation.

(1) Participant Observation :

The participant observation means watching the events or situation or activities from inside by taking part in the group to be observed. He freely interacts with the other group members, participates in various activities of the group, acquires the way of life of the

observed group or his own, studies their behaviour or other activities not as an outsider but by becoming a member of that group.

Goode and Hatt define participant observation as "the procedure used when the investigator can go disguise himself as to be accepted as a 'member of the group". So in this kind of observation the observer has to stay as a member in the group he wants to study.

According to P.V. Young, "the participant observer using non-controlled observation, generally lives or otherwise shares in the life of the group which he is studying".

Some of the examples of studies using the method of participant observation are : W.F. White's study of Cornville social and Athletic Club and P.V. Young's study ofMolokan people. The famous studies of Margaret Mead on primitive societies were also based on participant observation.

For the success of participant observation it is essential that the respondents being studied should not have any doubt about the intention of the research worker. A fruitful result of participant observation is very much dependent upon the resourcefulness, tactfulness, personality manners and wit of the research worker.

Advantages of Participant Observation

The following are the merits of participant observation:

(1) Observation of natural behaviour :

The natural behaviour of the respondent can be studied by participant observation. When a group knows that they are going to be observed by a stranger, they feel conscious, uncomfortable and therefore naturality in their behaviour and activity is lost. But in case of the participant observation, the respondents do not know that they are being observed. So their behaviour is not constrained by the conscious feeling of being observed by a stranger.

(2) Closeness with the group

In participant observation, the observer has a very good rapport with the respondents. He has a very close primary relationship with the group members. Because of this he can participate in all activities from a close angle and thus can better interpret the situation than a non-participant observer.

(3) Studying the real character:

Often in order to study the actual behaviour, the group research requires close participation and contact with the group members. Through participant observation the observer can make an intensive and inclusive study of the group and can gain into the real character of such group.

(4) Better Understanding:

In participant observation the observer can better understood the feeling of the respondents than an outsider. For example, a person who is actually living in a slum area can realise the feeling and hardship of the slum dwellers in a better way than an outsider.

(5) Participation provides opportunity to learn more about an event

The chief advantage of participant observation is that in it the observer gets an opportunity to interact with the group regarding various activities of them. He can thus learn the significance of these activities that are actually not open for observation. For example, if an observer participates in a religious ceremony of a tribe, viz. "Chaitra Parba" of Gadaba tribe, he can not only observe different aspects of the ceremony but also clear his doubt by asking various questions to the group members or learn more about that ceremony by discussing with the group in ""this regard. It is generally easier for the respondent to describe about the event on right occasion than before or after it.

Disadvantages of Participant Observation

In spite of above advantages of participant observation it has also many disadvantages. The following are the disadvantages of the participant observation.

(1) Lack of objectivity:

By becoming members of a group and participating very closely in it, the observer may lose his objectivity. His emotional and sentimental association with the group kills his impartiality and unbiased analysis. He may develop some soft corner for those group members and because of this; he may often justify their evil activities as just activity.

(2) Often close association brings biased interpretation:

Because of his close association and emotional participation with the group members the researcher creates a special position for himself in that group. He may be influenced or pleased by this and begins to support them blindly. Due to this he observes the things from his own personal point of view rather than scientific point of view.

(3) Misses important issues due to familiarity

Due to much familiarity many crucial events appear to the participant observer as of little or no significance. Therefore, he misses many of the important issues. But a stranger pays much attention even to a small thing, as this appears' new to Kim.

(4) Limited range of experience

In participant observation the observer is confines himself to a particular group. So his experience becomes very deep, but the range of his experience becomes very limited.

(5) Involvement in groupism

The active participation and proximity of the observer with the group may involve him in quarrels and group factionalism. He can not avoid taking side of one faction. But if he does so, he' loses his status as an impartial observer whom everybody is ready to co-operate. So it destroys the very purpose of the research and the researcher finds it very difficult to get proper information from the group.

(6) Limits of participant observation

There are certain situations in which the participant observation is not possible. For example, it is not possible to observe criminals or prisoners.

(2) Non-Participant Observation

When the observer observes the group passively from a distance without participating in the group activities, it is known as non-participant observation.

Here he does not try to influence them or take part in the group activities.

However, purely non-participant observation is extremely difficult. One cannot penetrate into the heart of a matter without proper participation in it. One really cannot imagine a kind of relationship, when the researcher is always present but never participates. This situation is hardly conducive for both the observer and the group. A combination of both participant and non-participant method is sometimes selected. The observer actively participates in some of the ordinary activities and observes passively from distance in others. Many sociologists therefore treat a non-participant observation in practice as only a quasi-participant observation. It is easier for the observer to perform both the roles than to disguise himself completely.

- **Disadvantages of Non-participant**

Observation:

The following are the disadvantages of non-participant observation.

(1) Subjectivity; in non-participant observation the observer does not have a clarity about certain events on activities. He can

not clear his doubts by asking various questions to the group members. Therefore he has to simply understand and interpret what he sees. This lack of understanding may make some of his findings biased and coloured by his personal prediction, belief and pre-conception.

(2) Inadequate observation: The observer can observe only those events which take place in front of him. But that is not enough and only a part of the phenomena as a vast range of information required for the research. He can know many things about the group when he participates in the group and interacts with the group members.

(3) Unnatural and formal information: The members of a group become suspicious of a person who observes them objectively. In front of an outsider or stranger they feel conscious and provide only some formal informations in an unnatural way. It creates bias and what the observer collects is not actual or normal thing but only formal informations.

(4) Inconvenience to the respondents : The members of a particular group always feel uncomfortable when they know that their behaviour is critically analysed by an outsider. Therefore in some cases the tribals do not allow an outsider to watch their socio-cultural activities. It is always better for a researcher to become a member of the group in order to learn much about it.

Advantages:

However, non-participant observation also has some advantages. Following are the merits of participant observation :

(1) Objectivity and neutrality

If an observer participates in the event actively and emotionally he may try to justify the evil things of the group as just things. In this frame of mind he cannot analyse the phenomena with neutrality. But in non-participant observation, the objectivity or neutrality can be maintained. The observer in this type of observation gives a detached and unbiased view about the group.

(2) Command respect and co-operation

In case of non-participant observation the researcher plays an impartial role. Therefore every member of the group gives him a special status and co-operate with his study.

(3) More willingness of the respondent

Often people donot feel shy to disclose their secrets, weaknesses or informal things to a stranger. But they always become reluctant to disclose these things to a known person.

(4) Careful analysis :

In participant observation because of the much familiarity with the events, sometimes the observer does not realize the significance of same events and neglects them. But in non-participant observation the researcher does not even miss a minute thing. He carefully judges the merits and demerits of each and every phenomenon under study.

(5) Freedom from groupism :

In non-participant observation the researcher always maintains his impartial status. His aloofness from petty conflicts helps him to carry his research work more smoothly.

MIXED OR QUASI-PARTICIPANT OBSERVATION

Many sociologists remark that there can be no pure non-participant observation. While observing in the field one has to be involved in the group activity. In the same way purely participant observation is also difficult, because there are certain activities, in which the observer can not participate. People do not allow an outsider to study and participate in their private activities. Therefore, a deliberate mixture of both participant and non-participant observation is always superior to either of the above two. It can bring a fruitful result for the researcher. This mixture of both the type of observation is called as mixed or quasi-participant observation. Goode and Hatt hold that in quasi-participant observation, at the beginning, the researcher works as a member of the group for some days and later on quasi-participate in the main activities of the group and simultaneously carries on his observation programme. In this way the observation will be more successful and the researcher will not have to waste his valuable time unnecessarily by participating in the whole activities of the group.

3.2.4.1. NON-CONTROLLED OBSERVATION:

Observation in social science by and large is uncontrolled. When the observation is conducted in the natural surroundings or in their usual course and the respondents performs their activities without being guided by any outside force or influenced by the observer it is known as non-controlled observation.

In uncontrolled observation the observer just visits the place where the required event in going on. For example, he can go to a tribal area in order to observe the tribal cultural practices. According to P.V. Young, "in non-controlled observation we resort

to careful scrutiny of real life situations making no attempt to use instrument of precision or to check for accuracy of the phenomenon observed".

Non-controlled observations are usually not very reliable because we do not have proper method to verify its findings. Therefore it sometimes becomes a biased observation. The subject matter of uncontrolled observations are so subjective that various researchers may observe the same thing differently and draw different findings. The precision and accuracy of n on-con trolled observation is limited. One of the chief limitations of uncontrolled observation is that it is not appropriate to study the complex facts, because in this the observer does not have any control over the factors of it. In uncontrolled observation sometimes the observer misses many significant facts or causes ; even a highly complex phenomenon or a significant event does not appear so to the researcher.

However, in spite of all these limitations uncontrolled observation can not be rejected as a valid research technique. P.V. Young says, "Life situations which can be adequately studied under controlled and artificial conditions are relatively few".

Most of the social phenomena have to be observed in an uncontrolled way, because social events or situations can not be artificially induced into laboratory and we have to carefully watch them only when they occur.

3.2.4.2. CONTROLLED OBSERVATION

Controlled observation is developed to reduce the defects and difficulties of non-controlled observation. These observations provide better precision and objectivity.

The most significant advantage of the controlled observation is that with the help of it the researcher can arrive at valid generalization and understand the cause-effect relationship between different phenomena.

By it one can verify the bias, inadequate data and thereby maintain objectivity.

The control exercised in the observation may be visualized in a couple of ways :

(a) Control over the observed phenomena.

(b) Control over the observer.

(a) Control over the phenomena

In this type of observation, the subject is put to laboratory type of test with different controls. Usually this type of laboratory experiment is not adequate for social phenomena which are so complex, abstract and relative by nature. Still in order to attain objectivity, researchers try to examine these social phenomena with the help of various controls.

This kind of observation is very useful to study the behaviour of children. The observer provides various stimuli to the children and examines their reaction to these stimuli. Various studies aiming at examining the working condition also use controlled observation method.

(b) Control over the observer

Goode and Hatt say, "It is rather difficult to control the object under investigation but he can at least control himself. As we have discussed earlier that there are many social events or situations which can not be controlled or experimented by guided conditions, therefore it is always better that the observer should try to control and guide carefully his own interpretation, perception and analytical mind in order to avoid bias in social research.

The following are the various controls used for this purpose.

(1) Detailed observation plan

Observation is not haphazard. It is always systematic. Therefore the researcher should make a detailed plan for the observation. He should decide in advance what situation is to be observed in what sequence he has to proceed, which method is appropriate for his study etc. This will save his time, money and effort and the observation would be conducted each time under standardised identical conditions.

(2) Use of schedule

In observation method the researcher can also use schedule simultaneously. It is difficult on the part of any researcher to remember each and every significant information. Therefore he can use schedule in the form of various blank tables and fill the relevant information under different heads of tables. It helps in valid classification of data.

(3) Use of mechanical appliances

The observer can use a variety of mechanical tools to make valid observation. Mechanical devices like photographs, films, sound recorders etc. not only preserve the details of an incident but also reproduce it even after the actual observation process has come to an end. Various mechanical devices can be studied later on-and if the observer has failed to record some of the important events he may cheek it at that time. By using this, the researcher even can study a minute thing objectively at his leisure.

(4) Maps

Maps and other devices are used successfully in social research to control the observation situation. This describes the graphical view of an event or situation. Through it we can very easily show different aspects of a social situation or structure. This can also provide us the knowledge about the situation in which controlled observation is possible.

(5) Socio-metric scales

These scales are employed to afford to statistical analysis in observation. These scales render qualitative observations to mathematically precise numbers. Thus they present greater objectivity in observational analysis.

(6) Formulation of hypothesis

The formulation of hypothesis in advance is very much essential for any observation. It makes observation more systematic, pointed and focused. It specifies exactly what things are to be observed. Therefore hypothesis should be properly planned in an observation.

(7) Group observation

Another control over the subjectivity of the observer can be eliminated through team or group observation. A single observer's analysis may be biased but a group observer's perception can never be biased one. Therefore, sometimes a number of observers are employed for the observation of the some phenomena. Their records are compared after the observation. If any differences are noticed, the "same must be connected or a cogent reason for the differences is found out. Again there are some situations whose full meaning cannot be analysed unless they are viewed from various angles. For this inter-disciplinary approach can be employed and the knowledge of the experts of various branches like economics, history, anthropology, history, psychology etc. may be required to undertake the task of observation. This combined effort is highly valuable for getting reliable conclusions.

(8) Control group

In order to determine the cause-effect relationship and draw valid generalization, control group is used for observation. In order to understand the nature and value of a specific factor, the researcher has to stabilize all other factors and keep changing this specific factor. The factor which is being changed is technically called as the controlled group. By varying this factor the researcher learns the actual fact about it. This technique helps the researcher to study various interwoven causes operating in a situation separately. For example if we wish to analyse about the effect

of violent movies on children, we can expose some children to those movies and compare their post-film behaviour with non-viewing children having the same rank and status as that of viewing group. Thus control group is very much useful for comparing two types of group, their behaviour and to draw a valid generalisation about it.

3.2.5. LIMITATIONS OF OBSERVATION AS A METHOD OF SOCIAL RESEARCH

Observation, although a widely used method in social research, has its own limitations. According to **P.V. Young,** "not all occurrences are open to observation, can be observed when the observer is at hand, not all occurrences lend themselves to study by observational techniques".

The following are the chief drawbacks of the observation method:

(1) Some of the Occurrences may not be open to observation:

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

(2) Not all occurrences open to observation can be observed when observer is at hand

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

(3) Not all occurrences lend themselves to observational study:

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

(4) Lack of reliability:

Because social phenomena cannot be controlled or used for laboratory experiments, generalizations made by observation method are not very reliable. The relativeness of the social phenomena and the personal bias of the observer again create difficulty for making

valid generalization in observation. P.V. Young remarks that in observation, no attempt is made to use instruments of precision to check the accuracy of the phenomenon.

(5) Faulty perception:

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

(6) Personal bias of the observer

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

(7) Slow investigation:

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

(8) Expensive:

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

(9) Inadequate method:

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

(10) Difficulty in checking validity:

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

3.2.6. ADVANTAGES OF OBSERVATION

Observation is one of the most commonly used methods of data collection in social science. The following are the some of the important advantages of the observation.

(1) Simplest method:

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

(2) Useful for framing hypothesis

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

(3) Greater accuracy:

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

(4) An universal method

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

(5) Observation is the only appropriate tool for certain cases

Observation can deal with phenomena which are not capable of giving verbal information about their behaviour, feeling and activities simply for the reason that they cannot speak *e.g.* infants or animals.

Observation is indispensable for studies on infants who can neither understand the queries of the researcher nor express themselves clearly. In the case of animals observation is the only way out. For deaf and dumb persons, for serious cases of abnormality or mad

persons, for non-cooperative persons, for too shy persons and for persons who do not understand the language of researcher, observation will be the only appropriate tool.

(6) Independent of people's willingness to report

Observation does not require the willingness of the people to provide various information about them. Often some respondents do not like to speak about themselves to an outsider. Some people do not have time or required skill to provide an important information to the researcher. Although observation cannot always overcome such problems, still relatively speaking it requires less active co-operation and willingness of respondents. Observation is ever possible without the knowledge of the respondents.

3.2.7. USES AND IMPORTANCE OF OBSERVATION

Observation is probably the most refined method modern research techniques. Direct observation is recently regarded as one of the best tools of scientific procedure.

Goode and Hatt nave rightly remarked that "Science begins with observation and must ultimately return to observation for its final validation".

Observation is the basis of science. It is the pivot around which all scientific analyses revolve. It is undoubtedly the first procedure of science as all scientific data must originate in some experience or perception. As data must originate in some experience or perception. As a scientific tool it may range from the most casual and uncontrolled stage to the most scientific and precise ones, involving modern mechanical and electronic means. It can be made progressively more scientific to meet the needs of the particular situation. Observation is also a fundamental tool even at the most advanced levels of science.

Science has been popularly defined as an accumulation of systematic knowledge. So undoubtedly the goal or objective of science is to gain knowledge, to reveal the truth and explore the reality behind a phenomenon. Science has to make much valid generalization for the growth of knowledge. In order to accumulate systematic knowledge science has to follow certain scientific methods. In this regard Karl Pearson has rightly remarked that "There is no shortcut to truth; no way to gain knowledge of the universal except through the gateway of scientific method".

A branch of knowledge can be called science only when it can be studied through scientific method. Observation is regarded as the chief tool of scientific analysis. It is the primary procedure of science for collecting various information regarding any phenomenon.

Man's first knowledge of the universe around him starts with observation. When a new born baby opens his eyes, he finds an unknown world. His curiosity forces him to observe

them. He perceives them, interprets them, gives a meaning to each situation and in this way he learns many things. So by observation only a child begins to experience many things about his world. By looking at a situation he guesses certain things, verifies it and finally realizes it as a truth.

On the basis of observation we all form an idea ; whatever vague it may be. Everyday, in each moment an individual observes many events. Some are old, others are new. We learn it and accordingly predict certain things on the, basis of it. Then of course we proceed to find out the validity and reliability of that. They may be totally correct, only partly correct or may be altogether false, but they help us to gain knowledge. This guess on prediction about certain things is nothing but called as simple hypothesis. Scientifically it is a proposition which can be put to test for determination of its validity.

Without hypothesis the research is unfocussed, a random empirical wandering. It is the necessary link between theory and investigations which leads to discovery of additional knowledge.

Observation not only helps the scientist to draw the valid hypothesis, but also to test their validity. Thus observation helps the scientist to make valuable scientific laws, generalizations or discoveries.

However, the task of the scientist never ends with formulating hypothesis. This is merely the beginning of scientific investigation. The final objective of any science is to make valid generalization. In order to fulfill this objective the scientist again requires observation. By taking recourse to careful, unbiased and objective observation a researcher can collect various information. Through different modern scientific and mechanical devices he has to check the accuracy and reliability of phenomena which he has observed and finally he can draw a valid generalization or establish an universal and objective fact. A simple or scientific discovery has come to this world only through this process.

Detailed discussion regarding the various tools for checking the objectivity, validity and accuracy of collected information has been given in the initial pages of his chapter but some guiding-rules may be summarized below.

(1) The aim and objectives of the observation must be known to the observer.

(2) The basis for the scientific judgment to be made by the observer must be clearly defined.

(3) To increase the range, depth, accuracy of his observations and guard against errors in perception, a scientist

requires a broad background of the field in which he works.

(4) To sharpen his perceptive power, he has to acquire abundant practice in the art of examining the phenomena with an alert and questioning mind.

(5) Since emotional and intellectual biases can present accuracy of observation, the observer should make a impartial study of the phenomena. Whenever possible, he should repeat observations to see whether the results will be the same on each occasion or not.

(6) Because human biases may introduce errors, the observer should use the mechanical devices to collect data whenever possible.

(7) He should write his descriptions carefully in precise and concrete terms, so that no other interpretation can possibly be placed on his words.

(8) The recording should be accurate. While recording he should not leave any important information.

If all the above mentioned conditions will be fulfilled, there is no doubt that science can successfully begin with observation and finally make valid generalization, so as to embark upon objective facts or various discoveries on the basis of it.

3.3. SURVEY METHODS

The term survey is the combination of two words namely 'Sur' or 'Ser' and ´Veeir' or 'Vor' meaning 'Over and 'to see' respectively. Thus the term survey implies 'to oversee' or 'to look over'. In simple terms it may be construed as an observation. According to P.V. Young, "In general we may observe that social surveys are concerned with *(i)* the formulation of constructive programmes of reform and *(ii)* amelioration of current or immediate conditions of a social pathological nature which have definite geographical limits and definite social implications and significance, *(iii)* these conditions

can be measured and compared with situations which can be accepted as model". E.W. Burgess

considers social survey as "the scientific study of community conditions and needs for the purpose of presenting a constructive programme of social advance... a method of social introspection checked by statistical measurement and the comparative standards of the social expert". In the words of Mark Abrams "A social survey is a process by which quantitative facts are collected about social aspects of a community's composition and activities". According to Shelly M. Harrison "Social survey is a cooperative undertaking which applies scientific method to the study and treatment of current related social problems and conditions having definite geographical limits and bearings plus such a spreading of its facts, conclusions and recommendations as will make them as far as possible the common knowledge of the community and a force for intelligent coordinated action". An analysis of the above definitions of social survey reveals that P.V. Young has clearly delimited the area of survey to a definite geographical locality. She has also held that social surveys deal with some definite social problems and conditions. Similarly Burgess's definition lays stress on the constructive and progressive purpose of social survey. It is a well-known fact that an implicit or explicit purpose of all social surveys is social progress and betterment of a community, as conditions and causes suited to such a kind of transformation are found out in social surveys. Briefly speaking, all the foregoing definitions consider social survey as the investigation of social problems, conditions, structures, processes etc., so as to bring about change in the positive direction on the basis of collected facts.

3.3.1. Purpose of Social Survey

1. Collection of data about certain **definite aspects of a community.** The basic objective of Social Survey is the collection of facts regarding certain social aspects of a community so as to find out the causative factors responsible for community problems. Keeping in view the progress and betterment of a community, both qualitative and quantitative data are collected from a selected group of members of a community. Before the investigation is undertaken, the surveyor makes the aims and objectives of the survey clear in order to establish and justify the fact that in it only social aspects of a community are considered, which are different from the land use survey, the

political or the geographical surveys. As social surveys aim at social amelioration, both descriptive and statistical data are involved.

2. Political and utilitarian viewpoint. Whereas the social research may either be theoretical or applied, social survey is undertaken with the objective of a practical end. Whereas social research is mostly born out of natural curiosity and urge to know, the motivation of social survey is practical and utilitarian. Social survey is undertaken with the express purpose of social amelioration of current or immediate conditions of a social pathological nature. The main purpose of social survey being social welfare or betterment of community, its findings enable the government to take various measures to protect the underprivileged, depressed and the downtrodden. In this regard C.A. Moser remarks : "the sociologist should look upon the survey as a way, and a supremely useful one of exploring the fields of collecting data around as well as directly on the subjects of study, so that the problem is brought into focus and points worth pursuing are suggested."

3. The study and treatment of current social problems.

Social survey is a cooperative undertaking which applies scientific method to the study and treatment of current related social problems and the conditions and factors of social retardation and or backwardness. A close look at the history of social surveys reveals that the. main purpose of social survey has been the study of different problems in society in general and the socio- economic conditions of the underprivileged, depressed and working classes in particular, so as to cognize the cause of backwardness and chalk out plans for removal of these conditions. The underprivileged, the depressed and the working classes are mainly focused because the socio-economic level of those classes remains much lower, compared to other sections in a society. They are mostly vulnerable to unhygienic conditions of living, moral depravity, sexual polymorphic, epidemics, sexually transmitted diseases, unemployment, mal-nutrition etc. Accommodation problem and low moral-sense, exposure to drunken and immoral behaviour of the elders causes juvenile delinquency among the children of the working classes. Therefore, the main purpose of social survey has always been to improve their conditions, erase their blunted moral-sense and to protect them against exploitation.

In social surveys attempts are also made to find out the inter-relationship between different social problems such as, unemployment and poverty, illiteracy and unhygienic etc.

3.3.2 INTERVIEW

Interview as a technique of data collection is very popular and extensively used in every field of social research. The interview is, in a sense, an oral questionnaire. Instead of writing the response, the interviewee or subject gives the needed information verbally in a face-to-face relationship. The dynamics of interviewing, however, involves much more than an oral questionnaire. Interview is relatively more flexible tool than any written inquiry form and permits explanation, adjustment and variation according to the situation. The observational methods, as we know, are restricted mostly to non-verbal acts. So these are understandably not so effective in giving information about person's past and private behavior, future actions, attitudes, perceptions, faiths, beliefs thought processes, motivations etc. The interview method as a verbal method is quite significant in securing data about all these aspects. In this method a researcher or an interviewer can interact with his respondents and know their inner feelings and reactions. G.W. Allport in his classic statement sums this up beautifully by saying that "if you want to know how people feel, what they experience and what they remember, what their emotions and motives are like and the reasons for acting as they do, why not ask them".

Interview is a direct method of inquiry. It is simply stated as a social process in which a person known as the interviewer asks questions usually in a face to face contact to the other person or persons known as interviewee or interviewees. The interviewee responds to these and the interviewer collects various information .from these responses through a very healthy and friendly social interaction. However, it does not mean that all the time it is the interviewer who asks the questions. Often the interviewee may also ask certain questions and the interviewer responds to these. But usually the interviewer initiates the interview and collects the information from the interviewee.

Interview is not a simple two-way conversation between an interrogator and informant. According to P.V. Young, "interview may be regarded as a systematic method by which a person enters more or less imaginatively into the life of a comparative stranger". It is a mutual interaction of each other. The objectives of the interviewer are to penetrate the outer and inner life of persons and to collect information pertaining to a wide range of

their experiences in which the interviewee may wish to rehearse his past, define his present and canvass his future possibilities. These answers of the interviewees may not be only a response to a question but also a stimulus to progressive series of other relevant statements about social and personal phenomena. In similar fashion, W.J. Goode and P.K. Hatt have observed that interviewing is fundamentally a process of social interaction", the interview two persons are not merely present at the same place but also influence each other emotionally and intellectually.

3.3.3. Objectives of Interview:

The following are some of the important objectives of interview method,

(1) Direct contact:

The first and foremost aim of the interview method is to establish a direct contact between the researcher and the interviewee, so that both can understand each other's feeling, attitude and needs. After the interviewer establishes a friendly relationship with the subject, certain type of confidential information may be obtained, that an individual might be reluctant to put in writing.

The interviewer can explain the purpose of his investigation, and can explain more clearly just what information he wants if the subject misinterprets the question, then the interviewer may describe it with a simple clarifying question and collect various information from them.

(2) Eliciting intimate facts:

In modern complex society, experiences are highly heterogeneous. Few people share a common lot, but their attitudes and values are quite varied. Many People can live within the protective wall of anonymity. There are many facts of personal life, one does not like to reveal. All other methods are not so effective in order to collect this intimate or personal information from a respondent which he does not want to share. But P.V. Young has rightly observed that interview is the most effective method through which the interviewer can penetrate into this protective mask and elicit these intimate facts. By establishing a rapport or a friendly relationship with the interviewee, the interviewer can gain his confidence and may be able to extract various confidential information from him.

(3) Establishing hypothesis:

Through the interview techniques the researcher may stimulate the subject to have greater insight **into** his own experiences, peculiar attitudes, outlooks, aspirations and thereby explore significant areas not anticipated by him. These new revelations help

him in forming new hypothesis about personal and social behaviour. P.V. Young says, "Every verbal response and non-verbal reaction may be an "eye opener" for a whole new train of thoughts. An answer may not only be a response to a question but also a stimulus to progressive series of other relevant statements about social and personal phenomena which might indicate cause-effect relationships and at times may lead to formulation of hypothesis regarding socio-personal interaction".

(4) Verification of unique ideas:

When a researcher elicits a novel idea about certain type of behaviour, it is always desirable to conduct interview with the concerned person and see how far the ideas elicited are true or valid facts. So one can examine its validity through interview method and safely conclude about it.

Various sociologists have remarked that the objective of interview is two-fold : *(i)* eliciting certain information from the interviewee^ which is known only to him and cannot be collected from any other source, *(ii)* psychological study of verbal and non-verbal behaviour under given circumstances. As regards the first objective, the interviewer clarifies the topic or area of study to the interviewee. Then interviewee narrates the experience of his life and his reactions pertaining to it. The interviewer listens to these descriptions carefully and tries to collect useful information out of it.

For the second objective the researcher plays more the role of a social psychologist than a sociologist. His attention is more centered on the attitude and expressions of the interviewee than the actual facts. It is basically to learn about what T.W. Adorno terms "levels of personality" of the interviewee.

INTERVIEW GUIDE

Apart from his skill, training and understanding of the problem, the interviewer must have a clear conception of just what information he needs. Then he has to prepare appropriate questions to extract the desired data. He has to clearly outline the best sequence of questions and stimulating comments that will make the respondent feel comfortable and stimulate his flow of conversation. All these require the help of an interview guide.

Interview guide is a brief written handbook of instructions prescribing an outline of different aspects of interview to be studied. According to P.V. Young, "when a prepared guide is judiciously used by the interviewer as a suggestive reference, it helps the conduct the interview in four ways:

(i) Focuses attention on salient points of the study.

(ii) Secures comparable data in different ^interviews by the same or by various interviewers and thereby maintains uniformity in interviews.

(iii) Gathers the same range of items essential in the analysis of data or in testing the hypothesis formulated.

(iv) Accumulate specific concrete details as a basis for quantitative studies of life histories.

However, she also remarked that the guide is not an oral questionnaire. The guide becomes a hindrance, instead of aid, if too much attention is paid to it. When too many questions are asked from the guide and fixed replies are expected the atmosphere of the interview becomes chocked and free self- expression vanishes and everything becomes artificial. Precautions should therefore be taken to see that interview guide does not become too much rigid or structured with too many details.

Guiding the interview is one of the important tasks of the interviewer. It is the duty of the interviewer to guide the interviewee as well as the whole process of interview. It is very important especially in narrative type of interview. In such type of discussion, sometimes the respondent becomes very emotional and he becomes so absorbed that he may drag the story to a point which is not relevant and very much detached from the matter. In that case it is the responsibility of the interviewer to bring the discussion to the original matter without hurting the emotions of the interviewee. Guiding an interview is a very delicate task which should be undertaken with utmost care. Sometimes if the interviewee will be stopped from what he wants to speak (even if that is irrelevant) he may lose interest in the process and may not respond the rest of the questions. Therefore the interviewer should deal all these matters very delicately. He should be a patient listener and should take care of the liking and taste of the interviewee rather than his own At the same time care should be taken so that no important point is slipped.

3.3.4. 4. INTERVIEW SCHEDULE

Schedule is one of the very commonly used tools of data collection in scientific investigation. P.V. Young says "The schedule has been used for collection of personal preferences, social attitudes, beliefs, opinions, behaviour patterns, group practices and habits and much other data". The increasing use of schedule is probably due to increased emphasis by social scientists on quantitative measurement of uniformly accumulated data.

Schedule is very much similar to questionnaire and there is very little difference between the two so far as their construction is concerned. The main difference between

these two is that whereas the schedule **is** used in direct interview on direct observation and in it the questions are asked and filled by the researcher himself, the questionnaire is generally mailed to the respondent, who fills it up and returns it to the researcher. Thus the main difference between them lies in the method of obtaining data.

Goode and **Hatt** say, "Schedule is the name usually applied to a set of questions which are asked and filled by an interviewer in a face to face situation with other person".

Webster defines a schedule as ("a formal-list, a catalogue or inventory and may be a counting device, used in formal and standardized inquiries, the sole purpose of which is aiding in the collection of quantitative cross-sectional data".

From the above discussion it is clear that in schedule method interview occupies a pivotal place and plays an indispensable role. So the success of schedule largely depends on the efficiency and tactfulness of the interviewer rather than the quality of questions posed. Because the interviewer himself asks all the questions and fills the answers all by himself, here the quality of question has less significance.

- Objectives of Schedule

P.V. Young has emphasized on the following objectives of the schedule. According to her a researcher "makes the schedule a guide, a means of delimiting the sense of his enquiry, a memory tickler, a recording device". These may be discussed elaborately.

(1) Aids to delimit the scope of the study:

The schedule always focuses on a definite item of inquiry, single and isolated subject, rather than a subject in general. The researcher asks about one item and writes down answers about it. Therefore schedule delimits and specifies the subject of inquiry. It also focuses the attention of the interviewer on salient points of the study.

(2) Acts as a memory tickler:

Another objective of schedule is to act as a memory tickler. In interview, the interviewer has to ask a wide range of questions to the respondent. While doing this he may forget to ask about some important aspects of the research problem and then may require going for the whole process again to collect that missing information. Usually most persons do not have good memories. Because of this there is always a possibility of missing certain important points. But in the schedule method an interviewer is not dependent upon his memory. Because the schedule is all planned, a formal written document of various questions which keeps the memory of the interviewer or observer refresh and keeps him reminded of the different aspects of the study which are to be observed. It also economizes the time, money and effort of the interviewer and observer.

(3) Helps in classification and analysis:

Another objective of schedule is to help in the work of tabulation classification and analysis. In case of narrative interview, the interviewee provides the information in a story form. It becomes very difficult afterwards to sort out the important points and classify them into different categories for purpose of further analysis.

The researcher may ask different types of questions. So the answers collected from the respondents are also varied. All these varieties of replies are classified under various heads. The schedule puts the entire thing in a structured form and facilitates in better tabulation and analysis.

(4) Standardized device:

Schedule also provides a standardized device for observation and interview. In a structured schedule every respondent gives reply to the same question, put in the same language, in the same sequence. So the entire process of interview takes place under standardized conditions. The data received from this is perfectly comparable and helps in an objective study.

Interview schedules are used for the interview purposes. These are the most common and important schedules used by many researchers to secure information from the interviewees. It consists of some standard questions which the interviewer asks to the respondents and then fills up the form after getting the information from the respondents.

3.3.4.1. STEPS OR PROCEDURES FOR FORMING A SCHEDULE:

A good schedule requires certain planning and preparation. In most of the cases schedule is combinedly used with the interview method. So this is more than a series of casual questions and their replies. Rather it is a dynamic, interpersonal experience that has to be done carefully. The researcher should follow certain procedures and take certain precautions before framing a schedule. Some of the steps which are to be followed by the researcher in this regard are:

(1) Knowledge about the different aspects of problem:

While framing 'the schedule the first step is to have proper knowledge about the different aspects of the problem- The researcher has to put a great deal of thoug ht into the selected research problem.

However, the following are the some of the necessary factors; one should take into consideration before framing a schedule on a particular topic.

(i) The researcher must have interest in the topic of research.

(ii) The nature of the problem or topic must have some social reference.

(iii) He should understand the problem thoroughly.

(iv) The problem must be defined in clear and explicit manner.

(v) The problem must be defined unambiguously, so that it will help to differentiate relevant data from irrelevant ones.

(vi) existing literature on the subject should be studied.

(vii) the problem under study should be split up into various aspects, the determination of these aspects depends upon the clear understanding of the problem.

Thus, for example, if the survey of the effect of family disorganisation upon the criminal behaviour of children is to made, then the different aspects of the problem would be the family background of the children, the parental relation, the socialisation process, family value, the authority structures etc. The researcher has to study all these aspects in complete details before framing a schedule.

(2) Knowledge about the information to be studied :

While framing a good schedule the second important step is to decide what information is necessary for a valid generalisation on each aspect of the problem. An extensive literature survey usually helps the researcher to get a proper knowledge about different aspects of the research problem. By studying the previous studies in the related field the researcher gets knowledge about relevant information needed for his current study. He can again subdivide each aspect of the topic. Thus, in the above illustration family value may be further subdivided into respect for morals, tolerance, fear of religion and authority, adjustment with others, learning manners, building character and personality etc. Required information may be collected about each of them. (3) Framing the actual questions :

The third step is the framing of the actual questions. This is the most essential part of the schedule and any error in it may invalidate the whole research study by providing biased, incorrect, incomplete or irrelevant information.

While framing the actual questions in a schedule the following are the some of sub-steps one must take into consideration.

(a) Nature of questions to be given

For framing a schedule there is no exclusive rules and regulation regarding the selection of nature of questions to be asked. It all depends on the nature of the

research topic, the skill of the researcher, kinds of respondents and other factors. The following points are some of the general guidelines about the nature of questions.

(i) Specific questions

A common error is to ask a general question when an answer on a specific issue is wanted. For example, if one is interested specifically in a canteen's meal prices and the quality of its service, the question "Are you satisfied or dissatisfied with your canteen ?" is unsatisfactory. In the above example, the general question was framed because it failed to specify the required frames of references. But when there are occasions when no required frame of reference is needed, the general question may be appropriate. However, the researcher should try to give specific questions to the respondents as much as possible. One way to make questions more specific is to frame them in terms of the respondent's personal experience rather than in general terms.

(ii) Simple language :

In choosing the language for a schedule, the population being studied should be kept in mind. The aim in question wording is to communicate with respondents as nearly as possible in their own language. A survey of the members of a particular profession, for instance, can usefully employ the profession's common technical terms. Not only such terms form a part of the informant's common language but they also normally have a single precise meaning. Technical terms and jargons are however obviously to be avoided in surveys of general population. The first principles in wording are that the questions should use the simplest words that will convey the exact meaning and that the phrasing should also be as simple and informal as possible.

It is not indeed enough to know that a word or phrase is commonly used, one must equally be sure that it is used in the same sense by all groups of respondents. Even a common word 'book' has different meanings in different parts of the country. A simple case is the 'book' which in some parts of the population is taken to include magazines. Hence while forming a schedule an interviewer should ask—"During the past week, roughly how many hours you spent reading books, I mean books, not magazines or. papers Clarity can still be further ensured by remembering that a simple question is more readily understood than a long complex one. So rather than relying on a single complex question, a series of simple questions should be asked. The number of such questions depends on the degree of simplicity required. Household composition is generally a complex subject. To present it in a simple way, a series of descriptive indices are required. The information can usually be best obtained by using a 'household box' on the

schedule in which the household members are listed together with their relevant characteristics *viz.* age, sex, marital status, working status, educational status etc.

(iii) Attention to be given to questions involving memory:

Most factual questions, to some extent, involve the respondent in calling information. His degree of success in presenting this accurately is thus a basic determinant of the quality of his response. With certain questions such as "Are you married, single or widowed?" there is no such problem, but with a large range of survey questions recalling information does bring a problem, the severity of which depends on what is to be recalled. Two factors of primary importance in memory are the length of time since the event took place and the event's importance to the respondent.

Even what the respondent considers insignificant is likely to be forgotten almost immediately and even the re-collection of significant events decreases as time elapses. Moreover, for events not forgotten completely, memory acts selectively, retaining some aspects and losing others, thus producing distorted images. For questions dealing with the past, serious attention must therefore be given to the respondent's ability to recall the required information accurately and to the ways by which they can be helped to do so.

(iv) Questions must be within the intellectual capacity of the respondent:

The questions included in the schedule should be within the respondents intellectual capacity to give answer. The researcher should not expect any reply which is beyond his informational scope. For example, an illiterate cannot reply about e-commerce, internet etc.

(v) Inter-relation of questions :

Various questions asked by the researcher should be inter-related with each other. They should be asked in a proper order, so that it will be systematic, interesting and continuous one.

(vi) Cross-checking questions ;

In **a** schedule the researcher should include certain questions for cross checking. It will provide a scope of verification to the researcher and he can check the incorrect or bias answers of the respondents

(b) Questions to be avoided:

Following types of questions should be avoided for a better response in schedule.

(i) Ambiguous questions:

Ambiguous questions are to be avoided at all costs. If an ambiguous word creeps in, different people will understand the questions differently and will in effect be answering differently to the some question. The following example is taken from a University Research Survey.

"Is your work made more difficult because you are expecting a baby?" The question was asked to all women in the survey, irrespective of whether, they were expecting a baby or not. What **then,** did a 'No' answer mean? Depending on the respondent, it might have meant—"No, I am not expecting a baby" or "No, my work is not made more difficult by the fact that I am expecting a baby". Such ambiguity has to be avoided in any social research; otherwise it **will** minimize the objectivity of research.

(ii) Double barreled questions

Ambiguity may also arise with double barreled questions, such as the following question on public transport, "Do you like travelling on trains or buses?" Respondents liking one and disliking the other would be in a dilemma in answering this question. Clearly it needs to be divided into two separate questions, each concerned with a single idea, in this case with a single mode of transport.

(iii) Vague words

Vague questions encourage vague answers. If the respondents are asked whether they go to the cinema regularly or occasionally, the meaning of their answers will be vague. (This common choice of alternatives is strictly illogical. Because the word "occasional" refers to frequency, the word 'regular' does not. However this may be the case where logic can give way to common usage). But the meaning can easily be made more precise, if the researcher will ask "How often these days do you go to the cinema ? Would it be nearer to twice a week or more often, once a week, once a fortnight, once a month, three or four times a year, less often, or do you never go these days?"

Vague words and phrases like 'kind-of, 'fairly', 'generally', 'often', 'many', 'much the same', 'on the whole', etc. should be avoided. If one asks—"What kind of house do you have" ? Without specifying a frame Of references, some people will answer that is semi detached, others that it is sub-urban, others that it is very pleasant and so on.

A similar type of vagueness occurs in 'why' questions. In answering the question "Why did you go to the cinema last night ? S ome respondents will say that they wanted to see that particular film, some that 'they did not want to stay at home', and others that 'the wife suggested it' or that 'they had not been since last week[1]. The word 'why' in this

question—as the phrase 'Kind of in the previous one—can mean so many different things and thereby produce a useless mixture of answers.

(iv) Leading or Suggestive questions;

Leading or suggestive type of questions should be avoided as they result in biased answers. A leading question is one which, by its content, structure or wording, leads the respondent in the direction of a certain answer. For example "You don't think do you?" as obviously leads to a negative answer and the question form like "Should not something be done about?" leads to a positive one. In addition to 'leading word', there is the risk that the general context of a question, the control of these preceding it and the tone of the whole schedule or interview can lead the respondent in a given direction and may bring bias in research. Therefore, while framing a schedule, the interviewer should try to avoid these types of leading questions as far as possible.

(v) Presuming question:

While forming a schedule, the researcher should not presume anything about the respondent. For example questions like "How many cigarettes you smoke a day?" or "How did you vote in the last election?" are best asked only after a 'filter question' has revealed that the respondent does smoke cigarettes and did vote in the last election. Without knowing this the researcher should not presume anything about the respondent. Otherwise the respondent may feel insulted and become reluctant to provide various information on research topic.

(vi) Hypothetical questions:

Hypothetical question like "Would you like to live in a flat?" Another kind of hypothetical question is "Would you like a more frequent bus service?" or "Would you like an increase in wages?" Such questions are unlikely to be of any value because the respondent is being asked if he would like something for nothing. It is hard to see how he could possibly say "No". If he did, it could be because he has taken into account some hidden factors of his own, or because he has failed to understand the question.

(vii) Personal questions *:*

Questions regarding personal, private or secret things of a respondent should be avoided unless they are relevant for the inquiry. People are usually reluctant to disclose their personal

matters regarding marital or sex life, various diseases etc. to a stranger.

(viii) Embarrassing questions :

Questions that may put the respondent in an embarrassing position should also be avoided. Subjects which people do not like to discuss in public create a problem to the schedule designer. Respondents are often embarrassed to discuss their personal matters, to give low-prestige answers, and to admit to socially unacceptable behaviour and attitudes. If, for instance, questions on sexual behaviour, frequency of taking bath, cheating in examinations or attitudes to communism were asked in the usual way, many respondents would probably refuse to reply and other would distort their answer.

One method of reducing the threatening nature of a question is to express it through a third person, instead of asking the respondent for his views. Here he can be asked about the views of others. An example from market research of an indirect question of this sort is—"Some women who use this cleanser find a lot of faults with it, I wonder if you can guess what they are objecting to". The purpose of this wording was to make the housewives feel free to criticize the product. The aim of such questions is to obtain the respondent's own views but he may, of course, answer the question asked, and give what he believes to be views of others. For this reason it is often advisable to follow the indirect questions.

There are several other indirect methods which can be useful in dealing with embarrassing topics. The respondent can, for instance, be shown a drawing of two persons in a certain setting with "balloons" containing speech coming from their mouth, as in comic strips and cartoons. One person's balloon is left empty and the position of that person and to fill in the missing words.

