Post Graduate Diploma in Agricultural Extension Management (PGDAEM)

AEM-105
Research Methods in Agricultural Extension
(2 Credits)

National Institute of Agricultural Extension Management
(An Organization of the Ministry of Agriculture, Govt. of India)
Rajendranagar, Hyderabad – 500 030, Andhra Pradesh, India
www.manage.gov.in
Published by
National Institute of Agricultural Extension Management,
Rajendranagar, Hyderabad – 500 030, Andhra Pradesh, India

First Published: 2007
Revised 2013
© MANAGE, 2007
All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means without permission in writing from the MANAGE.

Shri. B. Srinivas, IAS
Director General
National Institute of Agricultural Extension Management
(MANAGE), Rajendranagar, Hyderabad – 500 030,
Andhra Pradesh, India

Program Coordinators
Dr. S. Senthil Vinayagam, Director (Agril. Extn.) & Principal Coordinator (PGDAEM), MANAGE
Dr. K. Uma Rani, Deputy Director (Extn)
Dr. M.A. Kareem, Deputy Director (Agril. Extn)

Contributors (2007)
Dr. B.C. Muthayya, NIRD, Hyd
Dr. P. Satish Chandra, NIRD, Hyd.
Dr. C.S. Singhal, NIRD, Hyd.
Dr. T.D.S. Kumar, MANAGE, Hyd.
Dr. P. Kanaka Durga, MANAGE, Hyd.

Contributors (2013)
Dr. G.A.K. Kumar, P.S., Central Rice Research Institute, Cuttack
Dr. A.K. Mohanty, Senior Scientist and Principal Coordinator, ICAR Complex, Sikkim
Dr. R.M. Prasad, Associate Director of Research (Retd.), Kerala Agril. University
AEM-105
Research Methods in Agricultural Extension
(2 Credits)

Research Methods in Agricultural Extension

<table>
<thead>
<tr>
<th>Unit</th>
<th>Topic</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit - 1</td>
<td>Fundamentals of Research in Extension</td>
<td>5 – 14</td>
</tr>
<tr>
<td>Unit - 2</td>
<td>Identification of Research Problem and Formulating Research Design</td>
<td>15 – 32</td>
</tr>
<tr>
<td>Unit - 3</td>
<td>Literature Review and Developing Theoretical Orientation</td>
<td>33 – 42</td>
</tr>
<tr>
<td>Unit - 4</td>
<td>Sampling Methods</td>
<td>43 – 54</td>
</tr>
<tr>
<td>Unit - 5</td>
<td>Data Collection Methods</td>
<td>55 – 74</td>
</tr>
<tr>
<td>Unit - 6</td>
<td>Analysis of Data and Interpretation</td>
<td>75 – 108</td>
</tr>
<tr>
<td>Unit - 7</td>
<td>Report Writing</td>
<td>109 - 118</td>
</tr>
<tr>
<td>Unit</td>
<td>Topic</td>
<td>Pages</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Unit 1</td>
<td>Fundamentals of Research in Extension</td>
<td>5 – 14</td>
</tr>
<tr>
<td>Unit 2</td>
<td>Identification of Research Problem and Formulating Research Design</td>
<td>15 – 32</td>
</tr>
<tr>
<td>Unit 3</td>
<td>Literature Review and Developing Theoretical Orientation</td>
<td>33 – 42</td>
</tr>
<tr>
<td>Unit 4</td>
<td>Sampling Methods</td>
<td>43 – 54</td>
</tr>
<tr>
<td>Unit 5</td>
<td>Data Collection Methods</td>
<td>55 – 74</td>
</tr>
<tr>
<td>Unit 6</td>
<td>Analysis of Data and Interpretation</td>
<td>75 – 108</td>
</tr>
<tr>
<td>Unit 7</td>
<td>Report Writing</td>
<td>109 – 118</td>
</tr>
</tbody>
</table>
Unit 1

Fundamentals of Research in Extension

Structure

1.0 Objectives

1.1 Why do we conduct research?

1.2 How do we KNOW?

1.3 What is Science?

1.4 What are the assumptions of Science?

1.5 What is the aim of Social Science including Extension Education?

1.6 How important is research methodology?

1.7 How does research proceed?

1.8 What are the types of research?

1.9 Let us sum up

1.0 Objectives

By the end of this unit you will be able to understand

- Fundamentals of research methodology and
- Its application in field extension

1.1 Why do we conduct research?

As a post-graduate student of PGDAEM, you must be wondering why we conduct the research. You may also be bothering, ‘How research is conducted? ‘Am I suppose to invent something?.....‘Will I be able to cope with the jargons of research?....‘ and likewise, a number of doubts would arise in your mind. therefore, utmost important for
you to understand and know the reasons behind conducting research, in our social research or more specifically, extension research.

In fact, in the era of information explosion where information is available at the tip of the finger, it is difficult to visualize the importance of generating these information. Think of the time when the first *Homo sapien* was born. What all he knew about the world or the universe or the galaxy at that time? Not even what a three or four years old child of today knows. Probably, his world was restricted to his basic needs and his surroundings. Slowly but steadily, he started to KNOW.

### 1.2 How do we ‘KNOW’?

To know how do we know, we have to learn about the pioneer works of Charles S. Peirce (1839-1914). In his article on ‘Fixation of Belief’, Peirce explained the difference between doubt and belief. ....Doubt is an uneasy and dissatisfied state from which we struggle to free ourselves and pass into the state of belief which is a calm and satisfactory state which we do not wish to avoid, or to change to a belief in anything else.

There are four methods of fixing belief or say knowing something. The first is the *method of tenacity*. It means holding or tending to hold persistently to something, such as a point of view. We believe what we have always believed to be truth. Say for example, every child believes that his father is the greatest person on this earth. Peirce further explains....*The man who adopts this method of knowing will find that other men think differently from him, and it will be apt to occur to him, in some saner moment, that their opinions are quite as good as his own, and this will shake his confidence in his belief..... we shall necessarily influence each other’s opinions; so that the problem becomes how to fix belief, not in the individual merely, but in the community*. The second method takes care of fixing the belief in the community, which may be called the *method of authority*. This method has, from the earliest times, been one of the chief means of upholding correct theological and political doctrines, and of preserving their universal character. For example, many persons fix their belief based on the religious books such as the *Bible* for the Christians, the *Koran* for the Muslims and the Vedas for the Hindus.
However, no institution can regulate opinions of every individual. There are logical thinkers and they fix their belief through a superior and intellectual method of knowing; the method of intuition or a priori method. This method involves deductive reasoning, i.e. proceeding from a known or assumed cause to a necessarily related effect. It is derived without reference to particular facts or experience and made before or without examination; not supported by factual study. According to Pierce ....its failure has been the most manifest. It makes of inquiry something similar to the development of taste; but taste, unfortunately, is always more or less a matter of fashion, and accordingly metaphysicians have never come to any fixed agreement, but the pendulum has swung backward and forward between a more material and a more spiritual philosophy, from the earliest times to the latest.

.....To satisfy our doubts, therefore, it is necessary that a method should be found by which our beliefs may be determined by nothing human, but by some external permanency – by something upon which our thinking has no effect.

It must be something which affects, or might affect, every man. And, though these affections are necessarily as various as are individual conditions, yet the method must be such that the ultimate conclusion of every man shall be the same. Such is the method of science, the fourth method of fixing belief.

1.3 What is science?

It is amazing that many educated persons do not know what science is. Although the word science is so familiar, its meaning is not widespread. As a student, you may be interested in more definitions of science. According to Wikipedia, Science, in the broadest sense, refers to any system of objective knowledge. In a more restricted sense, science refers to a system of acquiring knowledge based on the scientific method, as well as to the organized body of knowledge gained through researches. Let us first understand what objective knowledge or objectivity in science is. It refers to the property of scientific
theories to make unequivocal predictions that can be tested by anybody. The results can be reproduced by others the systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles based on these facts therefore, unlike other three methods of knowing, the knowledge generated through method of science is amenable to testing.

