

# RESEARCH METHODS AND STATISTICS

V SEMESTER  
BA SOCIOLOGY  
COURSE  
(2011 ADMISSION)



**UNIVERSITY OF CALICUT**

**SCHOOL OF DISTANCE EDUCATION**

CALICUT UNIVERSITY P.O. MALAPPURAM, KERALA, INDIA - 673 635

285

**UNIVERSITY OF CALICUT**  
**SCHOOL OF DISTANCE EDUCATION**  
**BA SOCIOLOGY**  
**V SEMESTER**

**CORE COURSE**

**RESEARCH METHODS AND STATISTICS**

Prepared by:

Module I & II     Dr.Mahesh C.  
Asst. Professor,  
Department of Sociology,  
Z.G. College, Kozhikode

Module III & IV     Dr.Sara Neena T.T.  
Associate Professor,  
Department of Sociology  
Vimala College, Thrissur

Scrutinised by:     Dr.N.P.Hafiz Mohamad,  
'Manasam',  
Harithapuram  
Chevayur,  
Kozhikode-673 017

Lay out & Printing  
Computer Section, SDE

©  
Reserved

# **INDEX**

<b>MODULE</b>		<b>PAGE NO.</b>
<b>I</b>	<b>Statistical Techniques in Social Research</b>	<b>5</b>
<b>II</b>	<b>Sampling Techniques</b>	<b>16</b>
<b>III</b>	<b>Data Management And Presentation</b>	<b>27</b>
<b>IV</b>	<b>Report Writing</b>	<b>47</b>



# Module 1

## Statistical Techniques in Social Research

### 1.1. Statistics

#### Introduction

The subject- matter of statistics has to do a great deal with the science of statecraft. The very word 'statistics' is said to have been derived from, say, Latin 'status', Italian '*statista*', German '*statistik*' or French 'statistique' all referring to the political state. Obviously, only an organised and strong body of the state could venture into collection of statistics which in the past was mainly on population, its composition and wealth or poverty, as the case may be. Perhaps the earliest census of population was conducted in Egypt in connection with the raising of famous pyramids. Evidence of collection of population statistics in ancient India is found during the reign of Chandragupta Maurya. Land records were maintained by Todarmal during the reign of Akbar. Kautilya's 'Arthashastra' also provides statistical facts about State Administration in the country. Statistics in modern times is not a mere tool of State Administration; it has become a fact of day-to-day life.

'Statistics' is being used both as a singular noun and a plural noun. Initially, a distinction was drawn between these two uses. As a plural noun, it stood for data while as a singular noun, it represented a method of study based on analysis and interpretation of facts. But modern literature on the subject does away with any such distinction. Now 'statistics' can signify 'data' even when used as singular noun, in which case it would be treated as a group noun. The actual meaning of 'statistics' in each case shall be construed from the context in which it has been used.

Thus, the word 'statistics' may mean any one of the following:

- (i) Numerical statement of facts or simply data,
- (ii) Scientific methods to help analysis and interpretation of data,
- (iii) A measure based on sample observations.

But, only the first two of these, being more relevant to general purposes, are given greater prominence. The illustrations on the three possible uses are as follows:

- (i) Production statistics is compiled for judging the progress of a business firm (here statistics has been used for data).
- (ii) Statistics helps in simplification, analysis and presentation of data (here statistics has been used to represent statistical method).
- (iii) Statistics derived from a small representative group taken from the whole lot used for drawing inferences about the characteristics of the whole (here 'statistics' represents measures based on sample observations).

**Definitions**

Statistics has been defined either as a singular noun or as a plural noun in various ways by different authors.

**Statistics as Data (As a plural noun)**

Statistics as a plural noun stands for numerical figures collected. Some of the definitions of statistics describing it as quantitative facts are:

“Classified facts respecting the condition of the people in a State especially those facts which can be stated in numbers or in any other tabular or classified arrangements” - Webster

This definition emphasizes :

- (i) Facts, especially the numerical ones. the non-numerical facts obviously cannot be processed statistically.
- (ii) Facts which concern only the condition of the people in a State. Such an emphasis restricts the application of statistics to human sciences, but it is not unusual to make use in natural sciences as well.
- (iii) Classified and tabulated facts can only qualify to be included in statistics since a heterogeneous mass cannot be analysed and interpreted.

“... numerical statements of facts in any department of inquiry placed in relation to each other”. - Bowley

The chief features of the above definition are:

- (i) It emphasizes the numerical aspect of facts.
- (ii) It extends the application of statistics to any department of inquiry in human or the physical world.
- (iii) It emphasises the analytical aspect of study as against ‘classification and tabulation’ in the previous definition. It is necessary to bear in mind that no valid comparison is possible without proper arrangement of data and the use of statistical methods on such data.

“Statistics are measurements, enumerations or estimates of natural or social phenomena, usually systematically arranged, analysed and presented as to exhibit important inter-relationships among them.” - A.M. Tuttle

The above definition, though adequate and realistic, misses some of the chief characteristics of numerical statements as explained by the following.

“Aggregate of facts affected to a marked extent by multiplicity of causes, numerically expressed, enumerated or estimated according to a reasonable standard of accuracy, collected in a systematic manner, for a predetermined purpose and placed in relation to each other.” - Secrist

The above definition is of a comprehensive nature and deals with the various characteristics of statistics.

#### Characteristics of Statistics as Data

In order that numerical descriptions be called statistics they must possess the following

1. They must relate to the aggregate of facts. This means that a single fact, even though numerically stated, cannot be called statistics. For example, a solitary road accident cannot reveal anything unless such information is collected for a period and for different roads. Only then it can enable comparison over time or among places either on the basis of the aggregates, or on the basis of averages derived from the aggregate of facts.

It should then be, as stated by Wappaus, “aggregate of knowledge brought together for practical ends.” Another statistician W.I. King says. “The science of statistics is the method of judging collective national or social phenomenon from the results obtained by the analysis of an enumeration or collection of estimates.”

2. They are effected to a marked extent by a multiplicity of causes. As stated by A.L. Bowley, “Statistics is the science of measurement of social organism regarded as a whole in all its manifestations.” The business and economic phenomena are very complex. They are influenced by a large number of forces internal and external. Statistics helps in finding out the most potent and the most proximate cause out of the many that affect a business phenomenon. For example, the phenomenon of falling sales in a business firm may be caused by a general recession in business, lack of sales promotion effort, appearance of strong competitive forces, etc. If all the relevant facts are collected and analysed it is possible to determine the factors responsible for the decline in sales. Similarly, condition of a crop is affected by a number of factors like soil condition, use of fertilizers, rainfall, methods of cultivation, etc.

3. They are numerically expressed. In the words of Prof. H. Secrist, “The statistical approach to a subject is numerical. Things, attributes and conditions are counted, totaled, divided, subdivided and analysed.” This means that any facts, to be called statistics, must be numerically or quantitatively expressed.

Qualitative characteristics or attributes such as intelligence, beauty, etc., cannot be included in statistics unless they are quantified by assigning certain score as a quantitative measure of assessment. The intelligence quotient designed by Dr. Freud could be accepted as statistics of attributes. Similar techniques may be used to judge statistically various matters connected with the selection of personnel and to assess their ability.

4. They should be enumerated or estimated. The data may be obtained by counting or measurement or it may be estimated statistically when enumeration is not feasible or involves inordinate and high costs. It may be estimated by statistical methods. For example, the general quality of a product is estimated by experimental tests on small samples drawn from a big lot.

5. They should be collected with reasonable standard of accuracy. Data is collected only with a reasonable standard of accuracy. A high degree of accuracy as observed in accountancy or mathematics is not insisted upon in statistics, because first, a mass of data is involved and, secondly, the process of generalization can be achieved with a reasonable standard of accuracy only. However, certain rules are prescribed for rounding-up, etc, to achieve a fair degree of accuracy. There is also the consideration of having data in a suitable form for easy statistical treatment. Statistics is one of those sciences which are only indicative of a trend, it therefore is probabilistic rather than deterministic.... purposes to refer to the likelihood of certain events under certain given conditions and this purpose being only indicative and not conclusive only a reasonable standard of accuracy is prescribed.

6. They should be collected in a systematic manner. In the words of Secrist, “Stray and loose bits of quantitative information, hearsay and unrelated material gleaned here and there, from indiscriminate sources, having no common basis of selection, even when numerical cannot be termed as statistics.” The collection of data must be carefully and systematically done. A haphazard collection of figures may lead to erroneous conclusions. A good deal of caution is necessary in the collection of data. The statistical processing, however nicely done, would not yield the demonstrable or the desired results if the things are not systematically processed. It has been indicated later in this chapter how fallacious results are arrived at when information is collected by the use of improper methods.

7. They must be relevant to the purpose. The scope of purpose of inquiry should be stated clearly before actually conducting the inquiry. The definition of various terms, units of collection and measurement also help in ensuring that the data is relevant to the purpose in hand. For example, data on the physical personality will be irrelevant for considering ability of a person for an intellectual work. But surely, it will be relevant for selection into military service.

8. They should be placed in relation to each other. The main purpose of collection of facts and figures is to facilitate comparative study. In other words, statistics should be capable of comparison. For example, the statistics on yield of crop and condition of soil are related but these yields cannot have any relation with the statistics on the health of the people.

### **Statistics as a Method (As a singular noun)**

The word statistics as a singular noun stands for a body of methods known as statistical methods. Statistics is a method of obtaining and analysing numerical facts and figures in order to arrive at some decisions.

Statistics, in this context, has been defined as a science which provides tools for analysis and interpretation. These methods are applied on the data collected for the purpose of decision making in various fields of scientific inquiry. According to one definition, it is “the science which deals with the collection, classification and tabulation of numerical facts as a basis for explanation, description and comparison amongst phenomena.”



The above definition lists various activities for which statistics provides its tools. It emphasises more the descriptive character of statistics rather than the analytical or the inferential aspect of its tools. But, it is the latter type of roles that are more important in business decision making.

Another definition states “Statistics is the science and art of handling aggregate of facts-observing, enumeration, recording, classifying and otherwise systematically treating them.”

In the above definitions, statistics is considered both as a science and as an art because it provides not only the tools of analysis but also the precepts to determine the functioning of those tools thus facilitating the accomplishment of ultimate objectives of the inquiry.

Statistics can, more appropriately, be characterised as an applied science which helps in drawing inference about unknown events and relation on the basis of systematic analysis of the past experience. Even when statistics is not a body of substantive knowledge, it is at least a body of methods helping in the acquisition of scientific knowledge.

This is adequately corroborated by the following:

“..... Statistical methods include all those devices of analysis and synthesis by means of which statistics (data) are scientifically collected and used to explain or describe phenomena either in their individual or related capacities”. - Secrist

There are many more definitions.

“Statistics is the branch of scientific method which deals with the data obtained by counting or measuring the properties of population of natural phenomena.”

In this definition, ‘natural phenomena’ embraces all the occurrences of the external world, whether human or non-human.

Statistics is the science which deals with the methods of collecting, classifying, presenting, comparing and interpreting numerical data collected to throw some light on any sphere of inquiry.” - Seligman

“Statistics refers to the body of technique or methodology which has been developed for the collection, presentation and analysis of quantitative data and for the use of such data in decision making.” - Netter & Wassermass

“Modern statistics refers to a body of methods and principles that have been developed to handle the collection, description, summarization and analysis of numerical data Its primary objective is to assist the researcher in making decisions or generalizations about the nature and characteristics of all the potential observations under consideration of which the collected data form only a small part” - Lincon L. Chao

The following definitions emphasise in particular different tools of analysis:

“Statistics may be called the science of counting.” - Bowley

“Statistics may rightly be called the science of averages.” - Bowley

“Statistics is the science of estimates and probabilities.” - Boddington

Croxton and Cowden have given a simple yet comprehensive definition of statistics. According to them. “Statistics may be defined as the collection, presentation, analysis and interpretation of numerical data.” This definition brings into light the various phases in a statistical investigation.

From a logical analysis of the above definitions, it is clear that the methodology of statistics given by Croxton and Cowden is the most scientific and realistic one. It is this which will form the subject-matter of our study.

Statistics as data is like raw material to be processed through the use of appropriate statistical methods. It is only when this is subjected to such processing that the data can serve the desired purpose of analysis and interpretation of quantitative facts. The choice of methods will depend upon the nature of the data as well as the purpose for which it is collected. At the same time, the collection of facts too is under taken with an eye on the methods to be employed. There are statistical methods to guide us at every stage of statistical investigation, right from the planning stage to the stage of the drawing of final conclusions.

Without statistical treatment the raw data is not useful for decision making. On the other hand, unless suitable data is available even highly sophisticated statistical techniques will not yield reliable results. It is rightly said. “..no inference is better than the quality of data on which it is based.” Therefore, the data and the methods of study are complementary; the quality of analysis will depend on a suitable integration of the data for a given purpose. However, as a discipline of study the emphasis has to be on methods.