Another method is that of sentence completion; the respondent is given the beginning of a sentence and is asked to complete it, usually in a limited time to ensure spontaneity. Beison (1968) describes a study of randomly derived sample of London teenage boys on the sensitive subject of stealing. A variety of procedures were employed in this study to make it easier for the boys to admit that they had stolen things. In arrival at the interviewing centre a boy chose a false name to the interviewer, who knows him only by that name. After an extended initial phase the interview proceeded to the card-sorting technique by which the information on stealing was to be obtained. The interviewer and the boy sat on either side of a table, with a screen in between so that they could not see each other. Through a slot in the screen the interviewer passed to the boy a card on which one type of stealing *{e.g.* I have stolen cigarettes) was recorded. The boy was asked to put the card in a box labeled 'Yes' if he had ever done what was recorded on it and in a box labeled "Never" if not. This was repeated for 44 kinds of theft. At the end of this sorting stage, the, interviewer went through a procedure which tried, to reduce the force of a

boy's resistances and to strengthen his feeling of willingness, to admitting thefts. Then the boy was asked to resort all the cards he had put in the 'never' box. Finally he was asked for further details or each type of theft he had admitted. This detailed procedure elicited reports of many types of theft from many boys with, for example, 69% of boys admitting "I have stolen something from a shop", and 58% '(I have stolen money" at least once in their life.

(ix) Too long questions:

Too long questions are boring and respondents do not follow it easily. If a researcher feels the necessity of giving a long question then he should break it into some interrelated parts, so that it will be easier on the part of the respondent to reply it.

(x) Question causing suspicion:

Question which creates suspicion in the mind of the respondent like question about one's private relation, neighborhood relations, monthly income, accumulation of wealth etc. should be avoided as far as possible unless they are absolutely necessary.

(xi) Question on sensitive issue:

Question creates ill feeling for others or hurt one's emotion like "Are religious practices unscientific?" "Is Islam religion better than Hinduism?" etc. should be avoided.

(xii) Question against universally accepted norms:

Every society is having its own accepted normative structure. The members of that particular society always show respect to these norms. In case a schedule contains some questions which go against these accepted norms it creates dissatisfaction among the respondents. So these questions should be avoided by the researcher as far as possible.

(c) LANGUAGE:

While framing a schedule the researcher should be careful about the proper wording or language. Following types of word should be avoided as far as possible.

(i) ***Abbreviation:*** **In** order to reply a question the respondent should understand it clearly. A particular abbreviation given in schedule may be known to the researcher but the respondents may not understand it. Therefore, the researcher should try to avoid such abbreviations. If such abbreviations are used then its meaning and full form should be given in the schedule for the better understanding of the respondents.

ii) ***Value-loaded words:*** Words carrying different values *viz.* good and bad should be avoided as far as possible.

(iii) Native ***or unusual words:*** The researcher should try to avoid highly localized languages in his schedule. It is always better to use the words which can be understood by everyone.

(iv) ***Multi-meaning words:*** Words carrying different meanings should be avoided.

(d) Sequence of Questions

Although no hard and fast rule exists for giving a particular sequence, still in order to get a proper response the sequence of questions in a schedule needs to be planned. A proper sequence of questions may reduce the refusal rate and there is plenty of evidence that it may also influence the answer obtained. One can take into account the following factors for preparing a perfect sequence of questions. It is always good to start with a simple, general and broad questions about the topic and then to narrow down to the specific issues$_3$ using what is known as a "funner sequence" of question. Thus a general open question on the achievement of the present government may be the beginning of a sequence, then leading to specific questions on the government's action in the field of labour relations.

Unambiguous and uncontroversial questions should be included in the initial page of a schedule. Usually the researcher should start from a simple question and then proceed towards complex questions. Because if complex or ambiguous questions are included in the initial pages, the respondent may refuse to give interview. -*(iii)* At the start of the interview the respondent is unsure of himself and so the opening questions should be one to put him at ease and build up rapport between him and the interviewer. They should be interesting questions which he will have no difficulty in answering. These should not be on sensitive topics, otherwise he may refuse to continue with the interview.

(iv) Question seeking the advice of the respondents may be given in the beginning, so that the respondent will feel that his information is valuable and he will be more willing to extend his co-operation for rest of the interview.

(v) It is always better to divide the entire schedule into some sections and each section should deal with a particular topic.

(vi) The entire schedule should be regarded as a coherent unit. There should be a proper co-relation between each question and different sections of schedule. Various parts of the questionnaire should be arranged in such a way that they wouldn't be different from each other rather would make an entire schedule as an integrated whole.

(vii) The shift from one section to another section should be very natural or smooth. Sudden jumping from one topic to another may badly affect the response of respondents.

(e) Types of Questions :

(i) Open end question : The open form, open end or unrestricted type of questions call for a free response in the respondent's own words. The respondent is having much freedom.

Isere to provide his own response. No clues are provided. It probably provides for gathering depth of response. The respondent reveals his mind, provides his frame of reference with the reasons for his responses. This type of question is sometimes difficult to interpret, tabulate and summarize in the research report. When the respondent is allowed to give free response, his expression may take any unique direction which may not find any uniformity with other responses.

However they are used mostly in pilot studies to get an idea about the research area and the possible replies. The following are the some of the examples of open end questions :

Example—1 : What is your view about the current budget ?

Example—2 : Is it beneficial for the poor people ?

Example—3 : Here question 3 (b) is a typical open question not only in its form and content but also in that it opens the interview. It is often desirable to start the interview with an open question to get the respondent talking and to make him feel at ease.

3 (a) I am from the Survey Research Unit and we are trying to find out a few things about what people do in their spare time. Would you mind telling me, are there any things which you would like to spend more time on ?

Yes—1

No—2

Don't know—3

3 (b) If answer yes (1) to question 3 (a), what for instance ? (Tell in details).

(ii) Closed form of question : The questions that call for short, limited responses are known as restricted or closed form

of questions. They provide for marking ayes or no, a short response, or checking an item out of a list of given responses. It restricts the choice of response for the respondents. He has simply to select a response out of supplied responses and has not to frame responses in his own way. The following are the illustrations of closed form of questions.

Example 1 : Are you literate ? Yes/No. *Example* 2 : Are you a housewife ?

Yes/No.

Many of the questions here are opinion questions, in which the respondents are given choice between 'good' and 'bad', 'very bad', 'important', 'very important' and 'not at all important'. Such questions are very common in Opinion Research.

(Hi) ***Factual*** *question :* George A. Lundberg has mentioned this type of question. It requires certain information of facts from the respondent without any reference to his opinion or attitude about them.

(iv) ***Opinion question*** *:* This type of question collects data about one's opinion, attitude or preferences regarding some phenomena.

(v) ***Dichotomous*** *question :* When a question is given with only two possible alternative answers, that is called dichotomous questions For example, Do you belong to reserve category? Yes/No.

(vi) ***Multiple choice questions*** *:* These questions are otherwise known as cafeteria questions. These are just opposite of previously described dichotomous questions. In these questions the reply is not confined to two alternatives only, but to a number of possible alternatives. *For example* "What, according to you is an important cause of poverty in India ? *(a)* population growth (6) lack of education (c) lack of governmental initiative for its eradication *(d)* illness of people (e) lack of industry *if)* any other (specify).

(4) Content of Schedule

The fourth step in forming a schedule is to prepare the content of a schedule. It is nothing but the systematic structure of a schedule. The whole schedule may be divided into three parts *viz.* (a) Introductory part, (6) Main schedule and (c) Instructions to the interviewer/observer.

(a) ***Introductory part*** *:* This part includes introductory information about the schedule and its respondents.

In this opening part, the following type of information with regard to inquiry and respondent are sought : *k*

ii) Name of the survey with the name and address or its conducting authority.

(ii) Reference or case number.

(Hi) Name of the respondent, his address, age, sex, education, profession etc.

(iv) Place of interview. *(v)* Time and date of interview.

(&) ***Main Schedule*** *:* This is the main and vital portion of the schedule. It has to be prepared with great care. This part of schedule contains with different questions, columns, as well as blank tables where information supplied by the respondent has to be **filled.**

(c) ***Instruction to interviewer****:* In this part the field worker (interviewer or observer) who has to present the schedule and collect data is given elaborate instructions regarding the presenting the schedule and the method of interview. The field workers are given detailed instructions about the use of various units, technical terms, general method of fulfilling the schedule and the way in which the interview is to be conducted smoothly. Instructions are given in details in order to ensure uniformity of recording the responses.

(5) The fifth step is the general layout of the schedule.

The layout or physical design of the schedule is very important. If it is planned properly the interview will bring high response. A bad, unsystematic and a schedule without a proper layout may create errors frequently.

The following are some of the steps which may be taken to achieve this objective.

(i) **Size of Schedule** : Usually small sized schedule is preferred by the respondents because they can follow a small sized schedule easily. The schedule should not be too long as it may be difficult on the part of the respondents to spend their valuable time in it. Therefore, the length of schedule should be made in such a manner that it will be taking just less than half an hour in filling it up.

(ii) ***Paper*** *:* The paper used for printing of schedule should be of high quality. The letters printed on it should be visible clearly and must not be broken or the ink should not spread over the paper. If the paper is rough or of low quality, the letters printed will be poor in visibility and are liable to be broken. When the researcher fills it by ink and the ink may spread.

Therefore the paper printing must be of excellent quality. Under-economy, in this respect, may cause series of problem of response for schedule.

(Hi) ***Margin*** *:* The margin on the left should be about 1—"
and on the right it should be 1/2". This makes the schedule an attractive one. Besides the researcher can take some notes in this marginal space. Absence of margin may create problem for punching. Because punching without any margin may "destroy some of words.

{iv) **Spacing** : **In** between the questions, titles, subtitles and columns there should be reasonable space for noting the responses and demarcating one from the other.

(i>) ***Printing****:* A printed schedule is obviously more desirable as printing makes a schedule more attractive. But if the number of respondents is small or the researcher wants to reduce the cost of research, he can also use cyclostyled or type written schedule. However, in both these cases the schedule must be neat and free from over writing.

(vi) Use ***of picture****:* Sometimes the use of pictures in schedule influences the respondent in right manner and the respondent takes greater interest for giving replies. Therefore, it is desirable to insert suitable pictures whenever possible.

(6) Testing the Validity of Schedule

The last step for forming a schedule is testing the validity of schedule. After the schedule has been prepared the investigator should test it on a sample population in order to examine its validity, and to find out any discrepancies in it. Thus various mistakes, unsatisfactory or unnecessary things, can be located only when the schedule has been operated on trial basis. Thereafter if such things are noticed then the investigator can bring some changes to make the schedule more accurate. If all these steps will be taken into consideration then definitely the researcher can frame a qualitative and accurate schedule. By considering the above steps he can also be able to check problem of response in schedule. Except all these above discussed steps, according to P.V. Young, the essentials of a good schedule can be divided into two parts. They are : (a) accurate communication (6) accurate response.

Accurate communication is achieved when the respondents understand the questions clearly without any ambiguity. According to her, the basis of accurate communication or clear understanding is presentation of questions with proper wording. The researcher should try to frame the schedule with those words that will clearly carry the desired sense without any ambiguity.

Accurate response can be achieved when the researcher **will** get unbiased and true data from the respondents. An adequate length, attractive physical structure, clear wording,

right type of questions etc. can be taken into consideration in order to fulfill this objectives.

3.3.4.2. MERITS OF INTERVIEW SCHEDULE:

Schedule as a method of social research has been used extensively everywhere. Following are the chief merits of the schedule.

(1) Higher percentage of response:

In fact the return of the questionnaire depends exclusively on the goodwill of the respondents. If they have slightest doubt or suspicion they would not respond. But in the schedule the research worker is present, who can explain and clear the doubts and motivate the respondents to give replies. Therefore, getting responses in schedule method is much higher than other methods.

(2) Personality of research worker motivates respondents :

In case of schedule the research worker is takes the job of filling the schedule. He is a well educated and trained person in the concerned field. Here the researcher also knows the general habits, attitudes, behaviour, likes, dislikes etc. of the respondents. This prior knowledge about the respondents and his own presence in the field always help him to get proper responses from the respondents. With his pleasant personality and contact the researcher can also get very sensitive and personal data from the respondents.

(3) Personal contact:

Only a good researcher who is able to establish sympathetic contact with the respondents can better elicit proper responses. The researcher should create enthusiasm in the respondent towards the research topic. This can be done through a personal contact with the respondent. Schedule as a method of data collection provides enough scope for an interviewer to establish rapport or a healthy primary relationship with the respondents. By establishing rapport he comes to know about many details of respondent's life. His personal contact also helps him to collect many valuable information, private, secret and personal data from the respondents.

(4) Creates a proper atmosphere for getting response :

A proper atmosphere for research is very much essential in order to get correct replies. Through a schedule an interviewer may, initiate different kinds of informal and formal discussions, clear the doubts of the respondents, motivate them for participating in the entire research project.

(5) Schedule requires less time to be filled:

In case of questionnaire, it is sent to the respondents by post, the respondents take their own time to fill it. Sometimes they do not fill it. So it takes lot of time. The difficulties of bad writings, spelling mistakes etc. may lead to the problem of response in questionnaire. But in case of schedule it is filled by the investigator himself. He may follow any means or any shortcut method of recording. So the interview can take place quickly and data collected without any unnecessary delay.

(6) Human touch :

In other methods, the respondents do not feel a human touch because the researcher is not present there. Again sometimes reading a thing does not impress so much. But when the same is listened or spoken by experts, it becomes more effective. When the respondents listens to the same thing from an expert, they rather feel the importance of the issue and it greatly helps the researcher in eliciting a better response.

(7) Minimizes errors :

In case of schedule the answers are filled by the research workers who are socially and technically trained for this job. So they commit less error. But in case of questionnaire the responses are filled by the untrained respondents and therefore they are likely to commit more errors.

(8) Applicable to all :

Questionnaire is limited to the educated people. But schedule is applicable to all types of people.

(9) Sampling errors can be corrected :

Here the personal contact of the field worker may unmask any defect in the sampling method. Even if some significant cases are not covered in sampling, that can be included later at the time of actual interview with the help of a schedule.

(10) Left out questions can be discussed :

If any important questions have been left out in a questionnaire, the researcher may not be able to get any response for that question but if some important questions or important aspects of the topic have been left out in schedule, the same can be discussed by direct conversation with the respondents during the visit of the researcher.

(11) Doubts can be cleared :

In questionnaire if the respondent can not understand the meaning of a question, he either leaves it blank or fills that with some biased answers. But in schedule if there is

some doubt, it can be made clear by the interviewer. He can also check other biases through different cross examination or by asking various cross questions.

3.3.4.3. DEMERITS OF SCHEDULE

In spite of the above mentioned merits of the schedule method, it also has some limitations like any other method. The following are some of the important limitations or demerits of schedule method.

(1) Expensive

As compared to questionnaire, schedule is quite costly, The cost for a case is much higher in the schedule than in case of questionnaire. This is again a serious problem when the respondents are not found in a particular region but scattered over a wide spread area. To approach them all is always expensive for a researcher. Schedule needs a large number of field workers for the collection of information from the respondents. Extra staff may also be required to supervise other works. All these need a lot of expenditure, but the investigator with limited finance a faces lot of difficulties to meet these expenses.

(2) Time consuming

In comparison with the mailed questionnaire schedule is time consuming. Here the researcher has to spend much time for a respondent also.

(3) Adverse effect of personal presence

Whereas the personal presence is helpful in getting the response, it is also a source of bias in the interview. Sometimes the respondent may be influenced by the investigator and gives replies not of his own but provides only that answer which the interviewer wants.

(4) Less area

When the respondents are spread over a larger area, it is difficult on the part of an interviewer to contact each respondent for the interview. It requires a lot of time, money and energy of the researcher. So a researcher may face much difficulty to cover a large area through schedule method.

(5) Uncomfortable for respondents

Many people can write down certain facts on paper's but can not say them comfortably in the presence of others.

3.3.5. QUESTIONNAIRE

Questionnaire provides the most speedy and simple technique of gathering data about groups of individuals scattered in a wide and extended field. In this method, a questionnaire

form is sent usually by post to the persons concerned, with a request to answer the questions and return the questionnaire. According to **Goode** and **Hatt** "It is a device for securing answers to questions by using a form which the respondent fills in himself. According to **G.A. Lundberg** "Fundamentally the questionnaire is a set of stimuli to which illiterate people are exposed in order to observe their verbal behaviour under these stimuli".

Often the term "questionnaire" and "schedule" are considered as synonyms. Technically, however, there is a difference between these two terms. A questionnaire consists of a set of questions printed or typed in a systematic order on a form or set of forms. These form or forms are usually sent by the post to the respondents who are expected to read and understand the questions and reply to them in writing in the spaces given for the purposes on the said form or forms. Here the respondents have to answer the questions on their own. On the other hand schedule is also a form or set of forms containing a number of questions. But here the researcher or field worker puts the question to the respondent in a face to face situation, clarifies their doubts, offers the necessary explanation and most significantly fills their answers in the relevant spaces provided for the purpose.

Since the questionnaire is sent to a selected number of individuals, its scope is rather limited but within its limited scope it can prove to be the most effective means of eliciting information, provided that it is well formulated and the respondent fills it properly.

A properly constructed and administered questionnaire may serve as a most appropriate and useful data gathering device.

3.3.5.1. Characteristics of a Good Questionnaire ;

(1) It should deal with an important or relevant topic, so that it will motivate the respondents to give free and spontaneous response. Its significance should be carefully stated on the questionnaire itself or on its covering letter.

(2) It should elicit only those data which cannot be obtained from other sources like books, reports and records.

(3) Questionnaire should be as short as possible. It should be only long enough to get the essential data. If it is too lengthly the response will be poor. Long questionnaires are frequently thrown into the waste paper basket.

(4) It should be at the same time as much comprehensive

as necessary so that it does not leave out any relevant and crucial information.

(5) It should be as attractive as possible in appearance, systematically arranged and clearly printed.

(6) Direction should be clear and complete, important terms should be clarified, each question should deal with a single idea and the questions must be worded in as simple and clear manner as possible which will provide an opportunity for easy, accurate and unambiguous response.

(7) The questions should be of objective type and the researcher should not give any hints or suggestions in order to get his desired responses.

(8) Questions should be presented in a good psychological order proceeding from general to more specific responses. This order will help the respondent to organize his own thinking so that his answers are logical and objective.

(9) The opening question should create a favourable attitude before any progress is made towards the questions, which are a bit delicate or intimate.

(10) The questions included in a questionnaire should be within the respondent's intellectual capacity to give answer.

(11) Various questions given in a questionnaire must be asked in a proper sequence, so that those will be more systematic, interesting and continuous ones.

(12) The offending, embarrassing, ambiguous, double barreled, vague, suggestive, presuming, personal, hypothetical, sensitive and too long questions are to be avoided.

(13) As far as language is concerned, various abbreviations, value-loaded words, native or unusual words,

multi-meaning words etc. are to be avoided.

(14) In the beginning of the questionnaire the investigator should include simple questions which will put the respondent at ease and build up a healthy relationship between the investigator and the respondent.

(15) Questions seeking the advice of the respondent may be given in the beginning of a questionnaire.

(16) The general lay out or physical appearance of the questionnaire is very significant. Questionnaire without a descent layout and proper planning may bring problems of response frequently. Therefore, the paper used for printing of questionnaire should be of good quality. It should provide a proper margin. In between different questions, title and columns, there must be a reasonable space.

(17) Items should be arranged in specific categories to ensure easy and accurate responses and to help the respondent to give better justice to any category at a time.

(18) Descriptive adjectives and adverbs that have no agreed upon meanings, (for example, words like frequently, occasionally, rarely etc.) should be avoided as far as possible.

(19) Double negatives, for example, "Don't you disapprove of forced discipline?", should be avoided.

(20) The questions should carry adequate number of alternatives. Questions which carry inadequate number of alternatives, for example, "Are you married ? Yes/No"— should be avoided.

(21) A good questionnaire should facilitate easy tabulation, summary and interpretation of data. It is advisable to pre-construct a tabulation sheet, anticipating how the data will be tabulated and interpreted, before the final form of questionnaire is decided upon. This will help in avoiding ambiguity in the questionnaire.

3.3.5.2. Construction of questionnaire, Concept, types of question, question format and sequence of questions:

In case of questionnaire the greatest problem is that of response that arises because of the problem of clarity in questionnaire. **In** an **interview** or schedule the field worker is personally present to give a proper clarification or interpretation to the questions and different terms given in a schedule. The respondents can ask about their doubts and understand the meaning of it properly in schedule method. Usually they do not refuse to answer the questions put by the interviewer. But in case of questionnaire this facility is not available. Nobody is there to aid the respondent to interpret a question or to clarify the meaning of some important terms used in it.

One has to take much care in preparing the questionnaire, the construction of questions and the language to be used. A thorough testing of questionnaire is required before it is finally sent to the respondents. While preparing the questionnaire the researcher should remember that it is not meant for the respondent of high or average .intelligence, rather for the respondents of minimum intelligence. Therefore much care is needed in preparing questionnaire, its general physical outlook and the types of questions given in it, so as to justify it as a special schedule, *i.e.* although the researcher is not present in front of the respondents, the questionnaire will also explain the subject matter to the respondents in a simple and clear manner and facilitate a high response from them. In other words the questionnaire will be self explanatory.

However, in this respect the following are the important points which are to be considered by the researcher.

(1) **Importance of the problem under study or the formulation of the problem** :

The formulation of problem forms the starting point for developing the questionnaire. If the problem under study is an important one, a higher response is expected, but if it is a ordinary one, and which does not have any social relevance for the respondent's life, the response is likely to be low. The problem formulated for study should be relevant for the respondents. A researcher can expect higher response, if the problem under investigation directly focuses on the respondent's problem.

(2) **The type of information required** :

Very extensive bodies of data can not often be secured through the use of the questionnaire. The researcher should first find out the extent to which the desired data are already available in published reports and decide whether all or parts of the needed data

can be obtained through a formal questionnaire. In brief, the investigator must decide what aspects of the problem are to be dealt with in a particular study through a formal questionnaire.

(3) Securing help from the people who possess the experience in the related field:

The researcher should secure all the help he can, in planning and constructing his questionnaire. He should study other questionnaires and submit his questionnaire for a critical analysis to other members of his research organisation or his colleagues and especially to those who possess the experience of questionnaire construction.

(4) Thorough knowledge about **his hypothesis:**

He should thoroughly explore his hypothesis, experiences, the literature available in the concerned topic and other related field to frame his own questionnaire. **It will** help him to probe the crucial issues of his research problem in depth.

(5) Clear understanding of the objective of the study

He should attain a thorough knowledge of the field and a clear understanding of the objective of the investigation and of the nature of the data required for the investigation.

(6) A Proper Scrutiny:

The questionnaire should be properly scrutinized for technical defect, quite apart from biases and blind spots arising out of

personal values.

(7) **Numerous revision or pre-testing** :

Constructing a questionnaire calls for numerous revisions in which variations of the same question should be given for **an** experimental trial. The same question posed in different ways very frequently may bring out different responses. This trial can be done through a pre-testing or a pilot study. A pre-testing is necessary to find out how the questionnaire works and whether changes are necessary before applying the actual questionnaire. The pre-testing of elements provides a means for solving unforeseen problems in its administration in the field. It may also. indicate the need for addition or deletion of questions. Sometimes a series of revisions and pre testing is needed. After a pre-testing the final editing has to be done to ensure that every element has been scrutinized. Editing is required to make the questionnaire as clear and as easy to use as possible.

(8) Willingness of the respondent:

The questionnaire is effective only when the respondent is able or willing to express his reactions clearly. The respondents must be chosen carefully. The questionnaire should be sent only to those who possess the desired information and those who are likely to be sufficiently interested to respond. **Goode** and **Hatt** say, "the respondent will not ordinarily answer a questionnaire dealing with a subject with which he is unfamiliar, such as air travel, impending legislation or experience with a particular branch of the government or an occupation".

(9) Precision of the hypothesis:

The questionnaire is mostly useful when a considerable amount of exploratory work has narrowed the questions to be answered. What is more important here is the sharpness of the hypothesis. The more closely focused the hypothesis, the more effective is the questionnaire.

(10) Size of the questionnaire:

The questionnaire should be as short as possible. It should be only long enough to get the essential data. If the questionnaire is too lengthy the response is likely to be poor.

(11) Letter of Appeal:

A cover letter almost always accompanies the questionnaire, requesting it the respondent to provide his co-operation. In the researcher should explain what he is going to do, why he is doing it and for whom he is doing so? This appeal letter generally contains the name of the research institution carrying on research, the objective of research, the benefit that is likely to go for respondent or people in general. However the appeal should be brief, properly worded and impressive. Most of the respondents are not willing to read a long letter and lengthy appeal letter destroys its impact.

(12) Prestige of the research institute:

For the success of the questionnaire the prestige of the research organization matters a lot. Usually if the research organisation conducting the study is well reputed, responsible, and one scientific group, the response **is likely** to be high. People generally do not show their interest for giving response if the institution sanctioning the study is not known, reputed or of a doubtful integrity and in this case response is likely to be low.

In questionnaire method the cover letter should explain the character of the research organisation and its objective in a few sentences. It should also provide the impression of scientific competence, the address and telephone number of the research organization. Nothing should appear to be hidden, ambiguous and suspicious.

(13) **Purpose of the study** :

The researcher may include the purpose of the study in his introductory remarks. In questionnaire's cover letter it is essential to explain why research organisation requires such type of information from the respondents. However **Goode** and **Hatt** say, "the description of the purpose of questionnaire should be left out unless the respondent asks for an explanation of it".

(14) Appearance or a general layout of the questionnaire

While preparing a questionnaire a significant attention is required regarding the general layout or appearance of the questionnaire. Here unlike schedule the field worker is not present personally in the field to clear the doubts of the respondent. So this task has to be performed by the perfect appearance of the questionnaire itself.

(a) **Quality of paper:** The paper used for printing of the questionnaire should be of high quality. So that it will be durable and the letters printed on it -will be clearly visible. If the paper quality is low, the printing on it will not be visible and the *ink* may spread over it. Therefore the paper should be of good quality. Otherwise it may lead to a series of problems of response for the questionnaire.

(6) Spacing; between the questions, other titles and sub titles there must be proper space, so that the respondent can clearly and freely write his responses and it will be properly visible to the researcher also.

(c) **Margin:** A proper margin provides a better look to the questionnaire form. That apart, in order to keep the records systematically the researcher needs to punch and staple the questionnaire form. But if no proper space is provided in the questionnaire, the punching will destroy some of its written words.

Printing: A proper printing is obviously more desirable because it attracts the respondent for a better response. The questionnaire form should be typed or printed carefully. The printed letters must be clearly visible, neat and free from over writing. Otherwise the respondent will not understand the question properly and leave most of the questions without giving, any answer.

(e) **Use of pictures:** Often various relevant pictures should be inserted whenever possible in a questionnaire in order to attract the respondent for a better response. A less educated person may not understand a written question but by observing the picture he may understand it and respond to it.

(15) Language of the question:

Understandably great care is required in using the language of the questions. Various unknown abbreviations, multi meaning words should be avoided by the researcher, because these may be known to the researcher but respondent may not understand it. Again in the questionnaire the researcher is not present in the field. So here the respondent does not get a chance to clarify his doubt. In case he faces any difficulty in understanding some questions, he may not be able to answer those questions. So the language of questionnaire should be simple and unambiguous.

(16) Types **of questions:**

The ambiguous, double barreled, complex, suggestive, vague, sensitive, normative, hypothetical, personal and too long questions shou'd be avoided by the researcher as far as possible, because thes*e* questions may not bring a proper response from the respondents. Questions whose answers can be secured more accurately from other sources can be excluded.

(17) Sequence of the questions

It is essential to examine the order in which questions are to be asked. While framing questions the researcher should carefully consider the best sequence of the topics in a questionnaire. Questions should be arranged logically so as to determine the directions of the responses. It is always better to start with a simple, general, unambiguous and uncontroversial question and then proceed towards more complexes, specific and personal questions.

Question which might embarrass the respondent or question seeking secret information should be put at the end. The researcher should also give some closely related questions in order to measure consistency and for checking the reliability of responses.

There are some methods through which the researcher can check the problems of response. These are discussed below.

(18) Sending reminder:

The respondents are often slow to return completed questionnaires. To increase the number of returns, a vigorous follow-up procedure is necessary. In some cases a reminder may be sufficient but in extreme cases a telegram, phone call or personal visit may bring rapid responses.

(19) Inducements :

Some researchers are of the opinion that in order to get a proper response some kind of inducement is always required. Inducements are broadly divided into two types :

(a) Monetary inducement ; and (6) Non-monetary inducement.

(a) Monetary inducement: In questionnaire method some inducement in the form of money may be given to the respondents. The amount is generally very small. It may be sent to the respondents with the envelope containing the questionnaire form or the researcher may promise to give it after the successful return of the questionnaire. It is always better to send the money in advance than to promise payment on return of the questionnaire.

Another method of monetary inducement is to offer prizes by lottery. For example, the researcher may offer a few prizes on all accurate and completed forms on the basis of lottery.

(b) **Non-monetary inducement** : Often non-monetary inducements are more influential than the monetary inducements in order to get proper responses from the respondent. These may consist of the benefit that the respondent is likely to gain from the outcome of the study. If the respondents feel that the study is beneficial for them, they give more responses. Sometimes encouragement and motivation provided by the researcher works as non-monetary inducement for the respondent.

The researcher should praise the respondent for his valuable responses by saying— "Your information is needed by thousands of scholars attempting to solve today's problem", "You will be contributing to the advancement of science", "You will help to improve the education of thousands of students" etc. This will encourage him to provide more valuable information.

(20) Questionnaire through intermediaries:

Sometimes, the questionnaires instead of being sent directly to the respondents are sent to the head of a villages or an officer in charge of an organisation. He gets them filled up and returns them to the researcher. It has been observed that the percentage of response is very high when it is filled up through the intermediaries or the leader of the respondents. But it is also having some negetive consequence as the respondent may feel offended and give unauthentic reply.

(21) Proper time of sending the questionnaire:

The researcher should send the questionnaire in such a time that it will reach the respondent on week-end days. Usually the busy respondent gives reply when he is free. So the week-end days are probably the best days to utilise for responding to the questionnaire. But if the questionnaire reaches on the initial days of the week, it may be misplaced by the week-end. Therefore in order to get high response the questionnaire should reach on the week end days.

(22) Knowledge about the correct address of the respondent:

Sometimes the respondents are unable to send the answers because they do not'*get the questionnaire. If the researcher sends the questionnaire in wrong address, it may not reach the

respondent. Therefore, a proper knowledge about the respondent's address is very much essential for getting proper response.

It is difficult to estimate as to what percentage of responses can be considered adequate for an investigation. Importance of the project, quality of the questionnaire, nature of the project, nature of the group of the respondents selected, the duration and many other factors determine the proportion of responses that could not be considered adequate.

In case of interview or interview schedule, it is the investigator who motivates the respondents for better response. But in questionnaire nobody is present in the field and only its several pages of paper determine the course of response.

According to **Goode** and **Hatt,** "Only the papers are there to make his plea, and the researcher cannot count on any personal charm or social skill when the respondent opens the envelope".

Therefore the questionnaire maker, then must offer as impressive presentation as possible in order to get-adequate response. He must plan carefully and seek professional help before he sends out **ills** questionnaire.

If all the above discussed precautions are taken into consideration, the questionnaire can be thought of as a special schedule, which can enable to bring maximum possible response in spite of the absence of the investigator in the field.

3.3.5.3. Types of Questionnaire

There is a vast variety of questionnaires that has been classified in several ways. P.Y. Young has confined all the major types of **questionnaires** into three types *viz.* structured, unstructured and pictorial questionnaire.

(1) **Structured questionnaire:** According to P.V. Young, structured questionnaires are those which pose definite, concrete and pre ordained questions, *i.e.,* they are prepared in advance and not constructed on the spot during the questioning-period.

This questionnaire uses highly standardised techniques and a set of pre-determined questions. It includes both closed end and open ended questions. Closed questions are used when categorised data are required or when the researcher wants to make various classifications for his study. A simple example of closed end question is : "How many from your family are educated ?" One only/two/three/four/five or more than five. The respondent goes through all those given responses and chooses one which is true for his situation. The open end questions provide a proper space for the respondent to answer. According P.V. Young "The open end responses are free and spontaneous expression on the part of the informant who is not limited in his replies to a particular question posed on him". Here the subject can write freely and frankly their concrete views with no directions from the researcher, She also remarks that "the open-ended responses are used chiefly for intensive studies of a limited number of cases or for preliminary exploration of new problems and situations. At times, the respondent is asked to write a descriptive essay and express his view points, describe his relationships, attitudes, indicate his problems and report on events, without restrictions imposed as in the case of closed questions". An example of open end question is—"What are you thinking about the educational qualification of your family members ?"

Open ended questions are very much essential for exploratory research, for pilot study or where the responses to be collected are qualitative in nature.

The open-ended questions have also its demerits. Since it does not provide any direction and restriction for the respondents to answer, a wide range of answers is usually given, some meaningful and some meaningless, and obviously those meaningless, non-directed and irrelevant answers create some problems of classification and analysis. However in spite of these structured questionnaires are used in a wide range of research works both to initiate collection of primary data or to add and verify data previously collected.

(2) **Unstructured** questionnaire : P.V. **Young** says, "Unstructured questionnaires are frequently referred to as

"interview guides", also aim at precision and contain definite subject matters area, the coverage of which is required during the interview". This is characterised by a greater flexible approach in questioning the respondents. It is of non-directive type which involves relatively much less standardization of techniques and operation. Here the respondents have the freedom to express any event that seems significant to them, to give their own definition of an event or a situation and to narrate any particular incident of his life. In unstructured questionnaire the researcher is also having a greater freedom to ask any supplementary question of the respondents.

Such a type of questionnaire is very much useful for intensive studies and can be used as the major technique of collecting information in exploratory and formulative studies. But at the same time it is also having its own limitations. Such flexibility is not helpful for making a comparative study or to make different classifications and often it is more difficult and time consuming to analyse these non-directive responses.

(3) **Pictorial** questionnaire : Pictures have been used in some questionnaires in order to promote some interest and motivation among the respondents for answering the questions. It is useful for those respondents who are least educated. P.V.
Young- says that pictorial techniques have been used extensively **m** studies of social attitudes and prejudices in children.

3.3.5.4. DISADVANTAGES OF QUESTIONNAIRE

The questionnaire is probably the most used and the most abused data gathering device. It has been referred to as the lazy man's way of gathering information. As an method of data collection it is also having many limitations. Let us now turn to discuss the typical disadvantages or limitations of the questionnaire as compared to other major methods of collecting data for research.

(1) **Limited response** :

One of the major limitations of the questionnaire is that it can be applicable only to those respondents who have a considerable amount of education. It can neither be used for illiterate nor for semi-literate persons.

The questionnaire quite often fails to cover very busy and pre-occupied persons among the respondents, lazy and indifferent type of persons, the type of respondents who need to

conceal a lot about themselves, the easy-going and shirkers among the respondents, the persons who have a unreasonable contempt for research and reform and the persons who unnecessarily doubt the research worker's intentions, sincerity, devotion and commitment. These are the people who constitute a very important segment *of* the respondents to be covered in the collection of data, but they can be seldom caught. Thus questionnaires are hardly appropriate for a larger section of this type of **population.**

(2) **Lack of personal contact** :

As in case of questionnaire the researcher does not go to the field, he is not able to establish a proper personal relationship with the respondents. If the respondent fails to understand some of the technical terms or he has any doubt, there is nobody to clarify these technical terms or doubts. Even though the researcher tries in the best possible manner to make the questionnaire a simple, precise and convenient one, the aim and objective of the questionnaire can be much better explained personally than through any other means. Without the proper personal contact it is very difficult to motivate the respondent to fill up the questionnaire.

(3) **Poor response** :

In case of mailed questionnaire method, the proportion of return is usually low. The factors which are likely to affect the returns are : the layout of the questionnaire, its size, the organisation conducting the research work, the nature of appeal, the kind of respondents chosen for research, inducement for response etc. The causes of poor response has been discussed at length in the early part of this chapter.

(4) **Unreliability** :

The information collected through questionnaire cannot be said to be very much reliable or valid. If the subject mis-interprets a question or gives an incomplete or indefinite response, very little can be done to connect such response. As against this, in an interview there is always the possibility of rephrasing questions for further clarification. The questions can be repeated, with adequate elaboration if it is so required. But in questionnaire method there is no opportunity for repeating questions, explaining' them or clarifying the doubts for a particular response. Therefore, in it the validity of respondent's response can hardly be examined. The investigator here is not in a position to observe the gestures and expressions of the respondents. He can not cross check the inconsistencies or misrepresentation of the replies. So in questionnaire method, reliability of responses is very low...

(5) **Illegibility** :

Illegible handwriting of the respondent sometimes creates much difficulty for the researcher to understand the responses. Sometimes the respondents erase and overwrite too much. These create many difficulties in reading the answers,

(6) **Incomplete entries** :

Often most of the respondents fill up the questionnaire form very poorly. They sometimes leave out many questions altogether or fill in such a way that, it becomes very difficult on the part of the investigator to follow those responses. Other than this, there may be the problem of language, use of abbreviations and ambiguous terms etc. All these make a questionnaire an incomplete one.

(7) **Possibility of** manipulated **entries** :

In case of interview the investigator directly interacts with the respondents personally and intensively in a face to face situation. He can judge-a respondent, his attitude, understanding of the research topic and, if necessary, can ask some cross questions to correct various errors. So usually the respondent can not manipulate his answer. But in questionnaire it is very difficult to detect the errors of the respondents. Here the investigator does not have any facing to check the validity and reliability of the information. In the absence of the researcher, the respondents may supply manipulated information.

(8) Useless in depth-studies:

In questionnaire method, it is not possible on the part of the researcher to conduct an intensive or in-depth study of the feelings, reactions and sentiments of the respondents. All these require a healthy interaction of the researcher with the respondents. But in questionnaire method, the investigator is not present in the field, so nothing can be done to establish rapport with the respondent. Due to this lack of interaction with the respondent, the researcher cannot go into the details of the, respondent's life. So through questionnaire method one cannot conduct an in-depth study.

(10) Response front improper representative section of people :

(11) The respondents who return the questionnaires may not constitute a representative section of the entire group. Only mere responsible, research minded or those in favour of the issue may prefer to respond. Some of the important sections of the group may totally remain silent. This vitiates the final conclusions and findings.

(10) Lack of rapport with the subject:

There are many people who would not like to share any important information unless and until they are impressed about the rationale of the study and personality of the

investigator. The questionnaire does not provide for any opportunity to the investigator to establish rapport with the subject and this cannot attract the respondent for a better response.

(11) Not suitable for delicate issues:

Some of the research areas are so delicate, sensitive, intricate confidential in nature that it becomes difficult to frame questions on them. It is impossible to put down certain delicate

issues in writing.

3.3.5.5. ADVANTAGES

Owing to the above mentioned limitations, the questionnaire

method is unsuitable for the study of many social problems. Its use may bring the conclusion highly unreliable. But in spite of its limitations the questionnaire is regarded as the mos4 useful research tool. As an instrument of science, the -questionnaire has great potentialities when it is properly used. If it is eliminated, progress in many areas of research would be greatly handicapped. The following are the chief advantages of mailed questionnaire. .

(1) Economical: It is an economical way of accumulating information. It is economical both for the sender and for the respondent in time, effort and cost. The cost of conducting the study with the help of questionnaire method is very low. In questionnaire the researcher has to spend for paper printing and postage only. There is no need to visit each and every respondent personally. So it does not require high cost for conduct of the research.

(2) Wide coverage: It is probably the best method to collect information, compared to the other methods like interview or observation, when the sample population is spread over a large territory. It permits a nationwide or even international coverage.

Questionnaire makes it possible to contact with many people who could not otherwise be reached. It can cover a large group at the same time. Goode and Hatt say that when the researcher has to cover the group of respondents who are widely scattered, he can use the questionnaire in order to minimize the cost. For example, if the researcher wishes to poll the membership of the American Sociological Society, transportation costs for interviewing

would be excessive, both in terms of money and time. There may not be enough time to make the necessary interview. However, questionnaires could be distributed to all those members and information could be collected from them. This can be dons by a single researcher without the large funds otherwise required to hire an interviewing staff to carry out the interview.

(3) Rapidity : Replies may be received very quickly in questionnaire method. In this case there is no need to visit the respondent personally or continue the study over a long period. Therefore in comparison with other methods, the mailed Questionnaire is the quickest method.

(4) Suitable in special type of response: The information about certain personal, secret matters can be best obtained through questionnaire method. For example, information about sexual relationship, marital relationship, secret desires etc. can ,be easily obtained by keeping the names of the respondents anonymous.

(5) Repetitive information: Compared to other methods like schedule, interview or observation, questionnaire method is regarded as more useful and cheap, where the repetitive information has to be collected at regular interval.

(6) An easier method: Questionnaire is comparatively an easier method to plan, construct and administer. It does not require much technical skill or knowledge.

(7) It puts less pressure on the respondents: It puts less pressure on the respondents for immediate response. He can answer it at his own leisure, whereas interview or observation demands specific fixation of time and situation.

(8) Uniformity : It helps in focusing the respondent's attention on all the significant items. As it is administered, in a written form, its standardized instructions for recording responses ensure some uniformity. Questionnaire does not permit much of,-variation.

(9) Useful preliminary tool: Questionnaire may be used as a preliminary tool for conducting a depth study later on by any other method.

(10) Greater validity: Questionnaire has some unique merits as regards validity of information. In methods like interview and observation, the reliability of responses depends on the way the investigator has recorded them. Here they may present biased or prejudiced information of their own. But in questionnaire method, the responses given by the subjects are available in their own language and version. Therefore, it cannot be wrongly interpreted by the researcher.

(11) Anonymity: Questionnaire ensures anonymity to its respondents. The respondents have a greater confidence that they will not be identified by anybody for giving a particular view or opinion. They feel more comfortable and free to express their view in this method.

(12) Most flexible tool for data collection: Questionnaire is no doubt the most flexible tool in collecting both quantitative and qualitative information.

3.3.5.6. PERSONAL INTERVIEW AND MAILED QUESTIONNAIRE

Often questionnaire and interview schedule used in personal interview are treated as synonyms. Technically, however, it is desirable to differentiate between them. The following are some of the major differences between mailed questionnaire and interview schedule used in personal interview

(1) The main differences between the two is the fact that while the schedule used in personal interview and filled in by the field worker himself,

the questionnaire is generally mailed to the respondents who fills it up and return it to the researcher. Thus, the main difference between the two is regarding the method of collecting data through them.

(2) Large areas can be covered through questionnaire method, but it is very difficult to cover a large area by using personal interview method.

(3) Along with schedule, observation and interview method can be used. But mailed questionnaire can not be used in combination with other methods.

(4) In order to attract the attention of the respondent the

physical appearance of mailed questionnaire must be attractive, but this is not required for schedule in personal interview method. Because in schedule the responses are not filled in by the respondents. Here all the answers are filled by the investigator only.

The success of the questionnaire method depends on the quality of questions, length of questionnaire, attractiveness of the questionnaire etc., but in case of schedule it depends much upon -the honesty and the personality of the investigator.

(6) Questionnaire is impersonal in character, whereas schedule is personal in character.

(7) In comparison with personal interview, mailed questionnaire is very economical. It requires less money, time and energy.

(8) As far as coverage of population is concerned, mailed questionnaire is limited to only literate people, but schedule is applicable to all. In case of a schedule education of people is not a minimum requirement for giving responses because here the researcher himself fills the responses of the respondent.

(9) Appeal-letter is used in case of mailed questionnaire to motivate the respondent to send their responses. But it is not required for personal interview method. In this method, the researcher can directly interact with the respondent and can motivate them for giving their responses.

(10) As far as reliability is concerned the personal interview is reliable because the investigator is present in the field and influences the respondent to give the responses. But in mailed questionnaire method, reliability of information is very less. Because often the respondents do not send the filled in questionnaire form or send only an incomplete form to the researcher.

(11) Replies can be received very quickly in questionnaire method but the personal interview method requires more time to

collect the responses from the respondents.

(12) In comparison with the schedule used in interview, mailed questionnaire is more useful and economical, when the researcher has to go for a repetitive study.

(13) Questionnaire is comparatively an easier method to plan, construct and administer than personal interview.

(14) Questionnaire puts less pressure on respondents for immediate response than personal interview.

(15) Questionnaire can be used as a standardized device and uniformity can be ensured in questionnaire. But such uniformity cannot be ensured in personal interview method.