**Scientific Method**

In Science, knowledge is acquired using **Scientific method**. Scientific method is a body of techniques for investigating phenomena and acquiring new knowledge, as well as for correcting and integrating previous knowledge. It is based on gathering observable, empirical, measurable evidence, subject to the principles of reasoning. Scientific researchers propose specific hypotheses as explanations of natural phenomena, and design experimental studies that test these predictions for accuracy. These steps are repeated in order to make increasingly dependable predictions of future results. Theories that encompass wider domains of inquiry serve to bind more specific hypotheses together in a coherent structure. This in turn aids in the formation of new hypotheses, as well as in placing groups of specific hypotheses into a broader context of understanding.

Among other facets shared by the various fields of inquiry is the conviction that the process must be objective so that the scientist does not bias the interpretation of the results or change the results outright. Another basic expectation is that of making complete documentation of data and methodology available for careful scrutiny by other scientists and researchers, thereby allowing other researchers the opportunity to verify results by attempted reproduction of these results. This also allows statistical measures of the reliability of the results to be established. The scientific method also may involve attempts, to achieve control over the factors involved in the area of inquiry, which may in turn be manipulated to test new hypotheses in order to gain further knowledge.
1.4 What are the assumptions of science?

The scientific approach has been developed with a set of assumptions that cannot be proved. This will also help in understanding the superiority of scientific method over other methods.

Nature is orderly and regular. The basic assumption of scientific approach is that there exists a definite regularity and order in the nature. Lorenz’s works on the chaos theory depicts the same assumption. Another example could be that of atmospheric changes that lead to distinct pattern of climate - winter, summer, rainy season.

We can know nature. Human mind is capable of knowing nature and also itself. It can understand, comprehend, analyze, interpret, and infer.

Knowledge is superior to ignorance. The state of knowledge is a relative term. What we know today may be disproved tomorrow. The man has been restless, inquiring and soul searching because he knows the power of knowledge.

All natural phenomena have natural causes. Science assumes the existence of cause in nature for every effect. It rules out the possibility of metaphysical explanation until the science cannot account for the causation of the natural phenomena.

Nothing is self-evident. Science is not common sense. The truth must be objectively verifiable.

Knowledge is derived from the acquisition of experience. If science has to tell anything about the real world, then it should be empirical, that it must rely on perception, experience and observation.

1.5 What is the aim of Social Science including Extension Education?

Broadly speaking the aim of science itself is explaining natural phenomena. Similarly, the aim of social science is to explain social phenomena that occur in natural setting. When these explanations are applied to solve some societal problems, the branch of science is called applied social science. Extension Education, an applied social science, aims to explain social phenomena and solve problems related to development of individual
or the community at large. That is, the scientific explanations are used to predict and control the social phenomena for the greater interest of the society.

1.6 How important is research methodology?

Before we understand the importance of research, let us review the definition of scientific research presented. Now, if the research has to be systematic, controlled, empirical and critical, then we cannot go haphazardly but need to follow a systematic approach. We should have methods to create controlled situation. The results of the research should be empirical, that is, it should be capable of observation. When we say observation, it inherently means observable by our senses. Lastly, it should be based on the critically examined processes and results. In order to achieve these strictly laid down conditions, a standard methodology is required to conduct the research.

1.7 How does research proceed?

Fundamentally, the research is a process. It is cyclic in nature. Theory is the product of this process which also becomes input for future researches. The process of research starts with problem identification, setting objectives, formulating hypothesis, finalizing research design, determining measurement processes, collecting data, anglicizing data and finally making generalization. These generalizations help develop theory. It should be noted that each step is reciprocally related to the theory. As for example, previously established theory may help in identifying problem which in turn acts as the part of the process of developing or refining theory. It should be noted that the research is a process which
does not end. What do end are The efforts of an individual researcher. Is further carried forward by other researchers; hence we can say that the research goes on.

1.8 What are the types of research?

When we say types of research, we mean different ways of conducting research but having same ultimate goal. When earlier social researchers first started to use research method as means of knowing, (knowing with objectivity without any subjectivity), they started with anthropological type of research – wherein, they were interested to know what is where. That means exploring the study area. This type of research is called exploratory research. Exploratory research is often conducted because a problem has not been clearly defined as yet, or its real scope is as yet unclear. It allows the researcher to familiarize him/herself with the problem or concept to be studied, and perhaps generate hypotheses (definition of hypothesis) to be tested. The results of exploratory research are not usually useful for decision-making by themselves, but can provide significant insight into a given situation.

Once we were clear about ‘what is where’, the next question was ‘what is what’. In other words, describing different things (which were later on called as variables). This type of research is called descriptive research. There are three main types of descriptive methods: observational methods, case-study methods and survey methods.

Observational Method: Under this method, animal and human behavior are closely observed in natural or laboratory conditions which are also called naturalistic observation and laboratory observation, respectively. Naturalistic observation is
having greater ecological validity than laboratory observation. Laboratory observations are usually under controlled conditions, less time-consuming and cheaper than naturalistic observations. However, both naturalistic and laboratory observations are important for the advancement of scientific knowledge.

Case Study Method: Case study research involves an in-depth study of an individual or group of individuals. Case studies often lead to testable hypotheses and allow us to study rare phenomena. Case studies cannot be used to determine cause and effect, and they have limited use for making accurate predictions.

Survey Method: In survey method research, participants answer questions administered through interviews or questionnaires. After participants answer the questions, researchers describe the responses given. In order for the survey to be both reliable and valid it is important that the questions are constructed properly. Questions should be framed so they are clear and easy to comprehend.

It is important for you to understand that descriptive research methods can only describe a set of observations or the data collected. It cannot draw conclusions from that data about which way the relationship goes — Does A cause B, or does B cause A? This leads to another type of research called explanatory research. In case of explanatory type of research, we are interested to know what causes which kind of effect. This brings in the topic of Principle of Cause and Effect relationship. The cause and effect principle is based on four basic characteristics. These four characteristics are as follows:

1. Causes and effects are the same things.
2. Causes and effects are part of an infinite continuum of causes.
3. Each effect has at least two causes in the form of actions and conditions.
4. An effect exists only if its causes exist at the same point in time and space.

You might be surprised to know that cause and effect are same things. Actually, it is our perception based on which you may call something cause or effect. Following illustration will make the things clearer for you.
As you can see in the figure above, what is a cause for some event is an effect of some other cause. So, it is the perspective from which you look at it. In above case, it is a cyclic event, but it can be a linear cause and effect relationship. One can also see that other three basic characteristics also holds good in above example.

The fourth type of research is experimental research. One of the most important things for you to learn in this course is the difference between explanatory research (also called correlation research) and experimental research. Like explanatory research, experimental research concerns relationships between variables. Unlike explanatory research, however, experimental research provides strong evidence for causal interpretations. Here we will focus on the two most important features of experimental research.

1. Manipulation of an Independent Variable
2. Control of Extraneous Variables

One important feature that distinguishes experimental research from explanatory research is that instead of simply measuring two variables, the researcher manipulates one of them. This means that the experimenter actually changes the value of that variable in a systematic way. This variable, which is called the independent variable, is the one that the researcher believes is the cause. The other variable, which the researcher believes is the effect, is called the dependent variable.

The second feature that distinguishes experimental research from explanatory research is the control of extraneous variables. Extraneous variables are basically all variables other than those you are interested in for purposes of your research. To control extraneous variables means to keep their values or levels as similar as possible across the different values or levels of your independent variable.
The obvious advantage of experimental research is that it provides stronger evidence for causal claims. It does, however, have at least two limitations. The first is that sometimes you cannot do an experiment because you cannot manipulate the independent variable, either for practical or ethical reasons. The second limitation of experimental research is that sometimes controlling extraneous variables means creating situations that are somewhat artificial.

So far what we meant by experimental research is field experiments but there is yet another type of experimental research called – simulation or in our case social simulation. By definition simulation is the imitation of the operation of a real-world process or system over time. It aims to cross the gap between the descriptive approach used in the social sciences and the formal approach used in the hard sciences, by moving the focus on the processes/mechanisms/behaviors that build the social reality. The simulation research has advantage over the field experiments that it can manipulate any independent variable.

1.6 **Let us sum up**

There are four methods of knowing – method of tenacity, method of authority, method of intuition or a priori method and method of science.

Science is the systematic observation of natural events and conditions in order to discover facts about them and to formulate laws and principles based on these facts. Science implies two things, a body of knowledge as well as a system of methods.