Some distinctions between the two are given below:

Statistics as data	Statistics as a method
1. It is quantitative	It is an operational technique
2. It is often in the raw state.	It helps in processing the raw
3. It is descriptive in nature	It is basically a tool of analysis data
4. It provides material for processing data and helps in scientific methods and interpretation.	The Processing is done by the scientific unprocessed methods of analysis and not help in methods of analysis

**B: Functions of statistics.**

In the modern world statistical methods are of universal applicability. As a matter of fact there are millions of people all over the world who have not heard even a word about Statistics and yet who make use of statistical methods in their day to day decisions. Examples are many to show that human behaviour and statistical methods have much in common. In fact, statistical methods are so closely connected with human actions and behaviour that practically all human activity can be explained by statistical methods.

1. **Simplifies complexity:** Human mind is not capable of assimilating huge facts and figures. Statistical methods make these large number of facts easily intelligible and readily understandable. Statistics reduces a large mass of facts to a simple and single number like average. Classification and analysis are some of the methods of Statistics which simplify the complex nature of the data. Graphs and diagrams present unwieldy and complicated facts in the shape of attractive picture and diagrams. A layman can understand the significance of the data of seeing such diagrams and graphs.
2. **It presents facts in a definite and precise form:** Statement of facts conveyed in exact quantitative terms are always more convincing than vague utterances. Statistics presents facts numerically and thus gives definite and precise form to the data. this helps proper comprehension of what is stated.
3. **It helps condensation:** Statistics helps in condensing mass of data into a few significant figures. Average provides a single significant figure summarizing the entire data. Percapita income, for example, is the essence of the individual incomes.
4. **Statistics enables comparison of data:** Unless figures are compared with others of the same kind, they are of no use. Comparison is one of the main functions of Statistics. When we say that the price of a commodity has increased very much, the statement does not make the position very clear. But when we say that last year price was Rs. 10 but now it is Rs. 11 the comparison becomes easy. Statistics provides a number of suitable methods of comparison like ratios, percentages averages etc.
5. **It helps in formulating and testing hypothesis:** Statistical methods are extensively useful in formulating and testing hypothesis. they also help to develop new theories. Statistical test can be applied to test the hypothesis like whether a coin is unbiased, or whether vaccination is effective in preventing small pox, or whether advertisement has increased the sales volume etc. t-test, z-test, X<sup>2</sup>-test etc., are various tests of hypothesis commonly applied.
6. **It helps prediction (or forecasting):** Plans and policies of organizations are formulated well in advance of the time of their implementation. Knowledge of future trends is always necessary in framing policies and plans. Statistical methods provide helpful means of forecasting future events.
7. **It helps to forecasting:** Statistics can be used to forecast the future on the basis of past and present information.

### **Scope of Statistics:**

In the early period of development of Statistics, it had only limited scope. But in modern times, the scope of statistics has become as wide as to include in its fold all quantitative studies and analysis relating to any department of enquiry. It is used in all spheres of life such as social, economic and political fields.

In the field of economics it is almost impossible to find a problem which does not require an extensive use of statistical data. Modern age is an age of planning. Plans of economic development are constricted on the basis of statistical data.

Bankers, brokers and Insurance companies make use of statistical data for the study of business cycles, future trends, mortality rate etc..

Statistics comes to the field of Business management also. Statistics are very helpful to state as they help it in administration. Modern statistical data are being found increasingly useful in research in different fields.

#### 1. Importance of statistics in the field of commerce and business.

Statistics is an aid to business and commerce. It said that a business runs on estimates and probabilities. Statistics help business to forecast its expectations come to true. Modern statistical devices help business forecasting more precise and accurate.

Business men need statistics right from his business. He should have relevant facts and figures to prepare the financial plan of the proposed business.

#### **2. Uses of Statistics in Business Management**

One of the important functions of business management is to coordinate the activities of the various departments so as to secure maximum efficiency with minimum effort. To discharge this function efficiently, management should have adequate statistical data. Another function of management is marking wise decisions in the face of uncertainty. Modern Statistics has developed certain general principles and devices to deal with uncertainty wisely. Modern statistical tools of collection, classification, tabulation, analysis and interpretation of data, have been found to be an important aid in making wise decisions at various levels of managerial function.

The success of production programming both in the short as well as long period depends to a great extent on the quality of sales forecasts. Statistical methods can be applied to have better sales forecasts. Effective control on sales can also be exercised through regional allocations. Market research consumer preference studies etc. are some other methods of sales control, which make an extensive use of statistical tools.

#### **3. Importance of Statistics in Economics**

The Science of Economics is becoming statistical in its method. Statistical data are extensively used in all economic problems. As economic theory advances, uses of statistical methods also increase. The laws of Economics like Law of demand, Law of supply etc. can be considered true and established with the help of statistical methods. Statistics of consumption tell us about the relative strength of the desire of a section of people. Statistics of production describe the wealth of a nation. Distribution Statistics disclose the economic conditions of the various classes of people. Therefore statistical methods are necessary for proving economic laws. Marshall observes that “statistics are the straw out of which I like every other economist to make bricks” Bowley rightly observes that no student of Economics can pretend of complete equipment unless he is a master of the methods of Statistics.

Index number like Cost of living index number, Whole sale price index number etc. and analysis of time series are important statistical concepts in economic theory. The planning is based on the correct analysis of statistical data. Our five year plan is statistical methods. They are extensively used in Economics. Thus one can say that without Statistics many theories of Economics would have remained closed to mankind.

#### **4. Uses of Statistics in other fields**

**To the State (or Government):** Statistics are the eyes of government administration. Government has since long collected and interpreted data concerning the state. In fact, the word Statistics is originally derived from state. Conceptions of welfare state and increase in the duties and functions of state are reasons for the increase in the importance of Statistics in Government administration. For efficient administration, statistics are essential tools. The government have to collect statistical data whenever they adopt measures to achieve the objectives like reducing inequalities in the distribution of wealth, income etc. The Government administration is run through budgets which are formulated on the basis of statistics. The commissions and committees appointed by the Governments base their reports on statistics. A state, besides being an administrative body is a big commercial concern also. So it needs statistics to carry on these business works. Thus it can be said that statistics is the point around which government activities cluster.

**To Research:** Modern statistical methods and statistical data are being found increasingly useful in research in different fields. In the field of science, in the literary field, in the field of business activities, in the field of economic activities, research works are being undertaken with the help of statistical methods.

**In Planning:** Modern age is an age of planning. For the planning to be successful, statistical data are necessary. So planning cannot be managed without statistics. National sample survey scheme was primarily started to collect statistical data.

**For Bankers, brokers and Insurance Companies:** Statistical methods help various economic entities. A banker has to make statistical study of business cycles. The stock exchange brokers speculators and investors etc. have to rely on statistical data for precise forecasting. The Success of insurance company depends on the accurate basic data that it uses for the calculation of premium rates etc.

#### **Limitation of Statistics**

In spite of the fact that statistical methods have been universally applied to an increasingly larger number of fields, they suffer from certain limitations, which restrict their scope and utility.

Statistics, a branch of applied Mathematics, is regarded as mathematics applied to observational data. Conceivably everything dealing with the collection, processing, analysis and interpretation of numerical data belongs to the domain of statistics. However, the term 'statistics' is used in several ways. It denotes a compilation of data such as those found in the labour Gazette or say, Labour Statistics of the Labour Bureau published annually by the Government of India. The second meaning of the term statistics refers to the statistical principles and methods employed in the collection, processing, analysis and interpretation of any kind of data. In this sense, it is a branch of applied mathematics and helps us to know the complex social phenomena in a better way and lends precision to our ideas.

## **1.2 (A) Use of Statistical Methods in Social Research**

Statistics has patently two broad functions. The first of these functions is description and the summarizing of information in a manner so as to make it more usable.

The second function of statistics is induction, which involves either making generalizations about some 'population' on the basis of a sample drawn from this population or formulating general laws on the basis of repeated observations. The two functions of statistical methods can be easily understood by the following example. Suppose it is desired to study the problem of labour unrest in a particular area. The first thing to be done here will be to analyse the various causes of labour unrest and to study the impact of each one of these on the various categories of labour, viz., male workers and female workers or skilled labour and unskilled labour. This kind of analysis will give us an insight into the problem and we may be able to know from such an analysis many important things, e.g., that the involvement of male workers in strikes is much higher than that of the female workers or that the labour unrest in big industries is much higher than in small industries. Such an analysis may lead us to the conclusion regarding the incidence of labour unrest in the country and factors responsible for it. The former example illustrates the process of descriptive statistics whereas the latter, that of inductive statistics.

It is evident that knowledge of basic statistical concept and techniques is necessary for an intelligent understanding of the generality of life. Out of the welter of single events, social researchers seek general trends; out of the vast and confusing variety of individual characters; they continually search for the underlying group characteristics.

There are essentially two reasons why the expertise in statistics and the need to study statistics have grown enormously in the field of social sciences. One reason is that the huge amount of data collected by researchers needs simplification so as to render them capable of being commonly understood without much difficulty. The second and even more important reason is the increasing quantitative approach being currently employed in social science research.

Seemingly statistical considerations enter only at the analysis stage of the research process after the data have been collected, and near to the point in time when the initial plans for analysis are formulated and a sample is to be drawn. But this does not imply that a social researcher can plan and carry out his entire research without any knowledge of statistics and then hand over the data to the statistician for analysis. If a researcher lacked conversance in statistics the results of a costly research project would probably be disappointing if not unless.

Indeed, the problems that will be encountered in analysis and interpretation have to be anticipated at every stage in the research process and in this sense, statistical methods are involved throughout. This implies that statistics is a very useful tool for the social scientist. It is a much more useful tool for exploratory analyses than might possibly be imagined. Most social researches are based on highly tentative theoretical ideas. The variables that need to be controlled in the analysis or even the priorities and sequence of analysis-steps that should be followed are neither precise nor predetermined, researchers are generally awed by the complexity of data analysis no sooner a set of variables is introduced. In these circumstances especially, knowledge of the statistical methods becomes an invaluable tool for the social researcher in disentangling highly complex interrelationships.

The following are some important limitations of statistics:

1. Statistics does not study individuals. It deals with an aggregate of facts and ones not give any specific recognition to the individual items of a series. Individual items, taken separately, do not constitute statistical data and are meaningless for any statistical inquiry. For example, the individual figures of agricultural production, industrial output or national income of any country for a particular year are meaningless, unless these figures enable comparisons with similar figures for other countries and in the same country these are given for a number of years. W.I. King observes: “Statistics from their very nature of subject cannot and will never be able to take into account individual cases. When these are never be able to take into account individual cases. When these are important, other means must be used for their study.” Hence, statistical analysis is only for those problems where group characteristics are to be studied.
2. Statistics does not study qualitative phenomena. Being a science dealing with a set of numerical data, it is applicable only to the quantitative aspect of a problem. As such, qualitative phenomena like honesty, poverty, wisdom, etc., which cannot be expressed numerically, are not capable of direct statistical analysis. However, statistical techniques may be applied indirectly by first reducing the qualitative expressions into some quantitative terms. For example, the intelligence of a group of candidates can be studied on the basis of the scores assigned for various qualitative characteristics.
3. Statistical results are true only on an average. W.I. King writes, “Statistics largely deals with averages and these averages may be made up of individual items radically different from each other.” We known that statistical results reveal the average behaviour, the normal or the general trend; they are, therefore, useful for a general appraisal of a phenomenon and not for substitutions for any specific unit or event. Sometimes the average or trend indicated by statistics is applied to individual cases which are not proper. This may lead to what is called the error of false deduction. For example, the per capita income figure of Indians cannot be used for forming an idea about the income of an individual or a group without knowing the dispersion of income which would show the degree of variability in incomes of the units comprising the group.
4. Statistical laws are not exact. Unlike the laws of physical and natural sciences, statistical laws are only approximations and not exact. On the basis of statistical analysis of the problem, we can talk only in terms of probability and not certainty.
5. Statistics does not reveal the entire story. It only simplifies and helps the analysis of certain qualitative facts. But the real background of the data may not be reflected through these facts.
6. Statistics is liable to be misused. Perhaps the most important limitation of statistics is that it must be used by experts. As the saying goes “statistics one of the dangerous tool in the hands of the in experts”. Thus the use of statistics by the inexperienced and untrained person might lead to very fallacious conclusions.

## **Module II**

# **SAMPLING TECHNIQUES**

### **2.1. Introduction**

When secondary data are not available for the problem under study, a decision may be taken to collect primary data. The required information may be obtained by following either the census method or the sample method.

#### **(A) Census and Sample method**

Under the census or complete enumeration survey method, data are collected for each and every unit (person, household, field, shop, factory, etc., as the case may be of the population which is the complete set of items which are of interest in any particular situation. For example, if the average wage of workers working in sugar industry in India is to be calculated, then wage figures would be obtained from each and every worker working in the sugar industry and by dividing the total wages which all these workers receive by the number of workers working in sugar industry, we would get the figure of average wage. Some of the merits of the census method are:

- (i) Data are obtained from each and every unit of the population.
- (ii) The results obtained are likely to be more representative, accurate and reliable.
- (iii) It is an appropriate method of obtaining information on rare events such as areas under some crops and yield thereof, the number of persons of certain age groups, their distribution by sex, educational level of people, etc. This is the reason why throughout the world the population data are obtained by conducting a census generally every 10 years by the census method.
- (iv) Data of complete enumeration census can be widely used as a basis for various surveys.