(16) The researcher may get manipulated entries in mailed questionnaire method. Because here the researcher does not get an opportunity to cross check the responses of subjects, but in personal interview method the researcher can check manipulated answers of the respondents through direct interaction with them, by judging their attitude and by asking many cross questions to the respondents.

Electronic resources and its usage:

E-resources are an accepted means of information resources in the present information society; with the amount of new resources available, the need for adequate computer literacy and aptness in using the existing sources has become the need of the hour and the studies reveal its impact in terms of awareness and effective use of the available resources by the scientists of digital libraries of India.

Cost, the level of importance, and the use of e-resources have dramatically increased in the digital library environment; Web DB, e-book, e-journal, and other e-resources such as CD-ROM, DVD, and micro materials have become important sources in libraries.

Librarians of digital libraries should use new information technologies and new approaches to better serve their scientists in new ways of acquiring information. Libraries

should organize their services so that they bring their information resources closer to the busy scientists. They should acquire new skills and learn how to organize information for presentation on the Internet portals or by handheld devices. Additional drive for librarians in their efforts is the introduction of evidence-based practice. Librarians can teach scientists to search and critically evaluate information, thus helping in the process of their research. Librarians with experience in licensing and managing ERs are effective in rapidly diagnosing and remedying access problems.

3.4. Key Words

Primary source, Secondary source, Observation, Social survey, interview guide, Interview schedule, Questionnaire, Electronic resource,

3.5. References:

1. Whitley, Frederick, The elements of Research.
2. Goode and Hatt, Methods in social research.
3. Young.P.V., Social Survey and Social Research.
4. Kalton & Moser, Survey Method in Social Investigation.
5. Lundeberg, Social Research.

UNIT- IV PREPARATION OF RESEARCH PROPOSAL

STRUCTURE

1. OBJECTIVES

By studying this unit, you will able to

I. write a research proposal,

II. get an idea about the processing of data by editing and coding,

III. classify the data and present the data in the form of tables, graphs and charts,

IV. prepare a research report,

V. have a knowledge on the ethical aspects to be considered while conducting a research,

VI. know about different agencies involved in social research.

2. INTRODUCTION: WHAT IS A RESEARCH PROPOSAL?

A research proposal "*is a document that outlines how you propose to undertake your research studies*" (Mouton, 2001). Essentially, it outlines *what* we will research and *how* we will research it. The "what" part is called the *problem*; the "how" part we call the *plan*. A proposal describes a problem and sets out a logical, systematic plan to solve it. A proposal is an offer to produce or render a service to the potential buyer or sponsor. Thus, the research proposal presents a problem, discusses related research efforts, outlines the data needed and shows the research design. The following are some important usefulness of a research proposal:

- sponsor uses proposal to evaluate research idea,
- ensures the sponsor and investigator agree to research question,
- for newcomer, research proposal helps learning from others, and a
- completed proposal provides a logical guidance.

Writing the research proposal is the most difficult and the most important part of the entire research project. If we rush through this step, we will have a poorly conceived research problem and plan. This makes the rest of the study difficult. On the other hand, if we invest time and effort to produce a first class proposal, the rest of the study should fall into place.

The greatest value of a proposal is that it keeps the research project on course. It gives direction and focus to the project. It prevents us taking time consuming, energy sapping digressions. If we invest the time and energy at the start of our project to think through what we will research and how we will go about it, formulating a clear picture in our mind of our destination and the route we will take to get there, our journey should go without delay or detour.

For example, if we carefully and precisely formulate our research problem, keeping it narrow and focused and identifying which aspects to include and exclude, we dramatically reduce the amount of reading we need to do. This saves time. If our problem is vague and fuzzy, we will read five times as much because we have no clear yardstick to distinguish what we must read from what we need not read. Similarly, if we think through the essential steps we must take to solve the problem, we reduce the chance of wasting time gathering data not necessary to solve the problem.

In short, prepare a good proposal and our research will flow; prepare a poor one and it will flop. An old adage—"*failing to prepare is preparing to fail*"— certainly applies to research. Yet students often do a poor job of preparing their proposals. Why? Partly due to ignorance, that is, not understanding the nature of research well enough to plan the project. I suspect over-eagerness is an even bigger cause. In their haste to get on with "the real work", they slap together a poorly conceived proposal. Preparing a quality proposal is hard work. It

requires much reading and reflection. It is time-consuming. However, in the long run, it saves time. So, we should invest quality time in the proposal. Prepare it well. What we sow into the proposal, we will reap in the report.

If we are convinced of the value of preparing the proposal properly, we will be wondering what elements should go into a research proposal. Views differ from one researcher to another. In the next section, we will present some preferred breakdown of the elements that belong in a proposal.

3. THE ELEMENTS OF A RESEARCH PROPOSAL

A research proposal consists of two main parts: the *research problem* and the *research plan*. The first part, the research problem, addresses "*the what*" of the study; it describes the problem the researcher will attempt to solve. Part two, the research plan, focuses on "*the how*" of the study; it explains how the researcher will go about solving the research problem. Let us examine each part.

3.1 PART 1: THE RESEARCH PROBLEM

The first part of the proposal should state the research problem with the utmost focus and clarity. The problem the researcher will attempt to solve needs to be defined and delimited with such precision as to leave no confusion or ambiguity as to what the research is about and what it aims to accomplish. The more clearly and precisely the research problem is laid out, the more focus the research will have.

Our problem should be so clearly stated that anyone anywhere in the world (who reads English) may read it, understand it, and react to it without help. If the problem is not stated with such clarity and precision, then we are merely deceiving our self that we know what the problem is. Such self-deception will merely cause us trouble later on (Leedy, 1993).

What components should we include in our description of the research problem?

We should include these elements in this order:

1. The statement of the problem
    a. The main problem
    b. The key questions
    c. The hypotheses
2. The elucidation of the problem
    a. Delimitations of the study
    b. Definitions of key terms
    c. Presuppositions of the researcher
    d. Preliminary literature review
3. The value of the study

a. Theological value
b. Practical value

The statement of the problem: This part addresses the purpose of the investigation. It is here that we layout exactly what is being planned by the proposed research. We should begin our research proposal with a direct statement of the research problem. State the problem as a single sentence (at most a short paragraph). We may formulate it as a statement, a question or an objective. In a descriptive study, the objectives can be stated as the research question. The research question can be further broken down into investigative questions. In large research projects, the main problem is usually too large to solve without dividing into smaller units. In such cases, break the main problem down into 2-6 sub-problems; these are called as "*key questions*", sometimes these are called as "*objectives*". If the proposal is for a causal study, then the objectives can be restated as a hypothesis. A hypothesis is a calculated guess as what the answer to a research question will be. A hypothesis must be directly related to a research problem or question. Thus we may have a hypothesis for the main problem and, if we wish, one for each sub-problem or for some sub-problems.

The objectives module flows naturally from the problem statement, giving the sponsor specific, concrete, and achievable goals. It is best to list the objectives either in order of importance or in general terms first, moving to specific terms (i.e., research question followed by underlying investigative questions). The research questions (or hypotheses, if appropriate) should be set off from the flow of the text so they can be found easily. The research objectives section is the basis for judging the remainder of the proposal and, ultimately, the final report. We verify the consistency of the proposal by checking to see that each objective is discussed in the research design, data analysis, and results sections.

***The elucidation of the problem*:** If we begin with a direct statement of the research problem, there will be loose ends that need to be tied up. This section clarifies the nature of the research by explaining the delimitations, definitions and presuppositions of the study and by presenting a preliminary literature review.

In the case of small projects, literature review may not be in the form of a critical review of the literature, but this is often asked for and is a standard part of larger projects. In its simplest form, the review of literature is a list of relevant books and other sources, each followed by a description and comment on its relevance. The literature review should demonstrate that we have read and analyzed the literature relevant to our topic. From our readings, we may get ideas about methods of data collection and analysis. If the review is part of a project, we will be required to relate our readings to the issues in the project, and while describing the readings, we should apply them to our topic. A review should include only relevant studies. The review should provide the reader with a picture of the state of

knowledge in the subject. Our literature search should establish what previous researches have been carried out in the subject area. Broadly speaking, there are three kinds of sources that you should consult:

1. Introductory material;
2. Journal articles and
3. Books.

To get an idea about the background of our topic, we may consult one or more textbooks at the appropriate time. It is a good practice to review in cumulative stages - that is, do not think we can do it all at one go. Keep a careful record of what we have searched, how we have gone about it, and the exact citations and page numbers of our readings. We should write notes as we go along. We have to record suitable notes on everything we read and note the methods of investigations. We have to make sure that we keep a full reference, complete with page numbers. We will have to find our own balance between taking notes that are too long and detailed, and ones too brief to be of any use. It is best to write our notes in complete sentences and paragraphs, because research has shown that we are more likely to understand our notes later if they are written in a way that other people would understand. We should keep our notes from different sources and/or about different points on separate index cards or on separate sheets of paper. We will do mainly basic reading while we are trying to decide on our topic. We may scan and make notes on the abstracts or summaries of work in the area. Then do a more thorough job of reading later on, when we are more confident of what we are doing. If our project spans several months, it would be advisable towards the end to check whether there are any new and recent references.

The literature review section examines recent (or historically significant) research studies, company data, or industry reports that act as a basis for the proposed study. We should begin our discussion of the related literature and relevant secondary data from a comprehensive perspective, moving to more specific studies that are associated with our problem. If the problem has a historical background, we should begin with the earliest references. We have to avoid the extraneous details of the literature and have to do a brief review of the information, not a comprehensive report. We should always refer to the original source. If we find something of interest in a quotation, find the original publication and then we have to ensure our self that we understand it. In this way, we will avoid any errors of interpretation or transcription. We have to emphasize the important results and conclusions of other studies, the relevant data and trends from previous research, and particular methods or designs that could be duplicated or should be avoided. We have to clearly discuss, how the

literature applies to the study that we are proposing; show the weaknesses or faults in the design, discussing how we could avoid similar problems. If our proposal deals solely with secondary data, we have to discuss the relevance of the data and the bias or lack of bias inherent in it.

The literature review may also explain the need for the proposed work to appraise the shortcomings and informational gaps in secondary data sources. This analysis may go beyond scrutinizing the availability or conclusions of past studies and their data, to examining the accuracy of secondary sources, the credibility of these sources, and the appropriateness of earlier studies.

Under delimitations, we narrow the focus of our study by indicating what we will not research, that is, what we will exclude. The sections on definitions and presuppositions help us to understand the things we treat as "*givens*" (what we assume to be true) and how we use important terms. The literature review places our planned research in the context of related funding, helping readers to appreciate how our study relates to what others have done.

The value of the study: This section allows us to describe explicit benefits that will accrue from our study. The importance of "*doing the study now*" should be emphasized. Usually, this section is not more than a few paragraphs. If we find it difficult to write, then we have probably not understood the problem adequately. In such a case, we have to return to the analysis of the problem and ensure, through additional discussions with our sponsor or our research team, or by a reexamination of the literature, that we have captured the essence of the problem. This section also requires us to understand what is most troubling to our sponsor. If it is a potential union activity, we cannot promise that an employee survey will prevent unionization. We can, however, show the importance of this information and its implications. This benefit may allow management to respond to employee concerns and forge a linkage between those concerns and unionization. We must convince the sponsoring organization that our plan will meet its needs.

The first part of the proposal ends with a motivation for doing the study. We may motivate the study at a practical and/or theoretical level. This section provides an opportunity to introduce some essential background information by indicating how the proposed research will help to address social needs in a community. We may explain who should benefit from the research and how they might benefit.

Now let us examine the elements that constitute the second part of the proposal, the research plan.

3.2 PART 2: THE RESEARCH PLAN

Research follows a problem-solution format; so does the research proposal. The first part sets out the research problem; the second part presents a plan to solve it. The research plan contains thirteen sections: research design, qualification of researcher, budgeting, scheduling, etc..

4. The research design
    a. The model, data and sampling design
    b. The tools and techniques of data collection
    c. Steps for analysis
5. Qualifications of Researcher
6. Budgeting
    a. Financial
    b. Personnel
7. Time scheduling
8. Nature and form of results
9. Facilities and special resources
10. Project Management
11. The annotated bibliography
12. Appendices
13. Measurement instrument

The research design: We have studied the sponsor about what the problem is, what our study goals are, and why it is important for us to do the study. The proposal has presented the study's value and benefits. Our first decision is what kind of study is best suited to solve the research problem. For example, can we use a purely literary approach or do we need an empirical component? The design module describes what we are going to do in technical terms. This section should include as many subsections as needed to show the phases of the project. We should provide information on our proposed design for tasks such as sample selection and size, data collection method, instrumentation, procedures, and ethical requirements. When more than one way exists to approach the design, we have to discuss the methods we rejected and why our selected approach is superior.

The chosen design largely dictates the logic and structure of the study. In large projects, the proposal needs to include proposed timeframes; these serve as a progress agreement between researcher and financer or supervisor.

Keeping in view the above stated design decisions; we may split the overall research design into the following parts.

- *Sampling design,* which deals with the method of selecting items to be observed for the given study;

- *Observational design,* which relates to the conditions under which the observations are to be made;

- *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 analyzed;

- *Operational design,* which deals with the techniques by which the procedures specified in the sampling, statistical and observational designs can be carried out.

To assess the validity of a research project, one must know exactly how the study will be conducted. We should explain step-by-step how we intend to do the research. The best way of doing this is to work through our proposed study one section (or one sub-problem) at a time, describing the research tools (methods) we will deploy and indicating what data we will collect, how we will collect it and how it will be analyzed.

Qualifications of Researcher: In certain cases, it may be necessary to add a section indicating the qualifications of the researcher to do the research. For example, if the proposed study requires advanced analysis, the researcher needs to be competent to perform such analysis. If an empirical study depends on specialized psychometric testing, the researcher needs the skills and license to perform the testing.

This section should begin with the principal investigator. It is also customary to begin qualifications with the highest academic degree held. Experience in carrying out previous research is important, especially in the corporate marketplace, so a concise description of similar projects should be included.

Budgeting: The budget should be presented in the form the sponsor requests. For example, some organizations require secretarial assistance to be individually budgeted, whereas others insist it be included in the research director's fees or the overhead of the operation. In addition, limitations on travel, per diem rates, and capital equipment purchases can change the way in which you prepare a budget. Typically, the budget should be no more than one to two pages. Diagram below shows a format that can be used for small contract research projects. Additional information, backup details, quotes from vendors, and hourly time and payment calculations should be put into an appendix if required or kept in the researcher's file for future reference. The budget statement in an internal research proposal is based on employee and overhead costs. The budget presented by an external research organization is not just the wages or salaries of their employees but the person-hour price that the contracting firm charges.

The detail presented may vary depending on both the sponsors' requirements and the contracting research company's policy. (See Table 1 in the Appendix)

One reason why external research agencies avoid giving detailed budgets is the possibility that disclosures of their costing practices will make their calculations public knowledge, reducing their negotiating flexibility. Since budget statements embody a financial work strategy that could be used by the recipient of the bid to develop an independent work plan, vendors are often doubly careful. The budget section of an external agency's proposal states the total fee payable for the assignment. When it is accompanied by a proposed schedule of payment, this is frequently detailed in a purchase order. Unlike most product sale environments, research payments can be divided and paid at stages of completion. Sometimes a retainer is scheduled for the beginning of the contract, then a percentage at an intermediate stage, and the balance on completion of the project. It is extremely important that we retain all information we use to generate our budget. If we use quotes from external contractors, get the quotation in writing for our file. If we estimate time for interviews, we have to keep explicit notes on how we made the estimate. When the time comes to do the work, we should know exactly how much money is budgeted for each particular task. Some costs are more elusive than others. We should not forget to build the cost of proposal writing into our fee. Publication and delivery of final reports can be a last minute expense that can easily be overlooked in preliminary budgets.

Time Scheduling: **O**ur schedule should include the major phases of the project, their timetables, and the milestones that signify completion of a phase. For example, major phases may be

- exploratory interviews,
- final research proposal,
- questionnaire revision,
- field interviews,
- editing and coding,
- data analysis, and
- Report generation.

Each of these phases (named as A, B, C,….) should have an estimated time schedule and people assigned to the work. It may be helpful to us and our sponsor if we chart our schedule. We can use a *Gantt chart*. Alternatively, if the project is large and complex, a *critical path method (CPM)* of scheduling may be included. In a CPM chart the nodes represent major milestones, and the arrows suggest the work needed to get to the milestone.

More than one arrow pointing to a node indicates all those tasks must be completed before the milestone has been met. Usual1y a number is placed along the arrow showing the number of days or weeks or months required for that task to be completed. The pathway from start to end that takes the longest time to complete is called the *critical path*, because any delay in an activity along that path will delay the end of the entire project. An example of a CPM chart is shown in Figure 1 in the Appendix.

Nature and Form of Results: Upon finishing this section, the sponsor should be able to go back to the problem statement and research objectives and discover that each goal of the study has been covered. One should also specify the types of data to be obtained and the interpretations that will be made in the analysis. If the data are to be turned over to the sponsor for proprietary reasons, make sure this is reflected.

Alternatively, if the report will go to more than one sponsor, that should be noted. This section also contains the contractual statement telling the sponsor exactly what types of information will be received. Statistical conclusions, applied findings, recommendations, action plans, models, strategic plans, and so forth are examples of the forms of results.

Facilities and Special Resources: Often, projects will require special facilities or resources that should be described in detail. For example, a contract exploratory study may need specialized facilities for focus group sessions. Computer assisted telephone or other interviewing facilities may be required.

Project Management: The purpose of the project management section is to show the sponsor that the research team is organized in a way to do the project efficiently. A master plan is required for complex projects to show how the phases will all be brought together. The plan includes

- The research team's organization
- Management procedures and controls for executing the research plan.

Examples of management and technical reports

- Research team relationship with the sponsor
- Financial and legal responsibility
- Management competence

Tables and charts are most helpful in presenting the master plan. The relationships between researchers and assistants need to be shown when several researchers are part of the team. Sponsors must know that the director is an individual capable of leading the team and being a useful liaison to the sponsor. In addition, procedures for information processing, record control, and expense control are critical to large operations and should be shown as part of the management procedures. The type and frequency of progress reports should be

recorded so the sponsor can expect to be kept up-to-date and the researchers can expect to be left alone to do research. The sponsor's limits on control during the process should be delineated. Details such as printing facilities, clerical help, or information-processing capabilities that are to be provided by the sponsor are discussed. In addition, right's to the data, the results, and authority to speak for the researcher and for the sponsor are included. Payment frequency and timing are also covered in the master plan. Finally, proof of financial responsibility and overall management competence are provided.

Annotated bibliography: For all projects that require literature review, a bibliography is necessary with at least 20 good entries. The majority of the entries should be recent scholarly works. Avoid sources that are out-of-date and those classified as "*popular*" instead of "*academic*"; do not clutter the bibliography with irrelevant books (those not directly related to our topic) or with online articles. We must show that we know what the major works on the topic are. The annotations show that we are familiar with a source's content and understand its relevance to our study. We always use the bibliographic format required by the sponsor. If none is specified, a standard style manual (e.g., Kate L. Turabian, A Manual for Writers of Term Papers, Theses, and Dissertations; Joseph Gibaldi and Walter S. Achtert, MIA Handbook for Writers of Research Papers; or the Publication Manual of the American Psychological Association) will provide the details necessary to prepare the bibliography. Many of these sources also make suggestions for successful proposal writing.

Appendices:

(a) ***Glossary of terms -*** A glossary of terms should be included whenever there are many words unique to the research topic and not understood by in general community. This is a simple section consisting of terms and definitions. Also, we define any acronyms that we use, even if they are defined within the text.

(b) ***Measurement Instrument -*** For large projects, it is appropriate to include samples of the measurement instruments if they are available when we assemble the proposal. This allows the sponsor to discuss particular changes in one or more of the instruments. If exploratory work precedes the selection of the measurement instruments we will not use this appendix section. Other Any detail that reinforces the body of the proposal can be included in an appendix. This includes researcher vitae, budget details, and lengthy descriptions of special facilities or resources.

4. PRECAUTIONS WHILE WRITING THE RESEARCH PROPOSAL

We should take utmost care with the preparation of our research proposal. Based on the proposal, our funding agency(ies) or supervisor will make a decision as to whether we are capable of conducting serious research. Our proposal needs to make a positive impression.

Sloppiness in the presentation of our proposal sends the wrong message. Nobody will agree to provide fund to such kind of proposal.

What are the important errors to avoid when preparing a research proposal?

Here is a checklist of questions to consider:

- *Did we carefully check the spelling and grammar?* Even if we have to write the proposal in a second or third language, there is no excuse for the kinds of spelling, typing and grammar errors that can be corrected by using the spelling and grammar checking functions on our word processor. We should proof read our proposal several t imes before we submit it.
- *Does our proposal conform to the requirements of the institution/ funding agency?* Find out if the institution/ funding agency has set requirements for the following:
    (a) line spacing, font type and size, margins, etc.
    (b) the components of the proposal; and
    (c) referencing and bibliography. We have to make sure about our proposal conforms to all the institution's/ agency' s requirements.
- *Did we write the name of the qualification correctly?*
- *Is our language precise and modest?* We should be modest in our claims. We should not promise more than we can deliver. Support all claims with evidence. We should clearly state the objectives and their usefulness. It is our responsibility to be so precise that we leave no room for readers to misunderstand our proposal.

We have to pay attention to detail when we prepare a research proposal. We should not be careless or sloppy. The research proposal is the most important part of writing a thesis. It needs to convince the sponsoring / funding agency (a panel of professors) that we are capable of doing independent research.

5. EVALUATING THE RESEARCH PROPOSAL

Proposals are subjected to formal and informal reviews. The *formal method* has some variations, but its essence is described as follows. Before the proposal is received, criteria are established and each is given weights or points. The proposal is evaluated with a checklist of criteria in hand. Points are recorded for each category reflecting the sponsor's assessment of how well the proposal meets the category's established criteria. Several people, each of

whom is assigned to a particular section, typically review long and complex proposals. After the review, the category scores are added to provide a cumulative total. The proposal with the highest number of points will win the contract. The formal method is most likely to be used for competitive government, university, or public sector grants and also for large-scale contracts. Small-scale contracts are more prone to informal evaluation.

With *informal evaluation*, the project needs, and thus the criteria, are well understood but not usually well documented. In contrast to the formal method, a system of points is not used and the criteria are not ranked. The process is more qualitative and impressionistic. In practice, many items contribute to a proposal's acceptance and funding. Primarily, the content discussed above must be included to the level of detail required by the sponsor. Beyond the required modules, there are factors that can quickly eliminate a proposal from consideration and factors that improve the sponsor's reception of the proposal. First, the proposal must be neatly presented. Although a proposal produced on a word processor and bound with an expensive cover will not overcome design or analysis deficiencies, a poorly presented, unclear, or disorganized proposal will not get serious attention from the reviewing sponsors. Second, the proposal's major topics should be easily found and logically organized. The reviewer should be able to page through the proposal to any section of interest. The proposal also must meet specific guidelines set by1he sponsoring company or agency. These include budgetary restrictions and schedule deadlines. A fourth important aspect is the technical writing style of the proposal. The problem statement must be easily understood. The research design should be clearly outlined and the methodology explained. The importance/befits of the study must allow the sponsor to see why the research should be funded.

6. DATA PROCESSING

Once the data have been collected by using suitable methods and techniques, the researcher turns his focus of attention on its processing before their complete meanings and implications can be understood. At the same time emphasis has to be given on the quality or reliability of data. Because, a poorly designed data collection demolishes entire effort of research. In this Chapter, we will discuss about one of the most important stages of the research process, i.e. data processing and analysis. It leads us to derive logical conclusions based on data. Also it helps to ensure that data will be offered in an appropriate form, and in turn promotes the idea that the archiving and sharing of qualitative data are part of good research technique and practice.

Under the techniques of data processing, we consider the following activities:

(a) the establishment of response categories,

(b) editing,

(c) coding,

(d) preparation of master chart.

Analysis of data can be viewed as the ordering, the breaking down into constituent parts, and the manipulating of data to obtain answers to the research question or questions underlying the research project. The competent analysis of research-oriented data requires a blending of art and science, of intuition and informal insight, of judgment and statistical treatment, combined with a thorough knowledge of the context of the problem being investigated. Some of these qualities can only be acquired be experience, while others are heavily dependent on the native abilities of the analyst. Still others can be acquired through education and training (Green *et. al*., 1999).

A researcher has to design his plan of action and management for each and every stage of the research process. Therefore, a good researcher makes a perfect design for processing and analysis of data.

To some researchers data processing and analysis is not a very serious activity. They feel many times that data processing is a job of computer assistants. As a consequence, they have to be contended with the results given by computer assistants which may not help them to achieve their objectives. To avoid such situations, it is essential that data processing must be planned in advance and instructed to assistants accordingly.

6.1 THE ESTABLISHMENT OF RESPONSE CATEGORIES

When we have a large volume of data, it should be grouped into various categories or classes. It is most important to specify response categories at the earlier stage as it will simplify our analysis. We have to consider alternative responses in detail for clear interpretation. For this, the interviewers should be provided with more detailed instructions and guidance, which will ensure us consistent interpretations and thereby reduce complications in editing problems. It also reduces both the processing errors as well as cost.

Sometimes the response categories are established at the early stage for multiple choice questions set in the questions, but on the other hand it is done after the collection of data in the cases when we have free-answer, or open-end, questions, unstructured interviews, and projective techniques in the questionnaire. Here, we may get varieties of responses to

each question. For example, in a survey on 'The scope of primary education in rural area' we may ask

"What is your feeling towards the education of your child in a joint family?"

to a parent, we may get varieties of answers.

The selection of categories is controlled by both the purposes of the study and the nature of responses. We can classify the responses under the following conditions:

1. *Similarity of responses within categories.* The responses under each category should be sufficiently similar, that they can be treated as homogeneous depending upon the purpose of study.
2. *Differences of responses between categories.* There should be enough differences between different categories, so that it is sufficient to deduce important distinctions for the characteristics under study.
3. *Mutually exclusive categories.* Different categories of responses should be unambiguously defined of specified such that any particular response can be classified into one and only one category.
4. *Categories should be exhaustive.* The categories should be so formed that it should include all responses.

6.2 EDITING OF DATA

After collection of filled in questionnaires and specifying response categories, editing of entries becomes necessary and unavoidable for making subsequent steps simpler. Editing is the process of reviewing the data to ensure maximum accuracy and unambiguity. It ensures us about the consistency and uniformity in treatment of data. Many a times, a researcher or the assistants either miss entries in the questionnaires or enter responses, which are not legible. This sort of discrepancies can be resolved by editing the schedule meticulously. It is beneficial to start the process of editing at the early stage of data collection as it will point out misunderstandings of instructions, errors in recording, and other problems at a stage when it is still possible to eliminate them from the later stages of the study. Early editing has the additional advantage of permitting the questioning of interviewers while the material is still relatively fresh in their minds.

Each data collection form should be edited to ensure about the following requirements:

1. *Legibility of entries.* The questionnaires or the schedule should be filled legibly. When and kind of indistinctness is detected, it should be immediately clarified from the interviewer or it should be inferred from other answers. If it could not be clarified by any means, it i s better to avoid that form.
2. *Completeness of entries.* In a fully structured form, all the entries should be there. In case there is an absence of entry, it may occur due to the negligence of interviewer or due to the nonresponse of the respondent. So, an attempt must be made to identify the reason of missing.
3. *Consistencies of entries.* As the case with two watches that show different times, an entry inconsistent with another put us into trouble to decide which one is correct. For example, in a house hold survey, it was found that a family is reporting that they are not using cooking gas, but within their household expenditure, it was found that they have purchased a cooking gas in a month. These cases should be resolved by cross checking and re-interviewing.
4. *Accuracy of entries.* An editor should keep an eye on repetitive response patterns to detect any kind of inaccuracies in data. It mostly occurs due to interviewer bias or dishonesty.

6.3 CODING OF DATA

By coding we mean the transformation of raw data into quantitative or numerical data. It helps for analysis and interpretation systematic. So, coding of data involves assigning of numbers to each response of the question. The purpose of giving numbers is to translate raw data into numerical data, which may be counted and tabulated. The task of researcher is to give numbers to response carefully. The coding method depends upon the nature and type of question because, we get different responses according to type of question. For, example in a multiple choice type question (close-end question), the respondent chooses a response from the given set of responses but in an open-ended question, the respondent is free to answer. He/she answers according to his/her own choice. Therefore, a close-end question is already coded before the data collection (pre-coding) but coding of open-end questions is to be done after the editing of data (post-coding).

Coding is an important activity that requires a lot of expertise and experience. Improper coding will lead to a very poor analysis and vague result. So, the entire effort will go in vain. For example, in case of attitude scales, researcher has to keep in mind, the

direction or weightage of responses. For example, a response 'strongly agree' is coded as '5' , 'agree' is coded as '4', 'neither agree nor disagree' coded as '3', 'disagree' coded as '2' and 'strongly disagree' coded as '1'.

Again, the matrix questions have to be coded taking in to consideration each cell as one variable. For example, if the column of matrix represents gender, namely, 'male' and 'female' and row represents literary status namely 'illiterate' and 'literate'. Then the coding process will be as follows:

00 - Male and illiterate (variable name: *malilit*)
10 - Female and illiterate (variable name: *femilit*)
01 - Male and literate (variable name: *malit*)
11 - Female and literate (variable name: *femlit*)

The variable names as well as coding system should be done depending upon the software to be used to analyze the data, otherwise, it will not work or fit into the software.

The codes assigned to each response are to be recorded separately in a book known as '*Code Book'*. So, a code book is a list of codes assigned to each response for the questions. In order to demonstrate the points discussed above a section of the code book is reproduced in Table 2 in the Appendix.

Guidelines to keep in mind while coding

- ***Identification variables***. Provide fields at the beginning of each record to accommodate all identification variables. Identification variables often include a unique study number and a respondent number to represent each case.
- ***Code categories.*** Code categories should be mutually exclusive, exhaustive, and precisely defined. Each interview response should fit into one and only one category.
- ***Preserving original information*** . Code as much detail as possible. Recording original data, such as age and income, is more useful than collapsing or bracketing the information.
- ***Closed-ended questions***. Responses to survey questions that are pre-coded in the questionnaire should retain the coding scheme in the machine-readable data to avoid errors and confusion.
- ***Open-ended questions***. For open-ended items, investigators can either use a predetermined coding scheme or review the initial survey responses to construct a coding scheme based on major categories that emerge.

- ***Check-coding***. It is a good idea to verify or check-code some cases during the coding process — that is, repeat the process with an independent coder. For example, if more than one code is assigned to an interview response, this highlights problems or ambiguities in the coding scheme. Such check-coding provides an important means of quality control in the coding process.
- ***Series of responses***. If a series of responses requires more than one field, organizing the responses into meaningful major classifications is helpful. Responses within each major category are assigned the same first digit. Secondary digits can distinguish specific responses within the major categories. Such a coding scheme permits analysis of the data using broad groupings or more detailed categories. (S ee above for coding of matrix questions).

6.4 PREPARATION OF MASTER CHART

Once the code book is prepared, our next step is to prepare a master chart. a master chart is a tabular representation of all the information collected from all the respondents. A typical master chart is displayed as in Table 3 in the Appendix.

After preparing the master chart, we have to enter data into the computer. By doing so, we can check the wrong entries in the computer by comparing 'data listing' as a computer output and master chart. In the other way, after coding we can directly enter data into the computer also, but there is no scope for pointing out data entry error.

7. CLASSIFICATION AND TABULATION OF DATA

After the preparation of master chart, it is necessary to break this into different parts which are useful for analysis according to the objectives of survey. This is the most important part for starting the analysis of data. Here, we shall study the process of classification and tabulation of data.

7.1 CLASSIFICATION OF DATA

The data is to be classified according to the objectives and hypothesis of the study is performed with the help of frequency distributions. Reclassification is a process to rearrange responses with the help of statistical techniques, which helps researcher to justify the tabulation. Once, the responses are coded, these codes are summated and re-classified may be

as '*high*,' '*medium*' and '*low*'. For example, the income groups can be reclassified as '*high*,' '*medium*' and '*low*' income groups as follows:

Income Groups	Codes	Reclassification
Upto Rs. 15,000	1	Low
Rs. 15,000 to Rs.25,000 Rs. 25,000 to Rs.35,000	2 3	Medium
Rs. 35,000 or more	4	High

The basic principle in the process of classification or re-classification is that the categories thus obtained must be exhaustive and mutually exclusive. In other words, the categories have to be independent and not overlapping.

There are various types of classifications which are categorized into the following four types:

I. ***Qualitative Classification*** - When the classification is done according to some attributes or qualities (non-measurable characteristics), then it is called as qualitative classification. For example, classification according to gender, nationality, religion, occupation, etc.

II. ***Quantitative Classification*** – When the classification is done according to some variables or quantities (measurable characteristics), then it is called as quantitative classification. For example, classification according to age, monthly income, weight, height, etc.

III. ***Chronological Classification*** – When the classification of data is done in a chronological order, i.e., according to the time of occurrence, then it is called as a chronological classification. For example, numbers of students passing graduation with First Class from DDCE of Utkal University in different years, number of births in a village in different years, etc.

IV. ***Spatial Classification –*** When the classification of data is made according to different geographical areas like villages, Blocks, Tehsils, Districts, States or Countries or Regions, Climatic Zones, then it is called as a spatial classification.

7.2 TABULATION OF DATA

Tabulation is a process of presenting data in a compact form i.e., systematic presentation of data in rows and columns, in such a way that it facilitate comparisons and derive the type of relations present between variables. This also helps the researcher to perform statistical operation on the data to draw inferences. Tabulation can be generally in the form of uni-variate, bi-variate, tri or multi-variate tables. Accordingly, analysis proceeds in the form of uni-variate analysis, bivariate analysis and tri or multi-variate analysis.

A table consists of four parts as follows:

I. ***Title*** – The tit le of the table is self-explanatory describing the contents of the table. It is placed at the head of the table with a table number for ready reference.

II. ***Stub*** – The extreme left hand column containing the heading/ descriptions of the rows, with its own heading is known as the stub.

III. ***Captions*** – These are the heading/ descriptions of the columns and sub-columns other than the Stub in the upper part of the table.

IV. ***Body*** – It is the main part of the table except the Title, Stub & Captions. It contains the numerical information which are arranged in the table according to the description of rows and Columns given in stub and Captions.

At the end of the table, the *source* from which the contents of the table are derived should be specified. Consider the example given in Table 4 in the Appendix.

7.2.1 TYPES OF TABULATION

There are two types of tabulation:

a. Simple Tabulation
b. Complex Tabulation

a. Simple Tabulation

In these types of tabulation, the values of the items are placed just below the headings indicative of the characteristics. This type of tables is also called as One-way tables.

Table 4 given in the Appendix is an example of simple tabulation.

b. Complex Tabulation

In these types of tabulation, the values are representative of the characteristics shown both by the column and the row headings. These types of tables are also called as *two-way*

tables. Such types of tabulation are also called as ***Cross tabulation***. See Table 5 of the Appendix.

In the similar way, we can form three-way, four-way or manifold tables.

8. DIAGRAMMATIC PRESENTATION OF DATA

Textual presentation involving facts and figures do not catch our attention, as we always take less interest in going through these texts. So, an alternative way of representing facts and figures is the diagrams and pictures. Graphical representation of data is one of the most commonly used modes of presentation. Graphs can not only be made attractive, but they are also easy to comprehend and do not take much time to understand.

The graphical representation of data makes the reading more interesting, less time-consuming and easily understandable. The disadvantage of graphical presentation is that, it lacks details and is less accurate.

When creating graphic displays, we have to keep in mind the following points:

a. What I am going to represent?
b. Whom is to represent?
c. Is there any ambiguity in my presentation?
d. Am I able to explain all the things which are to be shown?

Keeping in view all these things, we have to choose an appropriate diagram or graph for the facts to be displayed. Therefore choosing a proper diagram is an important part of presenting the data. It attracts the audience as well as points on the focus of the output of survey.

8.1 TYPES OF GRAPHS, CHARTS AND DIAGRAMS

Several types of statistical/data presentation tools exist depending upon the type of data available. Some frequently used graphs, Charts and diagrams are discussed below.

8.1.1 Line Charts

It is the most widely used method of presenting statistical data, especially in business or any other sphere where data are collected over a long period of time. It is very much helpful in explaining the trend present in the series.

The graphical presentation of a time series is called as a ***Historigram*.**

Example – 1 : The data provided in Table 6 of the Appendix gives the import and export (in million tonnes) at Paradeep port during the period 200-01 to 2011-12. A line diagram representing this data is given in Figure 2 of the Appendix.

8.1.2 Bar Chart

Bar charts are used to graph the qualitative data. The bars do not touch, indicating that the attributes are qualitative categories, variables are discrete and not continuous. There are different types of bar charts. So, a particular type of bar chart to be used depends upon type of data and comparison needed. Different types of Bar Charts are

I. **Simple bar charts** sort data into simple categories.

Example – 2 : The data in Table 7 of the Appendix gives the Non-Odia Medium Schools in Odisha during the year 2012. A simple bar chart is drawn in Figure 3 of the Appendix.

Here, different categories are taken in the horizontal axis. But, it can be also taken in vertical axis also.

When we draw a simple bar diagram by taking the deviations from a particular value, it is called as a **deviation bar diagram**.

II. **Multiple bar charts** divide data into groups within each category and show comparisons between individual groups as well as between categories. (It gives more useful information than a simple total of all the components.).

Example – 3 : The data in Table 8 of the Appendix gives the growth of primary schools and primary school teachers in Odisha during 2000 – 01 to 2011 – 12 .

A Multiple bar chart is drawn in Figure 4 of the Appendix.

III. **Stacked bar charts**, which, like grouped bar charts, use grouped data within categories. (They make clear both the sum of the parts and each group's contribution to that total.)

Example – 4 : The data in Table 9 of the Appendix gives the indoor and outdoor patients treated in Odisha during 2006 to 2011.

A Divided or Stacked bar chart is drawn in Figure 5 of the Appendix.

When we express each of the components as the parts from 100, i.e. in terms of the percentages of the total, it is called as a ***percentage bar diagram***. For the data in Example – 4, we calculate the percentages of values in Table 10 and the percentage bar diagram is drawn in Figure 6.

8.1.3 Pie-Charts

Pie chart is often used in newspapers and magazines to depict budgets and other economic information. A complete circle (the pie) represents the total number of measurements. The size of a slice is proportional to the relative frequency of a particular category. For example, since a complete circle is equal to 360 degrees, if the frequency for a category is f and the total frequency is N, then the slice assigned to that category is $(\frac{f}{N} \times 360)$ degrees.

Example – 5: The data in Table 11 gives the number of un-electrified villages in some districts of Odisha as on 31st march, 2012. The Pie-Diagram is shown in Figure 7.

Bar and pie charts can be used in defining or choosing problems to work on, analyzing problems, verifying causes, or judging solutions. They make it easier to understand data because they present the data as a picture, highlighting the results. This is particularly helpful in presenting results to team members, managers, and other interested parties. Bar and pie charts present results that compare different groups. They can also be used with variable data that have been grouped. Bar charts work best when showing comparisons among categories, while pie charts are used for showing relative proportions of various items in making up the whole (how the "pie" is divided up).

8.1.4 Histogram

Histograms are used to graph absolute, relative, and cumulative frequencies. It is graphical representation of frequency distribution in the form of rectangles with class intervals as bases and the corresponding frequencies as heights, there being no gap between any two successive rectangles. A histogram is a diagram which represents the class interval and frequency in the form of a rectangle.

Example – 6: Draw the histogram for the following frequency distribution.

Variable:	10–20	20–30	30–40	40–50	50–60	60–70	70–80
Frequency:	12	30	35	65	45	25	18

The Histogram is shown in Figure 8.

8.1.5 Frequency Polygon

In a frequency distribution, the mid-value of each class is obtained. Then on the graph paper, the frequency is plotted against the corresponding mid-value. These points are joined by straight lines. These straight lines may be extended in both directions to meet the X - axis to form a polygon.

Also, we can draw a frequency polygon by joining the mid-points of the tops of the adjoining rectangles of a Histogram. The mid-points of the first and the last classes are joined

to the mid-points of the classes preceding and succeeding respectively at zero frequency to complete the polygon.

Example – 7: For the data given in Example – 6, a frequency polygon is shown in Figure 9.

8.1.6 Frequency Curve

In a frequency curve, we take the values or class-Intervals along the X- axis and the frequencies along the Y – axis. Then we plot the points and by joining these points by a smooth curve, we get a frequency curve.

Example – 8: For the data given in Example – 6, a frequency curve is shown in Figure 10.

8.1.7 Ogives or Cumulative Frequency Curves

Ogive is also used to graph cumulative frequency. An ogive is constructed by placing a point corresponding to the upper end of each class at a height equal to the cumulative frequency of the class. These points then are connected. An ogive also shows the relative cumulative frequency distribution on the right side axis. A less-than ogive shows how many items in the distribution have a value less than the upper limit of each class. A more-than ogive shows how many items in the distribution have a value greater than or equal to the lower limit of each class.

Example – 9: For the data given in Example – 6, both the cumulative frequency frequency curves are shown in Figure 11.

8.1.8 Pictogram

Rows of picture symbols of same size representing the comparative magnitudes of the data constitute a Pictogram. Pictograms are used to represent time series data and statistical data classified by attributes. A pictogram is shown in Figure 12.

9. REPORT WRITING

The research and its adequacy are examined on the basis of research proposal, research summary, research abstract and the research report. The contribution of the study is judged on the basis of research report. Thus the research report is the key aspect of the research and hence the researcher has to present the report.

A research report is a written document on a particular topic relating to a particular research problem, which conveys information and ideas and may also make recommendations. Reports often form the basis of crucial decision making. Inaccurate, incomplete and poorly written reports fail to achieve the objectives and purpose. A good report can be written by keeping the following features in mind:

1. All points in the report should be clear to the intended reader.
2. The report should be concise with information kept to a necessary minimum and arranged logically under various headings and sub-headings.
3. All information should be correct and supported by evidence.
4. All relevant material should be included in a complete report.

9.1 STRUCTURE OF A RESEARCH REPORT

The research report comprises of title page, preliminary pages, and the body followed by the references and the appendices. The abstract is typed on a separate page and there is no page breaks between sections in the body. The body comprises of the introduction, rationale of the study, limitations, review of related literature, the sample, methods, tools and techniques, analysis and interpretation, results, conclusion and suggestions. The research report is typed in double space on white $8\ ^{1}/_{2} \times 11$ inch paper with 12 pitch typeface. The margins are set to one inch on all sides. Every page has a header in the upper right corner with the running header right-justified on the top line and the page number right-justified and double-spaced on the line below it. The research paper disseminates the new contribution in the specific area. It avoids the repetition of the problem and saves time, energy and money.

Structure of a Research Report

TITLE PAGE	Full title of the report Name of the researcher Name of the unit of which the project is a part Name of the Institution/ Organization and Logo Date / Year
PRELIMINARY PAGES	*Background* (History (if any) behind project) *Foreword:* It is usually one page write up or a citation about the work by any eminent / popular personality or a specialist

	in the given field of study. *Preface:* It is again one/two pages write up by the author stating circumstances under which the present work is taken up, its importance, major dimensions examined and intended audience for the given work. The author gives his signature and address at the bottom of the page along with date and year of the work. *Acknowledgement:* Author thanks people and organization who helped during the project. *Executive Summary / Summary:* (sometimes called *abstract*) A condensed version of a report – outlines salient points, emphasizes main conclusions and (where appropriate) the main recommendations. *List of Contents:* An at- a – glance list that tells the reader what is in the report and what page number(s) to find it on. Headings and sub-headings used in the report should be given with their page numbers. *List of Tables:* An at- a – glance list that tells the reader what tables are in the report and what page number(s) to find it on. *List of Appendices:* An at- a – glance list that tells the reader what are the appendices in the report and what page number(s) to find it on.
BODY	*Introduction:* Author sets the scene and states his/ her intentions. This presents an overview of the whole report. *Aims and Objectives* *Review of Literature* *Methodology:* Work steps; what was done – how, by whom, when? *Results / Findings* *Analysis and Discussion*

	Limitations of Research *Conclusions* *Recommendations*
REFERENCES	A section of a report, which provides full details of publications mentioned in the text, or from which extracts have been quoted. We may use the pattern prescribed by American Psychological Association (APA style), or any other standard pattern recognized internationally. For example, Author, A. A. (Year). *Title of Book*. Location: Publisher**.** Author, A. A., & Author, B. B. (Year). *Title of book*. Location: Publisher.
APPENDIX ***(APPENDICES)***	We may not need an appendix, or may need several. It includes a blank copy of questionnaire, data or calculations, not given in the body, that are necessary, or useful, to get the full benefit from the report. There may be maps, drawings, photographs or plans.