The aim of science/social science itself is explaining natural phenomena. The purpose of research is to discover answers to questions or problems through the application of scientific procedures. It is a systematic, controlled, empirical and critical investigation of hypothetical propositions about the presumed relations among natural phenomenon.

Basically, there are four types of research – exploratory, descriptive, explanatory and experimental.
Unit 2

Identification of Research Problem and Formulating Research Design Structure

Structure

2.0. Objectives
2.1. Why do we need concepts in Extension research?
2.2. How to decide the research problems?
2.3. Why and how to set research objectives?
2.4. How to select variables?
2.5. What relationship exists between variables?
2.6. What are the different types of variables?
2.7. Is it necessary to formulate hypothesis?
2.8. What will make possible the testing of hypothesis?
2.9. Which research design should be used and why?
2.10. Examples of research designs
2.11. Let us summing up

2.0 Objectives

By the end of this unit you will be able to understand

- The concept
- Consideration in selecting research problem
- How to state research objectives
- How to identify variables
- How to formulate hypothesis and
- What is research design
2.1 Why do we need concepts in Extension research?

The scientific knowledge can be proved by reasoning and experience. This means that extension researchers operate at two distinct but interrelated levels - conceptual-theoretical and observational-empirical. Extension research is the outcome of the interaction between these two levels. The science begins with forming concepts to describe empirical world.

When an extension researcher wants to communicate to another researcher, s(he) will require certain connotations to use so that both can exchange their ideas. The concepts are such established connotations. If we say ‘adoption’, it arouses certain abstraction in the mind of the researcher. It provides a point of view - a way of looking at it empirically. The concepts help us in classification and generalizing in extension research. Lastly, the concepts serve as building blocks of theories and thus explanations and predictions. Therefore, if one wants to conduct research in extension then, it is must to transit from conceptual level to observational level. We may take the example of Education. As concept, we know that education is a lifelong process - only its form may change from time to time. Now at observational level, we boil it down to the number of the years of formal education the respondent has successfully undergone. The concept(s) at observational level is popularly explained as operational definition.

2.2 How to decide the research problems?

Please recall cycle of research process presented in previous unit. In this process, research problem, variable, relation and hypothesis are the most important and basic elements of research. They help
transform an idea into concrete research operations. Commonly, the research problem is selected on *ad hoc* basis which is the most vital and basic element. The extension students often select problem that has been studied earlier by other researcher, and pursue their research by changing the crop or locale or sampling unit of the previously studied problem.

Instead of following this approach you should, rather, think of what caused the need to do the research (problem identification). The question that you should ask is: Are there questions about this problem to which answers have not been found up till the present?

In fact, research originates from a need that arises. You should make a clear distinction between the PROBLEM and the PURPOSE. The problem is the aspect the researcher worries about, thinks about, wants to find a solution for. The purpose is to solve the problem, i.e. find answers to the question(s). If there is no clear problem formulation, the purpose and methods are meaningless. An extension student should keep the following in mind:

- Outline the general context of the problem area.
- Highlight key theories, concepts and ideas current in this area.
- What appear are the underlying assumptions of this area?
- Why are these identified issues important?
- What needs to be solved?
- Read around the area (subject) to get to know the background and to identify unanswered questions or controversies, and/or to identify the most significant issues for further exploration.

### Three criteria of good problem or problem statement

1. Problem should express a relation between two or more variables.
2. The problem should be stated clearly and unambiguously in question form.
3. Problem statement should be such as to imply possibilities of empirical testing.
The research problem should be stated in such a way that it would lead to analytical thinking on the part of the researcher with the aim of possible concluding solutions to the stated problem. **Research problems can be stated in the form of either questions or statements.** The research problem should always be formulated grammatically correct and as completely as possible. You should bear in mind the wording (expressions) you use. Avoid meaningless words. There should be no doubt in the mind of the reader what your intentions are. Demarcating the research field into manageable parts by dividing the main problem into sub-problems is of the utmost importance.

**Examples**

1. Is education related to adoption of hybrid rice variety?
2. Will demonstration of rice-cum-fish culture increase the knowledge of small and marginal farmers?
3. How much information flows through mass media about banned chemicals to the farmers?
4. Does extension literature has effect on attitude of farmers towards clean milk production?

**EXERCISE for Students**

Formulate five problem statements from your situation at work. Please write it on paper for future use.

**2.3 Why and how to set research objectives?**

Having identified the research problem, you need to set your research objectives. You may ask why it is so important. Ok, now think that you need to go to Delhi (This is already a set objective) and you will not start randomly traveling in any direction. What is the probability that you will reach Delhi; it may be a fraction of a million chances. But, Based on your objective to reach Delhi, insert and ultimately, you will reach. So you should understand that an important part of managing your research project is setting your own clear and achievable objectives. This will give you direction as it will help you focus on the things that matter to your research and future plans.
Examples

Take the same examples of problems.

PROBLEMS

1. Is education related to adoption of hybrid rice variety?
2. Will demonstration of rice cum fish culture increase the knowledge of small and marginal farmers?
3. How much information flows through mass media about banned chemicals to the farmers?
4. Does extension literature have effect on attitude of farmers towards clean milk.

5. RESPECTIVE OBJECTIVES

1. To study the factors affecting adoption of hybrid rice variety.
2. To assess the increase in knowledge of small and marginal farmers on rice cum fish culture due to selected extension interventions.
3. To analyze information flows through mass media about banned chemicals to the farmers.
4. To study the effect of extension literature on attitude of farmers towards clean milk production.

You may use following words to suitably state your research objectives. This is not an exhaustive list.

SMART: A tool to evaluate your objectives

SMART can help you critically evaluate the objectives you have set. For every objective ensure it meets the following criteria.

Specific: in both meaning and focus.

Measurable: you should be able to measure.

Achievable: do not set objectives which cannot be achieved with available resources

Realistic: make sure that you are being realistic.

Time limited: set deadlines and 'milestones', times when you will sit down and reflect on and
EXERCISE for Students

Write five objectives using any five words from above list.

Facilitator may randomly check the objectives set by the objectives. See that it adheres to SMART criteria.

2.4 How to select variables?

Research problems are conveyed with a set of concepts. In order to move from conceptual to empirical level, we convert concepts into variables by mapping concepts into a set of values. Our concepts will eventually appear as variables in hypothesis to be tested. Let us consider the example of education, the concept. Now when we map, we give values to illiterate, primary education, secondary education and so on as 0, 1, 2, 3, etc....
As a student of extension research, we generally select variables from the set of variables reported in past researches/thesis reports. In fact, if we can define the research problem properly, we can get the list of the variables to be selected from the concepts used in the research problem statements and sub-problem statements. Since the research problem statements are supposed to be based on rigorous review of literature, the variables to be studied are also based on the recourses of review of literature. Unfortunately, either review of literature is ignored or is conducted at a later stage of the research process. So you should rigorously review related literature to select suitable variables that will help further your research.

EXAMPLE

Let us consider objective 1 stated earlier in this unit and see what variables we can take for our study. Let’s write it again for the sake of convenience.

Objective 1

• To study the factors affecting adoption of hybrid rice variety.

Now review literature and find out what previous researchers have reported. To mention a few, these variables can be education, mass media exposure, technological characteristics and so on.

EXERCISE for Students

Find out more variables that may be affecting adoption of hybrid rice variety through review of literature.

2.5 What relationship exists between variables?

We know that scientific explanations and predictions involve relating the phenomena to be explained to other explanatory phenomena. In terms of variables, we want to explore the relationship between dependent variable and independent variables. Say for example, how education is related with income. When we say related, it means there is something common to both variables. Something that varies together in
both variables is common to them. Statistically, it is called covariance. Once the covariance is established, the question of direction and magnitude arises. Whether two variables vary in the same direction i.e. positively correlated, or in different directions i.e. negatively correlated. As an extension researcher, we are further interested to know the magnitude of change. How much change occurs in one variable with unit change in other variable? i.e. regression. However, in order to establish the relationship between two variables, we need to first formulate hypothesis.

2.6 What are different types of variables?

Variables can be classified in several ways. Some commonly accepted classifications are presented below.

i) Dependent and independent variables

Dependent variable may be defined as the phenomenon or characteristic hypothesized to be the outcome, effect, consequence or output of some input variables. Its occurrence depends on some other variable which has preceded it in time. On the other hand, there are characteristics hypothesized to be the input or antecedent variables called independent variables. There are presumed to cause the dependent variable and is selected, manipulated or measured prior to measuring the outcome or dependent variable.