However, despite these advantages the census method is not very popularly used in practice. The effort, money and time required for carrying out complete enumeration will generally be very large and in many cases cost may be so prohibitive that the very idea of collecting information may have to be dropped. This is more true of underdeveloped countries where resources constitute a big constraint. Also if the population is infinite or the evaluation process destroys the population unit, the method cannot be adopted.

Sampling is simply the process of learning about the population on the basis of a sample drawn from it. Thus, in the sampling technique instead of every unit of the universe is studied and the conclusions are drawn on that basis for the entire universe. A sample is a subset of population units. The process of sampling involves three elements:

- (a) Selecting the sample,
- (b) Collecting the information, and
- (c) Making an inference about the population.



The three elements cannot generally be considered in isolation from one another. Sample selection, data collection, and estimation are all interwoven and each has an impact on the others. Sampling is not haphazard selection-it embodies definite rules for sample has been selected.

Although much of the development in the theory of sampling has taken place only in recent years. The idea of sampling is pretty old. Since times immemorial people have examined a handful of grains to ascertain the quality of the entire lot. A housewife examines only two or three grains of boiling rice to know whether the pot of rice is ready or not. A doctor examines a few drops of blood and draws conclusion about the blood constitution of the whole body. A businessman places orders for material by examining only a small sample of the same. A teacher may put questions to one or two students and find out whether the class as a whole is following the lesson. In fact there is hardly any field where the technique of sampling is not used either consciously or unconsciously.

It should be noted that a sample is not studied for its own sake. The basic objective of its study is to draw inference about the population. In other words, sampling is a tool which helps to know the characteristics of the universe or population by examining only a small part of it. The values obtained from the study of a sample, such as the average and dispersion, are known as 'statistics'. On the other hand, such values for the population are called 'parameters'.

### **Theoretical basis of sampling**

On the basis of sample study we can predict and generalize the behaviour of mass phenomena. This is possible because there is no statistical population whose elements would vary from each other without limit. For example, wheat varies to a limited extent in colour, protein content, length, weight, etc., it can always be identified as wheat. Similarly, apples length, weight, etc., it can always be identified as wheat. Similarly, apples If the same tree may vary in size, colour, taste, weight, etc., but they can always be identified as apples. Thus we find that although diversity is a universal quality of mass data, every population has characteristic properties with limited variation. This makes possible to select a relatively small unbiased random sample that can portray fairly well the traits of the population.

There are two important laws on which the theory of sampling is based:

Law of Statistical Regularity', and

Law of Inertia of large Numbers'.

Law of Statistical Regularity

This law is derived from the mathematical theory of probability. In the words of King : The law of statistical regularity lays down that a moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristics of the large group." In other words, this law points out that if a sample is taken at random from a population, it is likely to possess almost the same characteristics as that of the population. This law directs our attention to one very important point, that is, the desirability of choosing the sample at random.

By random selection we mean a selection where each and every item of the population has an equal chance of being selected in the sample. In other words, the selection must not be made by deliberate exercise of one's discretion. A sample selected in this manner would be representative of the population. If this condition is satisfied it is possible for one to depict fairly accurately the characteristics of the population by studying only a part of it. Hence, this law is of great practical significance because it makes possible a considerable reduction of the work necessary before any conclusion is drawn regarding a large universe. For example, if one intends to make a study of the average height of the students of Delhi University it is not necessary to measure the heights of each and every student. A few students may be selected at random from every college, their heights measured and the average height of university students in general may be inferred.

It should be noted that the results derived from sample data may be different from that of the population. This is for the simple reason that the sample is only a part of the whole universe. For example, the average height of the students of Delhi University may come out to be 160 cm. by census method whereas it may be 159 cm. or 161 cm. for the sample taken. It should be just a coincidence if the height comes out to be exactly 160 cm. under both the methods. However, there would not be much difference in the results derived if the sample is representative of the universe.

### **Law of Inertia of Large Numbers**

This law is a corollary of the law of statistical regularity. It is of great significance in the theory of sampling. It states that, other things being equal larger the size of the sample, more accurate the results are likely to be. This is because large numbers are more stable as compared to small ones. The difference in the aggregate result is likely to be insignificant, when the number in the sample is large, because when large numbers are considered the variations in the component parts tend to balance each other and, therefore, the variation in the aggregate is insignificant. For example, if a coin is tossed 10 times we should expect equal number of heads and tails, i.e., 5 each. But since the experiment is tried a small number of times it is likely that we may not get exactly 5 heads and 5 tails. The result may be a combination of 9 heads and 1 tail, or 8 heads and 2 tails, or 7 heads and 3 tails. If the same experiment is carried out 1,000 times the chance of 500 heads and 500 tails would be very high, i.e., the result would be very near to 50 per cent heads and 50 per cent tails. The basic reason for such likelihood is that the experiment has been carried out a sufficiently large number of times and possibility of variation in one direction compensating for others in a different direction is greater. If at one time we get continuously 5 heads, it is likely that at other time we may get continuously 5 tails, and so on, and for the experiment as a whole the number of heads and tails may be more or less equal. Similarly, if it is intended to study the variation in the production of rice over a number of years and data are collected from one or two States only, the result would reflect large variations in production due to the favourable factors in operation. If, on the other hand, figures of production are collected for all the States in India, it is quite likely that we find little variation in the aggregate. This does not mean that the production would remain constant for all the years. It only implies that the changes in the production of the individual States will be counterbalanced so as to reflect smaller variations in production for the country as a whole.

### **Essentials of Sampling**

If the sample results are to have any worthwhile meaning, it is necessary that a sample possesses the following essentials:

(i) Representativeness: A sample should be so selected that it truly represents the universe otherwise the results obtained may be misleading. To ensure representativeness the random method of selection should be used.

(ii) Adequacy: The size of sample should be adequate, otherwise it may not represent the characteristics of the universe.

(iii) Independence: All items of the sample should be selected independently of one another and all items of the universe should have the same chance of being selected in the sample. By independence of selection we mean that the selection of a particular item in one draw has influence on the probabilities of selection in any other draw.

(iv) Homogeneity: When we talk of homogeneity we mean that there is no basic difference in the nature of units of the universe and that of the sample. If two samples from the same universe are taken, they should give more or less the same unit.

### **(B) Advantages and Limitations of Sampling**

#### **Merits**

1. It saves time, because fewer items are collected and processed. When the results are urgently required, this method is very helpful.
2. It reduces cost only a few and selected items are studied in sampling. So there is reduction in cost of money and reduction in terms of man-hours. It is advantageous to underdeveloped countries.
3. More reliable results can be obtained because (a) there are fewer chances of sampling statistical errors. If there is sampling error, it is possible to estimate and control the results. (b) Highly expert and trained persons can be employed for scientific processing and analysing of relatively limited data, and they can use their high technical knowledge and get more accurate and reliable results.
4. Sampling provides more detailed information: As it saves time, money and energy, we can collect more detailed information in a sample survey.
5. Sampling method is sometimes the only method possible. If the population under study is infinite, sampling method is the only method to be used.

For example, to test the breaking strength of bricks manufactured in a factory, under census method, all the bricks would be broken in the process of testing. There would be no bricks left for use. Thus, census method is impracticable. Also if the population under investigation is infinite, sampling is the only possible solution.

6. Administrative convenience: The organisation and administration of sample survey are easy.
7. More scientific: the method has full justification for the expenditure involved.
8. The degree of accuracy obtainable in this method is higher than that in the census method.

It is very important to note that the aim of sampling studies is to obtain maximum information about the phenomena under study with least sacrifice of money, time and energy. The purpose of sampling is to get information about the population from the sample. For example, a doctor examines few drops of blood and draws conclusions about the whole blood. We can obtain a large variety of information about the phenomena to which the sample relates. And, this helps us to have an idea about similar information relating to the universe. For example, when we go to the market we examine a sample of rice (a handful of rice) from the lot, form an idea about the quality and decide whether the quality is acceptable or not. Another example, we meet a person for a while, talk with him and form opinion about his character. In all these examples, we adopt sampling technique. Our knowledge, our attitude and our actions are based to a large extent on samples. The theory which helps us in studying samples is known as theory of sampling. Logic theory of sampling is the logic of induction i.e. from the study of a sample, one tries to infer about the population. Thus, the aim of sampling studies is to obtain the best possible values of the parameters. The population measures for example, mean, standard deviation etc. called parameters, while the measures obtained from sample are called statistics.

### **Shortcomings**

1. Illusory conclusion: If a sample enquiry is not carefully planned and executed, the conclusions may be inaccurate and misleading.
2. Sample not representative: To make a representative sample is taken from the universe; the result is applicable to the whole population. If the sample is not representative of the universe, the result may be false and misleading.
3. Lack of experts: As there is lack of experts to plan and conduct a sample survey, its execution and analysis, the results of the sample survey are not satisfactory and trustworthy.
4. Sometimes the sampling plan may be complicated and requires more money, labour, time than a census method.
5. There is organisational problem in sample investigation.
6. Personal bias: There may be personal biases and prejudices with regard to the choice of technique and drawing of sampling units.
7. Choice of sample sizes: If the size of the sample is not appropriate then it may lead to untrue characteristics of the population.
8. Conditions of complete coverage: If the information is required for each and every item of the universe, then a complete enumeration survey is better.

Though the shortcomings are there, yet sample investigation is very useful, provided there is a scientific selection. Frederick F. Stephen writes, “Samples are like medicines. They can be harmful when they are taken carelessly or without adequate knowledge of their effects. We may use their results with confidence, if the applications are made with due restraint. It is foolish to avoid or discard them, because someone else has misused them and suffered the predictable consequences of his folly. Every good sample should have a proper label with instructions about its uses. Further Prof. Chou states, Sampling is a simple process of learning about the population on the basis of sample drawn from it. In sampling, a few representative items are selected and studied; and on the basis of results, generalisations regarding the universe are made.

## **2.2. Types of Sampling**

There are many methods of sampling. The Choice of method will be determined by the purpose of sampling. The various methods can be grouped under two groups:

### **1. Random Sampling Method (Probability Sampling)**

(a) Simple or unrestricted random sampling

(b) Restricted Random Sampling

- (i) Stratified sampling
- (ii) Systematic sampling
- (iii) Cluster sampling

### **2. Non-Random Sampling (Non-probability Sampling)**

- (a) Judgment or purposive sampling.
- (b) Quota sampling
- (c) Convenience sampling
- (d) Snow-ball sampling

#### **1. Random Sampling (Probability sampling)**

A random sample is one where each item in the universe has an equal chance of known opportunity of being selected. According to Dr. Yates, “Every member of a parent population has had equal chances of being included”. According to Harper, “A random sample is a sample selected in such a way that every item in the population has an equal chance of being included.”

##### **A. Simple random sampling**

It is a technique in which sample is so drawn that each and every unit in the population has an equal and independent chance of being included in the sample. Several methods have been adopted for random selection of the sample. They are:

(i) Lottery Method. This is the most popular and simplest method. In this method, all the items of the universe are numbered on separate slips of paper of same size, shape and colour. They are folded and mixed up in a drum or container. A blindfold selection is made. The required number of slips are selected for the desired sample size. The selection of items thus depends on chance. For example, if we want to select 5 students, out of 50 students, then we must write the names of all the 50 students on slips of the same size and mix them; then we make a blindfold selection of 5 students. This method is also called unrestricted random sampling, because units are selected from the unrestricted random sampling, because units are selected from the population without any restriction. This method is mostly used in lottery draws. If the universe is infinite, this method is inapplicable. There is a lot of possibility of personal prejudice if the size and shape of the slips are not identical.

(ii) Table of Random Numbers: As the lottery method cannot be used, when the population is infinite, the alternative method is that of using the table of random numbers.

There are several standard tables of random numbers. But the credit for this technique goes to Prof. L.H.C. Tippett (1927). The random number table (taken from the British Census Report) consists of 10,400 four-figured numbers giving in all  $10,400 \times 4 = 41,600$ . There are various other random numbers. They are Fisher and Yates (1938) comprising of 15,000 digits arranged in twos, Kendall and B.B. Smith (1939) consisting of 1,00,000 digits grouped in 25,000 sets of 4 digit random numbers, Rand Corporation (1955) consisting of 2,00,000 random numbers of 5 digits each, etc.

### **Merits**

1. Scientific method: there is less chance for personal bias.
2. More representative: when the size of the sample increases, it is representative of the population, as the Law of Inertia of large Numbers and the Law of Statistical Regularity beings to operate.
3. Sampling error can be measured.
4. Theory of probability is inapplicable, if a sample is random.
5. this method is economical as it saves time, money and labour

### **Demerits**

1. This requires a complete list of the population but such up-to-date lists are not available in many enquiries.
2. If the size of the sample is small, then it will not be a representative of the population.
3. When the distribution between items is very large, this method cannot be used.

**B. Restricted Random Sampling**

(i) Stratified sampling: When the population is heterogeneous or of different segments of strata with respect to the variable or characteristic under study, then it is stratified. First the population is divided into a number of sub-groups or strata. Each stratum is homogeneous. A sample is drawn from each stratum at random.

There are two types of stratified random sampling. They are proportional and non-proportional. In the proportional sampling, equal and proportionate representation is given to sub-groups or strata. If the number of items is large in the population, the same will have a higher size and vice versa.

In disproportionate or non-proportionate sample, equal representation is given to all the sub-strata regardless of their existence in the population.