9.2 TYPES OF REPORTS

Reports may be categorized broadly as *Technical Reports* and *General Reports* based on the nature of methods, terms of reference and the extent of in-depth enquiry made etc.

Two types of report formats are described below:

Technical Report - A Technical report mainly focuses on methods employed, assumptions made while conducting a study, detailed presentation of findings and drawing inferences and comparisons with earlier findings based on the type of data drawn from the empirical work.

General Reports - General reports often relate popular policy issues mostly related to social issues. These reports are generally simple, less technical, good use of tables and charts. Most often they reflect the journalistic style.

9.3 ESSENTIALS OF A GOOD REPORT

Good research report should satisfy some of the following basic characteristics:

I. **STYLE -** Reports should be easy to read and understand. The sentences used should be simple, clear and unambiguous.

II. **LAYOUT -** A good layout reduces the communication gap between the reader and researcher.

III. **ACCURACY –** A repost should be accurate. Every statement should be based on the facts and figures, otherwise, it may mislead the reader or the management of the concerned organization in the decision making process.

IV. **CLARITY – T**he technical terms used in the report should be clearly defined and explained.

V. **READABILITY –** There are several factors helpful to enhance the patience while reading a report. These factors include attractive appearance, clear and direct style, etc.

VI. **REVISION -** When first draft of the report is completed, it be checked for spelling and grammatical errors.

VII. **REINFORCEMENT -** Reinforcement usually gets the message across. This old adage is well known and is used to good effect in all sorts of circumstances e.g., presentations - not just report writing.

9.4 PRESENTATION OF A REPORT

There are various issues present for the presentation of a research report. These include:

1. Purpose of the report and focal issues
2. Stakeholders and their areas of interest
3. Mode and media of presentation
4. Extent of coverage and depth
5. Time, Place and cost associated

10 LIMITATIONS OF RESEARCH

Depending upon the research problem, the two sections on methods and limitations might be combined. These parts of the proposal describe the procedures in the investigation, as well as their limitations.

- What kinds of materials will be used?
- Are they readily available?
- Can we read written materials if they are not in English?
- Are translations available?

- What kinds of difficulties will there be in sampling or collecting physical evidence?
- What standards of certainty can be expected?
- Is our sample size adequate?
- Shall we need statistics? If so, which statistics?

These are the issues we should address and answer here.

Especially for empirical research, we might wish to treat these sections separately, describing in detail the proposed subjects, apparatus, and procedures. Our proposal will be judged not only on the basis of its hypothesis but also on the proposed methods of data collection. Shall we be using questionnaires? Observation? If an original measuring device is to be used, a draft should be included, accompanied by a statement of the intended sampling procedure and a plan for establishing its reliability and validity. Note that we shall also need a backup plan if the new measuring device does not prove to be a viable assessment. Once we have collected the data, how will we analyze it? We must explicitly show how the data collected will test the proposed hypothesis. We have to make sure about thinking out all the steps involved in our research before we begin.

Without exception, all research is limited in several ways. There are internal or formal limitations, such as the materials and procedures used, the ways in which critical terms are defined, the scope of the problem explored and of the applicability of the results. And there are external limitations as well, governed by constraints upon one's time or pocketbook; the inability to travel to special collections, museums, or libraries, or to speak or read other languages; or to consider an evolving political situation beyond a certain date. These limitations should be acknowledged; indeed, identifying them may help us to focus our topic. However, problems such as time and money difficulties do not relieve us of the responsibility of designing a study that can adequately test your hypothesis and measure its results. Proposals that include no mention of limitations suggest that the candidate has not really gone beyond a superficial consideration of the subject. This section of the proposal, therefore, will require considerable thought. But close attention now to these and related questions will save you much time and discomfort in later stages of research and writing.

11 AGENCIES INVOLVED IN SOCIAL RESEARCH

There are various Government, autonomous and Non-government organizations involved in social research. Some of these important organizations are listed as follows:

1. **Indian Council of Social Science Research (ICSSR)** was established in the year of 1969 by the Government of India to promote research in social sciences in the country. It supports a network of 27 ICSSR research institutes, including:
    - Institute for Social and Economic Change (ISEC), Bangalore
    - Institute of Public Enterprise, Hyderabad
    - Institute of Studies in Industrial Development, New Delhi
    - Govind Ballabh Pant Social Science Institute, Allahabad
    - Centre for Studies in Social Sciences, Calcutta
    - Centre for Policy Research (CPR), New Delhi
    - Centre for Multi-Disciplinary Development Research
    - Centre for Women's Development Studies (CWDS), Delhi
    - Madras Institute of Development Studies
2. **Centre for Urban Economic Studies (CUES)** was established as a research centre by the Calcutta University under the special assistanc e programme of University Grants Commission in the Department of Economics with urban economics as the thrust area. For the last quarter of a century, it has established itself as a premier multi-disciplinary research centre of urban studies in India. It has been engaged in different type of collaborative activities with other organisations like Indian Council of Social Science Research, Anthropological Survey of India.
3. **Sardar Patel Institute of Economic and Social Research** is a leading Social Science Research Institute in India. It is an autonomous organisation registered as a Trust and Society under the relevant statutes in 1965. With the valuable cooperation of the Gujarat Government, businessmen, and industrialists of Gujarat, the Institute started academic activities in 1969. Government of India recognised it as a National Institute in 1975 and, consequently, the Indian Council of Social Science Research (ICSSR), New Delhi, started sharing the maintenance grant on a 50:50 basis with the Gujarat government.
4. **Centre for Social Research** was founded in 1983 is a non-profit, non-governmental organisation based in New Delhi with a mission is to empower the women and girls of India, guarantee their fundamental rights, and increase understanding of social issues from a gender perspective. It operates on local, national and regional levels in an effort to enhance the capacities of

individuals, communities and institutions for creating a humane, equitable and gender-just society.

Institute of Social Change and Development: The Omeo Kumar Das Institute of Social Change and Development (OKDISCD), originally called the Institute for Social Change and Development, came into being on March 30, 1989 in Guwahati under the joint initiatives of the Government of Assam and the Indian Council of Social Science Research (ICSSR), New Delhi. It undertakes research on problems and processes of social transformation and development of Assam and other States of North-East India.

12 ETHICAL CONSIDERATIONS

The term '*ethics'* derives from the Greek word '*ethos'*, meaning '*character*.' Research ethics provides guidelines for the responsible conduct of social research. In addition, research ethics educates and monitors researcher conducting research to ensure a high ethical standard. To consider the ethical dimension of our research, we have to ask our self some important points:

- Moral principles guide our research,
- Ethical issues influencing our selection of a research problem,
- Ethical issues affecting research design,
- Responsibility towards research subjects,
- Ethical issues/dilemmas behind the publication of research findings,
- Benefit to participants in the study.

Research ethics are a set of principles about how researchers and research organizations should conduct themselves when dealing with research participants, other researchers and colleagues, the users of their research and society in general. Particularly relevant to the social sciences are ethics associated with projects involving human participants, including conducting surveys, focus groups and the use of secondary data. Typical considerations include:

- Recruiting study participants and informed consent,
- Keeping data secure and confidential,
- Making procedures, methods and findings transparent so that they can be assessed,

- Safety and risk.

After all, we have to obey guidelines and codes of conduct relevant to the research being conducted.

12.1 IMPORTANCE OF ETHICS IN RESEARCH

Knowing what constitutes ethical research is important for all people who conduct research projects or use and apply the results from research findings. All researchers should be familiar with the basic ethical principles and have up-to-date knowledge about policies and procedures designed to ensure the safety of research subjects and to prevent sloppy or irresponsible research, because ignorance of policies designed to protect research subjects is not considered a viable excuse for ethically questionable projects. Therefore, the duty lies with the researcher to seek out and fully understand the policies and theories designed to guarantee upstanding research practices. Research is a public trust that must be ethically conducted, trustworthy, and socially responsible if the results are to be valuable. All parts of a research project – from the project design to submission of the results for peer review – have to be upstanding in order to be considered ethical. When even one part of a research project is questionable or conducted unethically, the integrity of the entire project is called into question.

In the following, we note down some ethical standards to be considered while conducting a research:

- In order to respect and cause no harm to the participants,
- As a sign of respect for other researchers and those who will use the research,
- It is a professional requirement particularly in some disciplines and failure to do so may result in disciplinary procedures,
- It is a requirement to obtain funding,
- Failing to conduct research ethically could be embarrassing or result in research (or the researcher) being dismissed or rejected by the research community.
- Research involving human beings, including using questionnaires and focus groups, must be passed by an Ethics Committee whose job it is to confirm that the research conforms to a set of ethical guidelines.

If ethics are considered, this should make sure that the work is acceptable to the research community and other users of the research results.

12.2 ISSUES IN SOCIAL RESEARCH

The following are some of the important issues involved in a social research:

1. ***Authorship***: - It is the process of deciding whose names belong on a research paper. In many cases, research evolves from collaboration and assistance between experts and colleagues. Some of this assistance will require acknowledgement and some will require joint authorship.
2. ***Plagiarism***: - It is the act of passing off somebody else's ideas, thoughts, pictures, theories, words, or stories as your own. If a researcher plagiarizes the work of others, they are bringing into question the integrity, ethics, and trustworthiness of the sum total of his or her research. In addition, plagiarism is both an illegal act and punishable, considered to be on the same level as stealing from the author that which he or she originally created.
3. ***Peer Review***: - It is the process in which an author (or authors) submits a written manuscript or article to a journal for publication and the journal editor distributes the article to experts working in the same, or similar, scientific discipline. The experts, otherwise called the reviewers, and the editor then enter the peer review process. The process involves the following:
    a. Reviewers and editors read and evaluate the article
    b. Reviewers submit their reviews back to the journal editor
    c. The journal editor takes all comments, including their own, and communicates this feedback to the original author (or authors).

    The peer review process seldom proceeds in a straight line. The entire process may involve several rounds of communication between the editor, the reviewers, and the original author (or authors) before an article is fully ready for publication.

4. ***Conflicts of Interest: -*** It arise when a person's (or an organization's) obligations to a particular research project conflict with their personal interests or obligations. For example, a university researcher who involved in violence against women in his/ her family, is obligated to report truthful and accurate data, but he might be conflicted if faced with data that would hurt his/ her family status. Conflicts of interest are particularly important to examine within the context of social research because research subjects may be particularly vulnerable to harm.

A researcher should attempt to identify potential conflicts of interest in order to confront those issues before they have a chance to do harm or damage. If conflicts of interest do exist, then the objectivity of the researcher and the integrity of the research results can be questioned by any person throughout the research review phase. It is therefore imperative to address conflicts of interest up front and discuss how to combat potential lack of objectivity, before the research is called into question.

5. ***Data management:*** - In respect to research ethics, references three issues: 1) the ethical and truthful collection of reliable data; 2) the ownership and responsibility of collected data; and, 3) retaining data and sharing access to collected data with colleagues and the public. Each issue contributes to the integrity of research and can be easily overlooked by researchers. Oftentimes, researchers will downplay the importance of data management because the details can be time consuming and they assume they can "figure it out" as they go along. It is not adequate research practice to assume issues involved in data collection will work themselves out on their own. Instead, a clear, responsible, ethically sound, and carefully outlined plan for data management is required at the beginning of research to prevent all manners of conflicts and inappropriate research methods.

   Ethical data collection refers to collecting data in a way that does not harm or injure someone. Harm and injury could range from outright physical injury to harmful disclosure of unprotected confidential health information. In comparison, truthful data collection refers to data that, once collected, are not manipulated or altered in any way that might impact or falsely influence results.

   Assigning and ensuring responsibility for collecting and maintaining data is one of the most important ethical considerations when conducting a research project.

6. ***Research misconduct*** is the process of identifying and reporting unethical or unsound research. The United States' Office of Scientific and Technology Policy (OSTP) defines misconduct, and its components, as follows:

   Research misconduct is defined as fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results.

   - Fabrication is making up data or results and recording or reporting them.

- Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
- Plagiarism is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit.
- Research misconduct does not include honest error or differences of opinion.

7. ***Research with Human Subjects:*** The issues concerning research with human subjects involve topics ranging from voluntary participation in research to fair selection and justice. This variety makes the topics surrounding research ethics with human subjects a challenging but important charge.

   Respect for Persons – Informed Consent. The Belmont Report of 1979 outlines the three requirements for informed consent. The first requirement is that information disclosed to research participants must include, "research procedure, their purposes, risks and anticipated benefits, alternative procedures (where therapy is involved), and a statement offering the subject the opportunity to ask questions and to withdraw at any time from the research." The second requirement for informed consent is comprehension. The concept of comprehension requires researchers to adapt information to be understandable to every participant. This requires taking into consideration different abilities, intelligence levels, maturity, and language needs. Finally, the third requirement for informed consent is voluntariness. Informed consent can be neither coerced nor improperly pressured from any participant.

   Respect for Persons – Privacy and confidentiality. Privacy and confidentiality are very important components for research involving human subjects. People have a right to protect themselves, and information gathered during research participation could harm a person by violating their right to keep information about themselves private. The information gathered from people in research studies has a unique potential to be particularly embarrassing, harmful, or damaging.

   Risk benefit and beneficence. Beneficence is a principle used frequently in research ethics. It means, "doing good." Social research strives to do good by studying and understanding the society and its relative changes to uncover information that may be used to help others through designing policies for development.

Justice. Particular interest has been paid to prevent the overburdening of some populations in order to apply research findings to other groups. Populations under consideration with particular potential for exploitation may include minority groups, women, mentally impaired individuals, children, financially disadvantaged individuals, disadvantaged people living in third world countries, prisoners, the deceased, employees, ***etc.***

12.3 ETHICAL GUIDELINES

The following is a rough and general summary of some ethical principles that various codes address:

Honesty

Strive for honesty in all scientific communications. Honestly report data, results, methods and procedures, and publication status. Do not fabricate, falsify, or misrepresent data. Do not deceive colleagues, granting agencies, or the public.

Objectivity

Strive to avoid bias in experimental design, data analysis, data interpretation, peer review, personnel decisions, grant writing, expert testimony, and other aspects of research where objectivity is expected or required. Avoid or minimize bias or self-deception. Disclose personal or financial interests that may affect research.

Integrity

Keep your promises and agreements; act with sincerity; strive for consistency of thought and action.

Carefulness

Avoid careless errors and negligence; carefully and critically examine your own work and the work of your peers. Keep good records of research activities, such as data collection, research design, and correspondence with agencies or journals.

Openness

Share data, results, ideas, tools, resources. Be open to criticism and new ideas.

Respect for Intellectual Property

Honor patents, copyrights, and other forms of intellectual property. Do not use unpublished data, methods, or results without permission. Give credit where credit is due. Give proper acknowledgement or credit for all contributions to research. Never plagiarize.

Confidentiality

Protect confidential communications, such as papers or grants submitted for publication, personnel records, trade or military secrets, and patient records.

Responsible Publication

Publish in order to advance research and scholarship, not to advance just your own career. Avoid wasteful and duplicative publication.

Responsible Mentoring

Help to educate, mentor, and advise students. Promote their welfare and allow them to make their own decisions.

Respect for colleagues

Respect your colleagues and treat them fairly.

Social Responsibility

Strive to promote social good and prevent or mitigate social harms through research, public education, and advocacy.

Non-Discrimination

Avoid discrimination against colleagues or students on the basis of sex, race, ethnicity, or other factors that are not related to their scientific competence and integrity.

Competence

Maintain and improve your own professional competence and expertise through lifelong education and learning; take steps to promote competence in science as a whole.

Legality

Know and obey relevant laws and institutional and governmental policies.

Animal Care

Show proper respect and care for animals when using them in research. Do not conduct unnecessary or poorly designed animal experiments.

Human Subjects Protection

When conducting research on human subjects, minimize harms and risks and maximize benefits; respect human dignity, privacy, and autonomy; take special precautions with vulnerable populations; and strive to distribute the benefits and burdens of research fairly.

13 SUMMARY

A research proposal "*is a document that outlines how you propose to undertake your research studies*". It research proposal presents a problem, discusses related research efforts, outlines the data needed and shows the research design. A research proposal governs and directs a research project. This is the most difficult and most important part of many research endeavors. In the long run, the time and energy invested in conceptualizing the project pays dividends.

Research proposals have two main parts—the research problem and the research plan, the what and the how. First it describes a problem; then it presents a plan to solve it.

A research design includes the tasks such as sample selection and size, data collection method, instrumentation, procedures, and ethical requirements. So, we may split the overall research design into four parts as *Sampling design, Observational design, Statistical design,* and *Operational design.*

Specifying the Budget for different expenditures, man power design and time scheduling are three most important aspects to be considered while writing a research proposal.

We should use the bibliographic format required by the sponsor. If none is specified, a standard style manual (e.g., Kate L. Turabian, A Manual for Writers of Term Papers, Theses, and Dissertations; Joseph Gibaldi and Walter S. Achtert, MIA Handbook for Writers of Research Papers; or the Publication Manual of the *American Psychological Association*) will provide the details necessary to prepare the bibliography.

We have to take utmost care while preparing a research proposal and should take sufficient steps to avoid errors in the research proposal because, it forms the basis on which different funding agencies will decide whether we are capable of doing independent research or not.

Proposals are subjected to *formal and informal* reviews.

Once the data have been collected, the researcher turns his focus of attention on data processing and analysis. It leads us to derive logical conclusions based on data. At the same time emphasis has to be given on the quality or reliability of data. Because, a poorly designed data collection demolishes entire effort of research.

Under the techniques of data processing, we consider the following activities:

(a) the establishment of response categories,
(b) editing,
(c) coding,
(d) preparation of master chart.

A researcher has to design his plan of action and management for each and every stage of the research process. Therefore, a good researcher makes a perfect design for processing and analysis of data.

By coding we mean the transformation of raw data into quantitative or numerical data. It helps for analysis and interpretation systematic. The codes assigned to each response are to be recorded separately in a book known as '*Code Book'*.

After coding is over, the data is to be classified according to the objectives and hypothesis of the study is performed with the help of frequency distributions. There are various types of classifications which are categorized into the four types: *Qualitative Classification, Quantitative Classification, Chronological Classification and Spatial Classification.*

After classification is over our next step is the tabulation of data. Tabulation is a process of presenting data in a compact form i.e., systematic presentation of data in rows and columns, in such a way that it facilitate comparisons and derive the type of relations present between variables. A table consists of four parts as *Title, Stub, Captions and Body.*

After tabulation of data, we have to consider it presentation. Textual presentation involving facts and figures do not catch our attention, as we always take less interest in going through these texts. So, an alternative way of representing facts and figures is the diagrams and pictures. Graphical representation of data is one of the most commonly used modes of presentation. Some common graphical presentations are line diagram, bar diagram, Histogram, Frequency polygons and frequency curves, ogives and pie-diagram.

A research report is a written document on a particular topic relating to a particular research problem, which conveys information and ideas and may also make recommendations. The research report comprises of title page, preliminary pages, and the body followed by the references and the appendices. Reports may be categorized broadly as *Technical Reports* and *General Reports* based on the nature of methods, terms of reference and the extent of in-depth enquiry made etc.

Research ethics are a set of principles about how researchers and research organizations should conduct themselves when dealing with research participants, other researchers and colleagues, the users of their research and society in general.

There are some ethical principles that various codes address. These includes ***Honesty,*** Objectivity, Integrity, Carefulness, Openness, Respect for Intellectual Property, Confidentiality, Responsible Publication, Responsible Mentoring, Respect for colleagues, Social Responsibility, Non-Discrimination, Competence, Legality, Animal Care and Human Subjects Protection.

There are various Government, autonomous and Non-government organizations involved in social research. Some of these important organizations are Indian Council of Social Science Research (ICSSR) , Institute for Social and Economic Change (ISEC), Institute of Public Enterprise, Institute of Studies in Industrial Development, Govind Ballabh Pant Social Science Institute, Centre for Studies in Social Sciences, Centre for Policy Research (CPR), Centre for Multi-Disciplinary Development Research, Centre for Women's Development Studies (CWDS), Madras Institute of Development Studies, Centre for Urban Economic Studies (CUES), Sardar Patel Institute of Economic and Social Research, Centre for Social Research and Institute of Social Change and Development.

14 KEY WORDS

Research Report, Research Proposal, Research Process and Research Design, Sampling Design, Budgeting, Data Preparation, Coding, Classification and Tabulation of Data, Diagrammatic Representation of Data, Ethics in Research.

15 SELF-ASSESSMENT QUESTIONS

1. What do you mean by a research proposal? Discuss different parts of a research proposal.
2. How can you plan for conducting a research? Discuss different steps of planning.
3. Explain the concepts of cost and human resource budgeting and time scheduling in a research proposal. Give suitable examples.
4. Discuss the precautions to be taken while writing a research proposal.
5. Discuss, how can you evaluate a research proposal?
6. What do you mean by data processing? Outline differe nt techniques involved in data processing.
7. What do you mean by Coding of data? What is its importance in research? Discuss the techniques of coding the data.
8. What do you mean by classification and tabulation of data? Discuss different types of classification. Outline the parts of a table and different types of tables with suitable examples.
9. What do you mean by pictorial representation of data? Why it is necessary? Discuss different types of pictorial representation of data.
10. What do you mean by a research report? Discuss the structure of a research report.
11. What is a research report? Discuss different types of research report. What are the essentials of a good report?
12. What do you mean by research ethics? Discuss different issues in research in ethical view.
13. Discuss the ethical guidelines for a researcher.
14. Discuss about different agencies engaged in social research.

16 FURTHER READINGS

Creswell, J. W. (2006). Qualitative inquiry and research design: Choosing among five traditions (2nd ed.). Thousand Oaks, CA: Sage Publications.

Dash, P.R. (2011) Research Methodology with SPSS, (1st ed.)Vrinda Publications (P) Ltd., New Delhi.

Leedy, P.D. (1993). Practical Research: Planning and Design (5th ed.). New York: MacMillan.

Mouton, J.(2001). How to succeed in your master's and doctoral studies. Pretoria, South Africa: Van Schaik.

Rubin, A., Babbie, E., & Lee, P.A. (2008). Research methods for social work: Custom edition prepared exclusively for San Jose State University. Belmont, CA: Wadsworth/Thomson Learning.
Paul D. Leedy, Practical Research: Planning & Design, Prentice Hall
Bell, J. (1993), Doing your Research Project, Oxford University Press.
Howell, DC. (1997). Psychology for statistics. Duxbury Press.
Alvesson, M. and Sköldberg, K. (2000) Reflexive Methodology: New Vistas for Qualitative Research. London: Sage.
Reinharz, S. (1992) Feminist Methods in Social Research. Oxford: Oxford University Press.
Sayer, A. (2000) Realism and Social Science. London: Sage.
Williams, M. and May, T. (1996) Introduction to the Philosophy of Social Research. London: UCL Press.
Buchanan, E., & Ess. C. (2009). Internet research ethics and the Institutional Review Board: Current practices and issues. Computers and Society, 39 (3) 43-49.
Karpf, D. (2012). Social science research methods in internet time. Information, Communication,
& Society, 15(5), 639-661.
Markham, A. (2012). Fabrication as ethical practice: Qualitative inquiry in ambiguous internet
contexts. Information, Communication, and Society, 5(3), 334-353.
Zimmer, M. (2010). But the data is already public: On the ethics of research in Facebook. Ethics & Information Technology, 12(4), 313-325.
Singer, Eleanor. 1978a. Consequences for response rate and response quality in social surveys. American Sociological Review 43:144-62.

APPENDIX 1: STATISTICAL TABLES AND FIGURES

Table 1: Budget Example: Research Program Budget

Budget Items	**Charge**	**Rate**	**Total Days**
A. Salaries			
1. Research director	Rs. 9,000	Rs 300/hr	30 hours
2. Associate	Rs 3,000	Rs 200/hr	15 hours
3. Research assistants (2)	Rs 12,000	Rs 30/hr	400 hours
4. Secretarial (1)	Rs 4,000	Rs 20/hr	200 hours
Sub Total			**Rs 28,000**
B. Other costs			
5. Employee services and benefits			
6. Travel	Rs 10,000		
7. Office supplies	Rs 2,000		
8. Telephone	Rs 1,000		
9. Rent			
10. Other equipment			
11. Publication and storage costs	Rs 10,000		
Sub Total	**Rs 23,000**		

C. Total of direct costs	Rs 20,000
D. Overhead support	Rs 10,000
E. Total funding requested	**Rs 81,000**

Table 2: Code Book

Qestion Number	Variable Number	Information Desired	Variable Name	Response	Code	Column Number	Remark
		Respondent Number					
1		Name					
2	1	Age	age	Actual		1	
3	2	Gender	gender	Male Female	0 1	2	
4	3	Marital Status	marsta	Married Unmarried Widow Divorced	1 2 3 4	3	
4	3	Educational Qualification	ednql	Primary Secondary Higher Secondary Graduate Post- graduate Professional	1 2 3 4 5 6	3	
5	4	Employment Sector	emplcat	Unemployed Self Employed Private Government	1 2 3 4	4	
6	5	Employment level	emplev	Managerial / Officer Supervisor / Clerical Worker/ Peon	1 2 3	5	
7	6	Salary / Wages per Month	salary	Upto Rs. 15,000 Rs. 15,000 to Rs.25,000 Rs. 25,000 to Rs.35,000 Rs. 35,000 or more	1 2 3 4	6	
8	7	Years of Experience	exper	Actual		7	

		in the current Job					
9	8	Bonus given to you	bonus	Yes No	1 0	8	
10	9	Have promotion	prom	Yes No	1 0	9	

Table 3: MASTER CHART

Respondent Number	Age	Gender	Marital status	Educational Qualification	Employment Sector	Employment level	Salary / Wages per Month	Years of Experience in the current Job	Bonus given to you	Have promotion
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	41	0	1	5	2	1	3	15	0	0
2	52	1	1	4	3	3	2	5	1	1
3	37	0	1	4	2	1	4	2	0	0
4	45	0	1	2	2	1	3	3	0	0
5	58	0	1	4	3	1	4	5	1	1
6	27	1	2	3	3	2	3	1	1	1
7	35	1	2	5	3	2	2	1	1	1
8	45	0	1	5	4	3	4	3	1	1
9	50	1	3	2	3	2	3	2	1	1
10	29	1	2	3	4	2	4	1	1	1
11	36	0	2	4	3	2	3	3	1	1
12	39	0	1	3	3	3	2	2	0	1
13	48	0	1	6	4	3	2	4	0	1
14	58	1	1	5	4	2	3	5	1	1
15	60	1	4	6	4	3	2	9	1	1

Table – 4: Gross State Domestic Product (GSDP) at Current from 2005-06 to 2012-13

YEAR	GSDP (Rs. In lakh)
2005-06	8509649
2006-07	10183947
2007-08	12927445
2008-09	14849071
2009-10	16294643
2010-11 (P)*	19446479
2011-12 (Q)*	21589944
2012-13 (A)*	25874409

Source: *Economic Survey, Odisha: 2012-13, pp. 29-30.*

(P = Provicinal Estimate, Q = Quick Estimate, A = Advance Estimate)

*Subject to Revision

Table 5: Percentage Distribution of Debt at all-India level among Indebted Rural Labour Households by Purpose (2004-05)

Purpose of debt	Households		
	With Cultivated Land	Without Cultivated Land	All Households
Household Consumption	23.5	34.5	29.4
Marriages & other Ceremonial Expenses	22.1	21.1	21.6
Purchase of land and construction of building	16.6	19.9	18.4
Productive purposes	27.2	17.5	22.0
Repayment of debt	3.1	2.5	2.8
Other purposes	7.5	4.5	5.8
Total	**100.0**	**100.0**	**100.0**

Source: Rural Labour Enquiry Report on Indebtedness among Rural Labour Households (Round Of N.S.S.) 2004-05, Government of India, Ministry of Labour and Employment, Labour Bureau, Shimla/Chandigarh, 2010, pp. 38

Table – 6: Traffic through Paradeep Port during 2000-2001 to 2011-2012.

Year	Import (in million tonnes)	Export (in million tonnes)
2000 – 01	6.85	13.05

2001 – 02	6.66	14.47
2002 – 03	6.99	16.91
2003 – 04	6.71	18.60
2004 – 05	8.44	21.66
2005 – 06	11.42	21.69
2006 – 07	13.66	24.86
2007 – 08	16.85	25.59
2008 – 09	19.89	26.52
2009 – 10	29.18	27.83
2010 – 11	31.22	24.81
2011 – 12	35.99	18.26

Source: Economic Survey, Odisha, 2012 – 13.

Table - 7: Non Odia Medium Schools, 2012

School Type	**Number**
CBSE (English medium)	187
ICSE (English medium)	85
Sanskrit Tolls	314
Navodaya Vidyalaya	30
Madrasa	197

Source: Economic Survey, Odisha, 2012 – 13, p. 281

Table – 8: The Growth of Primary Schools and teachers in Odisha during 2000 – 01 to 2011 – 12 .

Year	**No. of Primary Schools (in thousands)**	**No. of Primary School Teachers (in thousands)**
2000 – 01	42	115
2005 – 06	46	115
2006 – 07	47	114
2007 – 08	48	124
2008 – 09	50	125
2009 – 10	53	138
2010 – 11	54	136
2011 – 12	55	133

Source: Economic Survey, Odisha, 2012 – 13

Table – 9: Indoor and Outdoor patients treated in Odisha during 2006 to 2011 .

Year	No. of indoor patients (in thousands)	No. of outdoor patients (in thousands)	Total patients treated
2006	183.9	3238.8	
2007	68.5	1296.9	
2008	121.1	1712.8	
2009	180.2	2109.4	
2010	48.9	2056.0	
2011	429.6	1322.6	

Source: Economic Survey, Odisha, 2012 – 13, p. 288

Table 10: Percentage of Indoor and Outdoor Patients from 2006 to 2011

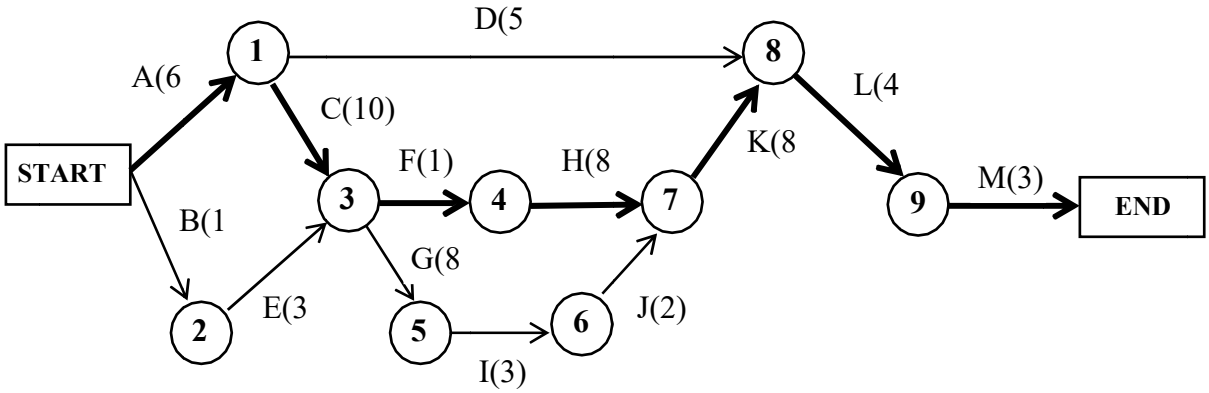
Year	Percentage of indoor patients (in thousands)	Percentage of outdoor patients (in thousands)
2006	183.9	3238.8
2007	68.5	1296.9
2008	121.1	1712.8
2009	180.2	2109.4
2010	48.9	2056.0
2011	429.6	1322.6

Table 11: Number of Un-electrified villages in Some Districts of Odisha as on 31st march, 2012

Name of the Districts	Number of un-electrified villages
Rayagada	1135
Koraput	1021
Kandhamal	872
Kalahandi	736
Malkangiri	735

Source: Economic Survey, Odisha, 2012 – 13, p. 254

Figure 1: CPM Chart showing the Entire Research Project



Critical Path: START → A → C → F → H → K → L → M : **END**
Time to Complete: $6 + 10 + 1 + 8 + 8 + 4 + 3 = 40\ Months$

Figure 2: Line Diagram showing the import and export at Paradeep Port.

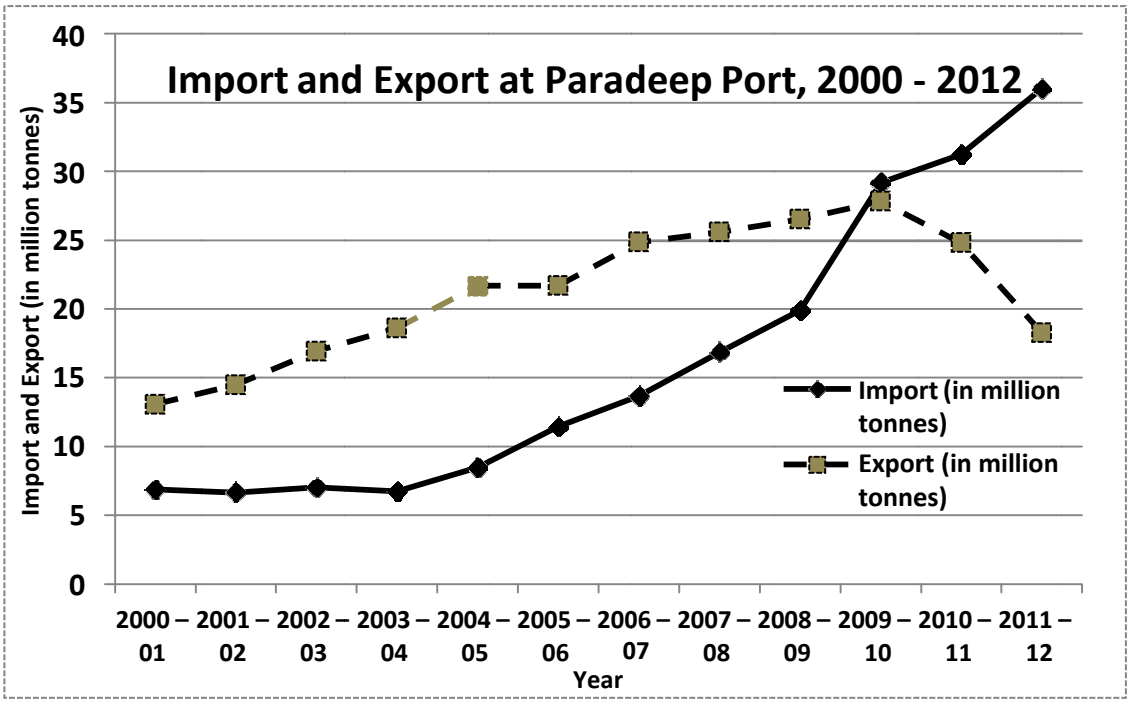


Figure 3: Simple Bar Chart showing the Non-Odia Medium Schools in Odisha during 2012.

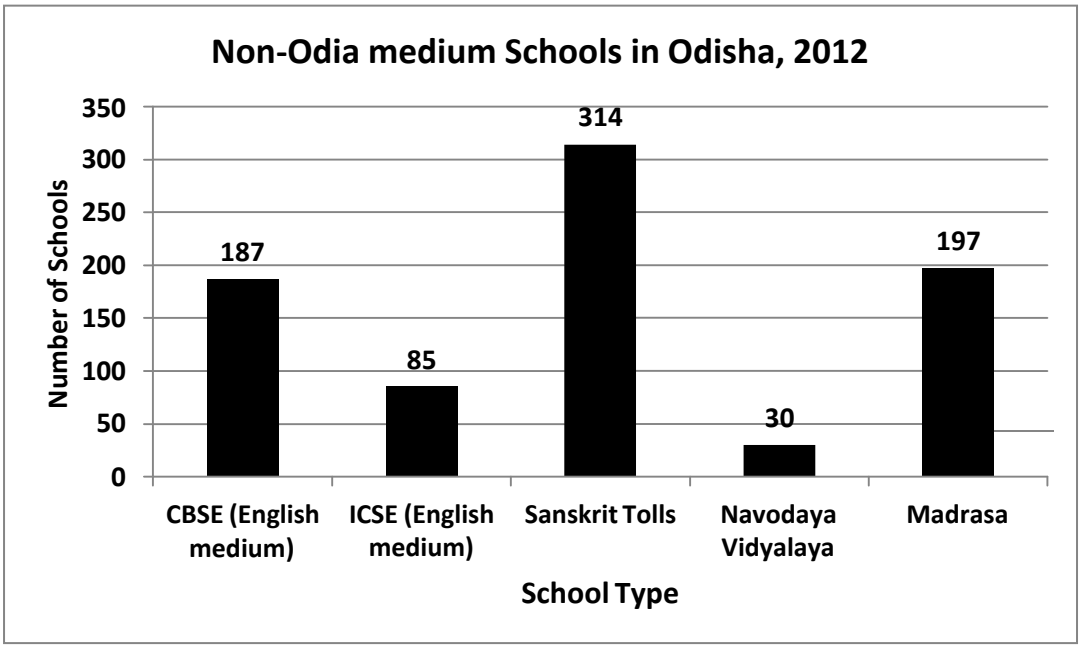


Figure 4: Multiple Bar Diagram Showing the Number of Primary Schools and School Teachers

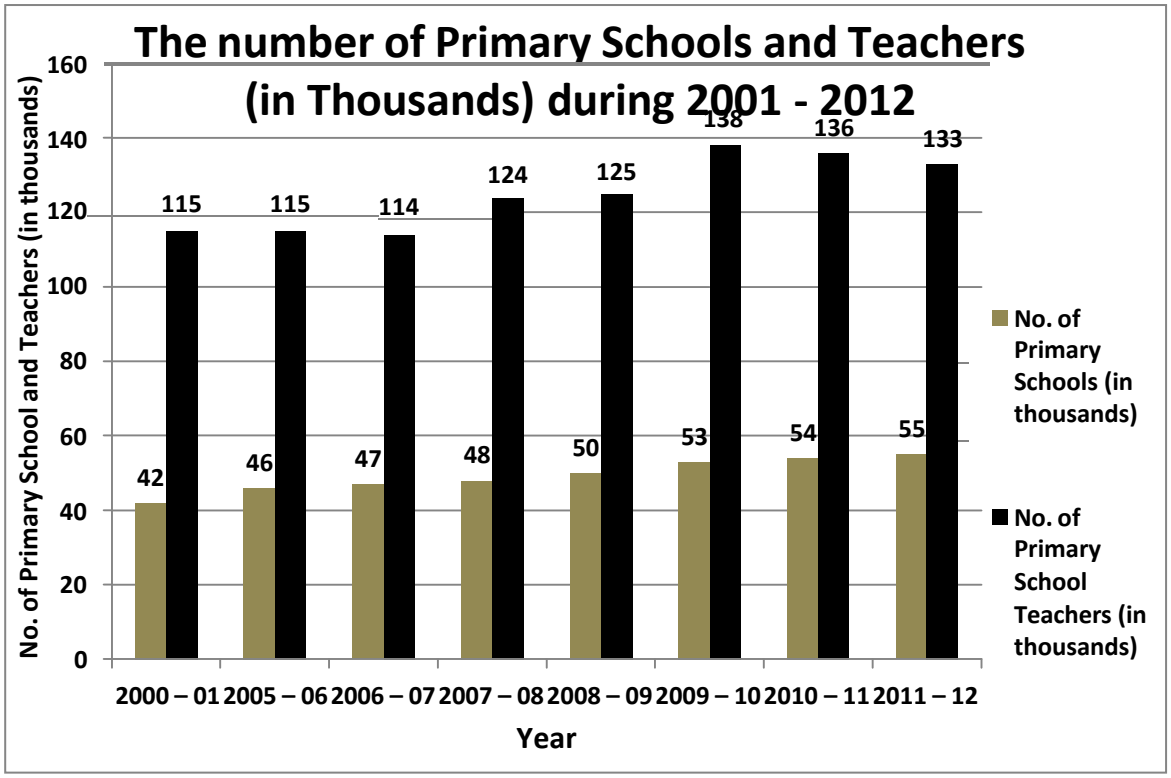


Figure 5: A Divided bar Diagram Showing Number of Indoor and Outdoor Patients Treated (in thousand) in Odisha during 2006 to 2011.

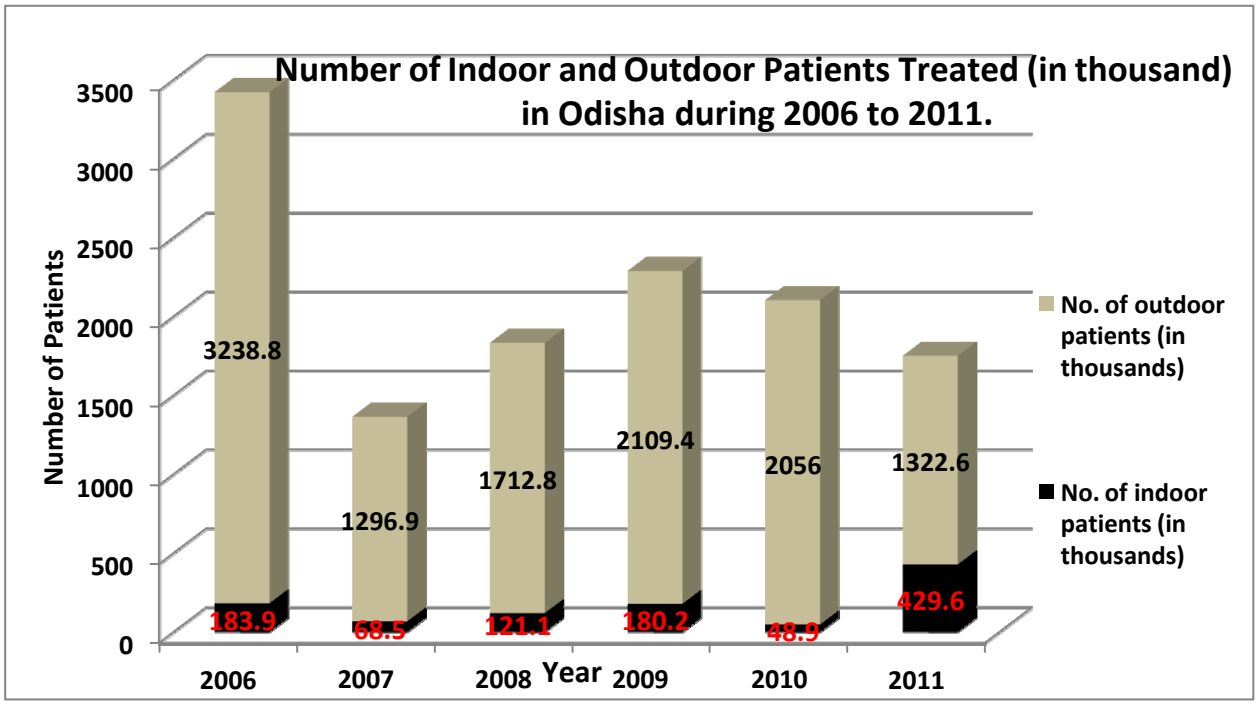


Figure 6: Percentage Bar Diagram Showing the Indoor and Outdoor Patients during 2006 to 2011.