For example, if one wants to study the effect of religion upon attitude towards family planning, the individual may take several religious groups and study their attitude towards family planning. By this, the researcher may be able to predict which religious group has a favorable attitude or unfavorable attitude towards family planning. Here, the religious groups constitute the example of independent variables and the attitude towards family planning constitutes the example of dependent variable.

A variable can be dependent in one study and independent in another. For example, level of adoption is generally recognized as a dependent variable. It may also be treated
as an independent variable if it is intended to study the contribution of adoption in enhancing income.

**ii) Moderator variables and control variables**

The moderator variables are special types of independent variables which are hypothesized to modify the relationship between the dependent and independent variables. Age, intelligence etc. are examples of moderator variables.

Control variables are those which may affect the relationship between the independent and dependent variables, and which are controlled (effects cancelled out) by eliminating the variable, holding the variable constant or using statistical methods.

The difference between a control and a moderator variable is that the effects of the control variables are minimized, eliminated or held constant while the effects of the moderator variables are studied. Since both control and moderator variables are independent variables, it is up to the researcher to determine the independent, moderator and control variables depending on the study.

**iii) Intervening variables**

A variable which is hypothesized to exist, but cannot be observed and is presumed to occur to explain the relationship between the independent and dependent variables is called intervening or hidden variable.

According to Kerlinger (1973), the constructs, which are non-observable, have been called intervening variables. It can neither be seen, nor heard nor felt, can be inferred from the behavior. ‘Hostility’ is inferred from, presumably hostile or aggressive acts. Motivation is an intervening variable. Motivation is a construct invented, by man to account for presumably ‘motivated’ behavior.
iv) Qualitative and quantitative variables:

The qualitative characters refer to those which cannot be manipulated after the research is started and which consist of categories that cannot be ordered in magnitude. It refers to quality or characteristic or attribute and hence known as *attributive* characters. Characteristics, such as color, race, sex, religion etc. are of qualitative type. Since qualitative characters cannot be ordered in magnitude, their precise measurements are not possible. However, we may obtain frequencies (a quantitative variable), corresponding to different categories of opinion (a qualitative character), by assigning values in order.

The quantitative variables refer to those variables which are composed of categories that can be ordered in magnitude i.e. it may exist in greater or smaller amounts. Examples of quantitative variables are age, income, size of land holding, size of group, intelligence, length of experience in the cultivation of a particular crop, adoption quotient etc. With the quantitative variables, precise measurements are possible because they can easily be ordered in terms of increasing or decreasing magnitude.

A variable which is manipulated by the researcher is *active* variable. Examples of active variables are reward, punishment, methods of teaching etc. There are variable which can be categorized as attributive variable as well as active variable. Anxiety is one such variable. Anxiety can be manipulated by giving a set of instructions to the subjects. In this case, it becomes an active variable. Since anxiety can be measured with the help of a scale or test, it constitutes the example of an attributive variable.

v) Continuous and discrete variables

Quantitative variables can be further divided into two categories—continuous variables and discrete variables. A continuous variable is one which is capable of being measured in any arbitrary degree of fineness or exactness. Age, height, intelligence, income, level of adoption etc. are some examples of continuous variable. Such variables can be measured in the smallest degree of fineness.
The discrete variables are those variables whose values can be determined by counting (D’Amoto, 1970). These are not capable of being measured in any arbitrary degree of fineness or exactness because the variables contain a clear gap. For example, the number of members in a group may be 10, 15 or 20 and so on. A discrete variable consists only of whole numbers and fractional values such as $10\frac{1}{2}$, $15\frac{1}{2}$ or $20\frac{1}{2}$ cannot occur. The number of females in a particular district or State, the number of books in a library, etc are some examples of discrete variables.

**vi) Stimulus variable and response variable**

A stimulus variable is the condition or manipulation created by the researcher so as to evoke a response in an organism. The general classes of things the researchers observe that relate to the environment, situation or condition of stimulation are referred to as stimulus variables. The stimulus variables, also known as action variables, may be items like a slide show, a field day, method demonstration etc.

Any kind of behavior of the respondent is called response or behavioral variable. This refers to some action or response of an individual. It may also refer to the frequency with which a particular event occurs or it may be the scale value of a particular event. The responses of farmers after being exposed to a demonstration can be considered as a response variable.

**vii) Extraneous variables:**

Extraneous variables are those independent variables that are not related to the purpose of the study, but may affect the dependent variable. It is, therefore, essential that extraneous variables are controlled. Suppose, an investigator is interested in studying the efficacy of method of instruction on the achievement scores (dependent variable) of trainees group. The methods of instruction to be evaluated are lecture, seminar and discussion (independent variables). The investigator discovers that the achievement scores i.e. the dependent variable is positively correlated with intelligence (an extraneous variable) of the subjects (trainees). That is, trainees with high intelligence
score tend to score high on the achievement test and those who are low on intelligence score are low on the achievement test. Thus, the variable, intelligence (not of direct interest to the investigator) needs to be controlled because it is a source of variance, which may influence the achievement scores.

2.7 Is it necessary to formulate hypothesis?

In India most of the students’ researches do not report the hypothesis formulated in their studies. It raises a serious doubt whether the students formulate hypothesis for their research or not. In fact, setting up and testing hypotheses is an essential part of statistical inference. In order to formulate such a test, usually some theory has been put forward, either because it is believed to be true or because it is to be used as a basis for argument but has not been proved, for example, claiming that educated farmers would be quick to adopt certain technology.

In each problem considered, the question of interest is simplified into two competing claims / hypotheses between which we have a choice; the null hypothesis, denoted by $H_0$, against the alternative hypothesis, denoted by $H_1$. These two competing claims / hypotheses are not, however, treated on an equal basis: special consideration is given to the null hypothesis.

The hypotheses are often statements about population parameters like expected value and variance; for example $H_0$ might be that the expected value of the height of ten year old boys in the Indian population is not different from that of ten year old girls. A hypothesis might also be a statement about the distributional form of a characteristic of interest, for example that the height of ten year old boys is normally distributed within the Indian population.

The outcome of a hypothesis test may be "Reject $H_0$ in favor of $H_1" or "Accept $H_0". 
It is, therefore, essential to set hypothesis based on previous theory or experience. Otherwise, any set of variables when directly tested will give some result, though spurious.

Again, let us take the example of first objective for which we have identified some of the variables (Education, Mass Media Exposure and Technological Characteristics). So the hypothesis to be tested can be…

\[ H_0 = \text{There is no relationship between adoption of hybrid rice variety and education} \]

\[ H_1 = \text{There is a relationship between adoption of hybrid rice variety and education} \]

**EXERCISE for Students**

Write null and alternate hypothesis for other set of variables.

### 2.8 What will make possible the testing of hypothesis?

Basically, there are four characteristics that any hypothesis should possess to make it possible to be tested. They are *clear, value free, specific* and amenable to *empirical testing* with available research methods.

A hypothesis is clear if the conceptual definition and operational definition are carefully formulated. There is no scope of any ambiguity in a hypothesis. Each word in a hypothesis must convey clear meaning to its reader.

When we say hypothesis should be specific, that means the variables, their relationship and the situation should be specifically spelled while formulating the hypothesis, otherwise it will not be possible to test it.

An empirical hypothesis is one which can be measured empirically and can be tested using empirical procedure i.e. through observation and experimentation.

A hypothesis can only be value free if the researcher is aware about his/her personal bias and make them as explicit as possible. In scientific research there is no place for person bias.
2.9 Which research design should be used and why?

Once the extension researcher determines the objectives of the study, explicates its hypothesis and defines variables, s(he) confronts with the problems like... *Whom to study? What to observe? When will observations be made? How to collect data?* The research design serves as a *blueprint* of the study and provides solutions to these problems. One of the important purposes that this blueprint serves is to help draw causal inferences. The idea of causality is central to most of the scientific investigations. However, if X is cause of Y then merely a change in X is followed by a change in Y does not establish causal relationship. In practice, the demonstration of causality involves three distinct operations i) demonstrating covariance ii) eliminating spurious relations and iii) establishing the time order of the occurrence.