**Merits**

1. It is more representative.
2. It ensures greater accuracy.
3. It is easy to administer as the universe is sub-divided.
4. Greater geographical concentration reduces the time and expenses.
5. When the original population is badly skewed, this method is an appropriate one.
6. For non-homogeneous population, it may yield more reliable results.

**Demerits**

1. To divide the population into homogeneous strata, it requires more money, time and statistical experience which is a difficult one.
2. If proper stratification is not done, the sample will have an effect of bias. If different strata of population overlap, such a sample will not be a representative one.

(ii) Systematic sampling: It is also known as quasi-random sampling. A systematic sample is selected at random sampling. When a complete list of the population is available, this method is used. We arrange the items in numerical, alphabetical, geographical or any other order. If we want to select a sample of 10 students from 100 students, under this method Kth item is picked up from the sample frame and K is the sample interval.

K = Sampling interval

N = Size of universe

n = Sample size

in the above example  $k = 10$ . 10 is the sampling interval. Every 10th student will be taken as sample, i.e., 10th 20th, 30th, and so on.

**Merits**

This is simple and convenient. the time and work is reduced much. If we take care, the result will be a satisfactory one. It can also be used in infinite population.

**Demerits**

It may not represent the whole population. There is the element of personal bias of investigators.

(iii) Cluster sampling or multistage sampling. It is also called as sampling stages. It refers to a sampling procedure, which is carried out in several stages. The whole population is divided into sampling units, and these units are again divided into sub-units. This process will continue when we reach a lease number.

For example, we want to take 5000 students from Madhya Pradesh. We want to take 5000 students from Madhya Pradesh. We must take universities at the first stage, then the number of colleges at the second stage, selection of students from the colleges at the third stage etc.

**Merits**

1. It introduces flexibility in the sampling method.
2. It is helpful in large -scale survey where the preparation of list is difficult, time-consuming or expensive.
3. It is valuable in underdeveloped countries, where no detailed and accurate framework is available.

**Demerits**

It is less accurate than other methods.

2. Non-random sampling method (Non Probability Sampling)

(a) Judgment sampling. (Purposive or Deliberate). The investigator has the power to select or reject any item in an investigation. The choice of sample items depends on the judgments of the investigator. He has the vital role to pay in collecting the information. For example, if a sample of 5 students are to be selected from a B.Com. Class of 50 students for analyzing the habit of picture-seeing, the investigator would select 5 students who, according to his opinion are the representative of the class.

**Merits**

1. It is a simple method.
2. It is used to obtain a more representative sample.
3. It is very helpful to make public policies, decisions, etc. The executives and public officials use this method for their urgent problem.



**Demerits**

1. Due to individual bias the sample may not be a representative one.
2. It is difficult to get correct sampling errors.
3. The estimates are into accurate.
4. Its results cannot be compared with other sampling studies.

(b) Quota sampling: This sampling is similar to stratified sampling. It is used in the U.S.A. for investigating public opinion and consumer research. to collect data, the universe is divided into quota according to some characteristics. Each enumerator is then told to interview a certain number of persons who are his quota. The selection of sample items depends on personal judgment.

It is a stratified-cum-purposive sampling and thus has the advantages of both the methods. There is saving of time and money. If there are trained investigators, the sampling will give quite reliable results.

Personal prejudice and individual bias are there. It is not based on random sampling, and so sampling error cannot be estimated.

(c) **Convenience or Chunk Sampling.** Chunk is a convenient slice of a population which is commonly referred to as a sample. It is obtained by selecting convenient population units.

1. It is suitable when the universe is not clearly defined.
2. Sample unit is not clear.
3. Complete source list is not available.

A Sample obtained from automobile registration, telephone directories, etc., is a convenience sample. The results of this sampling cannot be representative. They are unsatisfactory. They are biased. But they are used for pilot studies.

(d)**Snow-ball Sampling:** In sociology and statistics research, snow-ball sampling (or chain sampling, chain-referral sampling, referral sampling) is a non-probability sampling technique where existing study subjects recruit future subjects from among their acquaintances. Thus the sample group appears to grow like a rolling snowball. As the sample builds up, enough data is gathered to be useful for research. This sampling technique is often used in hidden populations which are difficult for researchers to access; example populations would be drug users or sex workers. As sample members are not selected from a sampling frame, snowball samples are subject to numerous biases. For example, people who have many friends are more likely to be recruited into the sample.

What is snowball sampling? Snowball sampling uses a small pool of initial informants to nominate, through their social networks, other participants who meet the eligibility criteria and could potentially contribute to a specific study. The term "snowball sampling" reflects an analogy to a snowball increasing in size as it rolls downhill

Snowball Sampling is a method used to obtain research and knowledge, from extended associations, through previous acquaintances, "Snowball sampling uses recommendations to find people with the specific range of skills that has been determined as being useful." An individual or a group receives information from different places through a mutual intermediary. This is referred to metaphorically as snowball sampling because as more relationships are built through mutual association, more connections can be made through those new relationships and a plethora of information can be shared and collected, much like a snowball that rolls and increases in size as it collects more snow. Snowball sampling is a useful tool for building networks and increasing the number of participants. However, the success of this technique depends greatly on the initial contacts and connections made. Thus it is important to correlate with those that are popular and honorable to create more opportunities to grow, but also to create a credible and dependable reputation.

**Advantages**

1. Locate hidden populations: It is possible for the surveyors to include people in the survey that they would not have known.
2. Locating people of a specific population: There is no lists or other obvious sources for locating members of the population of specific interest.

**Disadvantages**

1. Community Bias: The first participants will have strong impact on the sample. Snowball sampling is inexact, and can produce varied and inaccurate results. The method is heavily reliant on the skill of the individual conducting the actual sampling, and that individual's ability to vertically network and find an appropriate sample. To be successful requires previous contacts within the target areas, and the ability to keep the information flow going throughout the target group.
2. Not Random: Snowball sampling contradicts many of the assumptions supporting conventional notions of random selection and representativeness. However, Social systems are beyond researcher's ability to recruit randomly. Snowball sampling is inevitable in social systems.
3. Vague Overall Sampling Size: There is no way to know the total size of the overall population.
4. Wrong Anchoring: Another disadvantage of snowball sampling is the lack of definite knowledge as to whether or not the sample is an accurate reading of the target population. By targeting only a few select people, it is not always indicative of the actual trends within the result group. Identifying the appropriate person to conduct the sampling, as well as locating the correct targets is a time consuming process which renders the benefits only slightly outweighing the costs.

## MODULE III

### DATA MANAGEMENT AND PRESENTATION.

#### 111.1 Nature of Statistical Data: Variables and Attributes.

In social research and theory, both variables and attributes represent social concepts. An attribute is defined as a characteristic of something or qualities. It is a concept or a construct expressing the qualities possessed by a physical or mental object. Variables, in turn, have what social researchers call attributes (or categories or values). Thus for example, male and female are attributes, and sex or gender is the variable composed of these two attributes. The variable occupation is composed of attributes such as agriculturist, teacher, doctor, driver etc. Social class is a variable composed of a set of attributes such as upper class, middle class and lower class. Attributes are the categories that make up a variable. In science and research, attribute is a characteristic of an object (person, thing, etc.). While an attribute is often intuitive, the variable is the operationalized way in which the attribute is represented for further data processing. In data processing data are often represented by a combination of *items* (objects organized in rows), and multiple variables (organized in columns).

VARIABLES	ATTRIBUTES
Age, gender, occupation, race, social class	Young, middle, old, male, female, doctor, farmer, Dravidian, Aryan Upper, middle, lower

Age is an attribute that can be operationalized in many ways. It can be dichotomized so that only two values - "old" and "young" - are allowed for further data processing. In this case the attribute "age" is operationalized as a binary variable. If more than two values are possible and they can be ordered, the attribute is represented by ordinal variable, such as "young", "middle age", and "old"

The "social class" attribute can be operationalized in similar ways as age, including "lower", "middle" and "upper class" and each class could be differentiated between upper and lower, transforming thus changing the three attributes into six like upper-upper, upper-middle, middle-middle, middle-lower, lower-lower. The relationship between attributes and variables forms the heart of both description and explanation in science. Sometimes the meanings of the concepts that lie behind social science concepts are immediately clear. Other times they aren't. The relationship between attributes and variables is more complicated in the case of explanation and gets to the heart of the variable language of scientific theory.

Variables on the other hand, are logical groupings of attributes. According to Kerlinger, “A variable is a property that takes on different values. A variable is something that varies...A variable is a symbol to which numerals or values are attached”.

Black and Champion define a variable as ‘rational units of analysis that can assume any one of a number of designated sets of values.’

A variable uses numerical values to measure an attribute. It is a quantity that expresses a quality in numbers to allow more precise measurement.

1. Qualitative research focuses primarily on the meaning of subjective attributes of individuals or groups.
2. Quantitative research primarily focuses on the measurement of objective variables that affect individuals or groups.

There are many different types of variables.

1. Independent variables constitute the resumed cause. They are introduced under controlled conditions during the experiment as treatments to which experimental groups are exposed.
2. Dependent variables are the presumed effect. They are measured before and after the treatment to see whether any change occurred.
3. Background variables are antecedents that affect the situation prior to the study. They can be observed and measured, but usually not changed.
4. Intervening variables are events between the treatment and the post test measurement that might affect the outcome.
5. Extraneous variables are variables that can be observed and which might affect the outcome during the study, but which cannot be controlled.
6. Alternative independent variables suggest causes different from the existing independent variable.

#### **TYPES OF VARIABLE**

A variable can be classified in a three different ways.

- Casual relationship
- The design of the study
- The unit of measurement

#### **FROM THE VIEWPOINT OF CAUSATION**

In studies that attempt to investigate a casual relationship or association, four sets of variables may operate.

1. Change variables, which are responsible for bringing about change in a phenomenon.
2. Outcome variables, which are the effects of a change variable.
3. Variables which affect the link between cause and effect variables.
4. Connecting or linking variables, which in certain situations are necessary to complete the relationship between cause and effect variables.

In research terminology change variables are called independent variables, outcome/ effect variables are called dependent variables, the unmeasured variables affecting the cause and effect relationship are called extraneous variables and the variables that link a cause and effect relationship are called intervening variables. Hence:

1. Independent variable – the cause supposed to be responsible for bringing about change in a phenomenon or situation
2. Dependent variable – the outcome of the change brought about by introduction of an independent variable.
3. Extraneous variable- several other factors operating in areal life situation may affect changes in the dependent variable. These factors, not measured in the study, may increase or decrease the magnitude or strength of the relationship between independent and dependent variables.
4. Intervening variable – sometimes called the confounding variable (Grinnell 1988:203) links the independent and dependent variables. In certain situations the relationship between an independent and a dependent variable cannot be established without the intervention of another variable. The cause variable will have the assumed effect only in the presence of an intervening variable
5. From the viewpoint of the study design
6. A study that examines association or causation may be a controlled or contrived experiment a quasi- experiment or an ex post facto or non-experimental study. In controlled experiments the independent (Cause) variable may be introduced or manipulated either by the researcher or by someone else who is providing the service. In these situations there are two sets of variables.
  - Active variables – those variables that can be manipulated, changed or controlled.
  - Attribute variables – those variables that can't be manipulated, changed or controlled, and that reflect the characteristics of the study population; For example, gender, education and income.

### **From the viewpoint of the unit of measurement**

From the viewpoint of the unit of measurement there are two ways of categorizing variables:

- Whether the unit of variable is categorical (as in nominal and ordinal scales or continuous in nature (as in interval and ratio scales);
- Whether it is qualitative (as in nominal and ordinal scales) or quantitative in nature (as in interval and ratio scales).

The variables thus classified are called categorical and continuous, and qualitative and quantitative. On the whole there is very little difference between categorical and qualitative, and between continuous and quantitative, variables.

**Categorical variables** are measured on nominal or ordinal measurement scales, whereas for continuous variables the measurements are made either on an interval or a ratio scale. Categorical variables can be of three types:

1. Constant
2. Dichotomous
3. Polytomous

When a variable can have only one value or category, for example taxi, tree and water, it is known as a constant variable. When a variable can have only two categories as in yes/no, good/bad and rich/poor it is known as dichotomous variable. When a variable can be divided into more than two categories, for example: religion (Christian, Muslim and Hindu); political parties (labor, liberal, democrat); and attitudes (strongly favorable, favorable, uncertain, unfavorable, strongly unfavorable), it is called a polytomous variables.

**Continuous variables** on the other hand, have continuity in their measurement; for example, age, income and attitude score. They can take on any value on the scale on which they are measured. Age can be measured in dollars and cents.

## **CLASSIFICATION AND TABULATION**

After the data are collected with the help of questionnaire, interview schedule, observation etc, they need to be properly tabulated and presented. This process helps the researcher to eliminate the unnecessary details and keeps only the relevant part of the whole collected. The procedure adopted for this purpose is known as the method of classification and tabulation.