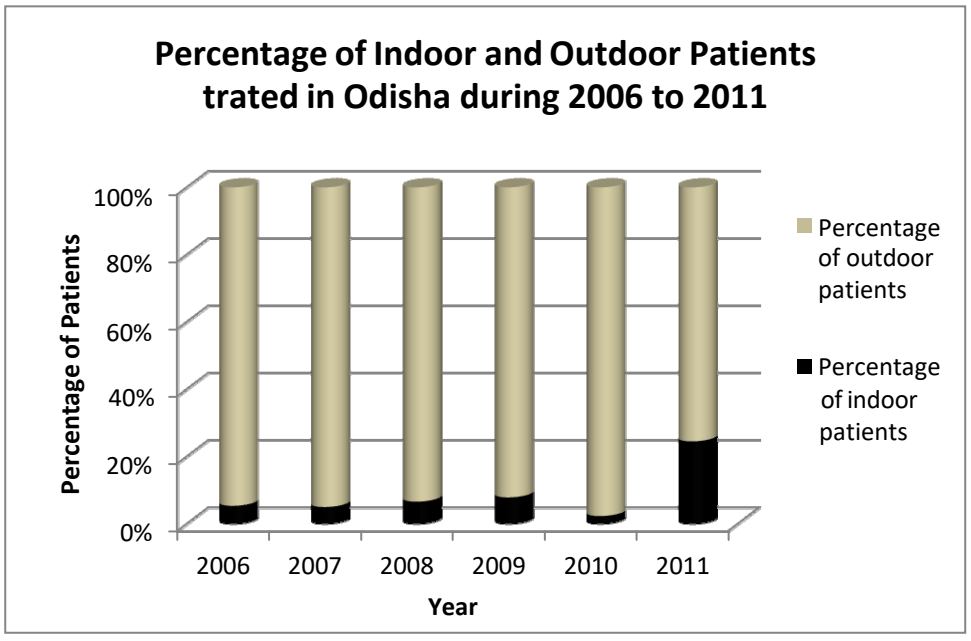


Figure 7: A Pie-Diagram Showing Number of Un-electrified villages in Some Districts of Odisha as on 31st march, 2012

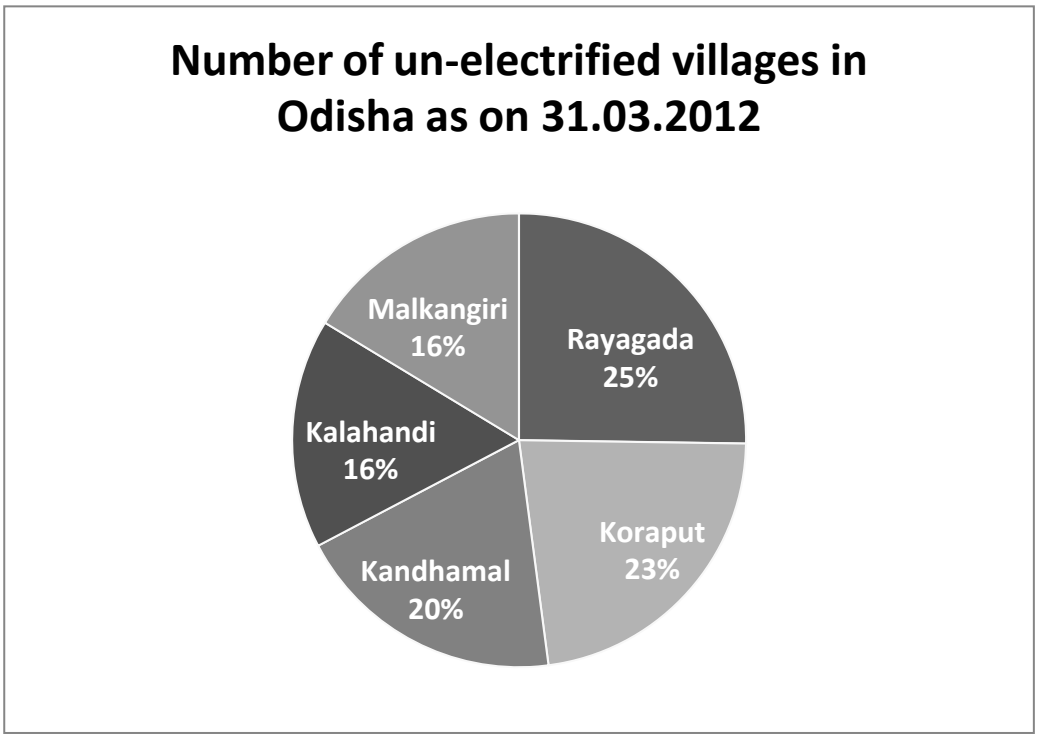


Figure 8: Histogram

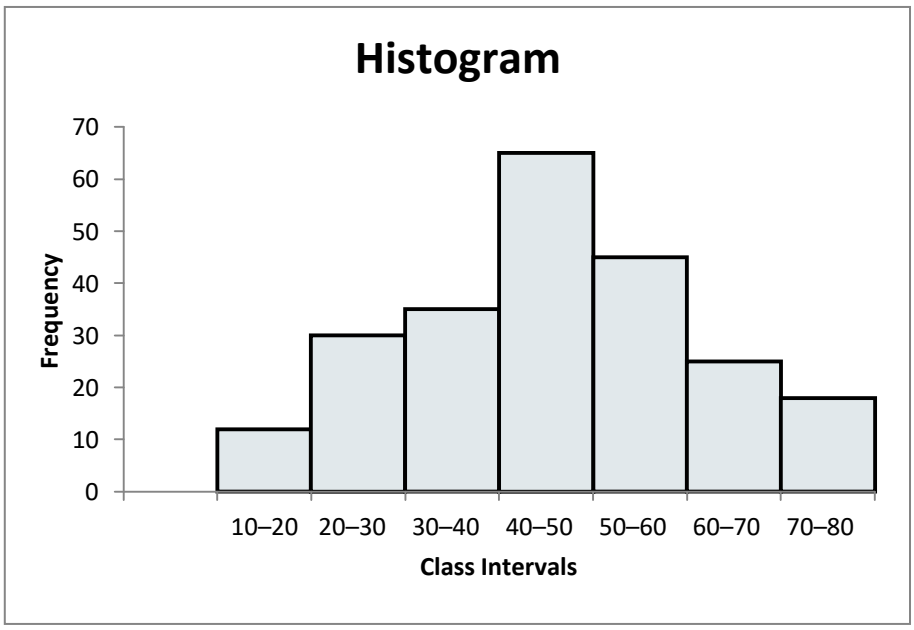


Figure 9: A Frequency Polygon

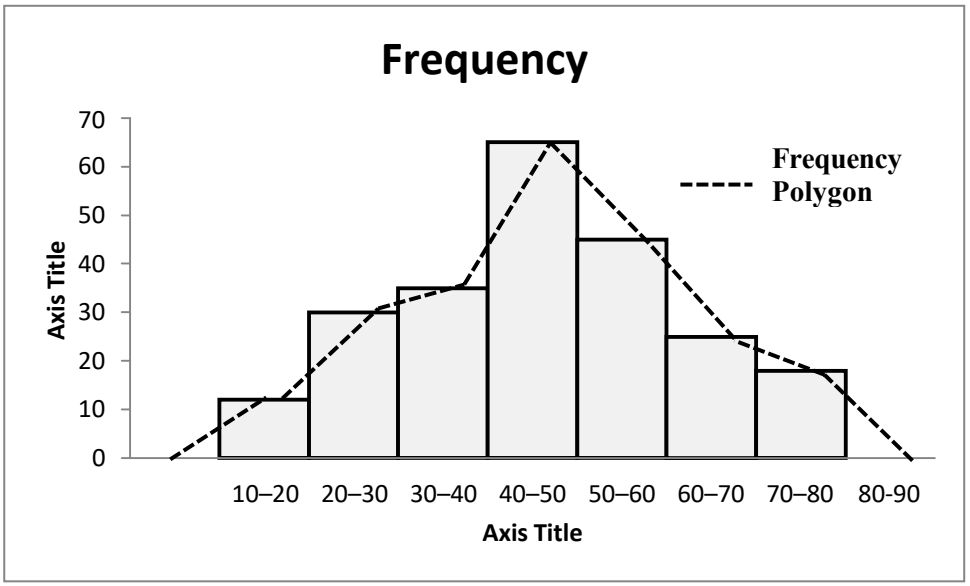


Figure 10: A Frequency Curve

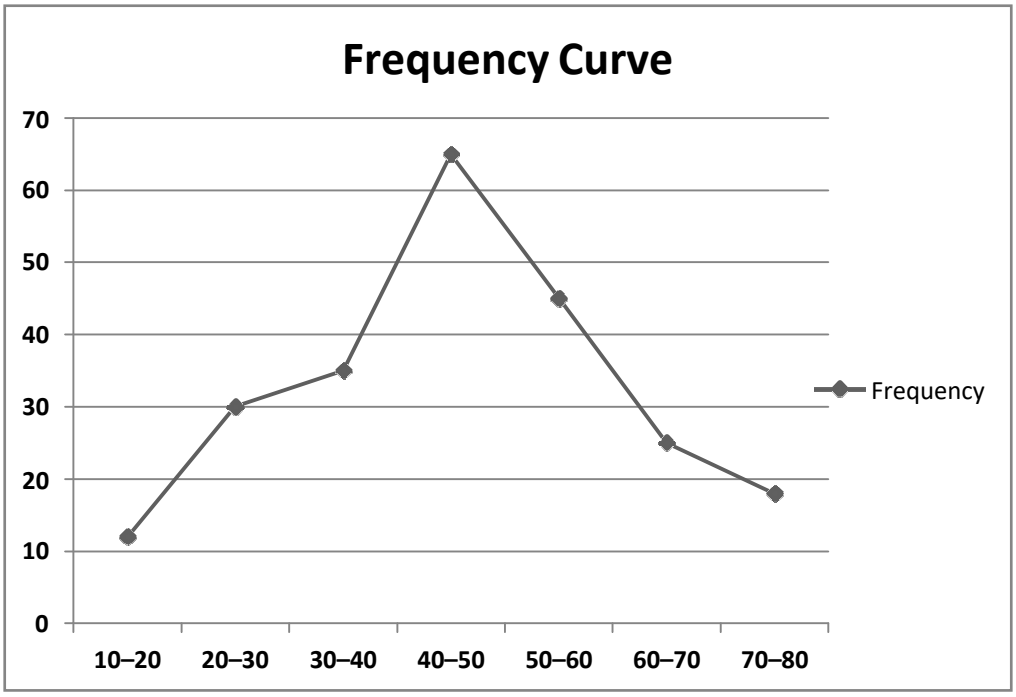


Figure 11: Cumulative Frequency Curves

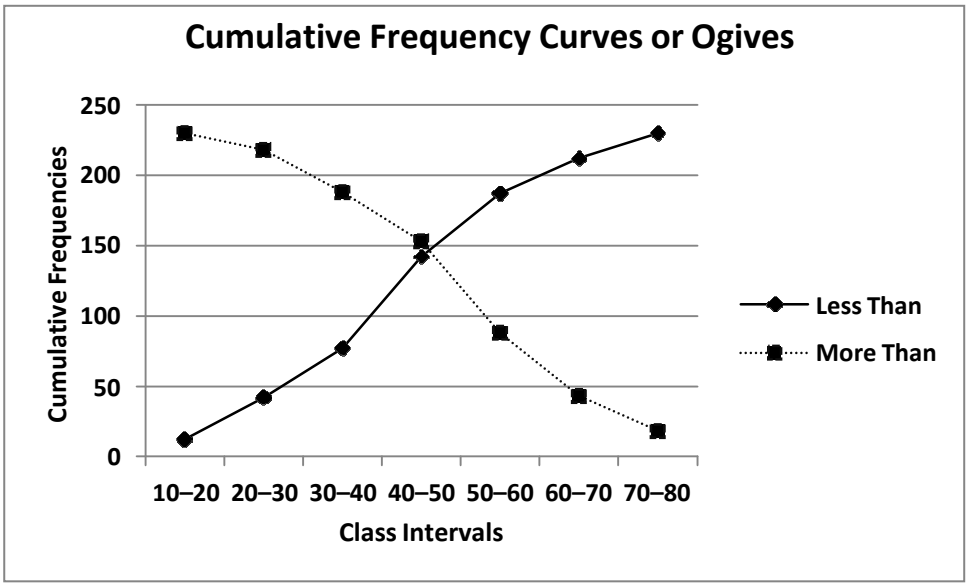
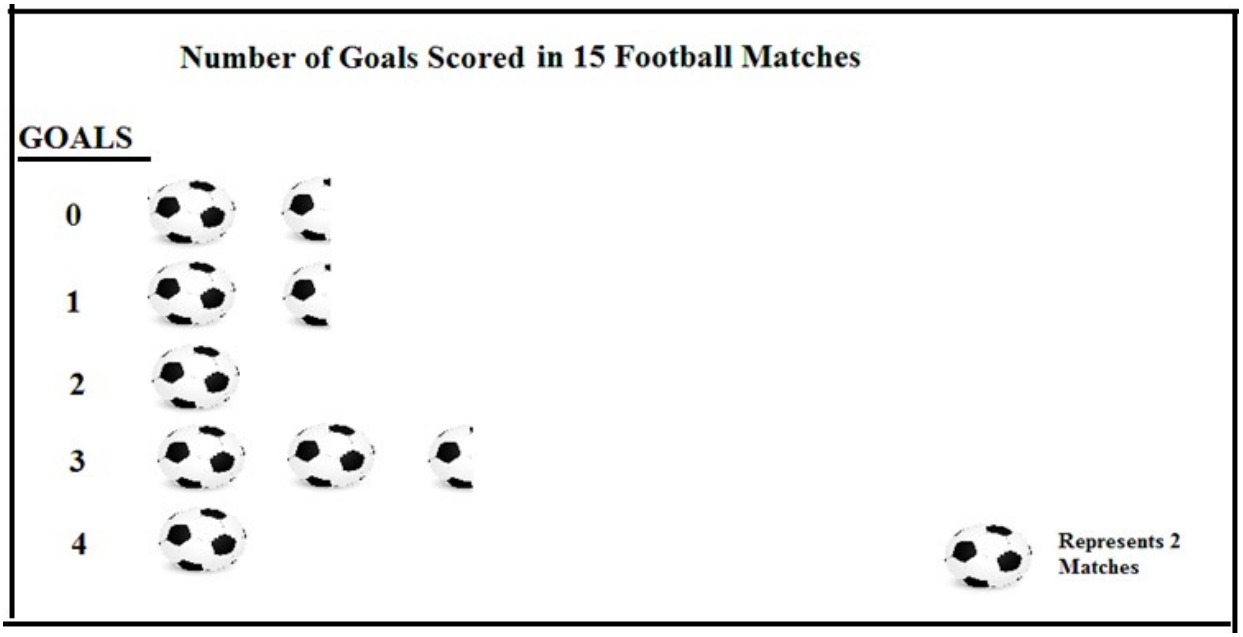


Figure 12: Pictogram showing the Number of goals Scored in 15 Football Matches



UNIT 5: SOCIAL STATISTICS

1. Introduction to Social Statistics

Statistics, as well as statistical analysis, is the key to the field of social sciences. Social scientists study the attitudes of groups of people, and this is often through the use of focus groups, questionnaires, and polls. These studies results in statistics, qualitative as well as quantitative. Statistics is important as it helps in problem solving. Social science or soft science studies the human aspects of the world. It includes an in-depth study and evaluation of human behavior by using scientific methods either in quantitative or qualitative manner. The subject matter of the soft science can be subjective or objective. In contrast to hard science or natural science reveals subjective, objective, inter-subjective and structural aspects of the society. In simple words, it measures the social developments of the society and finds the emerging drawbacks too. The work of social science is to watch where the society has been heading and what more can be done to benefit the entire race. The basic tool for providing a support to is social statistics.

1.1 Development and Importance of Statistics for Social Science

Nineteenth-century Americans worried about prostitution; reformers called it "the social evil" and warned that many women prostituted themselves. How many? For New York City, there are great differences between the numbers of prostitutes according different reports? Some reformers hoped that their reports of widespread prostitution would prod the authorities to act, but city officials' and police report counted only few prostitutes.Anti-prostitution reformers tried to use big numbers to arousepublic outrage. Big numbers meant there was a big problem. In response, the police countered that there were relativelyfew prostitutes—an indication that they were doing agood job. These dueling statistics resemble other, more recentdebates.

During Ronald Reagan's presidency, for example,activists claimed that three million Americans were homeless,while the Reagan administration insisted that the actual numberof homeless people was closer to 300,000, one-tenth what theactivists claimed. In other words, homeless activists argued thathomelessness was a big problemthat demanded additional governmentsocial programs, while the administration argued newprograms were not needed to deal with what was actually amuch smaller, more manageable problem. Each side presentedstatistics that justified its policy recommendations, and eachcriticized the other's numbers.

Statistics, then, can become weapons in political strugglesover social problems and social policy. Advocates of differentpositions use numbers tomake their points ("It's a big

problem!”“No, it’s not!”). And, as the example of nineteenth-century estimatesof prostitution reminds us, statistics have been used asweapons for some time.

In fact, the first “statistics” were meant to influence debates over social issues. The term acquired its modern meaning—*numeric evidence*—in the 1830s. The forerunner of statistics was called “political arithmetic”; these studies—mostly attempts to calculate population size and life expectancy—emerged in seventeenth-century Europe, particularly in England and France. Analysts tried to count births, deaths, and marriages because they believed that a growing population was evidence of a healthy state; those who conducted such numeric studies—as well as other, non-quantitative analyses of social and political prosperity—came to be called statists. Over time, the statists’ social research led to the new term for quantitative evidence: *Statistics*.

Early social researchers believed that information about society could help governments devise wise policies. They were well aware of the scientific developments of their day and, like other scientists, they came to value accuracy and objectivity. Counting and quantifying—offered a way of making their studies more precise, and let them concisely summarize lots of information.

Over time, social research became less theoretical and more quantitative. As the researchers collected and analyzed their data, they began to see patterns. From year to year, they discovered, the numbers of births, deaths, and even marriages remained relatively stable; this stability suggested that social arrangements had an underlying order, that what happened in a society depended on more than simply its government’s recent actions, and analysts began paying more attention to underlying social conditions.

By the beginning of the nineteenth century, the social order seemed especially threatened: cities were larger than ever before; economies were beginning to industrialize; and revolutions in America and France had made it clear that political stability could not be taken for granted. The need for information, for facts that could guide social policy, was greater than ever before. A variety of government agencies began collecting and publishing statistics: the United States and several European countries began conducting regular censuses to collect population statistics; courts, prisons, and police began keeping track of the numbers of crimes and criminals; physicians kept records of patients; educators counted students; and so on. Scholars organized statistical societies to share the results of their studies and to discuss the best methods for gathering and interpreting statistics. And reformers who sought to confront the nineteenth-century’s many social problems—the impoverished and the

diseased, the fallen woman and the child laborer, the factory workforce and dispossessed agricultural labor—found statistics useful in demonstrating the extent and severity of suffering. Statistics gave both government officials and reformers hard evidence—proof that what they said was true. Numbers offered a kind of *precision*: instead of talking about prostitution as a vaguely defined problem, reformers began to make specific, numeric claims.

During the nineteenth century, then, statistics—numeric statements about social life—became an authoritative way to describe social problems. There was growing respect for science and statistics offered a way to bring the authority of science todebates about social policy. In fact, this had been the main goalof the first statisticians—they wanted to study society throughcounting and use the resulting numbers to influence social policy.They succeeded; *statistics gained widespread acceptance asthe best way to measure social problems*. Today, statistics continueto play a central role in our efforts to understand theseproblems. But, beginning in the nineteenth century and continuingthrough today, social statistics have had two purposes, onepublic, the other often hidden. Their public purpose is to give anaccurate, true description of society. But people also use statisticsto support particular views about social problems. Numbersare created and repeated because they supply ammunition forpolitical struggles, and this political purpose is often hiddenbehind assertions that numbers, simply because they are numbers,must be correct. People use statistics to support particularpoints of view, and it is naive simply to accept numbers as accurate,without examining who is using them and why.

The lesson should be clear: statistics—even official statistics such as crime rates, unemployment rates, and census counts—are products of social activity. We sometimes talk about statistics as though they are facts that simply exist, like rocks, completely independent of people, and that people gather statistics much as rock collectors pick up stones. This is wrong. All statistics are created through people's actions: people have to decide what to count and how to count it, people have to do the counting and the other calculations, and people have to interpret the resulting statistics, to decide what the numbers mean. All statistics are social products, the results of people's efforts.

Once we understand this, it becomes clear that we should not simply accept statistics by uncritically treating numbers as true or factual. If people create statistics, then those numbers need to be assessed, evaluated. Some statistics are pretty good; they reflect people's best efforts to measure social problems carefully, accurately, and objectively. But other numbers are bad statistics—figures that may be wrong, even wildly wrong. We need to be able to sort out the good statistics from the bad.

1.2 Definition of Statistics

The origin of the word '*STATISTICS*' is indicated by the word itself, which seems to have been derived from the Latin word '*STATUS*' or from Italian word '*STATISTA'* or may be from the German word '*STATISTIK'*. The meaning of all these words is '*Political State*'. Every state administration collected and analyzed data in past regarding population to have an idea about the possible military strength and regarding material wealth to have an idea about possible source of finance to the state. Later on, data were collected for various purposes. But, the first analysis of data in a scientific manner was done by *Captain John Grunt* in his book "Observation on London Bills of Mortality" published in 1962. In India, the evidences of data collection can be seen in the *Kautilya'sArthasastra* (Chandragupta Maurya'a regime) and also in *Ain-i-Akbari* (Akber's period).

The tradition of collection of data and its use for various purposes is very old. The development of modern statistics as a subject is of recent origin. The actual developments held after 16^{th} Century.

Statistics is a subject that deals with the collection of data as well as the techniques of analysis and interpretation of data. The word 'Statistics' has been used in both *plural* and also in *singular* sense.

In *plural sense*, Statistics means a set of numerical figures or data. In *singular sense*, it represents a method for study and therefore, refers to statistical principles and methods developed for analysis and interpretation of data.

Statistics can be defined in different ways by different authors. But, these definitions can be broadly classified into two categories:

I. Definitions giving emphasis on Statistics as Data,
II. Definitions giving emphasis on Statistics as a Scientific Method.

I. Statistics as Data

Statistics used in the plural sense implies a set of numerical figures collected with reference to a certain problem under investigation. It may be noted that, any set of numerical figures cannot be regarded as Statistics. There are certain characteristics which must be satisfied by a given set of numerical figures in order that they may be termed as Statistics. Some of these definitions are as follows:

- ***A. L. Bowley*** – "Statistics are numerical facts in any department of enquiry placed in relation to each other."
- ***Yule and Kendall*** – "By Statistics we mean, quantitative data affected to a marked extent by multiplicity of causes."
- ***H. Secrist*** – "By Statistics we mean, aggregate of facts affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonably standard of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other."

II. Statistics as a Scientific Method

Statistics is a scientific method used in almost all the disciplines of natural as well as social sciences. Statistics as scientific method is divided into two categories:

a) Theoretical Statistics,
b) Applied Statistics.

a) Theoretical Statistics – It can be further subdivided into the following three categories:

- **Descriptive Statistics** – All those methods which are used for the collection, classification, tabulation, diagrammatic representation of data and the methods of calculating average, dispersion, correlation and regression, etc. are included in descriptive statistics.
- **Inductive Statistics** – It includes all those methods which are used to make generalizations about a population on the basis of a sample. It also includes forecasting techniques.
- **Inferential Statistics** – It includes all those methods which are used to test certain hypotheses regarding the characteristics of the population.

b) Applied Statistics – It consists of the application of statistical methods to practical problems. Design of sample surveys, techniques of quality control, decision making in business, etc. are included in applied statistics.

Statistics as a Science or as an Art

Science is a body of systematized knowledge developed by generalizations of relations based on the study of cause and effect. These generalized relations

are also called the laws of science. For example, laws in physics, chemistry, statistics, mathematics, etc. So, it is obvious that, Statistics is a science.

Now, we shall examine, it is an art or not. As we know, science is a body of systematized knowledge. But, how this know ledge is to be used for solving a problem is the work or an art. In addition to this, art also helps in achieving certain objectives and to identify merits and merits of methods that could be used. Since, Statistics possesses all these characteristics, it may be reasonable to say that, it is also an art.

1.3 Functions of Statistics

I. Presents facts in numerical figures.
II. Presents complex facts in a simplified form.
III. Studies relationships between two or more phenomena.
IV. Provides techniques for the comparison of phenomena.
V. Enlarges individual experiences.
VI. Helps in the formulation of policies.
VII. Helps in forecasting.
VIII. Provides techniques for testing of hypotheses.
IX. Provides techniques for taking decisions under uncertainty.

1.4 Scope of Statistics in Other Areas

It is difficult to imagine a field of knowledge which can do without statistics. The scopes of statistics provide precision to various ideas and also suggest possible ways to handle a problem in different fields. The scopes of statistics in some important areas are as under.

In Economics: Statistics is mostly used for studying various economic problems. It also provides guidelines for formulating different economic policies. Different economic principles like law of diminishing marginal utility, law of diminishing marginal returns are based on the generalizations of economic behavior basing upon a large number of individuals. Statistical methods are also used to derive relations between different economic variables such as relationship between demand and price of a commodity and many more.

In Management: With the increase of size of business and increasing competitions, different business organizations or firms using statistical information and statistical analysis for

arriving at a suitable business decision for improving their business. Different policy decisions are also taken by the owners or managers for developing their business in the present highly competitive markets. Important management decisions like amount of investment, deciding production schedules and volumes, developing marketing strategies, etc. are also decided basing upon the statistical principles. Banks are taking decisions regarding the amount of cash needed at hand per day to meet the customer demands.

In Science and Industry: Laws of science are based upon the generalizations of observations, which are derived by using statistical techniques. Theory of probability is used to obtain the accuracies and precisions in different scientific investigations. Again, in industry, different important decisions like location and its size and volume of production, planning of production, inventory control and management, quality control of products are also based on the statistical principles.

Planning without Statistics is a ship without Radar and Compass.

1.5 Limitations of Statistics

I. Statistics deals with numerical facts only. It is not suited to study of qualitative phenomenon. But, we can apply statistical techniques to these qualitative phenomenon indirectly by first reducing to a suitable precise measurement. For example the intelligence of a group of candidates can be expressed in terms of test scores.

II. Statistics deals only with groups or aggregate of objects and not with individuals.It does not permit any specific recognition for the individual items of a series. Accordingly, statistical analysis is suited to only those problems where group characteristics are to be reviewed.

III. Statistical results are true only on the average. On the basis of statistical analysis, we can talk only in terms of probability or chance ant not in terms of certainty. Statistical laws can never be regarded as universal truths.

IV. Statistical results are only approximately true.

V. Statistics are liable to be misused.

VI. Statistics must be used only by experts.

2. Frequency Distribution and Its Characteristics

In Statistics, we deal with data, which is a collection of observations on some characteristics expressed in numerical figures. These characteristics are of two kinds – measurable and non-measurable. Measurable characteristics are those which can be

numerically expressed in terms of some units, i.e., there is a definite scale to measure them. The measurable characteristic is known as variable. Age, weight, height, income etc., are examples of variables.

On the other hand, a non-measurable characteristic is a qualitative object and as such is incapable of numerical expression, i.e., there is no definite scale to measure these.Religion, Nationality, Occupation, Gender, Marital status, etc. are such non-measurable characteristics and these are, therefore termed as attributes.

A variable can be a continuous one or a discrete one. A continuous variable can assume its value for any real quantity within a specified interval. For example, in measuring the height of a group of people, the heights may have any value within a range from 140 cm to 180 cm. A discrete variable can assume only some isolated values. In most of the cases it assumes only integral values. For example, the number of children born to a couple, age at marriage (in years), number of earning members in a family, etc.

Data relating to the values of a continuous variable is called as continuous data and that relating to the values of a discrete variable is called as discrete data.

2.1 Definition and Importance

The values of a variable are obtained by observations or measurements. If we have a series of such values of a variable and a value is repeated more than once in the series, then the number of times a value is repeated in the series is called as the frequency of that value. If indicates how frequently that value occurs in the series. For example, in the data on the marks secured by the students in a class in Statistics, if the mark 75 is secured by 5 students, then the frequency of 75 is 5.

Frequency distribution of a variable is defined as the distribution of frequency over different values of the variable. When we write the frequencies against different values of a variable in the form of a table, it is called as a frequency table.

Let us consider mark secured by100 students in a class. These are written one by one. So, when we have to conclude

1. How many of them are passed in first division?
2. How many of them are passed in second division?
3. How many of them are passed in second division?
4. How many of then failed?

5. How many students scored more than 90 percent?

To answer these questions, we have to search the value one by one, which becomes a time consuming and tedious work. It becomes more complicated when the number of values increases. Therefore, an alternative technique is to construct a frequency distribution. So, frequency distribution reduces the volume of data without losing any important information.

2.2 Construction of a Frequency Distribution

The easiest method of organizing data is a frequency distribution, which converts raw data into a meaningful pattern for statistical analysis. Depending upon the nature of the values considered, there are two types of frequency distribution:

I. Ungrouped Frequency Distribution

II. Grouped Frequency distribution

2.2.1 Ungrouped Frequency Distribution

In this case, we first search the minimum value and the maximum value. Then we write the values one by one starting from the minimum to maximum (ascending order). (We can write in descending order also.) Next, we search the values one by one from the original series, and put tally marks (|) against the value (from the series written in ascending order) equal to a value in the original series. If we find the same value more than once, then each time we give a new tally mark. These tally marks are grouped in fives (𝍸) for easy counting. If we find a value 7 times, then for this value the tally marks will be (𝍸||) . After searching for all the values, we count the tally marks and write it against each value. This gives the frequency of different values.

For example, the marks of 100 students in Statistics in a certain examination are as follows:

75	82	39	65	48	92	19	76	84	56	49	83	25	44	59
65	68	74	83	14	29	32	46	49	92	84	65	47	29	53
48	62	75	94	26	47	56	39	44	52	46	74	24	46	75
54	49	75	64	25	34	56	84	71	50	82	40	36	95	45
25	86	75	91	67	59	60	76	63	70	81	76	49	60	76
49	56	72	49	65	82	46	58	72	73	46	59	45	65	31
50	70	14	46	38	46	75	80	60	56					

Here, minimum mark = 14 and Maximum mark = 95.

Table 2.1 Construction of Frequency Distribution for the Marks

Values	Tally Mark	Frequency	Values	Tally Mark	Frequency

14		2	55		0
15		0	56		5
16		0	57		0
17		0	58		1
18		0	59		3
19		1	60		3
20		0	61		0
21		0	62		1
22		0	63		1
23		0	64		1
24		1	65		5
25		3	66		0
26		1	67		1
27		0	68		1
28		0	69		0
29		2	70		2
30		0	71		1
31		1	72		2
32		1	73		1
33		0	74		2
34		1	75		6
35		0	76		4
36		1	77		0
37		0	78		0
38		1	79		0
39		2	80		1
40		1	81		1
41		0	82		3
42		0	83		2
43		0	84		3
44		2	85		0
45		2	86		1
46		7	87		0
47		2	88		0
48		2	89		0
49		6	90		0
50		2	91		1
51		0	92		2
52		1	93		0
53		1	94		1
54		1	95		1

Table 2.2 Frequency Distribution for the Marks

Marks	Frequency	Marks	Frequency	Marks	Frequency	Marks	Frequency
14	2	40	1	58	1	74	2
19	1	44	2	59	3	75	6
24	1	45	2	60	3	76	4
25	3	46	7	62	1	80	1
26	1	47	2	63	1	81	1
29	2	48	2	64	1	82	3
31	1	49	6	65	5	83	2
32	1	50	2	67	1	84	3
34	1	52	1	68	1	86	1
36	1	53	1	70	2	91	1
38	1	54	1	71	1	92	2
39	2	56	5	72	2	94	1
				73	1	95	1

2.2.2 Grouped Frequency Distribution

In the above, we have seen that the ungrouped frequency distribution reduces the volume of data by avoiding the writing of repeated values. But, it puts another difficulty of writing all the values starting from the minimum value to maximum value, which is time consuming and tedious task, when there is a large difference between the maximum and minimum values. In order to avoid this difficulty, we can form a grouped frequency distribution. Here, we form different groups of values, called as class intervals and find the frequency of these class intervals. The following are the steps of constructing a grouped frequency distribution:

1. *Specify the number of class intervals* . There is no specific rule for determining the number of class intervals. But, generally, between 5 and 15 class intervals are used. Note that the classes must be both mutually exclusive and all- inclusive. Mutually exclusive means that classes must be selected such that an item can't fall into more than one class and all - inclusive classes are classes that together contain all the data.
2. When all intervals are to be the equal width (size), the following rule may be used to find the required width of the class interval:
   W $= (L - S) / K$where: W = class width, = the largest value,
   S =the smallest value, = number of classes

Illustration - 1

Suppose the ages of a sample of 10 students are:

20.9, 18.1, 18.5, 21.3, 19.4, 25.3, 22.0, 23.1, 23.9, and 22.5

We select $K = 4$ and $\mathrm{W} = (25.3 - 18.1)/4 = 1.8$ which is rounded-up to 2. The frequency table is as follows:

Class Interval	Frequency
18—20	3
20—22	2
22—24	4
24—26	1

Illustration - 2

In case of the data given before about the marks of 100 students, suppose we have to form the class intervals of size 10, then the frequency distribution of marks is as follows:

Frequency Distribution of Marks

Class Intervals	**Tally Marks**	**Frequency**
10 – 20	𝍸	3
20 – 30	𝍸 ||	7
30 – 40	𝍸 ||	7
40 – 50	𝍸 𝍸 𝍸 𝍸 ||	22
50 – 60	𝍸 𝍸 |||	14
60 – 70	𝍸 𝍸 ||	13
70 – 80	|| 𝍸 𝍸 |||	18
80 – 90	𝍸 𝍸 |	11
90 – 100	𝍸	5
Total		100

This type of class intervals are called as exclusive type of class intervals. Here the class intervals are written in continuous form, *i.e.*, there is no gap between the class intervals. Therefore the frequency distribution is called as a continuous frequency distribution. In the upper limit of a class interval is equal to the lower limit of the next higher class interval. So, if an observation is equal to the upper limit or the lower limit of a class interval, then it should be classified into that class in which it is the lower limit. For example, if an

observation is 50, then it should be classified in the class interval 50 – 60 but not in 40 – 50. To avoid this problem, when we have only integral values, we can write the class intervals in the form 0 – 9, 10 – 19, 20 – 29, etc. This type of class intervals are called as inclusive type of class intervals. The frequency distribution is called as a discrete frequency distribution.

2.3 Comparison of Frequency Distributions

In social work research we come across different situations where we need to compare between two groups of values. For example, we can compare the age structure of males and females in a village, income of the people in two different communities, educational status (in terms of number of years of education) between two different income groups and also the gender difference with respect to educational status, and many more. In such cases, the graphical and the diagrammatical representation helps to some extent but we need to compare on the basis of some figures or the values or some measures, those should give a clear understanding about the differences and similarities prevailing between different groups. The common statistical measures those help to solve this problem are the *measures of central tendency* and the *measures of dispersion*.

The *measures of central tendency*give us an idea about the central values whereas the *measures of dispersion* give an idea about the heterogeneity or scattering-ness present in the data. Therefore these two types of measures play an important role in understanding the basic structure and nature of data.

2.4 Measures of Central Tendency

As we know that *the measures of central tendency give us an idea about the nature of central values of the series or the values those act as the representatives of the whole series*. For example, when we say the average income of the families in a community is Rs 15,000 per month *i.e.,*Rs 15,000 act as a representative of income status of the families living in that community. By seeing only this figure, one can easily guess the economic status of the families of that community; we need not observe the individual family incomes. We can think that the incomes of different families are more or less equal to Rs. 15,000 per month. *Thus, the measures of central tendency gives the value around which the whole series of data is assumed to be concentrated*.

Some common measures of central tendency are *arithmetic mean, median and mode*.

2.4.1 Arithmetic Mean

The *arithmeticmean* or simply *mean* of a set of values is defined as *the ratio of sum of all the values to the total number of values of the set*. It is also called as the *average* and denoted as $\bar{x}$ (read as x bar). This mean always lies between the maximum and minimum values of the set. For example, the ages of five students of a class are 8, 7.5, 7.2, 8.1, 7.9 years, then the mean of the ages of the students is

$$Arithmetic\ Mean = \bar{x} = \frac{8 + 7.5 + 7.2 + 8.1 + 7.9}{5} = \frac{38.7}{5} = 7.74\ years.$$

Here, we can say that the age of the students of the class is around 7.74 years.

Suppose, there are n values in a series or group with individual values $x_1, x_2, \cdots, x_n$, then the mean of this series is given by

$$Arithmetic\ Mean = \bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum}{n},$$

where $\sum x$ (read as summation x) stands for sum of all the values of the series, i.e., $\sum x = x_1 + x_2 + \cdots + x_n$.

When we have individual values with their respective frequencies, then we calculate the arithmetic mean by using the following formula:

$$\bar{x} = \frac{\sum(f \times x)}{\sum f} = \frac{\sum(frequency \times value)}{\sum(frequency)}$$

For example the following data gives the number of children born to 10 different couples.

Number of Children:	0	1	2	3
Number of Couples (Frequency):	2	4	3	1

Here, the average number of children =

$$\bar{x} = \frac{\sum(f \times x)}{\sum f} = \frac{2 \times 0 + 4 \times 1 + 3 \times 2 + 1 \times 3}{2 + 4 + 3 + 1} = \frac{13}{10} = 1.3 \approx 1\ child$$

But, sometimes the data is available in form of groups as in case of frequency distributions instead of individual values. In these cases these are calculated as follows.

Consider the income groups (Rupees per month) of 20 families randomly selected from a village is as follows. We have to calculate the mean, median and modal income of these families.

Income Groups:	0 – 10,000	10,000 – 20,000	20,000 – 30,000	30,000 – 40,000
Number of Families:	2	10	6	2

Calculation of Arithmetic Mean:

- Find the mid- values of the groups (class intervals) by taking the average of the lower and upper values of each group. These values are x.
- Find the product of the x values with the corresponding group frequencies, i.e., $(\mathrm{f} \times x)$. Find the total of these values, i.e., $\sum(\mathrm{f} \times x)$.
- Find the total frequency $\sum \mathrm{f}$.
- Calculate arithmetic mean

$$\bar{x} = \frac{\sum(\mathrm{f} \times x)}{\sum \mathrm{f}}$$

Income Groups	**Number of families**(f)	**Mid values** (x)	$(f \times \mathrm{x})$
0 – 10,000	2	5,000	10,000
10,000 – 20,000	10	15,000	1,50,000
20,000 – 30,000	6	25,000	1,50,000
30,000 – 40,000	2	35,000	70,000
Total	$\sum \mathrm{f} = 20$		$\sum(\mathrm{f} \times x) = 3{,}80{,}000$

$$Average\ Income\ of\ Families = \bar{x} = \frac{\sum(\mathrm{f} \times x)}{\sum \mathrm{f}} = \frac{3{,}80{,}000}{20}$$

$$= 19{,}000\ Rupees$$

Merits

- It is simple to calculate and easy to interpret.
- It is rigidly defined.

- It depends upon all the observations.
- It can be used for further mathematical treatment.
- It has less sampling variation than other measures of central tendency.

Demerits

- In spite of very simple calculation the arithmetic mean cannot be a suitable measure for all purposes. It can only be used in case of a measurable characteristics, but for non-measurable characteristics like honesty, I.Q., beauties', sweetness, etc., we cannot use arithmetic mean.
- Again, it sometimes gives very absurd results, for example, in calculating the average number of children born to couples in a village, it gives fractional values. Suppose there are five randomly selected couples from a village are selected and the number of children born to them is recorded as 3, 2, 0, 2, 1. So, the average number of children per couple is

$$\bar{x} = \frac{3 + 2 + 0 + 2 + 1}{5} = 1.6,$$

which is quite absurd.

- It is much more affected by extreme values.

2.4.2 Median

So, in order to avoid the drawbacks of arithmetic mean, we define another measure of central tendency, called as *median*. *The median of a series is defined as the value, which divides the series into two equal parts after arranging the values either in ascending or descending order*. In the above example, the median number of children born to the couple is calculated as follows.

Arranging the number of children in ascending order, we get: 0, 1, 2, 2, 3.

Median = The value present in the middle = 2 (third value)

So, the median number of children born to the couple is 2.

The value of the median depends upon the number of observations in the series, i.e., the number of observations is even or odd. After arranging in ascending order,

$$Median = \begin{cases} \left(\frac{n+1}{2}\right) th\ value, \text{ if } n \text{ is odd,} \\ The\ mean\ of\ \left(\frac{n}{2}\right) th\ and\ \left(\frac{n}{2}+1\right)\ th\ value, \text{ if } n \text{ is even.} \end{cases}$$

where$n$ is the number of observations. Suppose the age of 10 students of a school is 12, 11, 13, 14, 18, 16, 11, 15, 19 and 8 years. Here, the number of values $n = 10$ $(even)$.So, arranging in ascending order, we get 8, 11, 11, 12, 13, 14, 15, 16, 18, 19. So,

$$Median = The\ mean\ of\ (\frac{n}{2})\ th\ and\ (\frac{n}{2}+1)\ th\ value$$
$$= The\ mean\ of\ (\frac{10}{2})\ th\ and\ (\frac{10}{2}+1)\ th\ value$$
$$= The\ mean\ of\ (5)\ th\ and\ (6)\ th\ value = \frac{13+14}{2} = 13.5\ years.$$

When we have ungrouped frequency distribution, then the value of the median can be calculated as follows.

- Calculate the cumulative frequencies (less than type).
- The value whose cumulative frequency is just greater than equal to $(\frac{N+1}{2})$ is the median value, where $N =$ Total Frequency.

For example the following data gives the number of children born to 10 different couples.

Number of Children:	0	1	2	3
Number of Couples (Frequency):	2	4	3	1
Cumulative Frequency (less than)	2	$2+4=6$	$6+3=9$	$9+1=10$

Here, $N = 10$. So, $\frac{N+1}{2} = \frac{10+1}{2} = 5.5$.

Thus, the cumulative frequency just greater than equal to 5.5 is 6, which occurs for value 1.

So, median of this series is 1.

When we have grouped frequency distribution, then the value of the median can be calculated as follows.

Calculation of Median:

- ❖ Write the class intervals in continuous form.
- ❖ Find the cumulative frequencies (less than type) for different groups and find the total frequency $= \sum f$.

- Find the median class, i.e., the class interval containing the median. It is the class interval whose cumulative frequency is just greater than equal to $m = \frac{N}{2}$.
- Find the median value by using the following formula.

$$Median = L + \frac{m - c}{f}(L - U)$$

L =Lower limit of the median class

U = Upper limit of the median class

c = Cumulative frequency of the class interval preceding the median class

f =Frequency of the median class

We now calculate the median income for the previous example.

Income Groups	**Number of families**(f)	**Cumulative Frequencies**
0 – 10,000	2	2
10,000 – 20,000	10	$10 + 2 = 12$
20,000 – 30,000	6	$12 + 6 = 18$
30,000 – 40,000	2	$18 + 2 = 20$
Total	$\sum f = 20 = N$	

$$m = \frac{N}{2} = \frac{20}{2} = 10$$

The class interval whose cumulative frequency just greater than or equal to $m = 10$is $10{,}000 - 20{,}000$. So, $L = 10{,}000$, $U = 20{,}000$, $c = 2$, $f = 10$.

$$Median = L + \frac{m - c}{f}(L - U) = 10{,}000 + \frac{10 - 2}{10} \times (20{,}000 - 10{,}000)$$

$$= 18{,}000\ Rupees$$

Merits

- It is simple to calculate and easy to interpret.
- It is rigidly defined.
- It is less affected by extreme values.