The classical research design consists of three components: *comparison, manipulation* and *control*. These components are essential to establish that X, independent variable and Y, dependent variable are causally related. Comparison facilitates to demonstrate covariance, manipulation helps in establishing time order and control allows eliminating spurious relations.

Once the researcher is clear about the above three components of the research design and the degree to which comparison, manipulation and control is required for a particular research, a suitable research design can be chosen. There are three major types of designs, *viz. experimental, quasi-experimental* and *pre-experimental*. Experimental designs which are based on the principles of randomization allow comparison, manipulation, control and generalization. Quasi-experimental designs usually permit combinations of these components but not all of them. There is no possibility of randomization and manipulation. Pre-experimental designs include even fewer components than quasi-experimental.

Control is the most important feature of any experimental designs. Control is related to the internal validity of the research design. The internal validity of the
research design is affected by factors which can be classified into *intrinsic* and *extrinsic* factors. The extrinsic factors are called selection effects which results from biases in selection of the respondents. Whereas, the intrinsic factors are history, experimental mortality, maturation, instrumentation, testing, regression effect and factors that interact with selection effect.

Generally, extrinsic factors are controlled by matching and employing randomization in selection of respondents to experimental and control groups. The intrinsic factors are controlled by using a number of control groups as per the requirement of the study.

Another crucial aspect of social research besides internal validity is *external validity* of research designs. It is concerned with generalisability of the research findings. The findings of a study can be generalized to larger population or real life setting only if the sample selected is representative of the population thus sampling procedure followed in a research is key to generalisability.

### 2.10 Examples of research designs

**Only after experimental research design**

As the name suggests data is collected after the respondents are exposed to experimental variables like a demonstration of Zn application, a video on IPM, a radio talk or a telecasted programmed on gender issues. This design can be illustrated as follows.

1. No observation before exposure to ‘X’
2. Exposed to experimental variable ‘X’
3. Observation after exposure to experimental variable ‘X’

\[
\text{Effect of (X)} = Y - Y''
\]
The effect of experimental variable X on dependent variable Y can be ascertained by subtracting the value of dependent variable in control group (Y′′) from that of experimental group (Y). The value of Y′′, if any other than zero, is due to the effect of unintended independent variables. This helps us to eliminate effect of any other variables than intended variable X on the dependent variable and the result is exclusively the effect of X on Y.

*Ex post facto research design*

The *ex-post facto* design is a variation of the "after-only with control group" experimental design. The chief difference is that both the experimental and control groups are selected after the experimental variable is introduced rather than before. This approach eliminates the possibility that participants will be influenced by the awareness that they are being tested.

**Before-after experimental research designs**

The first and simplest type of before-after research design contains only experimental group without any control group. The following illustration will make it clear for you. This is as simple as only after but instead of taking observation of control group, the value of dependent variable before exposure to experimental variable is noted. This value acts as control. The effect of experimental variable X on dependent variable Y can be ascertained by subtracting the value of dependent variable before exposure (Y′) from the value of dependent variable after exposure (Y).

\[
\text{Effect of (X) = Y - Y'}
\]
The second type of before-after research design is with one control group, as you can see in the figure below. In this case, observations are made before the exposure to experimental variable as well as after the exposure. The data is collected from both experimental and control groups. You can see that now we have four values for dependent variable i.e. values for experimental group and control group before exposure and after exposure to experimental variable ‘X’.

Observation before exposure to ‘X’

\[ Y’, Y’’ \]

<table>
<thead>
<tr>
<th>Exposed to experimental variable ‘X’</th>
<th>No exposure to experimental variable ‘X’</th>
</tr>
</thead>
</table>

Experimental group  Control group

\[ Y \quad \quad Y’’ \]

Observation after exposure to experimental variable ‘X’

Effect of (X) = (Y – Y’) – (Y’’ – Y’’’)

Here, \( Y’ \) denotes value of dependent variable before exposure to ‘X’, which may be the effect of some unintended independent variable. So if we subtract \( Y’ \) from \( Y \) i.e. \( (Y – Y’) \), we get the effect of independent variable ‘X’ on dependent variable. But we cannot be sure if some other independent variable X’ may be in operation when the experimental group is being exposed to intended independent variable ‘X’. Let me take an example, if you want to see the effect of the demonstration on fertilizer application in rice that you are conducting you may take one group and measure the knowledge of fertilizer application before and after the demonstration and subtract before knowledge score from after knowledge score. But, there may be some radio talk being aired about fertilizer application during the demonstration period. Under the circumstances, if we have a control group then we can obtain the value of such unintended exposure by
subtracting before score from after score i.e. \((Y'' - Y''')\). Thereafter, if you subtract this value from \((Y' - Y'')\) then, you can get accurate measurement of the effect of experimental variable \(X\) on dependent variable \(Y\).

There are also more complex before-after research designs which you may not require to implement. In case you are interested to know further, then read suggested text books. Finally, you should choose the research design as per accuracy needed by you.

2.11 Let us sum up

We have examined the need of concepts in extension research, deciding the research problems, criteria of setting research objectives, method of select variables, importance of formulating hypothesis and testing it and suitability of research design for conducting meaningful research.
Unit 3

Literature Review and Developing Theoretical Orientation

Structure

3.0 Objectives

3.1 What is a literature review?

3.2 What is the purpose of literature review?

3.3 What a literature review can demonstrate?

3.4 What are the key features of a literature review?

3.5 What strategies should you follow for working bibliography to literature review?

3.6 How to organise your literature review?

3.7 What format to follow for writing references?

3.8 Why should we develop theoretical orientation?

3.9 Let us sum up

3.0 Objectives

*By the end of this course you will be able to understand*

- The importance of literature review for conducting research and
- The techniques for analyzing, synthesizing, comprehending and reporting literature review

3.1 What is a literature review?

Literature review is a continuous process on the part of the researcher and it starts even before finalizing research problem. The literature review conveys to your reader the existing knowledge and ideas on your research topic, and the strengths and weaknesses of these texts / experiments / studies. You should clearly organise literature review in line with the focus of your research (your objectives, research question or argument). You should understand that it is not a mere descriptive list or set of
summaries, but an evaluation of the research already conducted that provides a context to your writing.

Please do not confuse literature review with the reviewing or critical evaluation of political, social or economical conditions that is found in popular newspapers such as The Times of India, The Hindu or Deccan Chronicle. Even if little has been written about your specific topic of research, the literature review establishes how and why the gap you are filling may exist and the comparability of other studies.

### Literature Review

The selection of available documents (both published and unpublished) on the topic which contain information, ideas, data and evidence written from a particular standpoint to fulfill certain aims or express certain views on the nature of the topic and how it is to be investigated, and the effective evaluation of these documents in relation to the research being proposed.

- Chris Hart

#### 3.2 What is the purpose of Literature Review?

The purpose of literature review can be enlisted as follows. As you can see, these are self explanatory but for better understanding we will discuss each point in brief.

1. **Distinguishing what has been done from what needs to be done:** As you know that the research is an ongoing process, the present researchers need to take into account the work already done in the field of his/her interest. You cannot afford to rediscover what has already been discovered. Do not reinvent wheel. However, if you have a counterview, then you may go on to disapprove established theories or laws.

2. **Discovering important variables relevant to the topic:** This is one of the important purposes for any researcher. Identifying variables relevant to your topic by simply visualizing is not a cup of tea for every researcher. You can get a list of variables as well as tools to measure them. Then, you may go on to choose those variables which may be relevant to your research.
3. **Synthesizing and gaining a new perspective:** While you conduct the literature review, you will come across different views on your topic. After a systematic study of the available literature, you can synthesize it and gain an all together new perspective which the earlier researchers might not have reported.

4. **Identifying relationships between ideas and practice:** As a young researcher you may have ideas regarding research problem, methods, analysis and so on but what is the norm or what is the general practice by earlier researchers is of utmost importance. You need to relate ideas with practice.

5. **Establishing the context of the topic or problem:** No research problem exists in isolation, it exists in some or the other context. As a researcher you need to establish the context in which it exists. This will also help you while generalizing the results.

6. **Rationalizing the significance of the problem:** A good literature review will provide enough information to rationalize the significance of your problem. This will give answer as to why you should take up a particular problem.