### **Classification**

Classification is the process of arranging data in groups or classes on the basis of common characteristic. Data having common characteristics are placed in one class and in this way the entire data get divided into a number of groups or classes. Classification of data can be done on the following two types:

1. Classification on the basis of attributes.
2. Classification on the basis of class intervals.

#### **Classification on the basis of attributes.**

Attributes refer to the particular characteristics of the population. These attributes may be descriptive. The chief characteristic of the descriptive attributes are qualitative. They can not be measured in any numerical terms. Classification on the basis of attributes can be further divided into two;

##### **1. Simple classification**

##### **2. Manifold classification.**

In Simple Classification, only one attribute is considered and the universe is divided on that basis. But in

Case of manifold classification. the universe may be classified in to several groups on the basis of more than one attribute.

### **2. Classification on the basis of class intervals.**

Unlike descriptive characteristics, the numerical characteristics refer to quantitative phenomena which can be measured through some statistical units. When the goal of data classification is to arrange a set of data in to a useful form, frequency distribution provides a general approach to data classification.

In frequency distribution, raw data are represented by distinct groups called classes. The number of measurements in each class is called the class frequency. In this way, the data set which may be very large is considered into a smaller, more manageable set of numbers.

### **Tabulation**

It is apart of the technical process in the statistical analysis of data. The essential operation in tabulation in counting to determine the number of cases that falls into the various categories.

### **Frequency Table**

A table that displays the number and or percentage of units (people) in different categories of a variable. Frequency tables are the normal tabular method of presenting distributions of a single variable. Tabulation is a process of summarizing raw data and displaying them on compact statistical tables for further analysis. It involves counting the number of cases falling into each of the categories identified by the researcher. Tabulation can be done manually or through the computer. The choice depends upon the size and type of study, cost considerations, time pressures and the availability of software packages. Manual tabulation is suitable for small and simple studies.

### **Manual Tabulation**

When data are transcribed in a classified form as per the planned scheme of classification, category-wise totals can be extracted from the respective columns of the work sheets. A simple frequency table counting the number of 'Yes' and 'No' responses can be made by easily counting the 'Y' response column and 'N' response column in the manual worksheet table prepared earlier. This is a one way frequency table and they are readily inferred from the total of each column in the worksheet. Sometimes, the researcher has to cross tabulate two variables for instance the age group of vehicle owners. This requires a two way classification and cannot be inferred straight from the worksheet. For this purpose, tally sheets are used. This process of tabulation is simple and does not require any technical knowledge or skill.

Although manual tabulation is simple and easy to construct, it can be tedious, slow and error prone as responses increases.

### **Computerized tabulation**

Computerized tabulation is easy with the help of software package. The input requirement will be the column and row variables. The software package then computes the number of records in each cell of the row/column categories. The most popular package is the statistical package for social science (SPSS). It is an integrated set of programs suitable for analysis of social science data. This package contains programs suitable for analysis of social science data. This package contains programs for a wide range of operations and analysis such as handling missing data, recoding, variable information, simple descriptive analysis, cross tabulation, multivariate analysis and non parametric analysis.

### **Construction of frequency table**

Frequency tables provide a shorthand summary of data. The importance of presenting statistical data in tabular form needs no emphasis. Tables facilitate comprehending masses of data at a glance; they conserve space and reduce explanation and description masses of data at a visual picture of relationships between variables and categories. They facilitate summation of items and the detection of errors and they provide a basis for computations.

It is important to make a distinction between the general –purpose table and specific tables. The general purpose tables are primary or reference tables designed to include large amounts of source data in convenient and accessible form. The special purpose tables are analytical or derivate ones that demonstrate significant relationships in the data or the results or statistical information. Special purpose tables are found in monographs, research reports and articles and are used as instruments of analysis. In research, we are primarily concerned with special purpose tables.

### **Components of a Table**

#### **The major components of a table are:**

- A. Heading
  - (i) Table Number
  - (ii) Title of the table
  - (iii) Designation of units
- B. Body
  - (i) Stub Head: heading of all rows or blocks of stub items
  - (ii) Body Head: headings of all columns or main captions and their sub captions.
  - (iii) Field/body: the cells in rows and columns.
- C. Notations
  - (i) Footnotes, wherever applicable
  - (ii) Source, wherever applicable



**Principles of table construction:**

- (a) Every table should have a title and it should be placed above the body of the table. The title should represent a succinct description of the contents of the table. Table title should be clear and concise.
- (b) Every table should have a number. The number can be centered above the title. The table should run in a consecutive serial number.
- (c) The captions or column heading should be clear and brief. The units of measurements under each heading must always be indicated.
- (d) Any explanatory footnotes concerning the table itself are placed directly beneath the table.
- (e) If the data in a series of tables have been obtained from different sources, it is ordinarily advisable to indicate the specific sources in a place just below the table
- (f) Lines are always drawn at the top and the bottom of the table and below the captions.
- (g) Columns may be numbered to facilitate reference. All column figures should be properly aligned.
- (h) Columns and rows that are to be compared to one another should be brought close together.
- (i) Totals of rows should be placed at the extreme right column and totals of columns at the bottom. Different kinds of type spacing and identification can be used to emphasis the relative significance of certain categories.
- (j) Abbreviations and ditto marks should be avoided in a table.

**TABLES**

**Structure**

Tables are the most common method of presenting analyzed data. Tables offer a usual means of presenting large amounts of detailed information in a small space.

**A table has five parts.**

1. Title – this normally indicates the table’s number and describes the type of data it contains. It is important to give each table its own number. The tables should be numbered sequentially as they appear in the text. The procedure for numbering tables is a personal choice. The description accompanying the table number must clearly specify the contents of that table. In the description identify the variables about which information is contained in the table.
2. Stub – The subcategories of a variable, listed along the y-axis. According to the McGraw –Hill Style manual (longyear 1983:97), ‘The stub usually the first column on the left, lists the items about which information is provided in the horizontal rows to the right.’ The Chicago Manual of Style (1993:331) describes the stub as: ‘a vertical listing of categories or individuals about which information is given in the columns of the table.’
3. Column headings – the subcategories of a variable, listed along the x-axis (the top of the table).
4. Body
5. Supplementary notes or foot notes.

### **Types of tables**

Depending upon the number of variables about which information is displayed, tables can be categorized as univariate (containing information about one variable) also called frequency tables; bivariate( containing information about two variables) also called cross tabulations or polyvariate/multivariate (containing information about more than two variables)11.

### **Types of percentages**

The ability to interpret data accurately and to communicate findings effectively are important skills for a researcher. For accurate and effective interpretation of data, you may need to calculate other measures such as percentages, cumulative percentages or ratios. It is also sometimes important to apply other statistical procedures to data. The use of percentages is a common procedure in the interpretation of data. There are three types of percentage: row, column and total.

- Row percentage

Calculated from the total of all the subcategories of one variable that are displayed along a row in different columns, in relation to only one subcategory of the other variable.

- Column percentage

Calculated from the total of all the subcategories of one variable that are displayed in columns in different rows, in relation to only one subcategory of the other variable.

- Total percentage

This standardizes the magnitude of each cell; that is, it gives the percentage of respondents who are classified in the subcategories of one variable in relation to the subcategories of the other variable

### **Graphs/Charts/Diagrams**

In presenting the data of frequency distribution and statistical computations, it is often desirable to use appropriate forms of graphic presentation. In addition to tabular forms, graphic presentation involves use of graphics, charts and other pictorial devices such as diagrams. These forms and devices reduce large masses of statistical data to a form that can be quickly understood at a glance. The meaning of figures in tabular form may be difficult for the mind to grasp or retain. “Properly constructed graphs and charts relieve the mind of burdensome details by portrating facts concisely, logically and simply” They, by emphasizing new and significant relationships, are also useful in discovering new facts and in developing hypotheses.

The device of graphic presentation is particularly useful when the prospective readers are non-technical people or general public. It is useful to even technical people for dramatizing certain points about data for important points can be effectively captures in pictures than in tables. However, graphic forms are not substitutes for tables, but are additional tools for the researcher to emphasis the research findings.

Graphic presentation must be planned with utmost care diligence. Graphic forms used should be simple, clear and accurate and also be appropriate to the data. In planning this work, the following must be considered:

What is the purpose of the diagram?

What facts are to be emphasized?

What is the educational level of the audience?

How much time is available for the preparation of the diagram?

What kind of chart will portray the data most clearly and accurately?

### **Types and general Rules**

**The most commonly used graphic forms may be grouped into the following categories:**

- (a) Line Graphs or Charts
- (b) Bar Charts
- (c) Segmental Presentations
- (d) Scatter Plots
- (e) Bubble Charts
- (f) Stock Plots
- (g) Pictographs
- (h) Chesnokov Faces

**The general rules to be followed in graphic representations are:**

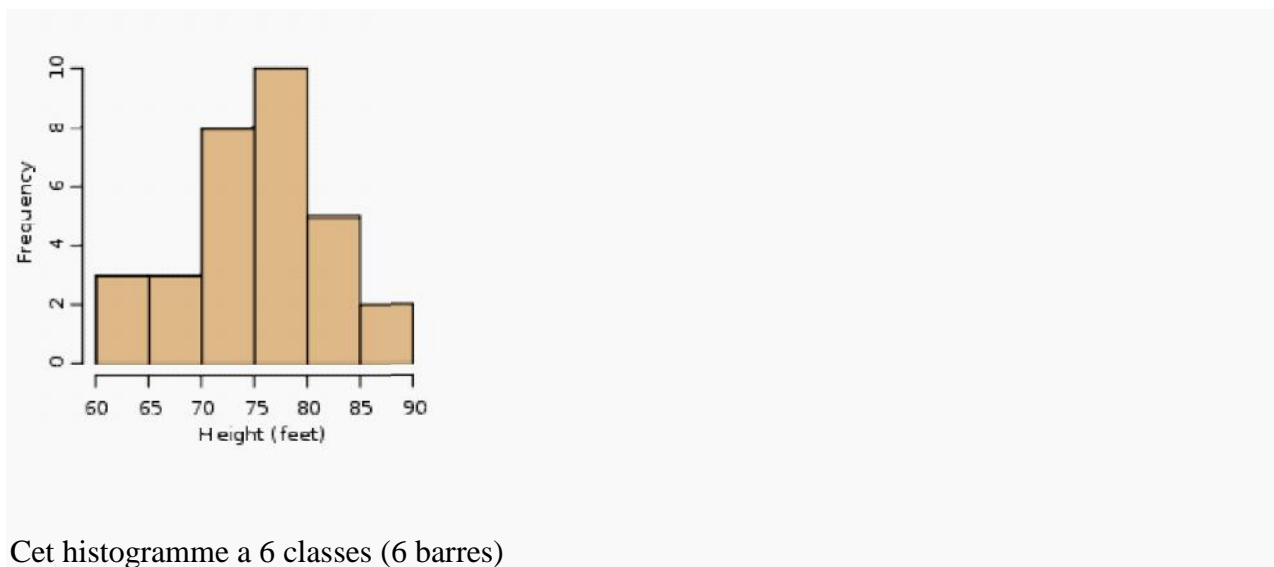
- (a) The chart should have a title placed directly above the chart.
- (b) The title should be clear, concise and simple and should describe the nature of the data presented.
- (c) Numerical data upon which the chart is based should be presented in an accompanying table.
- (d) The horizontal line measures time or independent variable and the vertical line the measured variable
- (e) Measurements proceed from left to right on the horizontal line and from bottom to top on the vertical.
- (f) Each bar or curve in the chart should be labeled.
- (g) If there are more than one curve and bar, they should be clearly differentiated from one another by distinct patterns or colours
- (h) The zero point should be used clearly differentiated from one another by distinct patterns or colours
- (i) The zero point should always be represented and the scale intervals should be equal .
- (j) Graphic forms should be used sparingly. Too many forms detract from rather than illuminate the presentation.
- (k) Graphic forms should follow not precede the related textual discussion.  
Histogram used for presenting a frequency distribution.

## HISTOGRAM

A histogram is drawn to represent relative frequency size of different groups. In a variable the variable is always taken on the 'x' axis and the frequencies are depending on it 'y' axis. Each class is represented by a rectangle which is proportional to its class interval. The distinction between a histogram and bar diagram lies in the fact that whereas bar diagram is one-dimensional, a histogram two-dimensional. In a bar diagram only the length of the bar is material and not the width, whereas in a histogram both the length as well as width are important. The histogram is commonly used for graphical presentation of a frequency distribution.

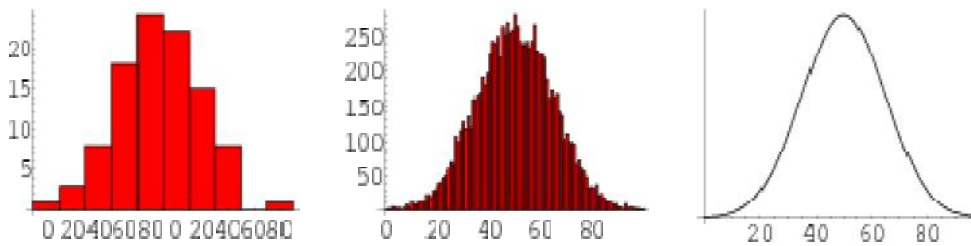
In statistics, a **histogram** is a graphical representation of the distribution of data. It is an estimate of the probability distribution of a continuous variable and was first introduced by Karl Pearson. A histogram is a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals. With an area equal to the frequency of the observations in the interval. The height of a rectangle is also equal to the frequency density of the interval, i.e., the frequency divided by the width of the interval. The total area of the histogram is equal to the number of data. A histogram may also be normalized displaying relative frequencies. It then shows the proportion of cases that fall into each of several categories, with the total area equaling 1. The categories are usually specified as consecutive, non-overlapping intervals of a variable. The categories (intervals) must be adjacent, and often are chosen to be of the same size. The rectangles of a histogram are drawn so that they touch each other to indicate that the original variable is continuous.