Demerits

- It is not based on all the values.
- It is much affected by sampling fluctuations in comparison to arithmetic mean.
- It is not suitable for algebraic treatment.
- It can only be calculated by arranging the values either in ascending or descending order, which is very difficult for large number of values.

2.4.3 Mode

Again, we sometime give emphasis on the most frequent value in the series. For example, in an election there are several contesting candidates and the candidate getting maximum votes wins the election. In previous example, the maximum couples have two children (two couples). In a shoe shop, there are shoes of different sizes available, but the shopkeeper stocks a shoe whose number has maximum demand. Therefore, we define another measure of central tendency, which is called as *mode*. *The mode of a series is defined as the most frequent value in the series, i.e., the value which has maximum frequency in the series.*

For example the following data gives the number of children born to 10 different couples.

Number of Children:	0	1	2	3
Number of Couples (Frequency):	2	4	3	1

Maximum frequency is 4, which occurs for 1, so mode is 1 child.

Calculation of Mode:

- ❖ Write the class intervals in continuous form.
- ❖ Find the maximum frequency f_1 .
- ❖ Find the modal class, i.e., the class interval containing the mode. It is the class interval whose frequency is maximum.

❖ Find the modal value by using the following formula.

$$Mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2}(L - U)$$

L =Lower limit of the median class

U = Upper limit of the median class

f_0 = Frequency of the class interval preceding the modal class

f_1 = Frequency of the modal class

f_2 = Frequency of the class interval succeeding the modal class

$$\Delta_1 = |f_1 - f_0| , \quad \Delta_2 = |f_1 - f_2|$$

We now calculate the modal income for the previous example.

Income Groups	**Number of families**(f)
0 – 10,000	2
10,000 – 20,000	10 (Maximum)
20,000 – 30,000	6
30,000 – 40,000	2
Total	$\sum f = 20 = N$

The class interval whose frequency is maximum$f_1 = 10$is 10,000 – 20,000. So, $L = 10{,}000$, $U = 20{,}000$, $f_0 = 2$, $f_1 = 10$, $f_2 = 6$.

$$\Delta_1 = |f_1 - f_0| = |10 - 2| = 8 , \quad \Delta_2 = |f_1 - f_2| = |10 - 6| = 4$$

$$Mode = L + \frac{\Delta_1}{\Delta_1 + \Delta_2}(L - U) = 10{,}000 + \frac{8}{8 + 4} \times (20{,}000 - 10{,}000) = 16{,}667\ Rupees$$

When the mode is *ill-defined*, we can calculate the modal value by using the empirical relation: $Mode = 3\ Median - 2\ Mean$.

From the above, we find that all these three measures of central tendency help us to understand about the nature of the central values in a series. Each one of these has their own

importance in different areas. They help us to summarize the whole series of values into a single figure. These measures can easily be calculated in case the individual values are given to us.

Merits

- It is easily understood by the common man.
- It is less affected by extreme values.
- It can be determined by inspection for an ungrouped frequency distribution.
- It can be calculated for a grouped frequency distribution with open-end classes.

Demerits

- It is not based on all the values.
- It is much affected by sampling fluctuations.
- It is not suitable for algebraic treatment.
- Sometimes it is not well defined.

2.4.4 Quartiles

The quartiles of a series are the values, which divide the series into four equal parts after arranging either in ascending or descending order. There are three quartiles for a series: First Quartile (Q_1), Second Quartile (Q_2) and Third Quartile (Q_3).

To find the quartiles, first arrange the data in ascending order and then the $^{(N+1)}/_{4}$ th value is the first quartile, the $^{(N+1)}/_{2}$ th value is the second quartile and the $^{3(N+1)}/_{4}$ th value is the third quartile, where N is the total number of observations.

Let a student has scored in Statistics from 8 class tests (out of 100) are given as 62, 65, 67, 70, 63, 58, 55, 50 respectively. We have to find out the quartile marks. Here, $N = 8$ and after arranging the marks in ascending order, we have 50, 55, 58, 62, 63, 65, 67, 70. So,

First quartile =Q_1= Value of the $^{(N+1)}/_{4} = {}^{9}/_{4} = 2.25$th item or weighted average of 2nd and 3rd value = (3×55+1×58) / 4=55.25.

Second quartile =Q_2= Value of the $^{(N+1)}/_2 = {}^9/_2 = 4.5$th item = Average of 4th and 5th value = (62+63) / 2=62.5.

Third quartile =Q_3= Value of the $^{3(N+1)}/_4 = {}^{27}/_4 = 6.75$th item or 3rd value or weighted average of 6nd and 7th value = (1×65+3×67) / 4=66.5.

When we have a frequency distribution, then the quartiles can be calculated by using the following steps.

Calculation of Quartile Deviation

- Write the class intervals in continuous form.
- Find the cumulative frequencies (less than type).
- Find the total frequency = N.
- Find $q_1 = {}^N/_4$and $q_3 = {}^{3N}/_4$.
- Find the quartile classes. First quartile class is the class interval whose cumulative frequency is just greater than equal to $q_1 = {}^N/_4$, second quartile class is the class interval whose cumulative frequency is just greater than equal to $q_2 = {}^{2N}/_4 = {}^N/_2$ and third quartile class is the class interval whose cumulative frequency is just greater than equal to $q_3 = {}^{3N}/_4$. So, the second quartile is the median.
- Find the quartile values:

$$Q_1 = First\ quartile = L_1 + \frac{q_1 - c_1}{f_1} \times (U_1 - \ )_1$$

L_1 =The lower limit of first quartile class.

U_1 =The upper limit of first quartile class.

f_1 = Frequency of the first quartile class

c_1 =Cumulative frequency of the class interval preceding the first quartile class.

$$Q_2 = Second\ quartile = L_2 + \frac{q_2 - c_2}{f_2} \times (U_2 - \ )_2$$

L_2 =The lower limit of second quartile class.

U_2 =The upper limit of second quartile class.

f_2 = Frequency of the second quartile class

c_2 =Cumulative frequency of the class interval preceding the second quartile class.

$$Q_2 = Second\ quartile = Median$$

$$Q_3 = Third\ quartile = L_3 + \frac{q_3 - c_3}{f_3} \times (U_3 - \ )_3$$

L_3 =The lower limit of third quartile class.

U_3 =The upper limit of third quartile class.

f_3 = Frequency of the third quartile class

c_3 =Cumulative frequency of the class interval preceding the third quartile class.

Income Groups	**Number of families(f)**	**Cumulative Frequencies (Less than type)**
0 – 10,000	2	2
10,000 – 20,000	10	12
20,000 – 30,000	6	18
30,000 – 40,000	2	20
Total	$\sum f = 20 = N$	

Here,$_1 = {}^{N}/_{4} = {}^{20}/_{4} = 5$, $q_2 = {}^{N}/_{2} = {}^{20}/_{2} = 10$

and$q_3 = {}^{3N}/_{4} = {}^{3 \times 20}/_{4} = 15$

So, the first quartile class is 10,000 – 20,000, the second quartile class is 10,000 – 20,000 and the third quartile class is 20,000 – 30,000.

So, L_1 =10,000 $,_1 = 20{,}000.$, f_1 =10 $,_1 = 2$; , $L_2 = 10{,}000$, $U_2 = 20{,}000.$, f_2 =10 $,_2 = 2$ and $L_3 = 20{,}000$, $U_3 = 30{,}000$, $f_3 = 6$, $c_3 = 10.$

$$Q_1 = First\ quartile = L_1 + \frac{q_1 - c_1}{f_1} \times (U_1 - \ _1) = 10{,}000 + \frac{5-2}{10} \times (20{,}000 - 10{,}000)$$

$$= 13{,}000$$

$$Q_2 = Second\ quartile = L_2 + \frac{q_2 - c_2}{f_2} \times (U_2 - \ _2)$$

$$= 10{,}000 + \frac{10-2}{10} \times (20{,}000 - 10{,}000) = 18{,}000$$

$$Q_3 = Third\ quartile = L_3 + \frac{q_3 - c_3}{f_3} \times (U_3 - \ _3)$$

$$= 20{,}000 + \frac{15-10}{6} \times (30{,}000 - 20{,}000) = 18{,}333.33$$

2.5 Measures of Dispersion

Consider the mark sheet of three students in eight tests in Statistics in a month as given below:

Tests	Full Marks	Marks Secured		
		Student A	Student B	Student C
1	100	62	35	12
2	100	65	59	60
3	100	67	87	95
4	100	70	85	97
5	100	63	74	85
6	100	58	67	65
7	100	55	50	49
8	100	50	33	27
Total Mark	800	490	490	490
Average Mark	100	61.25	61.25	61.25

Here, we have seen that all the three students have same average marks. But, can we say that all the three students are equally consistent? Definitely, the answer is “NO”. The region is clear, because the marks scored by Student A is different tests is more or less nearer to each other. He is homogeneous in all the tests. The Student B has secured more heterogeneous marks than Student A and his marks are not nearer to each other. But, the

Student C has very heterogeneous, i.e., has a wide difference of marks in different tests. Hence, we conclude that even though all the three students have same average marksbut the heterogeneity of marks is not same for all the three students. The marks of Student A are more homogeneous than the other two. Since, they have same average marks, but they have difference in heterogeneity or consistency or uniformity.

Therefore, the measures of central tendency fail to give an idea about the heterogeneity or variability among the values. Thus, we need some other measures which should give us an idea about the variability or scattering among the values. These measures are the measures of dispersion. So, *the measures of dispersion are the measures which give us an idea about the variability or scattering of values of a variable.*

Some common measures of dispersion are

I. Range,
II. Quartile Deviation,
III. Standard Deviation and
IV. Coefficient of Variation.

2.5.1 Range

*Range is the simplest measure of dispersion, which is defined as the difference between the maximum and the minimum value of a series.*So, for the Student A,

Range = Maximum Value – Minimum value = 70 – 50 = 20.

For the Student B, Range = Maximum Value – Minimum value = 87 – 33 = 54.

For the Student C, Range = Maximum Value – Minimum value = 97 – 12 = 85.

So, the heterogeneity of marks is more for Student C than Student A and B.

Merits and Demerits

Range is simple to understand and easy to calculate. But, it has many disadvantages. It is very much affected by the presence of an extremely high or low value. It is not based on all the values of the series. It cannot be calculated for the grouped frequency distributions with open – end classes.

2.5.2 Quartile Deviation

The range gives a rough idea about the scattering of data as it depends only upon the maximum and minimum value and not depends upon the intermediate values. So, its value may be same for two different series and it does not take into account the magnitude of the values. Therefore, we need another measure, which should depend upon intermediate values also. One such measure is the *quartile deviation*. So, the quartile deviation is a measure that depends upon the *quartiles* of a series. For the marks of Student A, it can be calculated that the first and third quartiles are 55.25 and 66.5 respectively (see section 2.4.4).

$$Quartile\ Deviation\ = \frac{Q_3 - Q_1}{2} = \frac{66.5 - 55.25}{2} = 5.625$$

Merits and Demerits

Quartile deviation is easy to calculate and its calculation depends on the first and third quartiles only. It is not based on all the values of the series. It can be calculated for a grouped frequency distribution with open-end classes.

2.5.3 Standard Deviation

Now, the quartile deviation does not depend upon all the values of the series, so we define another measure known as *standard deviation* that depends upon all the values of the series. *When there are n values in a series given by $x_1, x_2, \cdots, x_n$ then their standard deviation is the root mean square deviation of all the values, where the deviations are taken from the arithmetic mean of the series. It is denoted as σ (read as sigma).*

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

In the above example, we shall now find the standard deviation for the marks scored by Student A. For this, we shall first calculate the arithmetic mean as

$$\bar{x} = \frac{1}{n}\sum x = \frac{1}{8}(62 + 65 + 67 + 70 + 63 + 58 + 55 + 50) = \frac{490}{8} = 61.25$$

$$\frac{1}{n}\sum x^2 = \frac{1}{8}[(62)^2 + (65)^2 + (67)^2 + (70)^2 + (63)^2 + (58)^2 + (55)^2 + (50)^2]$$

$$= \frac{3844 + 4225 + 4489 + 4900 + 3969 + 3364 + 3025 + 2500}{8}$$

$$= \frac{30316}{8} = 3789.5$$

$$Standard\ deviation = \sigma = \sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2} = \sqrt{3789.5 - (61.25)^2} = \sqrt{37.9375}$$

$$= 6.1593$$

$$Variance = \sigma^2 = (6.1593)^2 = 37.9375$$

For a frequency distribution, the standard deviation can be calculated by using the following formula.

$$Standard\ deviation = \sigma = \sqrt{\frac{1}{N}\sum(\mathrm{f} \times x^2) - \bar{x}^2}$$

Where N is the total frequency, x is the value or mid-value of a class interval with respective frequency f.

Merits and Demerits

Standard deviation is the most important and widely used among the measures of dispersion. It is rigidly defined and based on all the values of the series. It is suitable for further algebraic treatment. It is less affected by sampling fluctuations than any other absolute measures of dispersion.

On the other hand, standard deviation is difficult to calculate and also difficult to understand. Its calculation is also very much complicated.

2.5.4 Coefficient of Variation

Again the standard deviation depends upon the units of measurement of the variable used, so it varies when the units of measurement varies. To avoid this difficulty, a more stable measure of variability independent of units of measurement of the variable is given by *coefficient of variation* and is given by

$$Coefficient\ of\ Variation = CV(x) = \frac{\sigma}{\bar{x}} \times 100 = \frac{S.D.}{Mean} \times 100$$

From the above the coefficient of variation for marks of Student A is

$$Coefficient\ of\ Variation = (\bar{x}) = \frac{.D.}{Mean} \times 100$$

$$= \frac{6.1593}{61.25} \times 100 = 10.06\ (approx.)$$

The above calculations for different measures of dispersion are used when the individual values are given. But, when the values are given in the form of class intervals and frequencies, we follow the following procedures in order to calculate different measures.

Consider the following illustration.

Illustration – 1.

Consider the income groups (Rupees per month) of 20 families randomly selected from a village is as follows. We have to calculate different measures of dispersion for the income of these families.

Income Groups:	0 – 10,000	10,000 – 20,000	20,000 – 30,000	30,000 – 40,000
Number of Families:	2	10	6	2

Calculation of Range

Range = Upper value of the last class interval – Lower value of the first class interval

= 40,000 – 0 = 40,000. Rupees

Income Groups	**Number of families(f)**	**Cumulative Frequencies (Less than type)**
0 – 10,000	2	2
10,000 – 20,000	10	12
20,000 – 30,000	6	18
30,000 – 40,000	2	20
Total	$\sum f = 20 = N$	

Here,$_1 = {}^{N}/_{4} = {}^{20}/_{4} = 5$ and $q_3 = {}^{3N}/_{4} = {}^{3 \times 20}/_{4} = 15$

So, the first quartile class is 10,000 – 20,000 and

the third quartile class is 20,000 – 30,000.

So, $L_1 = 10{,}000$, $_1 = 20{,}000.$, $f_1 = 10$, $_1 = 2$ and

$L_3 = 20{,}000$, $U_3 = 30{,}000$, $f_3 = 6$, $c_3 = 10.$

$$Q_1 = First\ quartile = L_1 + \frac{q_1 - c_1}{f_1} \times (U_1 - \ )_1 = 10{,}000 + \frac{5-2}{10} \times (20{,}000 - 10{,}000)$$

$$= 13{,}000$$

$$Q_3 = Third\ quartile = L_3 + \frac{q_3 - c_3}{f_3} \times (U_3 - \ )_3$$

$$= 20{,}000 + \frac{15-10}{6} \times (30{,}000 - 20{,}000) = 18{,}333.33$$

$$Quartile\ Deviation = \frac{Q_3 - Q_1}{2} = \frac{18{,}333.33 - 13{,}000}{2} = 2{,}666.665$$

Calculation of Standard Deviation

- Find the mid-values of the groups (class intervals) by taking the average of the lower and upper values of each group. These values are x.
- Find the product of the values with the corresponding group frequencies, i.e., $(f \times x)$and $(f \times x^2)$. Find the total of these values, i.e., $\sum(f \times x)$and $\sum(f \times x^2)$.
- Find the total frequency $\sum f$.
- Calculate arithmetic mean

$$\bar{x} = \frac{\sum(f \times x)}{\sum f}$$

- Calculate the standard deviation

$$Standard\ deviation = \sigma = \sqrt{\frac{1}{N}\sum(f \times x^2) - \bar{x}^2}$$

Income Groups	**Number of families**(f)	**Mid values** (x)	$(f \times x)$	$(f \times x^2)$
0 – 10,000	2	5,000	10,000	5,00,00,000
10,000 –	10	15,000	1,50,000	2,25,00,00,000

20,000				
20,000 – 30,000	6	25,000	1,50,000	3,75,00,00,000
30,000 – 40,000	2	35,000	70,000	2,45,00,00,000
Total	$\sum f = 20$		$\sum(f \times x) = 3{,}80{,}000$	$\sum(f \times x^2)$ =8,50,00,00,000

$$Average\ Income\ of\ Families = \bar{x} = \frac{\sum(f \times x)}{\sum f} = \frac{3{,}80{,}000}{20}$$

$$= 19{,}000\ Rupees$$

$$Standard\ deviation = \sigma = \sqrt{\frac{1}{N}\sum(f \times x^2) - \bar{x}^2}$$

$$= \sqrt{\frac{1}{20} \times 8{,}50{,}00{,}00{,}000 - (19{,}000)^2} = \sqrt{64000000}$$

$$= 8{,}000\ Rupees.$$

$$Coefficient\ of\ Variation = (\bar{x}) = \frac{S.D.}{Mean} \times 100$$

$$= \frac{8{,}000}{19{,}000} \times 100 = 42.105\ (approx.)$$

3. Judging about Linear Relation: Correlation

During the analysis of data relating to social work research, veryoftenwecomeacrosssituations in which our focus lies in two or more variables simultaneously and we observe that a change in one variable occurs due to a change in another variable. For example, age of the husband and age of the wife, income and expenditure of different families, age and weight of persons, number of years of marriage of the women and the age of their first child, etc.

The relationship between the variables creates an inquisitiveness to find out

I. A technique to develop a quantitative measure of the extent of relationship and
II. Finding an appropriate mathematical or statistical form between the two or more variables.

Thefirstobjectiveisfulfilledby the technique of *correlation* whereas the second one is fulfilled by the *regression* technique.

In case of two variables, if the increase or decrease in one variable produces a change in other variable, then the two variables are said to be correlated. *Correlation is a statistical technique which is used to measure a relationship between two variables.* The correlation between two variables can be studied by using the following two methods:

I. Scatter Diagram - Graphical Method
II. Karl Pearson's Method of Correlation Coefficient – Mathematical Method

3.1 Scatter Diagram

It is a graphical method of studying the relationship between the two variables$x$and y. Here, we take one of the variable(x) along X-axis and the other variable(y) along Y-axis and plot different sample points. The scattering of the points give an idea about the nature of correlation between the two variables. Generally the following standard types of scattering are observed.

I. Points are scattered on a straight line.

It is the case of perfect correlation. In this case, both the variables change in same direction and in same proportion. If the straight line faces upward, then it is the case of perfect positive correlation (Figure –3.1(a)) and if it faces downward, then it is the case of perfect negative correlation (Figure- 3.1(b))

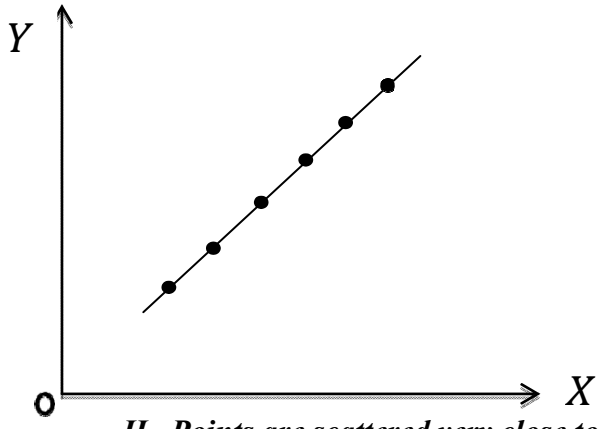


Figure 3.1 (a): Perfect Positive Correlation

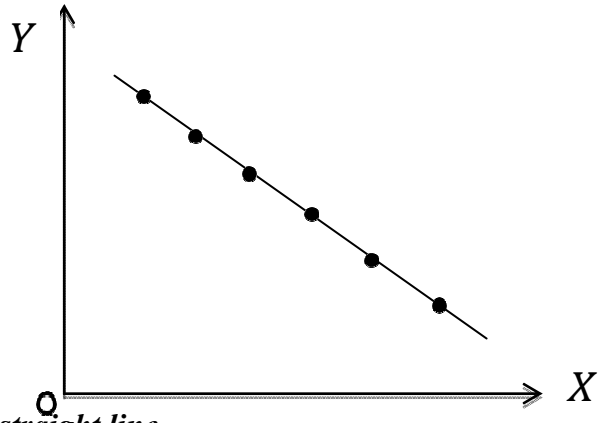
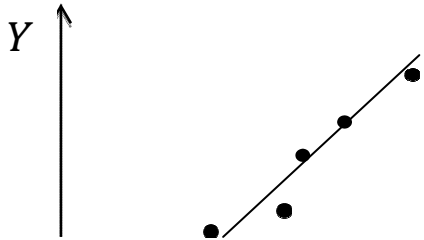
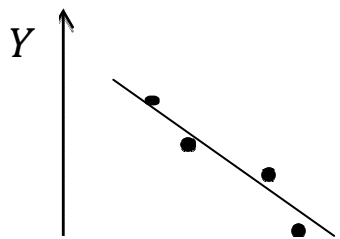


Figure 3.1(b): Perfect Negative Correlation

II. Points are scattered very close to a straight line.





It is the case of high degree of correlation. If the straight line faces upward, then it is the case of high degree of positive correlation (Figure – 3.2(a)) and if it faces downward, then it is the case of high degree of negative correlation (Figure- 3.2(b)).

III Points are scattered near to a straight line.

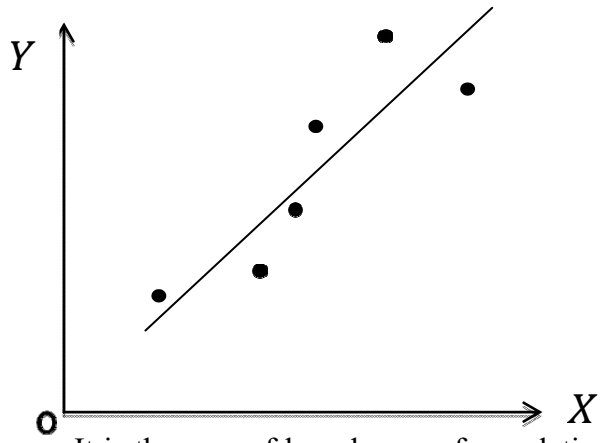


Figure 3.3 (a): High Degree of Positive Correlation

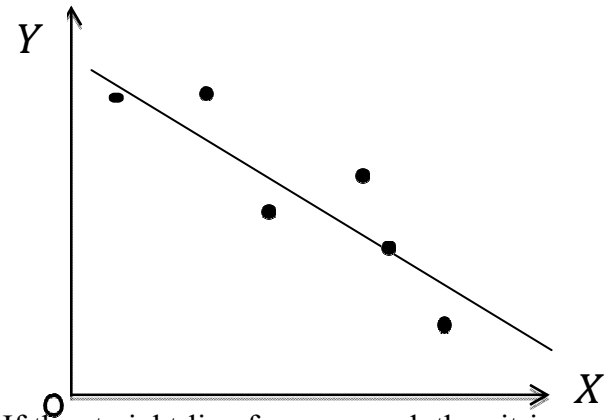


Figure 3.3(b): High Degree of Negative Correlation

It is the case of low degree of correlation. If the straight line faces upward, then it is the ur(it is (F

IV Points are scattered in the form of a disc.

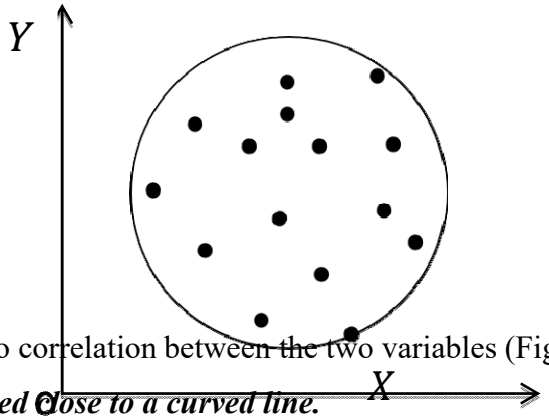


Figure 3.4: No Correlation

It indicates that there is no correlation between the two variables (Figure- 3.4).

V Points are scattered close to a curved line.

It is the case of curvilinear kind of non-linear relationship between the two variables. (Figure- 3.5)

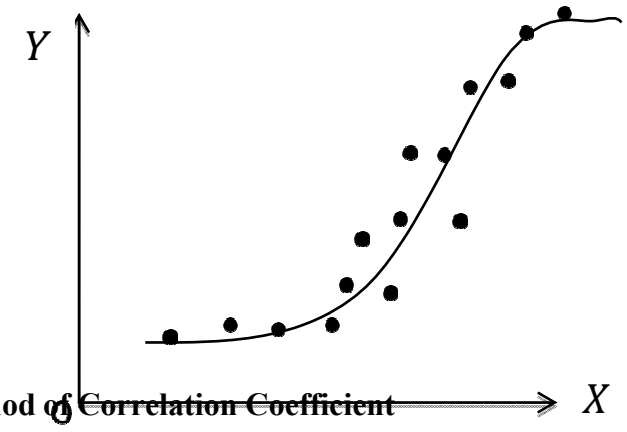


Figure 3.5: Curvilinear Correlation

3.2 Karl Pearson's Method of Correlation Coefficient

The scatter diagram is a simple method to study the relationship between two variables but it does not give any idea about the degree of relationship. So, we have to find out a measure that should give an idea about the extent of linear relationship between the two variables. For this Karl Pearson suggested a measure of extent of linear relationship between the two variables, which is known as 'Correlation Coefficient'. Suppose $(x_1, y_1), (x_2, y_2), \cdots \cdots \cdots, (x_n, y_n)$ be the n pairs of values on the two variables x and y, then the correlation coefficient between the two variables x and y is given by

$$r = \frac{Covariance\ between\ x\ and\ y}{Standard\ Deviation\ of\ x \times Standard\ Deviation\ of\ y}$$

Illustration 1.

A random sample of 10 couples is selected from a village and their age at marriage is recorded as follows. Calculate the value of Karl - Pearson's Correlation Coefficient between the age at marriages of the couples.

Age at marriage of the Husband (x):	32	28	34	33	36	27	30	32	38	35
Age at marriage of the Wife (y):	27	25	32	27	30	27	28	25	33	34

Solution

x	y	xy	x^2	y^2
32	27	864	1024	729
28	25	700	784	625
34	32	1088	1156	1024
33	27	891	1089	729

36	30	1080	1296	900
27	27	729	729	729
30	28	840	900	784
32	25	800	1024	625
38	33	1254	1444	1089
35	34	1190	1225	1156
$\sum x = 325$	$\sum y = 288$	$\sum xy = 9436$	$\sum x^2 = 10671$	$\sum y^2 = 8390$

From the above table, we have

$n = 10$ $\quad \sum x = 325$ $\quad \sum y = 288$

$\sum xy = 9436$ $\quad \sum x^2 = 10671$ $\quad \sum y^2 = 8390$

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

$$= \frac{10 \times 9436 - 325 \times 288}{\sqrt{[10 \times 10671 - (325)^2] \times [10 \times 8390 - (288)^2]}}$$

$$= \frac{94360 - 93600}{\sqrt{[106710 - 105625] \times [83900 - 82944]}}$$

$$= \frac{760}{\sqrt{1085 \times 856}} = \frac{760}{\sqrt{928760}} = \frac{760}{963.722} = 0.79$$

Here, the correlation coefficient between the age of the husband and age of wife at marriage is 0.79. So, it indicates that there is a positive correlation between the age at marriages of the husband and wife. So, more is the age at marriage of the husband, more will be the age at marriage of his wife.

3.3Mathematical Properties of Correlation Coefficient

(I) The value of the correlation coefficient (r) lies between -1 to $+1 i.e. -1 \leq r \leq 1$.

(II) Correlation coefficient (r) is independent of change of origin and scale *i.e.*, if we add or subtract a constant value from the x values and the y values then the value of the correlation coefficient remains unchanged. Similarly, if we multiply or divide and constant to the values of $x$and y, the also the value of the correlation coefficient remains unchanged.

(III) If two variables are independent, then the value of the correlation coefficient is zero.

Note : If $r = 0$, then the two variables may not be independent. For example: the values of $x$and y are given as follows.

x:	-2	-1	0	1	2
$y = x^2$:	4	1	0	1	4

Calculation of Correlation Coefficient

x	y	x^2	y^2	xy
-2	4	4	16	-8
-1	1	1	1	-1
0	0	0	0	0
1	1	1	1	1
2	4	4	16	8
$\sum x = 0$	$\sum y = 10$	$\sum x^2 = 10$	$\sum y^2 = 34$	$\sum xy = 0$

Here, $r = 0$, but $y = x^2$

So, x and y are not independent.

3.4 Co-efficient of Determination

The square of the correlation coefficient (r^2) is called as the coefficient of determination, because it gives the percentage of explained variance.

For example, if $r = 0.8$ then $r^2 = 0.64$ then 64% of the variation in variable 'y' depends upon 'x' and remaining 36% depends upon other factors. In a socio-economic survey, it was found that the correlation coefficient between the income (x) and expenditure (y) of different families is 0.8, that means 64% of the expenditure is affected due to the family

income and the remaining 36% depends upon other factors (may be number of family members, number of dependents in the family, number of school / college going children, etc.). Here, the value of r^2gives the percentage of explained variation. Therefore, the square of correlation coefficient is called as the coefficient of determination.

3.5 Interpretation ofCorrelation Coefficient

Value of *r*	**Interpretation**
$0 < r \leq 1$	Positive correlation
$r = +1$	Perfect positive correlation, i.e. x and y change in same direction and same proportion. Ex :$y = 2x$ x : -2 -1 0 1 2 y : -4 -2 0 2 4
$0.8 \leq r < 1$	Very high degree of positive correlation
$0 < r < 0.8$	Very low degree of positive correlation
$r = 0$	no correlation
$-1 \leq r < 0$	Negative correlation
$-0.8 < r < 0$	Very low degree of negative correlation
$-1 < r \leq -0.8$	Very high degree of negative correlation
$r = -1$	Perfect negative correlation, i.e. x and y change in opposite direction and same proportion. Ex ;$y = -2x$ x : -2 -1 0 1 2 y: 4 2 0 -2 -4

3.6Nature of Simple Correlation

The nature of simple correlation is either positive or negative. When both the variables change in same direction, then, it is known as a *positive correlation* and when they change in opposite direction, it is known as a *negative correlation*.

SPURIOUS OR NON-SENSICAL CORRELATION

When there is a systematic variation between the two variables but there is not any logical basis for this variation, then the correlation between then is known as a spurious correlation. For example the correlation between the weight of persons and their annual income, age of a farmer and production of rice, etc.

3.7 Types of Correlation

There are three types of correlation.

(i) Simple correlation;

(ii) Multiple correlation;

and (iii) Partial correlation.

(i) Simple Correlation

The linear relationship between two variables is known as simple correlations. For example, the correlation between income and expenditure, height and weight, etc. It is measured by Karl Pearson's correlation coefficient '***r***'.

(ii) Multiple Correlation

The correlation between a single variable on a group of other variables is known as multiple correlation. For example, if

x_1 = Monthly expenditure of a family,

x_2 = Number of dependents in the family,

x_3 = Monthly income of the family.

Then the correlation between x_1 on x_2 and x_3 is known as multiple correlation. It measures the effect of x_2 and $x_3$on x_1. It is denoted as $R_{1.23}$ and is given by

$$R^2_{1.23} = 1 - \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{23}r_{13}}{1 - r_{23}^2}$$

(iii) Partial Correlation

When, there are more than two variables then the correlation between any two of these variables at a constant level of other variables is known as partial correlation. In case of three variables, x_1, x_2, x_3, the correlation between x_1 and x_2 at a constant level of x_3 is known as partial correlation. For example, the correlation between monthly expenditure of a family (x_1) and number of dependents in the family (x_2) for the families having same monthly income (x_3).

It is measured by the partial correlation coefficient $r_{12.3}$ and is given by

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}},$$

where r_{12}, r_{23}, r_{13} denote the simple correlation coefficient between x_1 and x_2, x_2 and x_3, x_1 and x_3 respectively.

3.8 RankCorrelation (Without tie)

Karl Pearson's correlation coefficient formula is applicable in case of two variables. But, in practice, when we have two attributes like honesty, beauty, character, morality, etc. then we cannot use Pearson's correlation coefficient formula, because there is no definite scale to measure these qualities. In such situations we use Spearman's rank correlation coefficient formula to study the relationship between them.

For example, consider two attributes A and B. Suppose there are 'n' individuals. We can arrange these 'n' individuals, in order of merit according to these two attributes A and B separately. The position of an individual with respect to an attribute is called as the rank of the individual. Assuming that, all the individuals get different ranks, i.e., there is no tie, then these 'n' individuals gets the ranks from 1 to n with respect to two attributes A and B separately. Let $x_i$and y_i be the ranks of i^{th} individual due to the attributes A and B respectively (i= 1, 2, 2, ..., n). So, x_i and y_i are natural numbers ranging from 1 to n.

Individual No	:	1	2	3 ... i...	n
Rank due to A	:	x_1	x_2	$x_3 \ldots x_i \ldots$	x_n
Rank due to B	:	y_1	y_2	$y_3 \ldots y_i \ldots$	y_n

$$Rank\ correlati\ n\ coefficient\ =\ r = 1 - \frac{6\sum d^2}{(n^2 - 1)}$$

where, $d_{\text{i}} = x_{\text{i}} - y_{\text{i}}, \text{i} = 1, 2, \ldots, n.$

= difference of ranks of i^{th} individual.

Illustration 2

Calculate the rank correlation coefficient for the following data.

Students	:	1	2	3	4	5	6	7	8	9	10
Marks in Math	:	57	62	86	90	85	70	32	46	51	60
Marks in Stat	:	40	79	92	91	70	82	57	47	32	45

Solution

Students	Marks in Math	Marks in Stat	Rank in Math (x)	Rank in Stat (y)	$d = x - y$	d^2
1	57	40	7	9	-2	4
2	62	79	5	4	1	1

3	86	92	2	1	1	1
4	90	91	1	2	-1	1
5	85	70	3	5	-2	4
6	70	82	4	3	1	1
7	32	57	10	6	4	16
8	46	47	9	7	2	4
9	51	32	8	10	-2	4
10	60	45	6	8	-2	4
Total						$\sum d^2 = 40$

Here, $n = 10$ and $\sum d^2 = 40$.

So, Rank correlation coefficient = $1 - \frac{6\sum d^2}{(n^2-1)} = 1 - \frac{6\times40}{10(10^2-1)} \approx 0.36$, which is very low. Therefore we can conclude that there is a very little agreement between the ranks in Mathematics and rank in Statistics.

INTERPRETATION OF RANK CORRELATION CO-EFFICIENT

$r = 1, \text{Z}d = 0 \quad \Rightarrow \quad x_i = y_i, \quad i = 1, 2, ...n.$

i.e. the rank of i^{th} individual is same with respect to both the attributes. So, there is a perfect agreement between the ranks.

Illustration 4

Individuals : 1 2 3 .....*n*

Rank due to A: 1 2 3.....*n*

Rank due to B : 1 2 3.....*n*

$r = -1 \Rightarrow x_i + y_i = (n + 1)$.

i.e. the ranks are opposite to each other. Here, there is a perfect disagreement between the ranks due to A and B.

Illustration 5

Individual : 1 2 3 .... *n*

Rank due to A: 1 2 3 .... *n*

Rank due to B : *n* *n* -1 *n* - 2 1

4. Tests of Significance

Testing plays a major role in statistical investigation. Here, we are examining the truth or otherwise of the hypothesis (guess, claim or assumptions, etc.) about some feature about one or more populations on the basis of samples drawn from these populations. Generally, a

statistical hypothesis is a statement or a conclusion or an assumption about certain characteristics of populations which is drawn on a logical basis and it can be tested basing on the sample evidences. Test of hypothesis means either accept or reject the hypothesis under a valid reason. Test of hypothesis is based on test of significance. *The test of significance enables a researcher to decide either to accept or reject the statistical hypothesis.* For example, an NGO working for the development of rural poor claims that 70% of the families in a certain village have their monthly income below Rs 5,000 per month. In order to verify the claim is true or not, we have to check it on the basis of sample of families selected from that village. Anagency wants to verify the effectiveness of advertisement regarding the awareness towards adult literacy given through print media is less effective than audio-visual media or not. There are a wide range of areas relating to social work where we have to come across situations of arriving at a decision of accepting or rejecting the hypothesis. So, it is very much important to have knowledge about the logical basis of such decisions and it is provided by hypothesis testing, which is the objective of this chapter.

It is a usual procedure that samples are drawn from the population to obtain an estimate of population parameter which is in other word is called sample statistic. Estimate of population parameters thus obtained may or may not exactly match with true values. To take the sample statistic as the estimate of population parameter is involved with risk. So it is worthwhile to find whether the difference between the estimated value of the parameter or the true value is significantly different or it could have arisen due to fluctuation of sampling. For this reason a hypothesis is formulated and then tested for validity.

A ***statistical hypothesis*** is a quantitative statement about a population *i.e.,* more specifically about a population parameter. Test of hypothesis which is based on test of significance, one can find out whether it deserves acceptance or rejection.

4.1 Basic Concepts

In the below we discussed some basic concepts used in testing of hypothesis.

4.1. 1 Null and Alternative Hypothesis

Null hypothesis is the hypothesis of no difference and it is tested for possible rejection under the assumption that it is true. It is usually denoted by H_0.

Any hypothesis, which is complementary to null hypothesis is called *an alternative hypothesis*. It is denoted as H_1. For example, an anti-tobacco awareness programs were organized in a district by a Government agency at regular intervals. Now, the Government wants to know whether the programs are actually has an impact for the prevention of tobacco use or not after six months of campaigning. To verify this, we can use the testing of hypothesis procedure with the null hypothesis that

H_0 : There if no difference tobacco use in six months before the training and after the training,

against the alternative hypothesis

H_1: The average use of tobacco after the training is less than the average use of tobacco before the training.

Symbolically, we can write these two hypotheses as

$H_0 : \mu_1 = \mu_2$ against the alternative $H_1: \mu_1 > \mu_2$.

Where, μ_1 = Population Average use of tobaccobefore the training and

μ_2 = Population Average use of tobaccoafter the training.

Also, the Government wants to verify that thepercentage of adult literacy in some villages are meeting their targets or not in last year. So, we can set up the null and alternative hypotheses as

H_0 : There is no difference in average adult literacy (μ) and the target value (μ_0),

against the alternative hypothesis

H_1 : There is a difference between adult literacy (μ) and the target value (μ_0).

i.e. $H_0 : \mu = \mu_0$ against the alternative $H_1: \mu \neq \mu_0$.

We can generalize this as follows:

Suppose a researcher is interested to verify that for a population has mean (μ) is equal to $\mu_0$or not. Then we set up the null hypothesis as

$H_0 : \mu = \mu_0$

Then, the alternative hypothesis could be any one of the statements

(i) $H_1: \mu \neq \mu_0$ or (ii) $H_1 : \mu > \mu_0$ or (iii) $H_1 : \mu > \mu_0$.

The alternative hypothesis in (i) is known as a two tailed alternative and the test based on this known as *two tailed test*. The (ii) and (iii) are known as right tailed and left tailed alternatives respectively. In general (ii) and (iii) are known as one tailed alternatives and the test based on it is known as *one tailed test*. The setting of alternative hypothesis is very important since on the basis of this only, we decide whether a one tailed (right or left) or two tailed test is to be used in test of significance.

4.1.2 Errors in Testing of Hypothesis

In the procedure of testing of hypothesis, a decision is taken about the acceptance or rejection of null hypothesis. The possible decisions can be written in a tabular form.

	True State	
Decision	**H_0 is True**	**H_0 is False**

Accept H_0	Correct Decision	**Type II Error**
Reject H_0	**Type I Error**	Correct Decision

There is always some possibility of committing the following two types of errors in taking such a decision as

Type I Error : Reject the null hypothesis H_0 when it is true.

Type II Error : Accept the null hypothesis H_0 when it is false.

Now, we write α = Probability of committing Type I error

and β = Probability of committing Type II error.

The compliment of Type II error is called as the *power of the test* and is given by $(1-\beta)$ and the size of Type I error (α) is also called as *level of significance*. The level of significance is the quantity of risk, which can be readily tolerated in making a decision about the null hypothesis H_0. Usually the value of α is chosen depending upon the desired degree of precession and its value varies as 0.05 (for moderate precision) or 0.01(for high precision).

4.1.3 Critical Region

The decision regarding the acceptance or rejection of H_0 is taken basing on a function of sample observation. This function of sample values is called as a test statistic or simply statistic. The collection of all possible values of this test statistic is known as the sample space (***S***). Now, in order to take a decision, we divide this sample space into two regions i.e. ***acceptance region*** $(S - \mathrm{W})$***and rejection region***(W)(Figure 4.1).

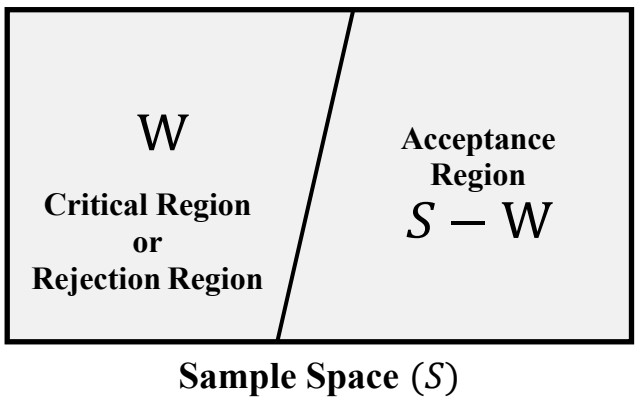


Figure 4.1 Sample Space and Acceptance region

When the value of the test statistic falls in the acceptance region ($S - W$), then we accept the null hypothesis and when the value of the test statistic falls in the rejection region or critical region (W), we reject the null hypothesis.

4.1.4. Degrees of Freedom

Number of independent observations are used to form the statistic is known as degrees of freedom (*d.f*) of that statistic. For example, consider the statistic with ***n*** independent observations, then its degree of freedom is ***n***. Let us consider a statistic, sample mean($\bar{x} = \sum_{i=1}^{n} x_i$). Since it is calculated from $S^2 = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, it has (n - 1) d.f. It loses one degrees of freedom due to one constraint $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.So, it is evident that the d.f. can be obtained by subtracting the number of constraints from the number of observations used to form the statistic.

4.1.5. One Tailed Test and Two Tailed Test

The concept of one tailed or two tailed test comes from the type of alternative hypothesis used in the test. Consider the following three set of null and alternative hypotheses formed to test the value of population mean.

(i)$\mathbf{H_0 : \mu = \mu_0, H_1: \mu \neq \mu_0}$

(ii) $\mathbf{H_0 : \mu = \mu_0, H_1 : \mu < \mu_0}$

(iii) $\mathbf{H_0 : \mu = \mu_0, H_1 : \mu > \mu_0}$

In the first case, the alternative hypothesis is in not equal to form, that indicates it includes both $\mathbf{\mu > \mu_0}$ as well as $\mathbf{\mu > \mu_0}$. So, the test can be called as a *two-sided test*. But, in the second and third case, it includes only one side (either greater than or less than). So, the tests in these two cases are called as *one tailed tests*.

In a two tailed test, the critical region lies in both the tails of the probability distribution curve of the test statistic (Fig. 4.2).