7. **Enhancing and acquiring the subject vocabulary:** Like any other subject, research has its own vocabulary and usages of phrases. You need to learn this for your report writing.

8. **Understanding the structure of the subject:** The area of research you may be interested has its own concepts, its interrelationship and structure which you need to understand before conducting research in that area.

9. **Relating ideas and theory to applications:** In real life situation, there is a triangulation of idea, theory and practice/ application and shown in the figure below. Literature review will help you identify specific triangulation as related to your subject.

![Triangulation Diagram](image)
10. Identifying the main methodologies and research techniques that have been used
11. Placing the research in a historical context to show familiarity with state-of-art developments

### 3.3 What a literature review can demonstrate?

The literature review places the topic of your research in a historical perspective, portraying when and how previous researches were undertaken by earlier researchers in past. What methodologies were followed to arrive at the results in a particular area? How the previous researches have helped in understanding the research problem in a better way? You should understand that research is an ongoing process and historical perspective provides a base or foundation to your current research problem.

You will also come across key landmark studies through selection of key sources and authors and understand the change in the course after these landmark studies. These studies will help you examine your own wisdom in selecting research problem and the proposed methodology you are going to employ.

Review of literature establishes context for your interest. It introduces and provides examples of the range of techniques and tools that can be used in analyzing your research idea. You can understand the nature and use of argument in research that others have used and develop your own skills in explaining and discussing the research idea.

Literature review is important for you to acquire an understanding of the topic. You can make out what has already been done to identify reasons for your own work. This amounts to showing that you have understood the main work in the area, its application and shortfalls. The literature review is part of your journey in becoming an expert in the field of study.
3.4 What are the key features of a literature review?

Undertaking a review of the body of literature is often seen as something obvious and easy to perform. In practice, many students do produce review of literature but of poor quality. These are nothing more than compilation of thinly annotated bibliography. We need to conduct a quality literature review. The key features of a quality literature review are:

- Breadth and Depth
- Rigour and Consistency
- Clarity and Brevity
- Effective analysis and Effective synthesis

A literature review with these quality can help answer following questions.

1. What are the key sources?
2. What are the key concepts, theories and ideas?
3. What are the epistemological and ontological grounds for the discipline?
4. What are the main questions and problems that have been addressed to date?
5. How is knowledge on the topic structured and organized?
6. What are the origins and definitions of the topic?
7. What are the existing standpoints?
8. What are the major issues and debates about the topic?

**EPISTEMOLOGY** – the branch of philosophy that deals with the nature, origin and scope of knowledge

**ONTOLOGY** – the study of being/existence
3.5 What strategies should you follow for working bibliography to literature review?

As you focus on your topic, you can develop a ‘working bibliography’: an on-going list of sources you consult. By including some additional information with your bibliographic records, this can help you write your literature review:

- How did I find this source? (Was it discussed or referenced in another source? Did my tutor/supervisor recommend it to me?)
- Summarize what was in it - approx. 5 keywords/themes
- Was it useful? What were its key/best features (a good glossary of terms, illustrations, case studies, historical context)? What were its limitations (out-of-date, limited practical examples or case studies, bias, narrow focus/too broad)?

<table>
<thead>
<tr>
<th>Tips</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jump offs</td>
<td>Start with books, handbooks and research overviews and review their references.</td>
</tr>
<tr>
<td>Track and map authors.</td>
<td>Track the names of key authors and locate their original works.</td>
</tr>
<tr>
<td>Use library guides if appropriate.</td>
<td>Use library guides of various universities. They are mostly online.</td>
</tr>
<tr>
<td>Record key definitions and their context.</td>
<td>Look for patterns and frameworks in what is written about a topic such as the context - social, political, historical.</td>
</tr>
<tr>
<td>Establish a personal search pattern.</td>
<td>Determine types of materials needed (books, primary sources, government documents, statistics, scholarly articles, opinion pieces, etc.) and how to access them. Become aware of your search patterns. Track your strategy including reference tools, databases, authors, questions, and search problems.</td>
</tr>
<tr>
<td>Make strategic use of journal index and search engine advanced search capabilities.</td>
<td>Work out the best way to search each journal index along with Google and Google Scholar. List search terms, subjects, and descriptors which are specific to each index. Find out how to narrow searches. Search key authors.</td>
</tr>
<tr>
<td>Use citation indexes.</td>
<td>Search key authors and specific works in a citation index to find who has cited them. See Web of Science; Google Scholar</td>
</tr>
<tr>
<td>Track your references.</td>
<td>Use RefWorks to manage references and create APA bibliographies. Learn the elements of APA: <a href="http://apastyle.apa.org/">http://apastyle.apa.org/</a> Path: Research Tools &gt;&gt; Find Journal Articles &gt;&gt; type Refworks in box &gt;&gt; set up your account</td>
</tr>
<tr>
<td>Talk to your librarian!</td>
<td>Make an appointment with your librarian to review your needs.</td>
</tr>
</tbody>
</table>
3.6 How to organize your literature review?

There are two ways of presenting Literature review, one knowledge-based and another argumentation. The key elements of these two approaches are presented below.

Knowledge-based elements

1. A description of previous work on the topic, identifying leading concepts, definitions and theories
2. Consideration of the ways in which definitions were developed and operationalised as solutions to problems seen in previous work
3. Identification and description of matter other researchers have considered important.

Argumentation elements

1. A description of what you find wrong in previous work on the topic
2. A proposal for action that might solve the problem – your research
3. An explanation of the benefits that might result from adopting the proposal
4. A refutation of possible objections to the proposal

3.7 What format to follow for writing references?

While writing citation, you are required to follow certain format. Worldwide, there are many variations in formats being followed by researchers for writing references. You may ask, in that case which format to follow. Well, it is not important which format you follow, rather it is important that you should follow the same format throughout your report. One of the types of formats for writing references of research paper, book, book chapter, thesis and other publications is given below. Please note the items that are written in bold or italics.

Research paper

<Author(s)'s name with Surname first>, <Year>, <Title of the research paper>, <Name of the Journal>, <Vol.No.(Issue No)>:<Page to Page>

**Book**

<Author(s)’s name with Surname first>. <Year>. <Title of the book>. <Publishers name>,<Place>.


**Book Chapter**

<Author(s)’s name with Surname first>. <Year>. <Title of the book chapter>. In:<Title of the book>.<(Eds.) Editors’ name>.<Publishers name>,<Place>.


**Thesis**

<Author’s name with Surname first>. <Year>. <Title of the thesis>. <University>,<Place>.


**Online resource**

Reference as above including online source address


**Reports**

<Anonymous>.<Year>. <Name of Report>. <Institute or organization name>.<Place>

3.8 Why should we develop theoretical orientation?

Theoretical orientation means orienting oneself with the underlying theoretical development around the subject and identifying the development that you want to bring through your research. It also helps the readers of your research report to understand the theoretical developments in your area of research and place your findings in the right perspective.

The output of the theoretical orientation is presented in the form of schematic diagram. For example, if your study is on potential entrepreneurs then you would like to understand the situation of a typical potential entrepreneur. You may come across the PEST factors that affect the potential entrepreneur. PEST stands for economic, social and technological factors. A schematic diagram may look like the one presented below. You will also like to discuss in detail how each factor is affecting the young potential entrepreneur.
Similarly, you will be able to develop your theoretical orientation of your area of research but you should follow below mentioned points.

1. Read original work.
2. Review the literature extensively
3. Explore yourself
4. Broaden your experience
5. Discuss with experts

3.9 Let us sum up

The key features of a quality literature review are:

- Breadth and Depth
- Rigour and Consistency
- Clarity and Brevity
- Effective analysis and Effective synthesis

A literature review with these quality can help answer following questions.