Histograms are used to plot the density of data, and often for density estimation: estimating the probability density function of the underlying variable. The total area of a histogram used for probability density is always normalized to 1. If the length of the intervals on the  $x$ -axis are all 1, then a histogram is identical to a relative frequency plot.



## Frequency Curve

---



A smooth curve which corresponds to the limiting case of a [histogram](#) computed for a frequency distribution of a continuous distribution as the number of data points becomes very large.

### Frequency Polygon and Frequency curve:

The frequency polygon of a grouped frequency distribution is constructed by joining by means of straight lines the points whose abscissas are the mid-points of the classes and the ordinates are the corresponding frequencies. Thus a frequency polygon can also be obtained from a histogram by joining the mid-points of the upper sides of the adjacent rectangles by means of straight lines.

To draw the frequency curve it is necessary first to draw the polygon. The polygon is then smoothed out keeping in view the fact that the area of the curve should be equal to that of the histogram.

Frequency polygons are a graphical device for understanding the shapes of distributions. They serve the same purpose as histograms, but are especially helpful for comparing sets of data. Frequency polygons are also a good choice for displaying *cumulative frequency distributions*.

To create a frequency polygon, start just as for histograms, by choosing a *class interval*. Then draw an X-axis representing the values of the scores in your data. Mark the middle of each class interval with a tick mark, and label it with the middle value represented by the class. Draw the Y-axis to indicate the frequency of each class. Place a point in the middle of each class interval at the height corresponding to its frequency. Finally, connect the points. You should include one class interval below the lowest value in your data and one above the highest value. The graph will then touch the X-axis on both sides.

### Definition of Frequency Polygon

- In a Frequency Polygon, a line graph is drawn by joining all the midpoints of the top of the bars of a histogram.

### More about Frequency Polygon

- A frequency polygon gives the idea about the shape of the data distribution.
- The two end points of a frequency polygon always lie on the x-axis.

### **Frequency Polygons:**

In laying out a frequency polygon, the frequency of each class is located at the midpoint of the interval and straight lines that connect the plotted points. If two or more series are shown on the same graph, the curves can be made with different kinds of ruling. If the total number of cases in the two series is of different size, the frequencies are often reduced to percentages. The frequency polygon is particularly appropriate for portraying continuous series. It is sometimes desirable to portray the data by a smoothed curve. The chart is then called a frequency curve. Frequency polygon gives an instant picture of a frequency distribution and shows whether the distribution is normal or otherwise. For example, the peak of occurs towards one end or the other, then the distribution is skewed.

Graphical display of the frequency table can also be achieved through a frequency polygon. To create a frequency polygon the intervals are labeled on the X-axis and the Y axis represents the height of a point in the middle of the interval. The points are then joined and connected to the X-axis and thus a polygon is formed. So, frequency polygon is a graph that is obtained by connecting the middle points of the intervals. We can create a frequency polygon from a histogram also. If the middle top points of the bars of the histogram are joined, a frequency polygon is formed. Frequency polygon and histogram fulfill the same purpose. However, the former one is useful in comparison of different datasets. In addition to that frequency polygon can be used to display cumulative frequency distributions.

#### ***How to Create a Frequency Polygon?***

As already mentioned, histogram can be used for creating frequency polygon. The X-axis represents the scores of the dataset and the Y-axis represents the frequency for each of the classes. Now, mark the mid top points of each bar of the created histogram for each class interval. One generally uses a dot for marking. Now join all the dots by straight lines and connect it with the X-axis on both sides. For creating a frequency polygon without a histogram, need to consider the midpoint of the class intervals, such that it corresponds to the frequencies. Then connect the points as stated above.

The following table is the frequency table of the marks obtained by 50 students in the pre-test examination.

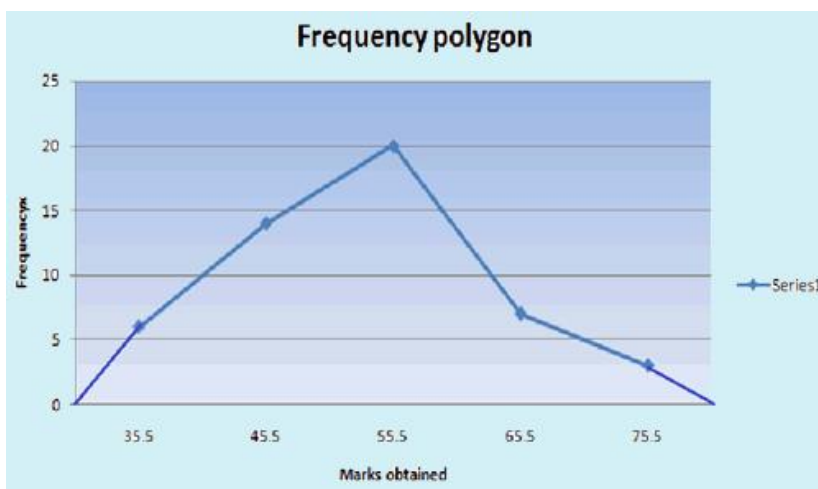
Frequency Distribution of the marks obtained by 50 students in the pre-test examination.

<b><i>Class Boundaries</i></b>	<b><i>Frequency</i></b>	<b><i>Cumulative frequency (Less than type)</i></b>
30.5-40.5	1	1
40.5-50.5	14	20
50.5-60.5	20	40
60.5-70.5	7	47
70.5-80.5	3	50
Total	50	

The labels of the X-axis are the midpoints of the class intervals. So the first label on the X-axis will be 35.5, next 45.5, followed by 55.5, 65.5 and lastly 75.5. The corresponding frequencies are then considered to create the frequency polygon. The shape of the distribution can be determined from the created frequency polygon. The frequency polygon is shown in the following figure.

$$A = \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$$

Fig 1: Frequency polygon of the distribution of the marks obtained by 50 students in the pre-test examination.



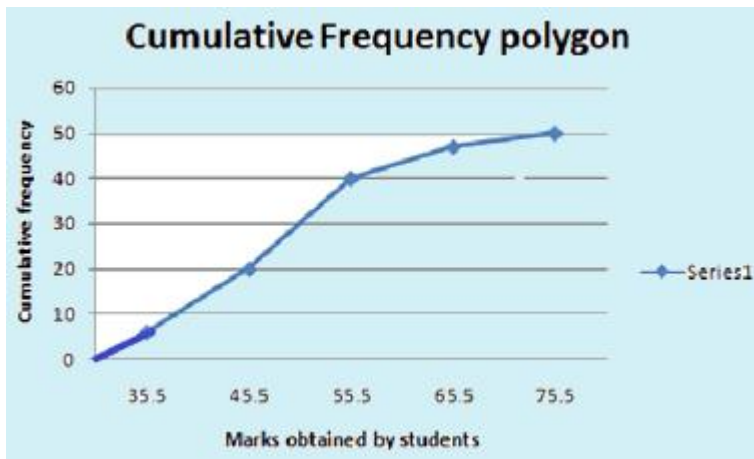
From the above figure we can observe that the curve is asymmetric and is right skewed.

**Cumulative Frequency Polygon:**

Cumulative frequency polygon is similar to a frequency polygon. The difference is that in creating a cumulative frequency polygon we consider cumulative frequencies instead of actual frequencies. Cumulative frequency of less than type is obtained by adding the frequency of each class interval to the sum of all frequencies in the lower intervals. In table 1 for example, the cumulative frequency for the class interval 30.5-40.5 is 6 since the sum of all frequencies in the lower intervals is 0. Again the cumulative frequency for the class interval 40.5-50.5 is 20 since the sum of all frequencies in the lower intervals is 14, i.e,  $6+14=20$ , so for the next interval it will be  $6+14+20=40$  and so on.

The following is the cumulative frequency polygon:

Cumulative Frequency polygon of the marks obtained by 50 students in the pre-test examination.



**Cumulative Frequency Curve (ogives):**

**Ogive**

The ogive is a line chart plotted on arithmetic graph paper from a cumulative frequency distribution that maybe cumulated upward or downward. It is useful in representing population, per capital income, per capital earnings etc. Converting the data of the distribution of the percentage of the total then cumulating the percentages and plotting the ogives on the same grid may give a useful comparison of two or more distributions. The differences in steepness and share of the ogives facilitate comparative observations.

**Cumulative Frequency** is the progressive total of the frequencies. It helps to find the median, quartiles and percentiles from large quantities of data organized into tables and graphs. To find the cumulative frequency, should add up the frequencies row by row.

An **Ogive** is a graph of the cumulative frequency. Interpreting quartiles, median and percentiles from a graph is more accurate than from a table.

To construct cumulative frequency curve or ogive it is necessary first to form the frequency table. Then the upper limits of the classes are taken as the x-coordinates and the cumulative frequencies as the y-coordinates and the points are plotted. The points are joined by a free hand smooth curve to give the ogive.



Example:

Draw a 'less than' ogive curve for the following data:

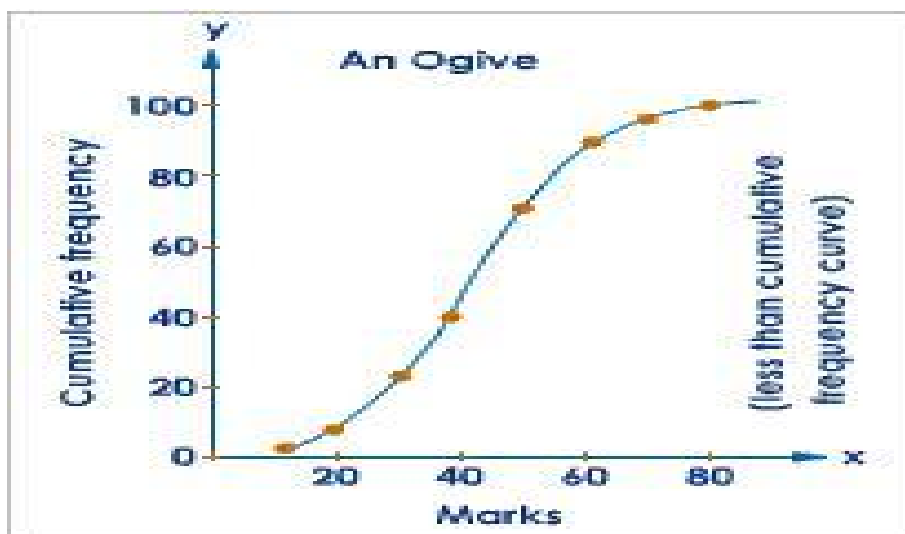
Marks	Frequency	Cumulative Frequency
0 - 10	2	2
10 - 20	8	10
20 - 30	12	22
30 - 40	18	40
40 - 50	28	68
50 - 60	22	90
60 - 70	6	96
70 - 80	4	100

To Plot an Ogive:

- (i) We plot the points with coordinates having abscissae as actual limits and ordinates as the cumulative frequencies, (10, 2), (20, 10), (30, 22), (40, 40), (50, 68), (60, 90), (70, 96) and (80, 100) are the coordinates of the points.
- (ii) Join the points plotted by a smooth curve.
- (iii) An Ogive is connected to a point on the X-axis representing the actual lower limit of the first class.

Scale:

X -axis 1 cm = 10 marks, Y -axis 1cm = 10 c.f.



## DIAGRAMS

Diagrams are among the most frequently used methods of displaying quantitative data. Their chief advantage is that they are relatively easy to interpret and understand. If one to work with nominal or ordinal variables, the bar chart and pie chart are two of the easiest methods to use.

### BAR CHARTS



Example of a bar chart, with 'Country' as the discrete data set.

These charts consist of either vertical or horizontal bars to represent variables. The length of the bar varies corresponding to the values of the variable. Bar charts are the most effective pictorial device for comparing data. The bars may be depicted in solid blocks or in patterns of dots, dashes etc. they may be of different forms: (1) linear or one-dimensional (2) a real or two dimensional and (3) cubic or three dimensional. The actual numerical values may be shown on the x-axis, or y axis, as the case may be, or at the immediate ends of the bars.

**vertical bar charts** consists of vertical bars or columns erected on the horizontal line and the valued of the bars are shown on the y axis. They are commonly used for presenting time series data.

**Horizontal bar charts** are commonly used for presenting qualitative and geographical distributions. They are also used for discrete quantitative distributions.

#### Component bar charts

This is employed to show comparisons involving two or more variables on a single chart. This, may consist of either horizontal or vertical bars. This type of chart shows not only variations in total values, but also components of the respective totals.