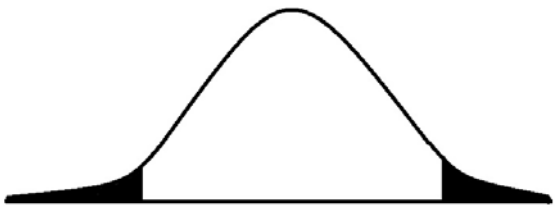


Figure 4.2 Two Tailed Test

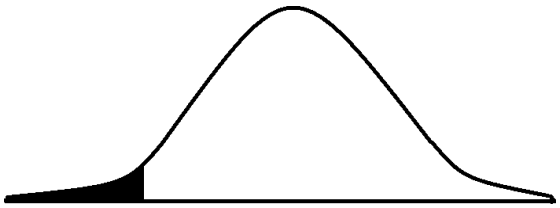


Figure 4.3 One Tailed Test – Left tailed

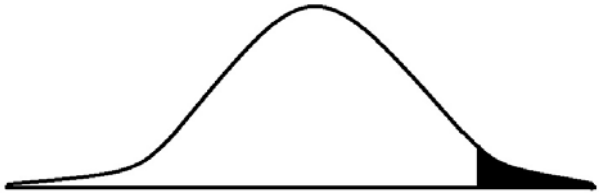


Figure 4.4 One Tailed Test – Right tailed

But, in case of a one tailed test, it lies to either in the left tail or in the right tail of the probability distribution curve of the test statistic (Figure 4.3 and 4.4) according as the alternative hypothesis is less than type or greater than type.

4.1.6 Standard Error

The standard deviation of sampling distribution of statistic is known as standard error (SE). If we calculate the standard deviation of sample means then it is called standard error of means.For example, consider a population of 5 children of a village with their ages is 5, 8, 6, 4, 6 years and we have to select a random sample of size 3 children from this population. When the second, fourth and fifthchildren are selected, the sample mean of their age is (8+4+6)/3=6 years. Similarly, if the sample units are first, third and fifth children, then the sample mean becomes (5+6+6)/3=5.67 years. So, when the sample changes, we get a different sample mean. Therefore, for each possible sample, we shall get a sample mean. So, by writing the sample means for all possible samples, we get the sampling distribution of sample mean and the standard deviation of this sampling distribution gives the standard error of the mean. In the following table, we calculate all possible sample means and its standard error.

Table 4.1: Calculation of Standard Error

Sample No.	Sample Units	Sample Values	Sample Mean
1	(1, 2, 3)	5, 8, 6	$\bar{x} = \frac{5+8+6}{3} = 6.33$
2	(1, 2, 4)	5, 8, 4	$\bar{x} = \frac{5+8+4}{3} = 5.67$
3	(1, 2, 5)	5, 8, 6	$\bar{x} = \frac{5+8+6}{3} = 6.33$
4	(1, 3, 4)	5, 6, 4	$\bar{x} = \frac{5+6+4}{3} = 5$

5	(1, 3, 5)	5, 6, 6	$\bar{x} = \frac{5+6+6}{3} = 5.67$
6	(1, 4, 5)	5, 4, 6	$\bar{x} = \frac{5+4+6}{3} = 5$
7	(2, 3, 4)	8, 6, 4	$\bar{x} = \frac{8+6+4}{3} = 6$
8	(2, 3, 5)	8, 6, 6	$\bar{x} = \frac{8+6+6}{3} = 6.67$
9	(2, 4, 5)	8, 4, 6	$\bar{x} = \frac{8+4+6}{3} = 6$
10	(3, 4, 5)	6, 4, 6	$\bar{x} = \frac{6+4+6}{3} = 5.33$

Standard Error of Mean = $SE\ (\bar{x})$

= Standard deviation of 6.33, 5.67, 6.33, 5, 5.67, 5, 6, 6.67, 6, 5.33

= 0.5709.

The standard error of the sampling distribution of means is equal to the ratio of standard deviation (**σ**) of the population and square root of sample size i.e.

$$Standard\ Error\ of\ Sample\ Mean = \frac{\sigma}{\sqrt{n}}$$

The following are some of the important uses of the standard error:

(i) The standard error provides an idea about the degree of precision of a sample or in other words it is an indication about the extent to which the sample is reliable.

(ii) It is used in test statistic to know whether a sample is drawn from a known population or not.

(iii) It is used to determine the size of the sample.

(iv) Standard error helps in determining the limits within which a parameter value is expected to lie. The sample distribution tends to Normal distribution, if the sample size is large.

4.2 Procedure for Testing of Hypothesis

STEP – 1: Specify the Null hypothesis (H_0) and the alternative hypothesis (H_1) .

STEP – 2: Fix a level of significance 'α' depending upon the accuracy desired in testing the hypothesis. Usually, the value of α is taken as 0.01 or 0.05 or 0.1

which is same as setting the confidence coefficient at 99% or 95% or 90% respectively.

STEP – 3: Select a sample of size n from the population and measure the study variable. Calculate the test statistic (**T**) basing on these sample values. If the sample size $n > 30$, then it is called as a large sample test(Z −Test). If the sample size $n < 30$, it is a small sample test (t −Test, 3^2 – Test).

STEP – 4: Find the degrees of freedom (**υ**) of the test statistic calculated in STEP - 3. For different probability distributions (like ***t, F***or 3^2) the statistical tables are available. So from the table, find the tabulated value of the statistic with required level of significance and value of degree of freedom.

STEP – 5: Compare the calculated value of the test statistic with the corresponding tabulated value.

If the Calculated Value of the test statistic is less than the tabulated value, we accept the null hypothesis H_0. Otherwise, we reject H_0.

STEP – 6: Write the conclusion on the basis of acceptance or rejection of the null hypothesis.

4.3 Large Sample Tests: Z − Test

When the sample size (n) is more than 30, almost all distribution tends to Normal distribution. So, the tests based on these samples are based on normal distribution i.e. the test statistic follows a normal distribution with zero mean and unit variance under the null hypothesis H_0. Usually, when the sample size is more than 30, it is called as a large sample and the tests are called as large sample tests. Some commonly used large sample tests are:

I. Test of significance for single mean
II. Test of significance of equality of two population means
III. Test of significance for single proportion
IV. Test of significance of equality of two population proportions

4.3.1 Test of Significance for Single Mean

Let a sample of size $n(> 30)$ is drawn from a population with unknown mean **μ** and known standard deviation **σ**. The observed sample values are $x_1, x_2, x_3, \ldots\ldots, x_n$. Now, can we consider that, this sample comes from a population with population mean μ_0.

Setting of Null Hypothesis

The null hypothesis is given by $\mathbf{H_0 : \mu = \mu_0}$against one of the alternative from the following three:(i)$\mathbf{H_1: \mu \neq \mu_0}$or(ii)$\mathbf{H_1 : \mu < \mu_0}$ or(iii) $H_1 : \mu > \mu_0$.

Calculation of Test Statistic

The Test statistic is given by

$$Z = \frac{\sqrt{(|\bar{x} - \mu_0|)}}{\sigma}$$

where $\bar{x} = \frac{1}{n}\sum x$ is the sample mean.

μ_0 is the hypothetical value of population mean, **σ**is the known population standard deviation, ***n*** is the size of the sample.

Under H_0, the test statistic **Z** follows a Normal distribution with Zero mean and unit variance. Let Z_{Cal} denote calculated value of the Z and Z_{Tab} is the tabulated value of Z with a level of significance α.

Finding the Tabulated Value

The following table gives the tabulated values of Z with different levels of significance **α**.

Table 4.2 Critical values of Z

		Level of Significance		
	Alternative Hypothesis	**α = 0.01**	**α =0.05**	**α =0.10**
One Sided Tests	**$H_1 : \mu < \mu_0$**	**-2.33**	**-1.645**	**-1.28**
	$H_1 : \mu > \mu_0$	**2.33**	**1.645**	**1.28**
Two Sided Tests	**$H_1 : \mu \neq \mu_0$**	**2.58**	**1.96**	**1.645**

Now, we can compare the calculated value Z_{Cal} with the tabulated value Z_{Tab} as follows and give a decision.

Decision

If $Z_{Cal} <$, Accept the null hypothesis $\mathbf{H_0}$ and if $Z_{Cal} \geq Z_{Tab}$, Reject $\mathbf{H_0}$.

Note: In case the population standard deviation is not known, we can calculate sample standard deviation (**s**) and use this in place of **σ**.

Illustration – 1.

The following data gives the sitting heights (in cm.) of 10 adults of an Indian tribe. Can we assume that the average sitting height of these tribes is 100 cm, when the population standard deviation is known as 53.21 cms? (Use %5 level of significance.)

84.33 85.01 84.35 86.21 85.72 86.12 85.91 90.35 95.42 92.86

Solution

Setting of Null Hypothesis –

The null hypothesis is given by

H_0: The average siting height of tribes is 100 cm., i.e., H_0: $\mu = 100$ cm.

against the alternative H_1: $\mu \neq 100$ cm.

Calculation of Test Statistic -

The sample size = n = 100

The hypothetical value of population mean = $\mu_0 = 100$ cm.

The sample mean =

$$\bar{x} = \frac{84.33 + 85.01 + 84.35 + 86.21 + 85.72 + 86.12 + 85.91 + 90.35 + 95.42 + 92.86}{10}$$

$$= \frac{876.28}{10} = 87.628$$

The population standard deviation is $\sigma = 53.21$ cm.

The test statistic is given by

$$Z = \frac{\sqrt{}(\bar{x} - \mu_0)}{\sigma} = \frac{\sqrt{10}(|87.628 - 100|)}{53.21} = 39.1237 = Z_{Cal}$$

Finding the Tabulated value -

Tabulated value of Z at 5% level of significance for two sided tests is $Z_{Tab} = 1.96$.

Decision -

So, the calculated value is much greater than the tabulated value. Hence, we reject the null hypothesis at 5% level of significance. Therefore, we can conclude that the average sitting height of the tribes cannot be considered as 100 cm.

4.3.2 Test of Significance of Equality of Two Population Means

In social work studies, we come in several situations where we need to compare between two means, for example, in comparing the age at marriage of male and female persons, for comparing the average number of children ever born in two different communities of tribes, for comparing the average time spent for learning between the students of two different schools, etc. Here, we are comparing two populations on the basis of their means. We draw sample from each population independently and on the basis of the samples we compare two population means.

Consider a population with an unknown mean μ_1 but known standard deviation σ_1. Let a sample of size n_1 is drawn from this population and the sample mean is found to be $\bar{x}_1$.There is another population with an unknown mean μ_2 but known standard deviation σ_2.

Let a sample of size n_2 is drawn from this population and the sample mean is found to be $\bar{x}_2$. It is summarized in the following table.

	Population - 1	**Population - 2**
Unknown Population Means	μ_1	μ_2
Population S.D. (Known)	σ_1	σ_2
Sample Size	n_1	n_2
Sample Means	$\bar{x}_1$	$\bar{x}_2$

We set up the Null hypothesis H_0 : $\mu_1 = \mu_2$

against one of the alternative from the following three:

(i) $\mathbf{H_1 : \mu_1 \neq \mu_2}$ or (ii) $\mathbf{H_1 : \mu_1 < \mu_2}$ or(iii) $H_1 : \boldsymbol{\mu_1 > \mu_2}$.

Test Statistic is given by

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

The test statistic **Z** follows a Normal distribution with Zero mean and unit variance. Let us denote Z_{Cal}as the calculated value of the Z and Z_{Tab} is the tabulated value of Z with a level of significance **α**.

Finding the Tabulated Value

The tabulated values of Z with different levels of significance $(\alpha)$are as given in Table4.2.

Decision

If $Z_{Cal} <$, Accept the null hypothesis $\mathbf{H_0}$ i.e. the two samples are from the populations with same population means and

if $Z_{Cal} \geq Z_{Tab}$, Reject $\mathbf{H_0}$i.e. there is a significant difference between the two population means.

Note: In many cases if population variances σ^2_1 and σ^2_2 are not known, we can use sample variances $s^2_1$and s^2_2as the estimates of σ^2_1 and σ^2_2 respectively.

Illustration – 2.

The following data gives the number of Hours different workers perform over-time duties in two different factories in a year:

	Factory 1	**Factory 2**
Number of sample workers	150	200
Average overtime (in Hours)	1240 Hours	1100 Hours
Standard Deviation	250 Hours	230 Hours

From the above information, can we conclude that there is a difference in average load on the workers in two factories or not?

Solution

Setting of Null Hypothesis : Let, μ_1 be the average life timeof the bulbs produced at Factory 1 and μ_2 be the average life timeof the bulbs produced at Factory 2. The null hypothesis is given by

H_0**:** The average lifetime of electric bulbs is same for both the factories. i.e., $H_0 : \mu_1 = \mu_2$

against the alternative $H_1 : \mu_1 \neq \mu_2$.

Calculation of Test Statistic: From the given data, we can write

	Factory 1	**Factory 2**
Number of sample workers	$n_1 = 150$	$n_2 = 200$
Average overtime (in Hours)	$\bar{x}_1 = 1240\ Hours$	$\bar{x}_2 = 1100\ Hours$
Standard Deviation	$\sigma_1 = 250$ Hours	$\sigma_2 = 230$ Hours

The test statistic is given by

$$Z = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{|1240 - 1100|}{\sqrt{\frac{(250)^2}{150} + \frac{(230)^2}{200}}} = \frac{140}{\sqrt{416.667 + 264.5}} = 5.364 = Z_{Cal}$$

Finding the Tabulated Value

Tabulated value of Z at 5% level of significance for two sided tests is $Z_{Tab} = 1.96$.

Decision : So, the calculated value is greater than the tabulated value. Hence, we reject the null hypothesis at 5% level of significance. Therefore, we can conclude that that there is a difference in average load among the workersof two factories.

4.3.3 Test of Significance for Single Proportion

Besides mean, we may be interested to compare the sample proportion with the population proportion to know they are same or differ significantly. For example, the proportion of bonded laborers in an interior tribal area of Odisha is about 50% of the total population. The proportion of child marriages occurred in an area is equal to 5% or less. The proportion of premature babies born to different mothers in different days in a hospital is about 10% of total births or more. The proportion of patients suffering from cancer due to tobacco habits is 70% or more, etc.

Let ***p*** denotes the sample proportion basing upon a sample size ***n*** and ***P*** denotes the population proportion (unknown).

Setting of Null Hypothesis –

The Null hypothesis H_0 :$p = P$

The alternative hypothesis is any one from

(i) $H_1 : p \neq P$ or (ii) $H_1 : p > P$ or (iii) $H_1 : p < P$

Calculation of Test Statistic -

Test statistic

$$Z = \frac{\sqrt{\pi}|p - P|}{\sqrt{PQ}} = Calculated\ value\ of\ Z$$

Where $Q = 1 - P$.

The test statistic **Z** follows a Normal distribution with Zero mean and unit variance. Let us denote Z_{Cal}as the calculated value of the Z and Z_{Tab} is the tabulated value of Z with a level of significance **α**.

Finding the Tabulated Value

The tabulated values of Z with different levels of significance **α** are as given in Table 4.2.

Decision

If $Z_{Cal} <$, Accept the null hypothesis $\mathbf{H_0}$ i.e. the sample is from the populationwith population proportion P and

if$Z_{Cal} \geq$, Reject $\mathbf{H_0}$i.e. there is a significant difference between the sampleproportion and population proportion.

Illustration – 4.

A local Tuberculosis center claims that 85% of all TB patientsare regularly continuing their treatment and get cured within two years after they are started taking medicines. Test this hypothesis at 5% level of significance if among 250 such TB patients in that area, 145 got cured within two years after they were started treatment?

Solution : Let ***p*** denotes the sample proportion of women got married within two years after they were hired and quit their jobs. So, *p = 145 / 250 = 0.58*

Setting of Null Hypothesis – The Null and alternative hypotheses are H_0 :*p = 0.85*and H_1 : *p ≠ 0.85.*

Calculation of Test Statistic -Test statistic

$$Z = \frac{\sqrt{n}|p - P|}{\sqrt{PQ}} = \frac{\sqrt{250}|0.58 - 0.85|}{\sqrt{0.85 \times 0.15}} = 33.48 = Calculated\ value\ of\ Z$$

Finding the Tabulated Value

The tabulated values of Z with 5% level of significance for two sided test is 1.96.

Decision

Here, the calculated value is greater than the tabulated value for 5% level of significance. So, we reject **H_0**i.e. there is a significant difference between the sample proportion and population proportion. Hence, the local Tuberculosis center's claim may not be true.

4.3.4 Test of Significance for Equality of Two Population Proportions

Let p_1 be the sample proportion considered from the population with proportion P_1 and p_2 be another sample proportion considered from the population with proportion P_2. Here, p_1 and p_2 are based on number of observation n_1 and n_2 respectively. For example, we are comparing the proportion of defective items produced by two machines is same or not.

Here we want to test equality of two population proportions.

Setting of Null Hypothesis –

The Null hypothesis H_0 :$P_1 = P_2$

The alternative hypothesis is any one from

(i) H_1 : $P_1 \neq P_2$ or (ii) H_1: $P_1 > P_2$ or (iii) H_1: $P_1 < P_2$

Calculation of Test Statistic -Test statistic

$$Z = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} - \frac{1}{n_2}\right)}} = Calculated\ value\ of\ Z$$

Where $P = \frac{n1p1+n2p2}{n_1+n_2}$ and $Q = 1 - P$.

The test statistic Z follows a Normal distribution with Zero mean and unit variance. Let us denote Z_{Cal} as the calculated value of the Z and Z_{Tab} is the tabulated value of Z with a level of significance **α**.

Finding the Tabulated Value: The tabulated values of Z with different levels of significance **α** are as given in the Table 4.2.

Decision

If $Z_{Cal}<$, Accept the null hypothesis H_0 i.e. both the samples are from the samepopulations with same population proportion.

Iif $Z_{Cal} \geq$, Reject H_0i.e. there is a significant difference between the twopopulation proportions. Hence, the samples are from different population proportions.

Illustration – 5.

In order to increase the activities of the Self Help Groups of rural areas, the Government introduced two types of schemes and these are to be implemented through two different NGOs. NGO – A took over the responsibility of producing, packaging and marketing of agricultural products and the other NGO – B took the responsibility of developing diary firms and marketing of dairy products through SHGs.They started individually awareness and training programs for different NGOs. Government believed that the scheme implemented byNGO – A is more effective than NGO – B. Two test areas with identical SHG characteristics are selected; NGO – A is used in one area and NGO – B is used in other area. In a random sample of 75 SHG who attended the training program by NGO – A, 20 started working. In another random sample of 90 SHG who attended the training program by NGO – B, 25 started working. Does this indicate that the scheme implemented byNGO – A is more effective than NGO – B, if 5% level of significance is used? Verify also at 10% level of significance.

Solution

Given that, n_1 = The number of SHG who attended the training program by NGO – A = 75 and n_2 = The number of SHG who attended the training program by NGO – B = 90. The number of SHG who started working by attending the training program by NGO – A is 20 and by NGO – B is 25.

So, sample proportions are $p_1 = \frac{20}{75} = 0.2667$ and $p_2 = \frac{25}{90} = 0.2778$

Setting of Null Hypothesis: The Null and alternative hypotheses are H_0 : $P_1 = P_2$ and H_1 : $P_1 > P_2$

This is a one tailed test.

Calculation of Test Statistic -

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{0.2667 \times 75 + 0.2778 \times 90}{75 + 90} = 0.273$$

$$Q = 1 - P = 1 - 0.273 = 0.727$$

Test statistic

$$Z = \frac{|p_1 - p_2|}{\sqrt{PQ\left(\frac{1}{n_1} - \frac{1}{n_2}\right)}} = \frac{|0.2667 - 0.2778|}{\sqrt{0.273 \times 0.727\left(\frac{1}{75} - \frac{1}{90}\right)}} = 0.529$$

$$= Calculated\ value\ of\ Z$$

Finding the Tabulated Value: The tabulated value of Z with 5% level of significance for one sided test is 1.645 and with 10% level of significance for one sided test is 1.28.

Decision: Here, the calculated value is less than the tabulated value for both 5% and 10% levels of significance. So, we accept H_0i.e. there is no significant difference between the effectiveness of both the schemes implemented.

4.4 Small Sample Tests

We know that for large sample $(n > 30)$ size, almost all distribution tends to Normal distribution. So, the entire large sample theories are based on the application of standard normal distribution. However, if the sample size is small, the distribution of various statistics cannot be approximated to Normal distribution. Usually, if sample size(n) is less than 30 we called it a small sample and the test is called as a small sample test. In this section, we discuss the small sample tests based on 't' distribution which is known as 't' test.

4.4.1 t −Test

There are various uses of t −test in small samples similar to Z − test in large samples i.e. for different tests of significance we use different 't' tests. Some uses of 't' tests are:

(i) To test the significant difference between the sample mean $(\bar{x})$ and the hypothetical value of the population mean (μ).

(ii) To test the significance difference between two sample means in case of two independent samples.

(iii) To test the significance difference between two sample means in case of two dependent samples.

(iv) To test the significance of an observed sample correlation coefficient.

4.4.1.1 To Test the Significant Difference Between the Sample Mean and the Hypothetical Value of The Population Mean . (Student'st −Test)

This test was originally due to *Sir William S. Gosset* of Ireland. But he published this in his pen-name '*Student*' which is popularly known as *Student's* $t-$*test*. This test is used to test if the sample mean $(\bar{x})$ differs significantly from the hypothetical value of population mean (μ_0) with following assumptions.

(i) Parent population is normal from sample is drawn.

(ii) Sample observations are independent of each other i.e. observations are randomly selected.

(iii) The population standard deviation (σ) is not known.

Consider a population with unknown population mean μ and an unknown variance. Let, μ_0 be the hypothetical value of the population mean. A random sample of size n is drawn from this population and $x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_n$ be the observed sample values.

Setting of Null Hypothesis: The null hypothesis is given by $H_0 : \mu = \mu_0$

against one of the alternative from the following three:

(i) $H_1: \mu \neq \mu_0$ or (ii) $H_1 : \mu < \mu_0$ or (iii) $H_1 : \mu > \mu_0$.

Here, we test to know whether the sample mean $(\bar{x})$ differs significantly from the population mean μor not.

Calculation of Test Statistic: The Test statistic is given by

$$t = \frac{\sqrt{(|\bar{x} - \mu_0|)}}{s}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean.

μ_0 is the hypothetical value of population mean.

s is an estimate of the population standard deviation,

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

n is the size of the sample.

The test statistic $t$follows a Student's $t-$ distribution with $(n-1)$ degrees of freedom.

Finding the Tabulated Value: Let, the calculated value of the test statistic is t_{Cal} and t_{Tab} be the tabulated value of t with $(n-1)$ degrees of freedom at levels of significance **α**. The tabulated value for t with different degrees of freedom at different levels of significance are given in the table entitled – "Critical values of t – distribution". Now, we can compare the calculated value t_{Cal} with the tabulated value t_{Tab} as follows and give a decision.

Decision: If $t_{Cal} <$, Accept the null hypothesis H_0 and if $t_{Cal} \geq t_{Tab}$, Reject H_0 at level of significance **α.**

Illustration – 5.

The following data gives the weight of 10 randomly selected pregnant females of a tribal community of Odisha during their 30 weeks of pregnancy. (in Kg. Weight):59.0, 57.8, 57.3, 57.0, 56.8, 57.1, 56.9, 58.2, 54.5, 56.5.

Test whether the mean weight of the females during 30 weeks of pregnancy may be taken as 57.5 Kg. wt.

Solution

Setting up the Null hypothesis

Let, μ = Mean weight of the females during 30 weeks of pregnancy

So, the null hypothesis is given by

H_0: μ = 575 Kg. Wt.

against the alternative H_1: μ ≠ 575 Kg. Wt.

Calculation of the Test Statistic

Here, the sample size = n = 10.

The sample mean = $\bar{x} = \frac{1}{n}\sum x$

$$= \frac{1}{10}[590 + 578 + 573 + 570 + 568 + 571 + 569 + 582 + 545 + 565] = 571.1$$

The sample mean square = $s^2 = \sum(x - \bar{x})^2$

$$= \frac{1}{10-1}[(590 - 571.1)^2 + (578 - 571.1)^2 + (573 - 571.1)^2 + (570 - 571.1)^2 + (568 - 571.1)^2 + (571 - 571.1)^2 + (569 - 571.1)^2 + (582 - 571.1)^2 + (545 - 571.1)^2 + (565 - 571.1)^2] = \frac{1}{9}[1260.9] = 140.1$$

$$\Rightarrow s = \sqrt{140.1} = 11.836$$

The test statistic is given by

$$t = \frac{\sqrt{n}|\bar{x} - \mu_0|}{s} = \frac{\sqrt{10}|571.1 - 575|}{11.836} = 1.042$$

$$= t_{Cal} = Calculated\ value\ of\ t$$

Finding the Tabulated Value

The tabulated value of t at 5% level of significance with 9 degrees of freedom is given by t_{Tab}= 2.26.

Decision

Here, the calculated value of t is less than the tabulated value. Hence, we accept the null hypothesis H_0. So, the mean weight of the females during 30 weeks of pregnancymay be taken as 575 Kg. Wt.

4.4.1.2 To Test the Significance Difference Between Two Sample Means in Case of Two Independent Samples

Consider a population with an unknown mean μ_1 with unknown standard deviation. Let a sample of size n_1 is drawn from this population and the sample mean is found to be $\bar{x}_1$.There is another population with an unknown mean μ_2 with unknown standard deviation. Let a sample of size n_2 is drawn from this population and the sample mean is found to be $\bar{x}_2$. It is summarized in the following table.

	Population - 1	**Population - 2**
Unknown Population Means	μ_1	μ_2
Sample Mean Square	s_1^2	s_2^2
Sample Size	n_1	n_2
Sample Means	$\bar{x}_1$	$\bar{x}_2$

Setting of Null Hypothesis

We set up the Null hypothesis H_0 : $\mu_1 = \mu_2$

against one of the alternative from the following three:

(i) $H_1 : \mu_1 \neq \mu_2$ or (ii) $H_1 : \mu_1 < \mu_2$ or (iii) $H_1 : \mu_1 > \mu_2$.

Calculation of the Test Statistic: The test statistic is given by

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where,

$$S^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)_2^2}{n_1 + n_2 - 2}$$

The test statistic t follows a Student's ***t***-distribution with $(n_1 + n_2 - 2)$ degrees of freedom. Let us denote t_{Cal}as the calculated value of the $t$andt_{Tab} is the tabulated value of $t$with $(n_1 + n_2 - 2)$ degrees of freedom at level of significance $\boldsymbol{\alpha}$.

Finding the Tabulated Value: The tabulated value for$t$ with different degrees of freedom at different levels of significance are given in the table entitled – "Critical values of t – distribution". Now, we can compare the calculated value t_{Cal} with the tabulated value t_{Tab} as follows and give a decision.

Decision

If $t_{Cal}<$, Accept the null hypothesis $\mathbf{H_0}$ and if$t_{Cal} \geq$

t_{Tab} , Reject $\mathbf{H_0}$ at level of significance **α.**

Illustration – 6.

Villagers in a village claim that, there are very high coconut trees in their village and in their neighbor village. They are facing a lot of problems as these are located nearer to their houses. Sometimes, coconuts are falling down on their houses and damaging the roofs. To verify the whether the average height of coconut trees of these two villages are considered same or not, two random samples were selected and the data is as follows.

	Village A	**Village B**
Number of coconut trees in the sample	15	20
Average height	1240 cm	1100 cm
Sample Mean Squares	950 cm	1130 cm

From the above information, can we conclude that there is a difference in average heights of coconut trees of two villages?

Solution

Setting of Null Hypothesis –

Let, μ_1 be the average height of coconut treesof Village A and μ_2 be the averageheight of coconut treesin Village B. The null hypothesis is given by

H_0: The average height of coconut trees is same. i.e., $H_0 : \mu_1 = \mu_2$

against the alternative $H_1 : \mu_1 \neq \mu_2$.

Calculation of Test Statistic: From the given data, we can write

	For Village A	**For Village B**
Number of coconut trees in the sample	$n_1 = 15$	$n_2 = 20$
Average Height	$\bar{x}_1 = 1240\ cm$	$\bar{x}_2 = 1100\ cm$
Standard Deviation	$s_1^2 = 950\ cm$	$s_2^2 = 1130\ cm$

$$S^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)_2^2}{n_1 + n_2 - 2} = \frac{(15 - 1)950 + (20 - 1)1130}{15 + 20 - 2} = 690.91$$

$$\Rightarrow S = \sqrt{690.91} = 26.285$$

The test statistic is given by

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{|1240 - 1100|}{26.285\sqrt{\frac{1}{15} + \frac{1}{20}}} = 15.594 = t_{Cal}$$

Finding the Tabulated value: Tabulated value of ***t*** at 5% level of significance for 15+20-2 = 33 degrees of freedom is t_{Tab} = 1.96.

Decision: So, the calculated value is greater than the tabulated value. Hence, we reject the null hypothesis at 5% level of significance. Therefore, we can conclude that that there is a difference in average height of coconut trees.

4.4.1.3 To Test The Significance Difference Between Two Sample Means in Case of Two Dependent Samples. (Observations Occur Pair-Wise) – Paired t– Test.

Sometimes there arise situations where samples are not independent. We get observation on each sample in two occasions i.e. before and after. From this it is evident that for each sample observation we get a pairs of observations. For example, the attendance of students in schools before and after the introduction of mid-day meal, the attendance of girl children in schools before and after an awareness program, to study the effectiveness of a drug for blood pressure we record the BP before and after the drug is taken, etc.

Let (x_1, y_1), (x_2, y_2), ... , (x_n, y_n) be n pairs of sample observation considering two occasions taken from a population. We consider following assumptions

(i) Parent Population from which the samples are drawn is Normal.

(ii) Sample observations are taken pair-wise.

(iii) Each pair of sample observation is independent to the other pair.

(iv) The population variances are not known.

Setting of Null Hypothesis: The Null hypothesis assumes that the mean population of difference values for each pair of observation is zero. If we define $d = x - y$ = the difference of values in a pair, then the population mean of this difference is $\mu_d = 0$. i.e.$H_0 : \mu_d = 0$

against one of the alternative from

(i) $H_1 : \mu_d \neq 0$ or (ii) $H_1 : \mu_d > 0$ or (iii) $H_1 : \mu_d < 0$

Calculation of Test Statistic: Here, the test statistic is given by

$$t = \frac{\sqrt{n}\bar{d}}{s_d}$$

where $\bar{d} = \frac{1}{n}\sum d$ and $s_d^2 = \frac{1}{n-1}\sum(d - \bar{d})^2$.

The test statistic ***t*** follows a Students ***t*** - distribution with $(n - 1)$ degrees of freedom.

Finding the Tabulated Value

Let, the calculated value of the test statistic is t_{Cal} and t_{Tab} be the tabulated value of t with $(n-1)$ degrees of freedom at levels of significance **α**. The tabulated values for t with different degrees of freedom at different levels of significance are given in the table entitled – "Critical values of t – distribution". Now, we can compare the calculated value t_{Cal} with the tabulated value t_{Tab} as follows and give a decision.

Decision

If $t_{Cal}<$, Accept the null hypothesis $\mathbf{H_0}$ i.e. there is no significant difference between the values before and after the occasions.

if$t_{Cal} \geq$, Reject $\mathbf{H_0}$ at level of significance **α.**

Illustration – 7.

A teacher claims that he has improved the standard of the students in Statistics within one year of his joining in a new school. In order to verify, whether the claim of the teacher is true or false, the school authorities collect the data on the percentage of students failed in the last annual examination (held before his joining) and this year annual examination (held after his joining) for the same classes. The percentage of failed students for 10 classes are given below.

Percentage of failed students

Before Joining: 12.25 11.97 12.15 12.08 12.31 12.28 11.94 11.8912.1612.04

After Joining :12.09 12.15 12.14 12.17 11.98 12.04 11.11 11.25 12.15 12.01

Can the authorities conclude that the teacher's perception is true?

Solution

Setting of Null Hypothesis –

If we define $d = x - y$ = the difference of percentage of failed students in a pair, then the population mean of this difference is $\mu_d = 0$. i.e. $H_0 : \mu_d = 0$,

against the alternative from $H_1 : \mu_d < 0$.

Calculation of Test Statistic -

X	***y***	***d = x - y***	$d - \bar{d}$ ***= d-0.207***	$(d-\bar{d})^2$
12.25	12.09	0.16	-0.038	0.001444
11.97	12.15	-0.18	-0.378	0.142884
12.15	12.14	0.01	-0.188	0.035344
12.08	12.17	-0.09	-0.288	0.082944
12.31	11.98	0.33	0.132	0.017424

12.28	12.04	0.24	0.042	0.001764
11.94	11.11	0.83	0.632	0.399424
11.89	11.25	0.64	0.442	0.195364
12.16	12.15	0.01	-0.188	0.035344
12.04	12.01	0.03	-0.168	0.028224
Total		$\sum d = 1.98$		$\sum(d - \bar{d})^2 = 0.94016$

$$Here, \bar{} = \frac{\sum d}{n} = \frac{1.98}{10} = 0.198$$

$$s_d^2 = \frac{1}{n-1}\sum(d - \bar{d})^2 = \frac{1}{10-1} \times 0.94016 = 0.1045$$

Here, the test statistic is given by

$$t = \frac{\sqrt{n}\bar{d}}{s_d} = \frac{\sqrt{10} \times 0.198}{\sqrt{0.1045}} = 1.937 = t_{Cal}$$

The above test statistic ***t***follows a Students ***t*** - distribution with $(n-1) = 10 - 1 = 9$ degrees of freedom.

Finding the Tabulated Value: The calculated value of the test statistic is $t_{Cal} = 1.937$ and $t_{Tab} = 2.26$ is the tabulated value of t with 9 degrees of freedom at levels of significance $\alpha = 0.05$.

Decision: Here, $t_{Cal}<$, so we accept the null hypothesis H_0 i.e. the teacher's perceptionis not true.

4.4.1.4 To Test The Significance of an Observed Sample Correlation Coefficient.(Fisher's t – Test)

This test is used to test population correlation coefficient (ρ (read as Rho)) is equal to zero or not with the following assumptions:

(i) The sample is based on n pairs of data drawn from a Bivariate Normal distribution.

(ii) Each pairs of observation are independent of each other pair.

Setting of Null hypothesis –

The null hypothesis is population correlation coefficient is zero i.e. $H_0 : \rho = 0$

against one of the alternative from(i) $H_1 : \rho$ G 0 or (ii) $H_1 : \rho > 0$ or (iii) $H_1 : \rho < 0$

The alternative hypothesis is chosen according to the need of the test.

Calculation of Test Statistic: Let (x_1, y_1), (x_2, y_2)... (x_n, y_n) are n pairs of observation from which the sample correlation coefficient 'r' is calculated. The test statistic is given by

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = Calculated\ value\ of\ t = t_{Cal}$$

which follows t −distribution with $(n-2)$ degrees of freedom.

Finding the Tabulated Value

Let, the calculated value of the test statistic is t_{Cal} and t_{Tab} be the tabulated value of t with $(n-2)$ degrees of freedom at levels of significance α.

Decision

If $t_{Cal}<$, Accept the null hypothesis H_0 i.e. there is no correlation between thetwo variables in the population i.e. two variables are uncorrelated.

If $t_{Cal}\geq$, Reject H_0 at level of significance α**.**

Illustration – 8.

A random sample of 27 students in a class shown that the correlation coefficient between their marks in Mathematics and Social Science subjects in the last annual examination is 0.6.Can we conclude that the two marks are correlated?

Solution

Setting of Null Hypothesis : The null hypothesis is population correlation coefficient between two marks is zero i.e. $H_0 : \rho = 0$ and the alternative hypothesis is $H_1: \rho$ G 0.

Calculation of Test Statistic: Given that the sample correlation coefficient between the speed and efficiency of waiters is *r* = 0.6. So, the test statistic is given by,

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0.6\sqrt{27-2}}{\sqrt{1-(0.6)^2}} = \frac{0.6\times 5}{\sqrt{1-0.36}} = 3.75 = Calculated\ value\ of\ t = t_{Cal}$$

Finding the Tabulated Value: The tabulated value of t at level of significance α = 0.05 for $(n-2) = 27 - 2 = 25$ degrees of freedom = 2.06 = t_{Tab}

Decision: Since t_{Cal} = 3.75 is greater than t_{Tab} = 2.06 the Null hypothesis is rejected i.e. we accept the alternative hypothesis i.e. $H_1: \rho$ G 0. Thus, we conclude that the two marks are not correlated in the population.

4.4.2 The Chi-Square Tests (^{2} − TESTS)

The test of chi-square is a widely used nonparametric test in social work researches. We pronounce χ^2 test as '*Kai square test'*. The chi-square test was developed by *Karl Pearson* in 1900. There are various types of chi-square tests are applied in various fields. But, we restrict ourselves into three common uses of chi-square tests.

(i) Test for population variance,

(ii) Test of goodness of fit, and

(iii) Test of independence of attributes.

4.4.2.1 Chi-Square Test for Population Variance

When we have to test that a random sample is drawn from a normal population with a specified variance or not i.e. we have test the null hypothesis that the population variance has a specified value or not, we can use this test.

Let $x_1, x_2, ..., x_n$ be a random sample of size $n$drawn from a population with unknown variance (σ^2).

Setting of Null hypothesis: The Null hypothesis states that the variance (σ^2) of the population from where this sample is drawn is equal to σ_0^2 (a specified value) i.e.$H_0 : \sigma^2 = \sigma_0^2$

against one of the alternative from

(i) $H_1: \sigma^2 \neq \sigma_0^2$ or (ii) $H_1: \sigma^2 < \sigma_0^2$ or (iii) $H_1: \sigma^2 > \sigma_0^2$

Alternative hypothesis is chosen according to the need of the test.

Calculation of Test Statistic: The test statistic is given by

$$3^2 = \frac{\sum(x-\bar{x})^2}{\sigma_0^2} = \frac{1}{\sigma_0^2}[\sum x^2 - \frac{(\sum x)^2}{n}] = \frac{ns^2}{\sigma_0^2}$$

$$= Calculated\ value\ of\ 3^2 = 3^2_{Cal}$$

Finding the Tabulated Value

Let, the tabulated value of 3^2 at α % level of significance with $(n-1)$ degrees of freedom is 3^2_{Tab} The tabulated values of 3^2 for different levels of significance can be obtained from the table – "Critical values of Chi – Square Distribution".

Decision

If $3^2_{Cal} \geq 3^2_{Tab}$ then reject the Null hypothesis H_0 i.e. the population variance significantly different from the specified value σ_0^2.

If $3^2_{Cal} < 3^2_{Tab}$ then accept the Null hypothesis i.e. the population variance may be equal to the specified value σ_0^2.

Note: (i) If sample size $n \leq 30$, then apply above test referring to the table value of χ^2.

(ii) If sample size is more than 30 ($n > 30$), then we use the Normal (Z) test i.e.by using (Fisher's approximation) $= \sqrt{23^2} - \sqrt{2n-1} \sim N(0,1)$. Here, decision will be taken by considering Normal table as in case of large sample test as in 4.2.

Illustration – 9.

A weighting machine manufacturer claims that the precision of the machines produced by his industry have a precision (which is measured by variance) of 0.10. In order to verify this claim, a sample of ten measurements was taken the same weighting machine with a fixed amount of goods and the weights are recorded as follows.

1.50 1.55 1.49 1.51 1.49 1.53 1.54 1.54 1.52 1.50

Carry out the test at 5% level of significance whether the belief of the owner is true or not.

Solution

Setting of Null hypothesis: We set up the Null hypothesis that the precision of the instrument be 0.15 i.e.$H_0 : \sigma^2 = 0.10(\sigma_0^2)$ against the alternative hypothesis $H_1 : \sigma^2 \neq 0.10$.

Calculation of Test Statistic

x	$x - \bar{x}$	$(x - \bar{x})^2$
1.50	-0.017	0.000289
1.55	0.033	0.001089
1.49	-0.027	0.000729
1.51	-0.007	4.9E-05
1.49	-0.027	0.000729
1.53	0.013	0.000169
1.54	0.023	0.000529
1.54	0.023	0.000529
1.52	0.003	9E-06
1.50	-0.017	0.000289
$\sum x$ $= 15.17$		$\sum(x - \bar{x})^2 = 0.00441$

Mean of the sample $\bar{x} = \frac{\sum x}{n} = \frac{15.17}{10} = 1.517$and $\Sigma(x - \bar{x})^2 = 0.00441$

The test statistic is given by

$$3^2 = \frac{\sum(x - \bar{x})^2}{\sigma_0^2} = \frac{0.00441}{0.10} = 0.0441$$

$$= Calculated\ value\ of\ 3^2 = 3^2_{Cal}$$

Decision : The tabulated value of 3^2 for $(n-1) = (10-1) = 9df$ at 5% level of significance is $3^2_{Tab} = 16.919$. So, $3^2_{Cal} < 3^2_{Tab}$ then accept the Null hypothesis i.e. the precession of the weighting machine may be equal to the specified value 0.10 (σ_0^2).

4.4.2.2 Chi-Square Test of Goodness of Fit

This test is used to test whether the set of observed values as the outcome of our experiment support the expected values, which based on any hypothesis or theory. By using chi-square test of goodness of fit we can know if the deviation between observed (experimental) value and expected (theoretical) values significantly different from each other or not. That is to say if the observed value do not differ significantly from the expected (Theoretical) values then we say the data to be good fit to the theory or hypothesis. For example, whether the chewing of tobacco is equally prevailing among males and females in a tribal community or not. To verify this belief, we can use a chi-square test for goodness of fit. Another example, suppose we want to verify that, male and females are equally represented in different societies of an area or not. To verify whether all classes of people are equally represented in different village committees of a block or not. So, this chi-square test has a wide range of applications in social work research.

Let there are *n* observed values i.e. $O_1, O_2, \ldots, O_n$ and for each observed value we obtain an expected value using a specific theory belief or any hypothesis. Let these are be E_1, E_2, ..., E_n respectively.

Setting of Null Hypothesis: We set up the Null hypothesis that there is no significant difference between observed (experimental) and expected (theoretical) values i.e. there is no discrepancy between the experimental value and expected value. i.e.H_0 :$O = E$against the alternative hypothesis H_1 : $O \neq E$(O = Observed value and E = Expected value)

Calculation of Test Statistic: The test statistic is given by

$$3^2 = \sum \frac{(O-E)^2}{E} = Calculated\ value\ of\ Chi-square = 3^2_{Cal}$$

Under the null hypothesis, this 3^2 follows a chi-square distribution with $(n-1)$ degrees of freedom.

Decision: Here the calculated value (3^2_{Cal}) will be compared with the tabulated value (3^2_{Tab}) with desired level of significance (α) with $(n-1)$degrees of freedomconsidered from the critical values for χ^2 distributiontable.

If $3^2_{Cal} \geq 3^2_{Tab}$, then we reject the Null hypothesis H_0, i.e. O ≠ E i.e. observed data do not fit to the expected values.

If $3^2_{Cal} < 3^2_{Tab}$, then accept the Null hypothesis H_0 i.e. O = E i.e. observed data is well fitted to the expected values experiment supports the theory.

Illustration – 10.

The following table gives the number of persons present in an awareness program about health and sanitationon various days of the week in a village. Find whether the attendanceof personsis uniformly distributed over week or not?

Days:	Sun	Mon	Tue	Weds	Thurs	Fri	Sat
No. of persons participated:	140	160	80	120	110	90	140

Solution

Setting of Null hypothesis –

We set up the Null hypothesis that $H_0 : O = E$ i.e., the observer data supports the hypothesis that the attendance of persons are uniformly distributed over the week.

The alternative hypothesis $H_1 : O \neq E$

Calculation of Test Statistic: Here the observed values are given and the expected values are calculated as under.

The total of observed values is 840. Since, it is uniform for all the days in a week, the expected values for the attendance of personswill be same for all the days. Thus, the expected value of attendance of persons for a day is 840 / 7 = 120. We now calculate the value of the test statistic from the following table.