1. What are the key sources?
2. What are the key concepts, theories and ideas?
3. What are the epistemological and ontological grounds for the discipline?
4. What are the main questions and problems that have been addressed to date?
5. How is knowledge on the topic structured and organized?
6. What are the origins and definitions of the topic?
7. What are the political standpoints?
8. What are the major issues and debates about the topic?
Unit 4

Sampling Methods

Structure

4.0 Objectives
4.1 Why should we follow sampling procedure?
4.2 How to define the population for the research study?
4.3 What should be the sample size?
4.4 What are the types of sampling design?
4.5 Which sampling design should I follow?
4.6 What is sampling error?
4.7 Let us sum up

4.0 Objectives

By the end of this course you will be able to understand

- The fundamental of population and sample
- The sample size and
- The sampling methods you should follow in your research

4.1 Why should we follow sampling procedure?

As we have already discussed, that most of the scientific researches focus on causality. Here, our interest is to establish whether X causes Y in the population. One way of establishing this fact is to collect information regarding X and Y from each unit of the population and analyze it. Actually, it is impractical to collect data from every unit of the population in a survey research or in a controlled experiment to establish this...
causation. Under the situation, inferences based on a subset of population would be quite accurate. This subset of population is called sample. However, it is imperative to have representative sample so that generalization about population parameters could be made from sample statistics. In other words, population parameter which is unknown is estimated from the sample statistics that can be measured. In order to arrive at accurate estimates of population parameters, the extension researcher should properly i) define the population, ii) determine suitable sample size and iii) decide the sampling design.

4.2 How to define the population for the research study?

Traditionally, one's membership in a population is self-defined, influenced by concepts such as race, ethnicity, gender, geographic location, origin, socioeconomic and educational status, language, culture or religion.

For research purposes, the definition of a population depends on the context of the study. For example, a worldwide study might refer to those individuals living in Asia (Asians), and to particular racial/ethnic groups (i.e., Asian Muslims, Aryans, and Magnolias) as subgroups. An economist might define a population as "individuals below poverty line in a country or state"; while an extension researcher might define a population as a group of people who cultivate some crop (say, rice) or rear some particular animal say, buffalo.

This ambiguity in the definition of a population is a problem for young researchers, but it is a reminder to you that any research must define the population under investigation very clearly from the outset. The fundamental purpose of defining a population is to enable the researcher to draw a representative sample based on which you want to bring out some conclusive result which can be generalized for the population.

For example, you want to study that in case of extension workers of Bihar whether the attitude towards computer is related with their knowledge about rice production.
technologies. Then you will define your population as all the extension workers in State Department of Agriculture in Bihar. And you would utilize a suitable sampling procedure to draw a representative sample. Remember that the population parameter of one population may vary from another population. So do not try to merge different populations whose characteristics are very different.

4.3 What should be the sample size?

Various misconceptions about suitable sample size are held by extension researchers. One is that the sample size should be a particular proportion, say 5 per cent, of the population; another is that the sample size should be about 2000. However, no such thumb rule is adequate. In fact, the size of the sample is dependent on the level of accuracy required or how large an error is acceptable. It also depends on the time, money and manpower available to collect data and conduct research. However, the sample size should not be less than 30, if you want to employ parametric tests such as correlation, regression, t-test, ANOVA, MANOVA etc.

4.4 What are the types of sampling design?

A sampling design should specify the population frame, sample size, sample selection, and estimation method in detail. The objective of the sampling design is to know the characteristic of the population i.e. population parameter that we have discussed earlier. So after having defined the population for your study and decided the sample size that you would like to draw from the population, you will be required to draw the sample from the population using suitable sampling method. In general sampling method can be classified into two broad categories-

a. Non-probability sampling
b. Probability sampling
As the term suggests, in case of probability sampling you know exact possibility of selecting each member (sampling unit) from the population, whereas, in later case the probability of selecting a member is unknown. Let me tell you that the probability sampling is more tedious and costly than non-probability sampling. You may question as to why we should follow probability sampling. The answer is simple. Only in case of probability sampling, the results obtained can be generalized and error can be estimated which is impossible in case of non-probability sampling. Remember, we want to estimate population parameter/characteristics from sample statistics.

a. **Non-probability sampling**

The advantage of non-probability sampling is the ease in which it can be administered. Non-probability samples tend to be less complicated and less time consuming than probability samples. If you have no intention of generalizing beyond the sample, you can go for one of the non-probability sampling method for desired information. There are three common types of non-probability sampling methods - convenience sampling, quota sampling, and judgmental sampling.

**Convenience sampling**

As the term suggests, convenience sampling is done at the convenience of the researcher. The example of convenience samples can be people you come across at the street, farmers who have come to visit your institute, farmwomen gathered to celebrate Farmwomen Day or students in a classroom. You will agree that we cannot ascertain the probability of being selected to such samples and therefore, none of the reliability or sampling precision statistics can be calculated. Researchers employ this sampling method when there is a crunch of time and money.

**Quota Sampling**

Quota sampling is often confused with two probability sampling method – stratified sampling and cluster sampling. You will be able to distinguish once we discuss probability sampling. Let us first understand this non-probability quota sampling
method. In this case, you select people non-randomly (please note non-randomly) according to some fixed quota. There are two types of quota sampling: proportional and non-proportional. The quotas are fixed based on some prior selected criterion. Some of the popular criteria are gender, age, education, income, land holding, caste, religion etc.

In case of **proportional quota sampling**, you should know the proportion of characteristics that exits in the population and accordingly you will draw the sample. Let us take the example of gender – male and female. Now, suppose there are 40 per cent females and 60 per cent males in the population according to census report, then for a sample of 100 individuals drawn using proportional quota sampling method will contain 40 females and 60 males.

However, knowing the proportion in a population may not be possible or feasible, in such cases a researcher goes for drawing equal quota from different sections, which is called **non-proportional quota sampling**. For example, you want to draw quota sample from the country based on religion. You know the proportions of Hindus (80.5%), Muslims (11.5%), Christians (2.3%), Sikhs (1.79%), Buddhists (0.8%), Jainism (0.4%), Judaism (0.000016%) and many other minor religions (less than 0.5% each) exist in India. It is not possible to draw proportional sample if sample size is 100, for religions having less than 1 proportion of the total population of India, however, as a researcher you don’t want to omit any religion on this basis from your sample, so you fix non-proportional quota from every religion and draw sample.

**Purposive sampling**

Purposive sampling is yet another form of non-probability sampling method, wherein, samples are drawn with some purpose in mind. Suppose you want to study the attitude of farmers towards rice transplanter, then you will select those farmers who possess or who uses rice transplanter for transplanting the rice seedlings. Again, as with all non-probability sampling methods, the degree and direction of error introduced by the researcher cannot be measured and the reliability of the estimates cannot be
calculated. Purposive sampling has its own advantages and limitations; therefore it should be employed only when it fulfills some purpose of your research.

a. Probability sampling

Basically, there are four types of probability sampling methods commonly used by researchers: these are simple random sampling, stratified random sampling, cluster sampling and systematic sampling. In case of all types of probability sampling methods, the degree and direction of error can be measured and the statistics that measure the precision of the estimates can be easily calculated. Please note that the simple random sampling method is the foundation for all other probability sampling methods.

Simple random sampling

The basic principle behind simple random sampling method is to assign a known probability to each unit of the population so that it may have equal chance of getting drawn in the sample. In order to draw a sample using simple random sampling, you need to prepare an exhaustive list of all the members of the population of research interest. Using random number generator programme or random number table, you can select individuals from the list. Suppose, there are 1000 members in the population and you want to draw a sample of five units, then you will have to get first five random numbers using random number generator programme or random number table. Say, the first five numbers without replacement are 0961, 0087, 0669, 0230 and 0425 then you will select those members of the population who have these serial numbers in the exhaustive list. If the population in the question is very large and dispersed in a bigger geographical area,
then simple random sampling method would be most costly and time consuming, impractical method of sampling.

*Stratified random sampling*

Stratified random sampling is used under following conditions

a. when there are smaller sub-groups that are to be investigated,
b. when you want to achieve greater statistical significance in a smaller sample and/or
c. when you want to reduce standard error.

First you have to divide the population into mutually exclusive and collectively exhaustive groups or strata and then draw simple random sample from each group/strata. The strata for dividing the population can be income groups, regions, progressiveness or any other such group. The basis of the grouping depends on the assumption that a population varies across groups with respect to the variable in question. Suppose, the researcher considers that the income of farmers varies with respect to regions then s(he) will divide the population based on the region and study the income of the farmers. This will help the researcher get statistically more precise result.

**Cluster Sampling**

Cluster sampling is somewhat similar to stratified random sampling, in a way that the population is divided into mutually exclusive clusters. The clusters are made in such
a way that they are representative of population. Further, simple random sample is drawn from any one randomly selected cluster. This is one stage cluster sampling. We can also follow the same process of cluster formation at various stages and draw simple random sample. This would be called as multi-stage cluster sampling.