#### Principles of designing bar charts

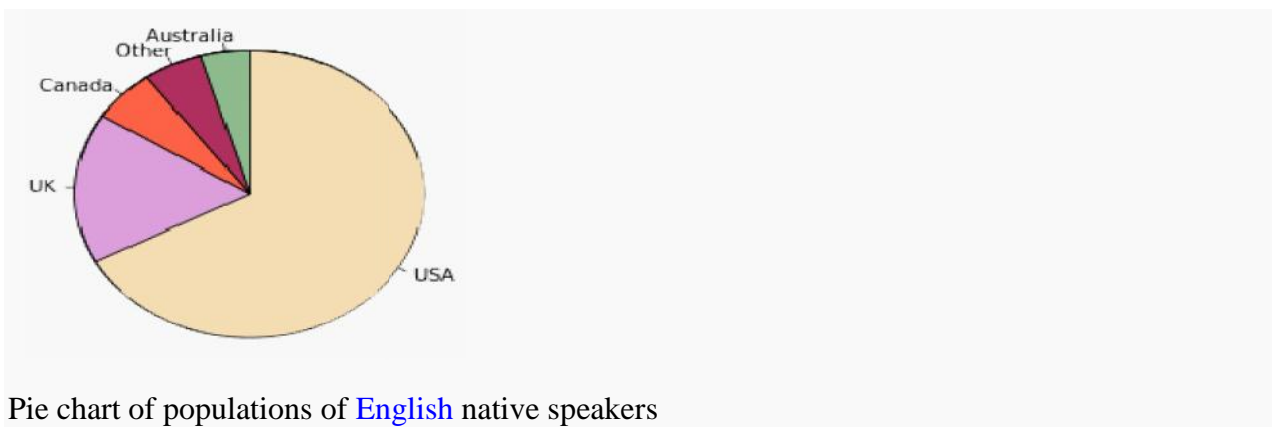
- The bars should be arranged in some systematic order: in chronological in presentation of time series; according to magnitude, starting with the largest, in other cases.
- The bars should be of uniform width and properly adapted to the over all size, proportion, and other features of the chart.

- A scale should be included in every bar chart. The number of intervals on the scale should be adequate for measuring distances but not too numerous to cause confusion. The intervals should be indicated in round numbers.
- The status or designations for the various categories of a bar chart should be clearly indicated to the left of the vertical base line.

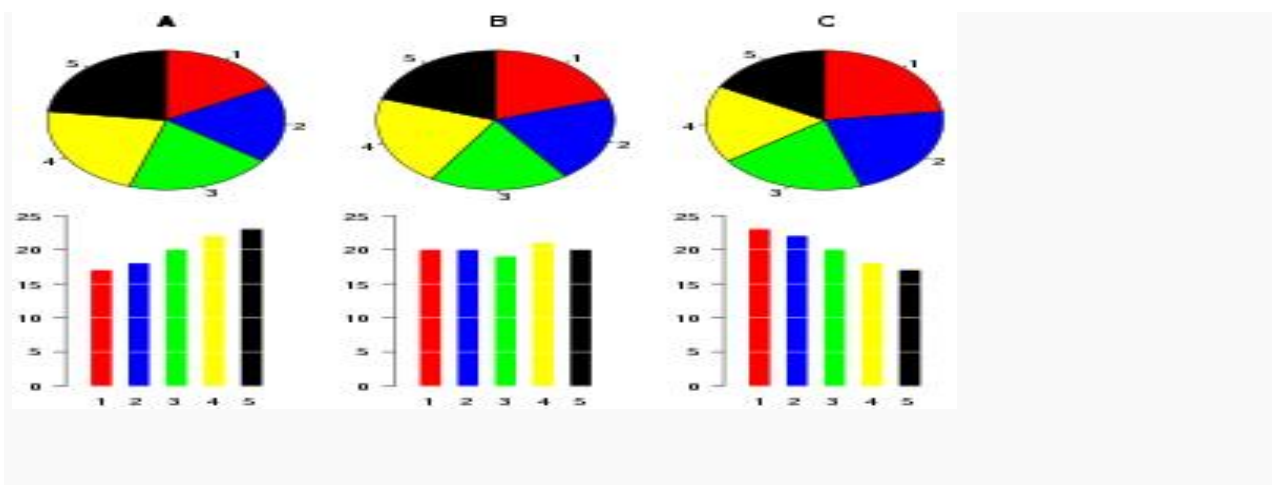
**Pie or circle charts**

**Pie chart** (or a **circle graph**) is a circular chart divided into sectors, illustrating numerical proportion. In a pie chart, the arc length of each sector (and consequently its central angle and area), is proportional to the quantity it represents. While it is named for its resemblance to a pie which has been sliced, there are variations on the way it can be presented. The earliest known pie chart is generally credited to William Playfair's *Statistical Breviary* of 1801.

Pie charts are very widely used in the business world and the mass media. However, they have been criticized, and many experts recommend avoiding them, pointing out that research has shown it is difficult to compare different sections of a given pie chart, or to compare data across different pie charts. Pie charts can be replaced in most cases by other plots such as the [bar chart](#).



Pie chart of populations of [English](#) native speakers



Three sets of [data plotted](#) using pie charts and bar charts.

The circle or pie chart is a component parts bar chart. The component parts form the segments of the circle. The circle chart is usually a percentage chart. The data are converted to percentage of the total; and the proportional segments, therefore, give a clear picture of the relationship among the component parts. The name of segment and its percentage are placed inside its own area. When a segment is too small, an arrow is drawn to it and the legend is placed outside, in a horizontal position. The pie chart is commonly used for presenting the sectoral distribution of national income, the cost structure of a firm or any other type of simple percentage distribution.

### **MEASURES OF CENTRAL TENDENCY**

Measures of central tendency encapsulate in one figure a value that is typical for a distribution of values. It is also known as Statistical average.

#### **MEAN**

There are three types of mean

1. Arithmetic mean
2. Geometric mean
3. Harmonic mean

Simple arithmetic mean

The arithmetic mean is most commonly used statistical average in the disciplines such as commerce, management, economics, etc. the arithmetic mean of series of data is the sum of all the values divided by the total number. If  $x_1, x_2, \dots, x_n$  are the 'n' values of the variate x. the arithmetic mean of these values which is denoted by  $\bar{x}$  is defined by

$$\bar{x} = \frac{(X_1 + X_2 + \dots + X_n)}{n}$$
$$= \frac{\Sigma x}{n}$$

Where  $\Sigma x$  stands for the sum of all observations.

#### **Advantages**

1. The arithmetic mean is the most familiar and widely used measure of central tendency. It is simple to understand and easy to calculate.
2. It is rigidly defined.
3. It acts as a single representative value of the whole data.
4. Its calculation depends upon all the values in the series.
5. It is suitable for algebraic treatment.
6. It is least affected by sampling fluctuations.
7. It is useful in further statistical analysis, i.e. useful in the computation of standard deviation, correlation, coefficient of skewness, etc.

### **Disadvantages**

1. It is very much affected by the presence of a few extremely large or small values of the variable.
2. Mean cannot be calculated, if a single item is missing.
3. Arithmetic mean from a grouped frequency distribution cannot be calculated unless some assumptions are made regarding the sizes of the classes.
4. Arithmetic mean has no significance of its own.
5. For non-homogeneous data, the average may give misleading conclusions.

### **Median**

Median is another measure of central tendency. This is the mid-point in a distribution of values. Unlike arithmetic mean, median is based on the position of a given observation in a series arranged in ascending order. Therefore, it is called positional average. It is unaffected by the presence of an extremely large or small value. Median of a given series is the value of the variable that divides the series into two equal parts. It can be calculated from a grouped frequency distributions with the open-end classes.

#### Calculation of Median

Ungrouped Data: The given values are arranged in order of magnitudes. Median is calculated by formula  $((n+1)/2)$ th item, N being the numbers of items.

When N is odd: When the number of observations is an odd number, the median will be calculated by the formula  $N - ((N+1)/2)$  the item, where 'N' is the numbers of items.

### **Advantages**

1. It is simple to understand and easy to calculate.
2. For an open-end distribution median gives a more representative value.
3. For a qualitative phenomena, median is the most suitable average.
4. It is not affected by the extreme values
5. It is rigidly defined.

### **Disadvantages**

1. It is not based on all the values.
2. It is much affected by sampling fluctuations in comparison to mean
3. It is not suitable for algebraic treatment.
4. Its calculation depends on the arrangement of the data in order of magnitude.
5. The formula for median depends on the assumption that the items in the median class are uniformly distributed, which is not very true.

## **Mode**

Mode is another measure of central tendency. It is also a positional average like the median. The mode is defined as the most frequently occurring value. In other words mode is that value of frequency distribution whose frequency is maximum. But for some frequency distributions mode may not be most frequent value. It is that value of the variate around which other items tend to concentrate most heavily. For some distributions the mode may not exist and even if it exists it may not be unique as there may be more than one mode. A distribution having only one mode is called unimodal, the distribution having two modes is called bimodal and the distribution having more than two modes is called multimodal. Mode is often used in business. In many situations mode is more suitable than mean or median. For example, when we speak of “most common wage”. We mean modal wage is the usage that the largest number of workers receive. In the case of a shopkeeper who sells shoes, he is interested in knowing the size of shoes which are commonly demanded. In such a situation, the mean would indicate a size that may not fit any person. Mode will give most common size of shoe which is most usually purchased by the customers.

### **Advantages**

1. Mode is easily understood.
2. It is used widely for market research.
3. In certain situations mode is the only suitable average, e.g., modal size of shoes, modal wages etc.
4. For the preference of consumers product, the modal preference is considered.
5. Mode can be calculated for open end classes also provided the closed classes are of equal widths.
6. It is not affected by extreme values.

### **Disadvantages**

1. It is not always possible to find the well defined mode.
2. It is not suitable for further algebraic treatment.
3. It is not based on all the items of the data.
4. The value of the mode is affected significantly by the size of the class interval.

## **MODULE IV**

### **REPORT WRITING**

#### **INTRODUCTION**

The research task is not completed until the report has been written. The most brilliant hypothesis, the most carefully designed and conducted study, the most striking findings, are of little import unless they are communicated to others.

Report writing is the final stage of the research.

The research report is a means for communicating our research experiences to others and adding them to the fund of knowledge.

#### **Meaning and purpose of a research report**

A research report is a formal statement of the research process and its results. The social scientist who reads a research report needs to be told enough about the study. So that he can place it in its general scientific context, judge the adequacy of the methods and thus form an opinion of how seriously the findings are to be taken and-if he wishes- repeat the study with other subjects. In order to give him the necessary information, the report must cover the following points;

- Statement of the problem with which the study is concerned.
- The research procedures; the study design, the method of manipulating the independent variable if the study took the form of an experiment, the nature of the sample, the data collection techniques, the method of statistical analysis.
- The results.
- The implications drawn from the results.

The purpose of a research report is to communicate to interested persons the methodology and the results of the study in such a manner as to enable them to understand the research process and to determine the validity of the conclusions. The aim of the report is not to convince the reader of the value of the result but to convey to him what was done, why it was done, and what was its outcome. It is so written that the reader himself can reach his own conclusions as to the adequacy of the study and the validity of the reported results and conclusions.

#### **Characteristics of a report**

A research report is a narrative but authoritative document on the outcome of a research effort. It presents a highly specific information for a clearly designated audience. It is non-persuasive as a form of communication. Extra caution is shown in advocating a course of action even if the findings point to it. Presentation is subordinated to the matter being presented. It is a simple, readable and accurate form of communication.

### **Functions of Research Report**

A well written research report performs several functions.

1. It serves as a means for presenting the problem studied, methods and techniques used for collecting and analyzing data ,the findings ,conclusions and recommendations in an organized manner.
2. It serves as a basic reference material for future use in developing research proposals in the same or related area.
3. A report serves as a means for judging the quality of the completed research project.
4. It is a means for evaluating the research's ability and competence to do research
5. It provides factual base for formulating policies and strategies relating to the subject –matter studied.
6. It provides systematic knowledge on problems and issues analyzed.

### **Essentials of a good report**

The following are said to be the essentials of a good report.

- Clarity and coherence.
- Writing correctly.
- Styled to the Reader's taste.
- Readability.
- Effective arrangement.

### **IV-2 TYPES OF REPORTS**

Research reports may be classified into (1)synopsis(2)research proposal (c)comprehensive report for the academic community .These types of reports vary from one another in terms of the degree of formality, physical form, scope, style and size.

#### **Report for the academic community**

This is a comprehensive full report of the research process and its outcome. It is primarily meant for academic community, i.e., scientists of the researcher's discipline and other researchers .It is formal long report covering all the aspects of the research process ;a description of the problem studied, the objectives of the study, methods and techniques used a detailed account of sampling, field and other research procedures, sources of data, tools and techniques of data collection, methods of data processing and analysis, detailed findings and conclusions and suggestions. There is also a technical appendix for methodological details, copies of measuring instruments and the like. It is so comprehensive and complete that the study can be replicated by others.



## **Synopsis**

This is a short summary of the technical report. It is usually prepared by a doctoral students on the eve of submitting, his thesis .Its copies are sent by the university along with the letters of request to the examiners invited to evaluate the thesis. It contains a brief presentation of the statement of the problem, the objectives of the study, methods and techniques used and an overview of the report. A brief summary of the results of the study may also be added. This synopsis is primarily meant for enabling the examiner-invitees to decide whether the study belongs to the area of their specialization and interest.