Days	***O***	***E***	***O - E***	$(O-E)^2$	$(O-E)^2/E$
Sun	140	120	20	400	3.333
Mon	160	120	40	1600	13.333
Tue	80	120	- 40	1600	13.333
Wed	120	120	0	0	0
Thu	110	120	- 10	100	0.083
Fri	90	120	- 30	900	7.50
Sat	140	120	20	400	3.333
Total	**840**	**840**	**0**		**40.915**

The test statistic is given by

$$3^2 = \sum \frac{(O-E)^2}{E} = 40.915 = Calculated\ value\ of\ chi-square = 3^2_{Cal}$$

Under the null hypothesis H_0, this 3^2 follows a chi-square distribution with $(- 1) = (7 - 1) = 6$ degrees of freedom.

Decision: The tabulated value of chi-square for 6 degrees of freedomat 5% level of significance $= 12.592 = 3^2_{Tab}$. Since, $3^2_{Cal} > 3^2_{Tab}$ we reject the null hypothesis i.e. the attendance of personsis not uniformly distributed among the days of week.

4.4.2.3 Chi-Square Test of Independence

In many situations we have to verify whether two attributes are independent or not. For example, one may be interested to know whether any dependency exist between gender and salary received by the employees of a company or smoking habits and suffering from cancer, age groups and mobile phone usage, occupation and internet use, different geographical areas and selling of a product, etc. Here we may setup a null hypothesis that the two attributes are independent of each other and the alternative hypothesis is they are dependent. This type test is called test of independence of attributes which can be done by help of *chi-square*.

The table indicating the classification of the sample according to two different attributes is called contingency table or cross tabulation table.

Let us consider two attributes A and B. Suppose A divided into 'm' classes i.e. A_1, A_2, ...A_i, ... A_m and B divided into 'n' classes i.e. B_1, B_2, ..., B_j, ... B_n. The classification in which the attributes are divided into more than two classes is called manifold classification and expressed is called manifold classification and expressed in a $m \times n$contingency table as shown below.

		Attribute - B						
Attribute - A		**B_1**	**B_2**	**…..**	**B_j**	**…**	**B_n**	**Total**
	A_1	***N_{11}***	***N_{12}***	**…..**	***N_{1j}***	**…**	***N_{1n}***	***R_1***
	A_2	***N_{21}***	***N_{22}***	**…..**	***N_{2j}***	**…**	***N_{2n}***	***R_2***
	….	**….**	**….**	**…..**	**….**	**…**	**….**	**….**
	A_i	***N_{i1}***	***N_{i2}***	**…..**	***N_{ij}***	**…**	***N_{in}***	***R_i***
	….	**…..**	**…..**	**…..**	**…..**	**…**	**…..**	**…..**
	A_m	***N_{m1}***	***N_{m2}***	**…..**	***N_{mj}***	**…**	***N_{mn}***	***R_m***
	Total	***C_1***	***C_2***	**…..**	***C_j***	**…**	***C_n***	***N = Grand Total***

Where N_{ij}, is the number of units or individuals possessed by both A_i (i^{th} level of attribute A) and B_j (j^{th} level of attribute B). R_i is the total possessed by A_i (i.e. row total) and C_j is the total number of attribute possessed by B_j(i.e. column total). Further , N is the total frequency which is equal to total of all the row totals as well as equal to total of all the column totals.

Setting of Null hypothesis: So we set up the Null hypothesis as

H_0: The two attributes A and B are independent,

against the alternative H_1 : The two attributes A and B are dependent.

Calculation of Test Statistic: The test statistic is given by

$$3^2 = \sum \frac{(N_{ij} - (N_{ij}))^2}{(N_{ij})} = Calculated\ value\ of\ chi - square = 3^2_{Cal}$$

Under the null hypothesis H_0, this 3^2 follows a chi-square distribution with $(n-1)\times(m-1)$ degrees of freedom.

Where N_{ij} is the observed frequency of the occurrence of both attributes A_i and B_j. $E(N_{ij})$ is the expected frequency of the cell where N_{ij}, lies i.e. expected number possessed by both attribute $A_i$and B_j.

The calculation of $E(N_{ij})$ is done by the formula given below.

$$(N_{ij}) = \frac{R_i \times C_j}{N} = \frac{Row\ Total\ of\ the\ row\ containing\ N_{ij} \times Column\ total\ of\ the\ column\ containing\ N_{ij}}{Total\ frequency}$$

Decision

Here the calculated value (3^2_{Cal}) will be compared with the tabulated value (3^2_{Tab}) with desired level of significance (α) with $(n-1)\times(m-1)$degrees of freedomconsidered from the critical values for χ^2 distributiontable.

If $3^2_{Cal} \geq 3^2_{Tab}$, then we reject the Null hypothesis H_0, i.e. two attributes are not independent.

If $3^2_{Cal} < 3^2_{Tab}$, then accept the Null hypothesis H_0 i.e. the two attributes are independent.

Illustration – 11.

The following data gives the smoking habit and the suffering from cancer in a random sample of 100 persons from an area. Can we say that the smoking habit and suffering from cancer are dependent upon each other?

	Suffering from Cancer		Total
Smoking Habit	**Yes**	**No**	
Smokers	45	8	**53**
Non-Smokers	15	32	**47**
Total	**60**	**40**	**100**

Solution

Setting of Null hypothesis: Our null hypothesis is

H_0: The smoking habit and suffering from cancer are independent upon each other, against the alternative

H_1 :The smoking habit and suffering from cancer are dependent upon each other.

Calculation of Test Statistic: Before calculating test statistic, we calculate the expected frequencies as follows.

	Observed Values			Expected Values		
	Suffering from Cancer	Not Suffering from Cancer	Total	Suffering from Cancer	Not Suffering from Cancer	Total
Smokers	45	8	53	$\frac{53 \times 60}{100} = 32$	$\frac{53 \times 40}{100} = 21$	53
Non-Smokers	15	32	47	$\frac{47 \times 60}{100} = 28$	$\frac{47 \times 40}{100} = 19$	47
Total	60	40	100	60	40	100

The test statistic is calculated as follows:

Values	N_{ij}	$E(N_{ij})$	$N_{ij} - E(N_{ij})$	$(N_{ij} - E(N_{ij}))^2$	$\frac{(N_{ij} - E(N_{ij}))^2}{E(N_{ij})}$
1	45	32	13	169	5.281
2	8	21	−13	169	8.048
3	15	28	−13	169	6.036
4	32	19	13	169	8.895

Total	**100**	**100**	**0**		28.260

The test statistic is given by

$$3^2 = \sum \frac{(N_{ij} - (N_{ij}\ ))^2}{(N_{ij})} = 28.26 = Calculated\ value\ of\ chi - square = 3^2_{Cal}$$

Under the null hypothesis H_0, this 3^2 follows a chi-square distribution with $(n-1)\times(m-1) = (2-1)\times(2-1) = 1$ degree of freedom.

Decision

The tabulated value of chi-square for 1 degree of freedomat 5% level of significance = $3.84 = 3^2_{Tab}$

Since, $3^2_{Cal} > 3^2_{Tab}$ we reject the null hypothesis i.e. the two attributes smoking habit and suffering from cancer are dependent upon each other.

4.4.2.4 Conditions for the Validity of Chi-Square Test

i. The frequencies used in chi-square test must be absolute, not relative in terms.
ii. Each of the observations making up the sample for this test must be independent of each other.
iii. Total number of observations used in chi-square test must be large. It should be at least 50.
iv. The observation collected for this test must be on basis of random sampling.
v. No assumption is made regarding population distribution and no parameter of populat ion distribution is required for this test. This test is solely based on sample data.
vi. The expected frequency of any item or cell must not be less than 5. If it is less than 5, the frequencies taking from adjacent items or cells should be pooled together in order to make it 5 or more than 5. This done by Yate's Correction method.(Given below)

4.4.2.5 Yates Correction

For a 2 × 2 contingency table if one of the cell frequency in less than 5, then the validity of test becomes questionable. Yates has given a procedure, which applied 2 × 2 contingency table for validity of test, known as Yates's correction. Let 2 × 2 contingency table is given by

2 × 2 Contingency Table

a	b	a + b
c	d	c + d
a + c	b + d	N = a + b + c + d

Where *a*, *b*, *c,* and *d* are sample observations and N = *a* + *b* + *c* + *d*.

When either *a* or*d* is the smallest and less than 5, the following table is prepared after adding 0.5 to *a* and *d* and subtracting 0.5 from *b* and *c*.

a +0.5	b – 0.5	a + b
c – 0.5	d +0.5	c + d
a + c	b + d	N = a + b + c + d

Similarly, if either *c* or *d* is smallest, then we add 0.5 to *c* and *d* and subtract 0.5 to *a* and*d*. Hence the new table is

a – 0.5	b + 0.5	a + b
c + 0.5	d – 0.5	c + d
a + c	b + d	N = a + b + c + d

By doing so, the total values will remain same. Then the test is carried out as usual. The calculated chi-square without Yates's correction is given by (for 2 × 2 contingency table) and the calculated chi-square after Yates's correction is given by

$$3^2 = \frac{N\left[|ad - bc| - \frac{N}{2}\right]^2}{(a+c)(b+d)(a+b)(c+d)}$$

Remarks

(i) If total frequency N is large the use of Yates's correction will make very little difference in the value of χ^2.

(ii) However it is advised to apply Yates's correction to every 2 × 2 contingency table, even if all cell frequency is more than 5.

Illustration – 12.

Twenty college students were classified according to their intelligence and economic conditions as in the following table. Test whether there is any association between intelligence and economic condition?

		Intelligence		
Economic Condition		Good	Excellent	Total
	Good	5	7	12
	Not Good	2	6	8
	Total	7	13	20

Solution

Setting of Null hypothesis: We set up the Null hypothesis as

H_0: The two attributes intelligence and economic conditions are independent, against the alternative

H_1 : The two attributes intelligence and economic conditions are dependent.

Calculation of Test Statistic: Before calculating test statistic, we calculate the expected frequencies as follows.

	Observed Values			**Expected Values**		
	Good	Excellent	Total	Good	Excellent	Total
Good	5 (– 0.5)	7(+ 0.5)	12	$\frac{12 \times 7}{20} = 4.2$	$\frac{12 \times 13}{20} = 7.8$	12

Not Good	2 (+ 0.5) (Smallest)	6 (– 0.5)	8	$\frac{8 \times 7}{20} = 2.8$	$\frac{8 \times 13}{20} = 5.2$	8
Total	7	13	20	7	13	20

The test statistic is calculated as follows:

Values	N_{ij}	$E(N_{ij})$	$N_{ij} - E(N_{ij})$	$(N_{ij} - E(N_{ij}))^2$	$\frac{(N_{ij} - E(N_{ij}))^2}{E(N_{ij})}$
1	4.5	4.2	0.3	0.09	0.021
2	7.5	7.8	- 0.3	0.09	0.012
3	2.5	2.8	- 0.3	0.09	0.032
4	5.5	5.2	3	0.09	0.017
Total	**20**	**20**	**0**		**0.082**

The test statistic is given by

$$3^2 = \sum \frac{(N_{ij} - (N_{ij}))^2}{(N_{ij})} = 0.082 = Calculated\ value\ of\ chi - square = 3^2_{Cal}$$

Under the null hypothesis H_0, this 3^2 follows a chi-square distribution with($n - 1$)×($m - 1$) =(2 – 1)× (2 – 1) = 1 degree of freedom.

Decision: The tabulated value of chi-square for 1 degree of freedomat 5% level of significance = 3.84 = 3^2_{Tab}. Since, $3^2_{Cal} < 3^2_{Tab}$ we accept the null hypothesis i.e. the two attributes intelligence and economic conditions are independent upon each other.

5. Application of Computers in Social Work Research: SPSS

SPSS (Statistical Package for the Social Sciences) has now been in development for more than thirty years. Originally developed as a programming language for conducting statistical analysis, it has grown into a complex and powerful application with now uses both a graphical and a syntactical interface and provides dozens of functions for managing, analyzing, and presenting data.

There arefive stages of growth in the institutional history of SPSS:

1968-1975: SPSS becomes a product.

In 1970 : *McGraw-Hill published the first SPSS user's manual*

In 1971 : *SPSS became a small software company*

In 1975 : *SPSS incorporated and the two founders,* ***Norman H. Nie*** *and* ***C. Hadlai Hull***

1975-1984: SPSS becomes a corporation.

July 2009 :*IBM announcement its agreement to purchase SPSS for $1.2 billion.*

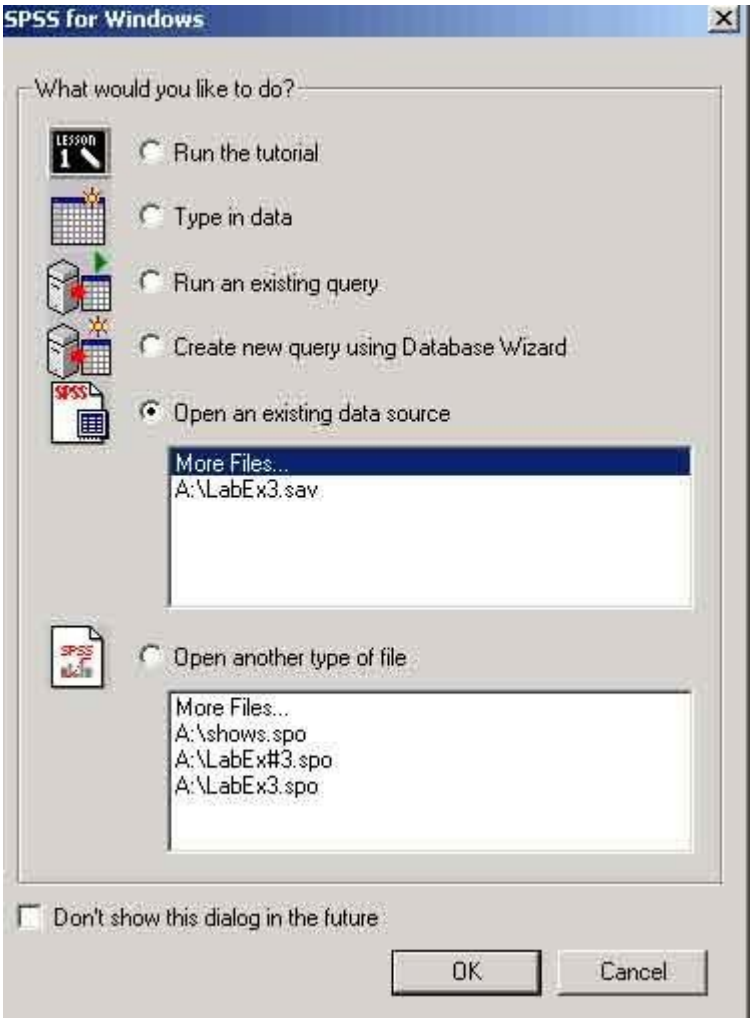
The main characteristics of SPSS are

- Comprehensive system for analyzing data
- Used to generate reports, charts, trendsand descriptive statistics.
- Makes statistical analysis.
- Statistical analysis becomes more convenient.
- No need for typing of command syntax

5.1 Basic Framework

Starting SPSS: **To start**

Windows **Start** menu ⟶Programs ⟶ SPSS for Windows⟶Get **SPSS** Data Editor window.



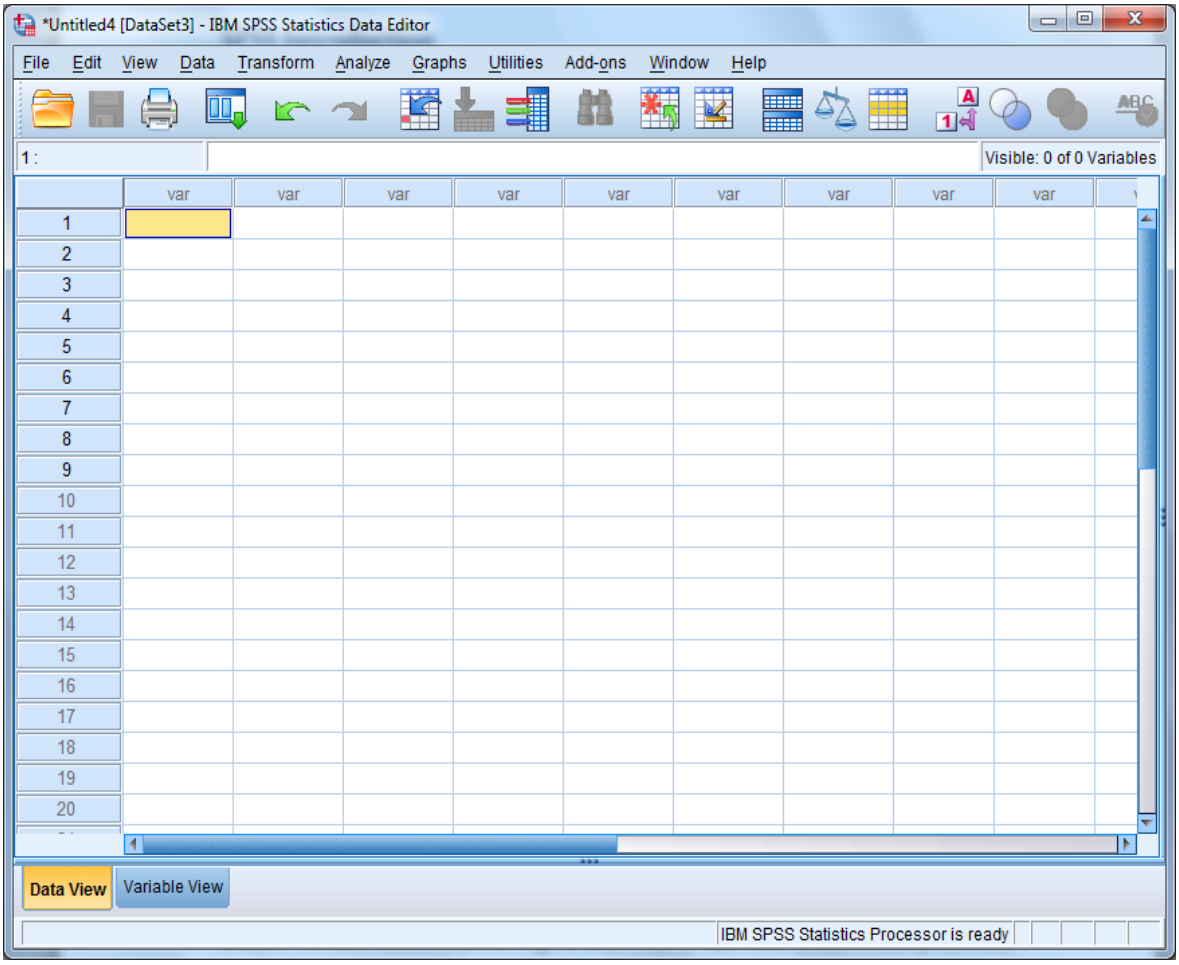
5.1.1 SPSS Data Editor MENU

SPSS is Menu driven. The main menu bar contains TEN menus.

SPSS MENU

1. File

- Create new files
- Opening existing files
- Read in files from other software programs
- Save files and print



2. Edit

- Modify or copy text from the output / syntax windows
- To search for and replace
- Other personal information

3. View

- To make status bar and tool bar active
- To change the characteristics of the window (change font style, size, displaying value labels)

4. Data

- Define variables
- Merging files, inserting, sorting and transposing variables and cases

- Selecting and weighting cases

5. Transform

- To create new variables from old variables

    1. By recoding

    2. By ranking cases

6. Analyze

- To make statistical analysis as required

7. Graphs

- To create different graphs

8. Utilities

- To display file and variable information
- To define and use different variable sets

9. Window

- To handle different attributes of various windows (arrange, select and control)
- To efficiently handle data, syntax, output and chart windows

10. Help

- To accesses different features of SPSS
- To accesses SPSS tutorial

5.2 Preparation of Data Files

The following steps are to be used while preparing a data file.

1. Naming the variable
2. Variable Levels
3. Value levels
4. Missing values

5. Variable type
6. Column Format
7. Measurement level

1. Naming the variable

- Must begin with a letter
- Cannot end with full stop (.) or underscore (_)
- Blanks and special characters cannot be used
- Within 64 characters (< 64 bytes)
- Cannot use reserved words (ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO, WITH, etc.)
- Names are not case sensitive

2. Variable labels

- Full description of the variable name.
- It is optional means of improving interpretation.

Variable Name	**Label**
Gender	Optional
Age	Optional
Allday	Desire for 24-hour shopping facilities
Area	Choice of shopping area
Cost	Amount spent on groceries per week

3. Value labels

- Data type is categorical or discrete type
- Used with alphanumeric codes

Variable Name	Label
Employee Id	Not applicable
Gender	1 = female 2 = male
Age	Not applicable
Allday	1 = would use 24 hour shopping 2 = would not use 24 hour shopping
Area	1 = nearby retail shops 2 = travel to nearer market 3 = travel to malls
cost	Not applicable

4. Missing values

- Missing value code is same as data type

5. Variable Type

- Default – Numeric type with two decimal places
- Other variable types – Date, currency, string
- Specify the digits after decimals

6. Column Format

- Adjust the column width
- Alignment within column (left, right, centre)

7. Measurement Level

- Nominal
- Ordinal

- Scale (Interval and Ratio)

5.3 Data Screening and Transformation

- **Necessity of data screening:**
  - Errors in data entry
  - Assessing normality
- **Necessity of variable transformation**
  - For normality

- **Necessity of Data transformation**
  - Collapsing continuous variables into categorical variables
  - Recoding negatively worded scale items
  - Replacing missing values

Errors in data entry

- Out of range values entered in entering attitude items. In order to detect out of range values in the data we can use frequency distribution technique as in the following steps.

`Analyze⟶Descriptive statistics⟶Frequencies⟶select the variables`into`Variable(s):`box

Assessing normality

- To access whether the data for a particular variable or variables follows a normal distribution or not, we use the following methods by SPSS.

Graphical Methods	**Statistical tests**
• Histogram	• Kolmogorov – Smirnov statistic and Shapiro-Wilk statistic

• Stem – and – leaf plot • Box plot • Normal probability plot • Detrended normal plot	• Skewness • Kurtosis

Assessing normality: For accessing normality of a variable we use the following steps.

- `Analyze` → `Descriptive statistics` → `Explore…`
- Select the variable to `Dependent list`
- Click `Plots…`
- Click on `Histogram, Normality plots with tests`
- Select `Factors levels` together in `Boxplots` display.
- Click on `Continue`
- `Display box`→ activate `Both`
- Click on `Options…` (opens `Explore: Options`)
- In `Missing values` box, select `Exclude cases pairwise`

Data Transformation: The following are the three procedures can be used in SPSS for transforming data.

• **Recode**	– To collapse a continuous variable – To recode negatively worded scale items – To replace missing values
• **Compute**	– To compute a new variable

	– To compute new variables based on certain conditions
• **Data Selection**	– To select first n number of cases in the data file for subsequent analysis

Recode - To collapse a continuous variable

- `Transform → Recode → Into Different Variables`
- Opens `Recode into Different Variables`dialogue box
- Select `Into Different Variables`
- `Variable List` (select the variable to recode into `Input Variable` -

`→Output Variable:` box

- `Name:` (Type new variable name) `→Change - Old and New Values…`
- Opens `Recode into Different Variables: Old and New Values`
- `Range→New Value→Add →Continue→OK`.

Data Selection

- `Data →Select Cases→`
- Select `box→If` condition is satisfied `→If…→SelectCases:If`
- Select the `variable`
- Type the `condition`
- `Continue → OK`

5.4 Descriptive Statistics

In order to find the descriptive statistics (Frequency Distribution – Histogram or Frequency Polygon – Measures of Central Tendency and dispersion) values using SPSS , we proceed with the following steps.

- `Analyze → Descriptive Statistics → Frequencies.`
- `Frequencies →Variable(s): box. → Display frequency tables→Statistics.`

- Frequencies: Statistics → select the statistics you require
- Continue → Charts… →Frequencies: Charts (select the charts you require)
- Continue →OK.

5.5 Correlation

The following measures of correlation can be obtained using SPSS.

- *Scatter Diagram*
- *Karl Pearson's correlation Coefficient* – Relationship between two continuous variables
- *Phi Coefficient* – Correlation between two dichotomous or categorical variables
- *Point-biserial Correlation* – Correlation between a continuous and a categorical variable
- *Rank Correlation* – Correlation between two attributes.

Assumptions

- Related Pairs – Observations must be from related pairs.
- Scale of measurement – Interval or Ratio
- Normality – Scores for each variable should be normal
- Linearity – Two variables must be linearly related. (Scatter Diagram)
- Homoscedasticity – Uniform clustering of values around the regression line (Scatter Diagram)

➢ **FOR GETTING A SCATTER DIAGRAM**

- Graphs → Scatter/dot…..→ Simple Scatter → Define→ opens the Simple Scatter sub-dialogue box.
- Select one variable and move the variable into Y-axis: box.
- Select the second variable and move the variable into the X-axis: box.
- OK.

➢ **TO FIND THE KARL-PEARSON'S CORRELATION COEFFICIENT**

- Analyze →Correlate →Bivariate… (open the Bivariate Correlations dialogue box)
- Select two variables to move into the Variables: box.
- Pearson from Correlation Coefficients box (Spearman's Correlation , Kendall's Tau-b)
- Test of Significance box →One-Tailed ratio button→Options
- Statistics pane, select Means and standard deviations→Continue → OK.

➢ **PARTIAL CORRELATION**

- Analyze→Correlate→Partial.
- Select two variables to the Variablespane.
- Select variables to the Controlling Forpane.
- Options (to select the statistics) →Continue→OK.

5.6 T-Tests

1. A One Sample t-test
2. An Independent Samples t-test
3. A Paired Samples (Repeated Measures) t-test

Assumptions

- Scale of Measurement – Interval or Ratio.
- Random Sampling – Values selected at random from the population.
- Normality – Values are normally distributed.

➢ **THE ONE-SAMPLE T-TEST**

- Analyze→Compare Means→One-Sample T Test…
- Select the variable you require into the Test Variable(s): box.
- In the Test Value: box type the mean score (μ_0) →OK.

- **PAIRED T – TEST (REPEATED MEASURE T – TEST)**
  - Analyze→Compare Means→Paired-Samples T Test…
  - Select the variables into Paired Variable: box →OK.
- **INDEPENDENT-SAMPLES T-TEST**
  - Analyze→Compare Means→Independent Samples T Test…
  - Select the test variables into the Test Variable(s): box.
  - Select the grouping variable to Grouping Variable: box.
  - Define Groups…
  - In Define Groups sub dialogue box, in the Group 1: box, type the lowest value of the grouping variable →TAB, enter the second value for the grouping variable in Group 2: box.
  - Continue→OK.

5.7 Analysisof Variance(ANOVA)

- **ONE-WAY BETWEEN GROUP ANOVA WITH POST-HOC COMPARISONS**
  - Analyze→Compare Means→One-Way ANOVA…
  - One-Way ANOVA dialogue box, select the dependent variable into the Dependent List: box.
  - Select the independent variable (Classes) into the Factor: box.
  - Options… (to open the One-Way ANOVA:Options)
  - Descriptive and Homogeneity-of-variance.
  - Continue and then Post Hoc… (to open One-WayANOVA: Post Hoc Multiple Comparisons sub-dialogue box) →(preferablyTukey) →Continue→OK.
- **TWO-WAY BETWEEN-GROUPS ANOVA**
  - Analyze→General Linear Model→Univariate…
  - Select the dependent variable into the Dependent Variable: box.
  - Select the independent variables into the Fixed Factor(s): box.
  - Options… (to open the Univariate: Options) →Descriptive statistics, Estimates of effect size, Observed power and Homogeneity tests→Continue →OK.
  - **To Plot The Cell Means**

– Graphs→Line… (to open Line Charts dialogue box) →Multiple box and ensure that the Summaries for groups of cases is selected →Define (to open Define Multiple Line: Summaries for Groups of Cases sub dialogue box)
– Line Represent box, select Other Statistic (e.g. mean).
– Select the dependent variable into the Variable: box.
– Select the independent variables with most levels into the Category Axis: box.
– Select the other independent variables into the Define Lines by: box. →OK.

5.8 Factor Analysis

➢ To Conduct a principal axis factor analysis

- Analyze→Data Reduction→Factor… to open Factor Analysis dialogue box.
- Select the variables required and click on  to move them to Variables: box.
- Click on Descriptives… to open Factor Analysis: Descriptives sub dialogue box.
- In the Statistics box, select Initial solution
- In the Correlation Matrix box, select the Coefficients, KMO and Bartlett's test of sphericityand Anti-image check boxes.
- Click on Continue.
- Click on Extraction… to open Factor Analysis: Extraction sub dialogue box.
- From Methods: drop-down menu, select Principal-axis Factoring.
- Ensure Correlation Matrix is selected in Analyze box.
- In the Extract box, ensure that the Eigenvalues over: is selected and *I*is displayed in the box.
- In the display box, ensure that the Unrotated factor solution and Scree plot check boxes are selected.
- Click on Continue→Rotation… to open Factor Analysis: Rotation sub dialogue box.
- In the Method box, ensure that the Rotated solution check box has been selected.
- Click on Continue.→Options… to open Factor Analysis: Options sub dialogue box.

- In the Coefficient Display Format box, select the Sorted by Size and Supress absolute values less Than: check boxes. Give a value here (may be 0.3)
- Click on Continue and then OK.

5.9 Cluster Analysis

- Analyze →Classify→Hierarchical Cluster
- Select the variables from the list on the left-hand side and transfer them to the box labeled Variablesby clicking on  .
- Click on Statistics to open Hierarchical Cluster Analysis: Statistics sub dialogue box.
- The main use of this dialog box is in specifying a set number of clusters. By default, SPSS will simply merge all cases into a single cluster and it is down to the researcher to inspect the output to determine substantive sub-clusters. However, if you have a hypothesis about how many clusters should emerge, then you can tell SPSS to create a set number of clusters, or to create a number of clusters within a range. For this example, leave the default options as they are and proceed back to the main dialog box by clicking Continue.
- Click on Method to open Hierarchical Cluster Analysis: Method sub dialogue box.
- Click on an appropriate Method (Choose Ward's method). In the below, under Transform Values→ selectZ - score.
- Click Continue
- In the main dialogue box, select Plots… and then click at Dendrogram.
- Click Continue
- Once back in the main dialog box, select the Savedialog box by clicking Save. This dialog box allows us to save a new variable into the data editor that contains a coding value representing membership to a cluster. As such, we can use this variable to tell us which cases fall into the same clusters. By default, SPSS does not create this variable. Suppose, we're expecting three so we could select Single solutionand then type 3 in the blank space. In reality, what we would normally do is to run the cluster analysis without selecting this option and then inspect the resulting dendrogram to establish how many substantive clusters lie within the data. Having done this, we could re-run the analysis, requesting that SPSS save coding values for the number of clusters that we identified.

6. Summary

The origin of the word '*STATISTICS*' is indicated by the word itself, which seems to have been derived from the Latin word '*STATUS*' or from Italian word '*STATISTA'* or may be from the German word '*STATISTIK'*.

In India, the evidences of data collection can be seen in the *Kautilya'sArthasastra* (Chandragupta Maurya'a regime) and also in *Ain-i-Akbari* (Akber's period).

Statistics is a subject that deals with the collection of data as well as the techniques of analysis and interpretation of data. The word 'Statistics' has been used in both *plural* and also in *singular* sense.

In *plural sense*, Statistics means a set of numerical figures or data. In *singular sense*, it represents a method for study and therefore, refers to statistical principles and methods developed for analysis and interpretation of data.

It is difficult to imagine a field of knowledge which can do without statistics. The scopes of statistics provide precision to various ideas and also suggest possible ways to handle a problem in different fields.

A variable can be a continuous one or a discrete one. A continuous variable can assume its value for any real quantity within a specified interval. A discrete variable can assume only some isolated values.

The number of times a value is repeated in the series is called as the frequency of that value. Frequency distribution of a variable is defined as the distribution of frequency over different values of the variable. When we write the frequencies against different values of a variable in the form of a table, it is called as a frequency table.

The *measures of central tendency* give us an idea about the central values whereas the *measures of dispersion* give an idea about the heterogeneity or scattering-ness present in the data.

Some common measures of central tendency are *arithmetic mean, median and mode*.The *arithmeticmean* or simply *mean* of a set of values is defined as *the ratio of sum of all the values to the total number of values of the set*. It is also called as the *average* and denoted as $\bar{x}$ (read as x bar).

The median of a series is defined as the value, which divides the series into two equal parts after arranging the values either in ascending or descending order.

The mode of a series is defined as the most frequent value in the series, i.e., the value which has maximum frequency in the series.

The measures of dispersion are the measures which give us an idea about the variability or scattering of values of a variable. Some common measures of dispersion are Range, Quartile Deviation, Standard Deviation and Coefficient of Variation.

Range is the simplest measure of dispersion, which is defined as the difference between the maximum and the minimum value of a series.

When there are n values in a series given by $x_1, x_2, \cdots, x_n$ then their standard deviation is the root mean square deviation of all the values, where the deviations are taken from the arithmetic mean of the series. It is denoted as σ (read as sigma).

Correlation is a statistical technique which is used to measure a relationship between two variables. The correlation between two variables can be studied by using the two methods:Scatter Diagram - Graphical Method and Karl Pearson's Method of Correlation Coefficient – Mathematical Method.

There are three types of correlation: Simple correlation; Multiple correlation and Partial correlation.

The test of significance enables a researcher to decide either to accept or reject the statistical hypothesis.A ***statistical hypothesis*** is a quantitative statement about a population *i.e.,* more specifically about a population parameter.

Null hypothesis is the hypothesis of no difference and it is tested for possible rejection under the assumption that it is true. It is usually denoted by H_0. Any hypothesis, which is complementary to null hypothesis is called *an alternative hypothesis*.It is usually denoted by H_1.

There is always some possibility of committing the following two types of errors in taking such a decision asType I Error (Reject the null hypothesis H_0 when it is true) and Type II Error (Accept the null hypothesis H_0 when it is false).The compliment of Type II error is called as the power of the test and is given by (1-β) and the size of Type I error (α) is also called as level of significance.

The standard deviation of sampling distribution of statistic is known as standard error (SE).Usually, when the sample size is more than 30, it is called as a large sample and the tests are called as large sample tests and if sample size(n) is less than 30 we called it a small sample and the test is called as a small sample test.

SPSS (Statistical Package for the Social Sciences) has now been in development for more than thirty years. Now it has grown into a complex and powerful application with now uses both a graphical and a syntactical interface and provides dozens of functions for managing, analyzing, and presenting data.

SPSS is Menu driven. The main menu bar contains tenmenus like File, Edit,View, Data, Transform, Analyze, Graphs, Utilities, Window and Help.

7. Key Words

Variable, Frequency distribution, measures of central tendency, measures of dispersion, Correlation, statistical hypothesis, population parameter, Null hypothesis, alternative hypothesis, power of the test, level of significance, standard error, large sample and small sample test, SPSS (Statistical Package for the Social Sciences).

8.Self-Assessment Questions

1. What do you mean by 'Statistics'? Discuss its development and use in social work research.
2. Discuss different definitions of Statistics. Outline the limitations of Statistics.
3. Discuss the scope and functions of Statistics.
4. A research without statistical planning is just like sailing a ship without knowing the coast. Justify.
5. Define a frequency distribution and explain how can you construct a frequency distribution with the help of an example?
6. The following data gives the number of children found in different households in a village. Construct an ungrouped frequency distribution for this data.

5	2	3	4	6	3	2	2	2	3
4	5	1	6	0	0	1	5	2	3
4	3	4	2	6	3	2	1	5	0
6	4	2	2	5	4	3	4	0	5
3	4	1	1	5	2	3	4	4	5

7. In a socio-economic survey in a rural area, it was found that there are some land less people living in these villages in this area and the following data shows the number of land less people in 30 villages in this area. Construct a grouped frequency distribution by taking the class intervals of size 3.

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

8. The expenditure of 1000 families is given as under:

Expenditure (× Rs.100):	40 – 59	60 – 79	80 – 99	100 – 119	120 – 139
No. of Families:	50	?	500	?	50

the mean and median of the distribution are both Rs. 87.50. Calculate the missing frequencies.

9. Define the measures of central tendency. Discuss different measures of central tendency with their relative merits and demerits.

10. Calculate the value of arithmetic mean for the data given in Question No. 7. Also find the value of arithmetic mean after constructing the frequency distribution. Comment on the results.

11. Calculate the value of median and mode for the data given in Question No. 7 after constructing the frequency distribution.

12. Define measures of dispersion. Discuss its importance. Discuss the measures of dispersion based on the quartiles.

13. The number of physically challenged children below 5 years of age recorded in seven Gram Panchayat offices is 2, 3, 4, 1, 0, 5, 3. Calculate the values of range, quartile deviation, standard deviation for the number of physically challenged children.

14. Calculate the value of different measures of dispersion for the data given in Question No. 7 after constructing the frequency distribution.

15. The following data gives the daily earnings of 250 shopkeepers in a village. Find the quartile deviation and coefficient of variation of daily earnings.

Daily Earnings (in Rupees)	Number of Shopkeepers
Less than 250	200
Less than 300	500
Less than 400	1250
Less than 500	3000

Less than 540	3750
More than 600	750
More than 700	250

16. Calculate the value of range and standard deviation for the marks of the Student B and Student C given in article 3 under *Measures of Dispersion* . Comment on the results. Compare these with those of Student A.

17. What is correlation and it s importance in social work research? Explain the graphical method of studying correlation.

18. The following data gives the age of the father and the age of their first child in years. Verify whether there exists any kind of statistical relationship between these two variables. Interpret the result.

Age of the Father :	42	57	39	50	47	38	44	60	34
Age of the first child :	50	25	6	15	12	2	6	28	1

19. Explain correlation coefficient. Discuss Karl Pearson's method of studying correlation.

20. Define correlation coefficient. Discuss it s mathematical properties. How can you interpret the value of a correlation coefficient?

21. What do you mean by the rank correlation? Discuss it importance. Explain the Spearman's rank correlation coefficient.

22. Explain with the help of examples, how can you interpret the value of rank correlation coefficient?

23. In a song competition the marks given by two judges out of 10 to seven participants. Can you conclude that there is the presence of any kind of agreements between the two judges?

Mark by Judge - I :	9	8	5	7	6	7.5	8.4
Mark by Judge – II :	8.5	9.2	7	7.5	5	8	9

24. What is test of significance? Discuss different steps of tests of significance.

25. Explain a statistical hypothesis and its importance in tests of significance.

26. Explain the following terms:

a. Critical Region
b. One sided and Two sided tests
c. Errors in testing of Hypothesis
d. Standard error
e. Level of significance
f. Large sample and small sample tests

27. The manufacturer of electric bulbs claims that the bulbs manufactured by their company shows an average life t ime of 2500 hours with a standard deviation of 230 hours. To verify the claim of the manufacturer is justified or not, a sample of 100 bulbs were taken and their average life time was found to be 2450 hours. Can we assume that the claim of the manufacturer is true? (Use %5 level of significance.)

28. Ten individuals are chosen at random from a normal population and their heights are found to be 63, 63, 67, 66, 72, 73, 68, 69, 65, 71 inches. Can we say that the mean height of the individuals of the population is 6 6 inches? (Use 5% level of significance.)

29. Two independent groups of 10 children were tested to find how many digits they could repeat from memory after hearing them. The results are as follows. Is the difference between the mean scores of the two groups significant?

Group A:	5	6	8	9	7	6	8	7	6	6	7	8
Group B:	8	5	6	7	5	7	8	5	9	6	7	7

30. The following data gives the number of persons in seven villages have the habit of chewing tobacco before and after an awareness program. Can you conclude that the awareness program was effective at 5% level of significance?

Before:	8	5	7	9	6	8	5
After:	5	6	7	6	7	6	6

31. A personnel manager claims that 75% of all single women hired for secretarial job get married and quit work within two years after they are hired. Test this hypothesis at 5% level of significance if among 200 such secretaries, 145 got married within two years after they were hired and quit their jobs?

32. In a large city A, 20% of a random sample of 900 school children had defective eye sight. In other large city B 15% of random sample of 1500 children has the same defect. Is there any significant difference between the eye sight defects between two cities?(Test at 95% level of significance.

33. A survey of 800 families with four children each revealed the following distribution. Is this result consistent with the fact that male anf female births are equally possible.

No. of Boys:	0	1	2	3	4
No. of Girls:	4	3	2	1	0
No. of Families:	32	178	290	236	64

34. The following data gives the average t ime (in minutes) of using mobile phones in a day among different age groups? Can we say that the mobile phone usage depends upon age groups? (Use 5% level of significance.)

	Mobile Phone Usage				
Age Groups	**Less than 1 Hour**	**1 Hour to 2 Hours**	**2 Hours to 3 Hours**	**3 Hours to 4 hours**	**4 Hours and more**
10 – 20	8	15	10	4	0
20 – 30	4	9	15	20	27
30 – 40	13	20	8	2	1
40 – 50	20	12	0	0	0
50 – 60	15	4	0	0	0
60 and above	18	25	0	0	0

35. What is SPSS? Discuss its development and its characteristics.

36. Explain SPSS data editor menu and functions of different menus.

37. Explain the preparation of data files using SPSS software.

38. Explain how can you detect the presence of error in data using SPSS?

39. What are the methods of testing for normality of data? Discuss how you can perform these tests in SPSS.

40. Discuss the steps for finding different descriptive statistics using SPSS software.

41. What are different types of correlation? Explain different steps involved in finding these correlation measures in a SPSS software.

42. What is a t – test? Explain different types of – tests performed using SPSS software.

43. Discuss the techniques of factor analysis and cluster analysis using SPSS software.

Further Readings

1. Aneshensel, C. S. (2002). Theory-Based Data Analysis for the Social Sciences. ThousandOaks,
2. Arrow, K. J. (1951). Social Choice and Individual Values. Wiley, New York.
3. Baron, R. M. and Kenny, D. A. (1986). "The Moderator-Mediator Variable Distinction inSocial Psychological Research: Conceptual, Strategic, and Statistical Considerations",CA: Sage Publications.
4. Creswell, J. W. (2006). Qualitative inquiry and research design: Choosing among five traditions (2nded.). Thousand Oaks, CA: Sage Publications.
5. Dash, P.R. (2011) Research Methodology with SPSS, (1sted.)Vrinda Publications (P) Ltd., New Delhi.
6. Dudewicz, E. J. and Mishra, S. N: Modern Mathematical Statistics. John Wiley.NewYork (International Student Edition).
7. Ferguson T. S. Mathematical Statistics.Academic Press.
8. Goon, A.M., Gupta, M.K. and Dasgupta, B: An Outline of Statistical Theory, Vol II., The World Press.
9. Howell, DC. (1997). Psychology for Statistics. Duxbury Press, Journal of Personality and Social Psychology, 51: 1173–82.
10. Kim, J. O., & Mueller, C. W. (1978).*Introduction to factor analysis: What it is and how to do it.* (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-013). Newbury Park, CA: Sage.
11. Leedy, P.D. (1993). Practical Research: Planning and Design (5th ed.). New York: MacMillan.
12. Lehmann, E. L: Testing Statistical Hypotheses. John Wiley.
13. Paul D. Leedy, Practical Research: Planning & Design, Prentice Hall
14. Pedhazur, E. J. and Schmelkin, L. P. (1991). Measurement, Design, and Analysis: An IntegratedApproach. Hillsdale, NJ: Lawrence Erlbaum.
15. Reinharz, S. (1992) Feminist Methods in Social Research. Oxford: Oxford UniversityPress.
16. Rubin, A., Babbie, E., & Lee, P.A. (2008). Research methods for social work: Custom edition preparedexclusively for San Jose State University. Belmont, CA: Wadsworth/Thomson Learning.
17. Sayer, A. (2000) Realism and Social Science. London: Sage. Zacks, S: Theory of Statistical Inference. John Wiley.