## **Research proposal**

All research endeavors in every academic and professional field are preceded by a research proposal. It informs academic supervisor or potential provider of a research contract of researcher's conceptualization of the total research process that he propose to undertake, and examines its suitability and validity. In any academic field, research proposal will go through a number of committees for approval. Certain requirements for a research proposal may vary from university to university, within a university from discipline to discipline, but what is outlined here will satisfy most requirements. A research proposal is an overall plan, scheme, structure and strategy designed to obtain answers to the research questions or problems that constitute one's research project. A research proposal should outline the following. Objectives, text hypothesis, reasons for selecting the study, area of study, variables under study, tool for study, analysis etc. hence it is important to study what constitute a research proposal? And how to write a good research proposal?

## **IV-1PLANNING REPORT WRITING**

### **Steps in planning Report Writing**

After the data analysis is over, report writing cannot be started abruptly .It requires careful pre-planning. This planning process involves the following considerations and steps:

As a research report is a means of communication, we have to consider some basic questions which determine the effectiveness of communication ,namely,' who' says 'what' to 'whom' in 'which way' and with 'what effect.'

**1. The Target Audience:** The first step in planning report writing is to determine the target audience. The form and style of report in and other aspects depend upon the type of the reader, for whom the report is intended. The identification of the target audience depends on who is the researcher and what is his intention

The target audiences, may be classified in to 1,the academic(or scientific )community, 2,the sponsors of research and 3,the general public.

**1.1 The academic or scientific community** will be the primary target audience in the following cases: (1) when the research is undertaken as an academic exercise for Master's degree, or M.Phil. Degree or Ph.D degree (in this case thesis evaluation committee will be the immediate target audience); (2) when a research student or social scientist plans to publish his research output in the form of a research monograph; (3) when a researcher plans to write research articles based on his research for publication in professional journals. (in last two cases, the referees, and the fellow scientists interested in the study will be the target audience).

**1.2 The Sponsors of Research** may consist of two categories: (1) research promotion bodies like Indian Council of Social Science Research, the University Grants Commission, and educational foundations which provide financial support to social scientists working in universities and colleges for undertaking research with a view to encouraging them to do researches; and (2) government department, industrial and other organizations which sponsor research for their *own use* in policy making and the like.

If the research is sponsored by a research promotion organization, the reporting has to follow its prevalent norms. In general a full-fledged technical report is expected, along with an abstract of the report. When the research is sponsored by an organization for its own use, it has to be reported according to its requirements. It is to be written as a private document, emphasizing the findings and recommendations rather than methodology.

**1.3 The general public** is viewed as a cross-section of the society. This lay audience may be interested in the broad findings and the implications of research studies on socio-economic problems. The reporting for this audience may be in the form of a summary report or an article written in non-technical journalistic language.

The communication characteristics, viz., the level of knowledge and understanding the information needs and the kind of language to which one is accustomed, are not the same. They vary from one type of audience to another. The preference and the requirements of different audiences differ widely and cannot be reconciled. Hence it is neither possible nor desirable to attempt to write one multipurpose report. A separate report tailored to the needs of each type of audience has to be written when there is a need for communicating to different types of audience.

**2. The communication characteristics of the audience:** The second step in planning report writing is to consider the selected audience's communication characteristics such as;

- What is their level of knowledge and understanding?
- What is the gap in knowledge on the subject between the readers and the writer?
- What is the kind of language –scientific or journalistic– that which the readers are accustomed?
- What do they need to know about the study?
- what is likely to be of interest to them?
- How can the needed information be presented best-verbal, a combined tabular or pictorial presentation?

These questions determine the scope, form and style of reporting. The underlying purpose of a report should be noted. The purpose of report is not communication with oneself, but communication with the target audience. Hence we must constantly keep in mind the needs and requirements of the target audience.

**The intended purpose of the report:** What is the intended purpose of the report? Is it meant for evaluation by experts for the award of a degree or diploma? Is it to be used as a reference material by researchers and fellow scientists? Or it is meant for implementation by a user-organization? This intended purpose also determines the type of the report and its contents and form of presentation.

**The type of report:** With reference to the intended use, the type of report to be prepared should be determined. When the researcher undertaken to fulfill the requirements of a degree or diploma or funded by a research promotion agency, the report is prepared as a comprehensive technical report. When it is sponsored by a user-organization, it is written as a popular or summary report.

**The scope of the report:** The next step is to determine the scope of the contents with reference to the type of the report and its intended purpose. For example, a research thesis or dissertation to be submitted for award of a degree or diploma should narrate the total research process and experience; the state of the problem, a review of previous studies ,objectives of the study, methodology, findings, conclusions and recommendations.

**The style of reporting:** What should be the style of reporting? Should it be simple and clear or elegant and pompous? Should it be technical or journalistic? These questions are decided with reference to the target audience. For a detailed discussion on style, see section 12.5 principles of writing, below.

**The format of the report:** The next step is to plan the format of the report, which varies according to the type of report.

**Outline/Tables of contents:** The final step in planning report writing is to prepare a detailed outline for each of the proposed chapters of the report. An outline lends cohesiveness and direction to report writing work. Until an outline is prepared, the researcher does not know that he has to do and how to organize the presentation.

## **BODY OF THE REPORT**

After the prefatory items the body of the report is presented. It is the major and main part of the report. It covers the formulation of the problem studied, methodology, findings and discussion and a summary of the findings and recommendations. In a comprehensive report, the body of the report will consist of several chapters.

## **1. Introduction**

This is the first chapter in the body of the research report. It is devoted for introducing the theoretical background of the problem, its definition and formulation .It may consist of the following sections.

**(a)Theoretical background of the topic:** The first task is to introduce the background and the nature of the problem so as to place it into a *larger context* to enable the reader to know its significance in a proper perspective. This section summarizes the theory or *conceptual framework* within which the problem has been investigated. For example, the theoretical background in the thesis entitled, "A Study of Social Responsibilities of Large-scale industrial units in India's constitution to establish an egalitarian society order ,the various approaches to the concept of social responsibility – property rights approach ,truseeship approach, letimacy approach , social responsibility approach – their implications for industries in the Indian context. Within this conceptual framework ,the problem was defined , the objectives of the study were set up, the concept of social responsibility was operationalised and the methodology of investigation was formulated.

Similarly,the theoretical background in another doctoral thesis on "the relationship between capital structure and cost of capital in larger cooperative undertakings ",deals with an overview of the financial management decisions ,the characteristics of cooperatives , the irrelevance of wealth maximization goal to cooperatives and the alternative 'capital-cost – minimization' objective of financial management in cooperatives .The problem under study was formulated within this conceptual framework.

**The Statement of the Problem:** In this section why and how the problem was selected are stated, the problem is clearly defined and its facets and significance are pointed out .

**Review of Literature :** This is an important part of the introductory chapter. It is devoted for making a brief review of previous studies on the problem and significant writings on the topic under study .This review provides a summary of the *current state of knowledge* in the area of investigation. Which aspects have been investigated ,what *research gaps* exist and how the present study is an attempt to fill in that gap are highlighted .Thus the underlying purpose is "to locate the present research in the existing body of research on the subject and to point out what it contributes to the subjects."

**The scope of the present study:** The dimensions of the study in terms of the geographical area covered, the designation of the population being studied and the level of generality of the study are specified.

**The Objectives of the Study:** The objectives of the study and investigative questions relating to each of the objectives are presented.

**Hypotheses:** The specific hypotheses to be tested and are started .The sources of their formulation may be indicated.

**Definition of Concepts:** The reader of a report is not equipped to understand the study unless he can know what concepts are used and how they are used. Therefore, the operational definitions of the key concepts and variable 's of the study are presented, giving justifications for the definitions adopted. How those concepts were defined by earlier writers and how the definitions of the researcher were an improvement over earlier definitions may be explained.

**Models:** The models, if any, developed for depicting the relationships between variables under study are presented with a review of their theoretical or conceptual basis. The underlying assumptions are also noted.

### **The Design of the study**

This part of the report is devoted for the presentation of all the aspects of the methodology and their implementation, viz., overall typology, methods of data collection, sample design, data collecting instruments methods of data processing and plan of analysis. Much of this material is taken from the research proposal plan. The revisions, if any made in the initial design and the reasons therefore should be clearly stated.

If *pilot study* was conducted before designing the main study, the details of the pilot study and its outcome are reported. How the outcome of the pilot study was utilized for designing the final study is also pointed out.

The details of the study's design should be so meticulously stated as to fully satisfy the criterion of replicability. That is, it should be possible for another researcher to reproduce the study and test its conclusions.

Technical details may be given in the Appendix. Failure to furnish them could cast doubts on the design.

- (a) **Methodology:** In this section, overall typology of research (i.e., experimental, survey, case study, or action research) used, and the data collection methods (i.e., observation, interviewing or mailing) employed are described. The sources of data, the sampling plan and other aspects of design may be presented under separate subheadings as described below.
- (b) **Sources of Data:** The sources from which the secondary and or primary data were gathered are stated. In the case of primary data, the *universe* of the study and the unit of study are clearly defined. The limitations of secondary data should be indicated.
- (c) **Sampling Plan:** The size of the universe from which the sample was drawn, the sampling methods adopted and the sample size and the process of sampling are described in this section. What were originally planned and what were actually achieved and the estimate of *sampling error* are to be given. These details crucial for determining the limitations of generalisability of the findings.

- (d) **Data- Collection Instruments:** The types of instruments used for data collection and their contents ,scales and their devices used for measuring variables, and the procedure of establishing their validity and reliability are described in this section  
How the tools were *pre-tested* and finalized are also reported.
- (e) **Field Work:** When and how the field work was conducted, and what problems and difficulties were faced during the field work are described under this sub-heading. The description of field experiences will provide valuable lessons for future researches in organizing and conducting their field work.
- (f) **Data Processing and Analysis Plan:** The method-manual or mechanical-adopted for data processing ,and an account of methods used for data analysis and testing hypotheses must be outlined and justified. If common methods like chi-square test, correlation test and analysis of variance were used ,it is sufficient to say such and such methods were used. If an unusual or complex method was used, it should be described in sufficient detail with the formula to enable the reader to understand it.
- (g) **An Overview of the Report:** The *scheme* of subsequent *chapters* is stated and the purpose of each of them is briefly described in this section in order to give an overview of presentation of the results of the study .
- (h) **Limitations of the Study:** No research is free from limitations and weaknesses. These arise from methodological weaknesses, sampling, imperfections, non-responses, data inadequacies, measurement deficiencies and the like. Such limitations may vitiate the conclusions and their generalisibility .Therefore a careful statement of the limitations and weaknesses of the study should be made in order to enable the reader to judge the validity of the conclusions and the general worth of the study in the proper perspective .A *frank statement of limitations* is one of the hallmarks of an honest and competent researcher.

### **Documentation**

The ethics of scholarship require proper acknowledgement of all source materials by the writer.

There are two alternative models for documenting sources of ideas and information:

1. Footnotes
2. References-cited format

### **Footnotes**

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

Footnotes serve several purposes

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

### **Reference cited format.**

The reference cited format of documentation consists of a single listing of research reference at the end of a paper or thesis. This is preferred in most scientific writings. It is most often headed 'Reference cited' or 'References. Sources are referred to in the text by author and year of publication and page number.

References are listed in alphabetical order at the end of the report/paper without numbering. They include page numbers, if page numbers are not mentioned in the references cited in the text

### **Bibliography**

The Bibliography is a list of reference relating to a topic or subject. It is presented at the end of the research project. It contains all the information found in a first footnote relating to a work. The bibliography lists in alphabetical order all published and unpublished references used by the writer in preparing the report. All books, articles and reports and other documents may be presented in one common list in the alphabetical order of their authors. Bibliography may be classified into following sections;

- Books
- Articles
- Reports
- Other documents

The Bibliography gives a list of materials relating to the topic under study as a ready reference to the reader. Bibliography listing should be done with proper format in order to serve its purpose. There are several well-established systems for writing a bibliography. The choice is dependent upon the preference of the discipline and university. In the most commonly used ones are (Longyear 1983:83)

- The Harvard system;
- The American psychological Association system;
- The American Medical Association System;
- The Modern Languages Association system;
- The footnote system

### **Examples of Bibliographical forms**

- Books:  
Kaplan, Abraham., *The Conduct of Inquiry: Methodology for Behavioural Science*, San Francisco: Chandler Publishing Co,1964.
- Journals (periodicals)  
Neale, W.C., “The Limitations of Indian Village Survey Data”, *Journal of Asian Studies*, 17, 1958, pp. 383-402.
- Report  
world Bank, *World development Report 1987*, Washington D.C, 1987

### **Appendices**

An appendices is a plural borrowed directly from Latin, is sometimes used, especially in scholarly writing, to refer to supplementary material at the end of a book. An APPENDIX gives useful additional information, but even without it the rest of the book is complete. An appendix contains supplementary material that is not an essential part of the text itself but which may be helpful in providing a more comprehensive understanding of the research problem and/or is information which is too cumbersome to be included in the body of the paper. A separate appendix should be used for each distinct topic or set of data and always have a title descriptive of its contents.

The following documents are included in appendix.

- Copies of data collection instruments, eg: interview schedule or questionnaire used for study.
- Complex or simple tables.
- Supporting documents
- Glossary of new concepts

An appendix follows the bibliography. If a study is of a major importance and is to be published in book or monograph form, the researcher also prepares an index in alphabetical order which follows the appendix.