Population divided into clusters and random sample drawn from randomly selected cluster

Population arranged in a sequence and every fifth member is systematically selected after selecting the first member randomly, thus drawing 20 % sample (30 members) from the population (150 members)
Systematic Sampling

Systematic sampling is often used in place of simple random sampling. In this method of sampling every \( n \)th member of the population is systematically selected for the sample. However, the first unit of sample is selected using simple random sampling method. For example, if you want to select every 10th member of a population that means you want to select 10 per cent of the population. So the if the population is 1000, then the first unit will be selected from the first 10 members. Suppose, the first member selected randomly is 2nd member then next member would be 12nd, followed by 22nd, 32nd, and so on till 992nd members, so that 100 members will be selected for the population.

4.5 Which sampling design should I follow?

Now, you must have understood various sampling designs and the question before you is ‘How to select an appropriate sampling design for your research?’ There can be many considerations before a researcher and the choice of best sampling design is not obvious.

The best sampling design is the one that most effectively meets your goals of your research. By this stage you might have already defined your population and sample size. Good researchers use the following strategy to identify the best sampling design.

- List the research goals which are usually some combination of accuracy, precision, and/or cost. Please differentiate between research goals and research objectives.
- Identify potential sampling designs that might be effective in achieving these goals.
- Test the sampling design which may best suit your goals. Standard error can be one of the criteria to choose sampling design. Also remind yourself of the advantages and disadvantages of each sampling designs.

Choose the sampling design that does the best job of achieving the goals.
4.6 What is sampling error?

Let me remind you that we go for sampling because we want to predict population parameters from sample statistics. Since the sample is a subset of the population, it is obvious that there may be some deviation in the value of sample statistics from the value of population parameter. This difference is called sampling error. For example, consider the land holding (in hectare) of farmers in a population (N=30) from where we draw six samples (n=5, each sample consisting of 5 members) so that whole population is divided into six population. Please take a note from the table on the following page that the population parameter, in this case population mean, is 5.65 but none of the sample statistics (sample means) are same as population parameter. This is because of the sampling error. Now, let me tell you that larger the sample size lesser is the sampling error. Let us enlarge the samples by adding samples 1 with 2, 3 with 4 and 5 with 6. Hence, the sample means of these three samples of 10 members each will be 6.05, 5.57 and 5.33 with sampling error -0.4, 0.08 and 0.32 respectively. These sampling errors are lesser than the sampling error of samples containing 5 members which are -0.93, 0.13, 1.13, -0.97, 1.43 and -0.79. So, we can observe that by increasing the sample size, the sampling error reduces.

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Land Holding (ha)</th>
<th>Samples (ha)</th>
<th>Sample Mean (ha)</th>
<th>Population Mean (ha)</th>
<th>Sampling error (ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1.5</td>
<td>1.5, 2.6, 9.3, 5.5, 14.0</td>
<td>6.58</td>
<td>(6.58 + 5.52 + 4.52 + 6.62 + 4.22 + 6.44)/6 = 5.65</td>
<td>5.65 - 6.58 = (- 0.93)</td>
</tr>
<tr>
<td>2.</td>
<td>2.6</td>
<td>3.3, 5.6, 7.4, 1.1, 10.2</td>
<td>5.52</td>
<td>5.65 - 5.52 = (+0.13)</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>9.3</td>
<td>2.2, 11.3, 6.8, 1.1, 1.2</td>
<td>4.52</td>
<td>5.65 - 4.52 = (+1.13)</td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>5.5</td>
<td>14.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>14.0</td>
<td>3.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td></td>
<td>5.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td></td>
<td>7.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.</td>
<td></td>
<td>10.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10.</td>
<td></td>
<td>2.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.</td>
<td></td>
<td>11.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.</td>
<td></td>
<td>6.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.</td>
<td></td>
<td>1.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14.</td>
<td></td>
<td>1.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
So you may like to take a very big sample for your study. But if your population is too large say, 100 billion, then it is not possible to take a 10% sample or so to reduce sampling error. Here, you need to weigh between preciseness of estimate and cost of research.

### 4.7 Let us sum up

In this unit, we have discussed the concepts of population and samples. The purpose of sampling is to make generalizations about population parameters from sample statistics.

Broadly, there are two methods of sampling – non-probability sampling and probability sampling. Non-probability sampling includes convenience sampling, quota sampling and purposive sampling. While, probability sampling includes simple random sampling, stratified sampling, cluster sampling and systematic sampling.

Sampling error is the difference in the value of sample statistics and population parameters. Sampling error can be kept low by taking larger samples.
Unit 5

Data Collection Methods

Structure

5.0 Objectives
5.1 Can you differentiate among data, information, knowledge and wisdom (DIKW)?
5.2 What do you mean by qualitative and quantitative data?
5.3 Why should we collect data?
5.4 What are different types of variables?
5.5 What are discrete and continuous variables?
5.6 What are independent and dependent variables?
5.7 What do you understand by moderating variables, intervening variables and extraneous variables?
5.8 How to measure the variables?
5.9 How do we record the data?
5.10 Let us sum up

5.0 Objectives

By the end of this unit you will be able to understand

- the concept of data and types of data
- the level of measurement
- different types of data collection methods and
- Case study method of data collection
5.1 Can you differentiate among data, information, knowledge and wisdom (DIKW)?

We often use these terms loosely but if you are able to understand the minute differences among these terms then it will be easy for you to understand the importance and purpose of data collection. First of all you should understand that wisdom contains knowledge, knowledge contains information, information contains data but visa versa is not true. Data is symbols assigned to some discrete and objective facts about an event. Data is raw which simply exists and has no significance in itself beyond its existence. It can exist in any form, usable or not. It does not have meaning of itself. Say for example, you ask a farmer if he has adopted Bt cotton. The answer you will get is yes or no. You go on asking this question to 120 farmers and then you assign ‘2’ for yes and ‘1’ for no. These 2s or 1s are data.

Information is derived from data by identifying relational connections and attaching meaning to data. In above example if you find out the number of 1s and 2s, this can be called as information. Say for example, if 90 farmers said yes and remaining 30 farmers said no then the information that we get is that ‘Out of 120 farmers, 90 farmers have adopted Bt cotton technology.’

Data could be converted into information using 5 main processes (Leibowitz, 2003):

- Condensation – items of data are summarized into a more concise form and unnecessary depth is eliminated;

According to Stonier (1993, 1997), data is a series of disconnected facts and observations. These may be converted to information by analyzing, cross-referring, selecting, sorting, summarizing, or in some way organizing the data. Patterns of information, in turn, can be worked up into a coherent body of knowledge. Knowledge consists of an organized body of information, such information patterns forming the basis of the kinds of insights
- Contextualization—the purpose or reason for collecting the data in the first place is known or understood;
- Calculation - data is processed and aggregated in order to provide useful information
- Categorization – is a process for assigning a type or category to data;
- Correction – is a process for removal of errors.

Knowledge, on the other hand, is the appropriate collection of information, such that its intent is to be useful. Knowledge is a deterministic process. You can generate knowledge when you process, organize or structure information in some way so that it can be applied or put into use. Suppose, you collect information about adoption of Bt cotton from different parts of Andhra Pradesh (AP) and after processing you are able to say that 75% of AP farmers have adopted Bt cotton. You have knowledge of level of adoption in AP with regards to Bt cotton. This can be categories as processed knowledge which results from synthesis of multiple sources of information over time. Second type of knowledge is procedural knowledge - the know-how which results from application of data and information. The third type of knowledge is propositional knowledge which is based on our belief. It is the reflective and/or the expressed content of what a person thinks that he or she knows. Note that the contents of our reflective and/or expressed thoughts are in the form of propositions (Zins, 2007).

Wisdom is insights and judgments formed on the basis of knowledge. It is your ability to identify truth, take wiser decisions and act accordingly. In our example of Bt cotton, as an extension functionary, you may like to warn farmers from possible downfall of the cotton price in the market and advise farmers to diversify so that risk is covered.
5.2 What do you mean by qualitative and quantitative data?

Now you know from earlier discussion that data is fundamental to any research. Any experimental observation is nothing but data. Number of adopters, number of participants in a demonstration, seed rate of rice used by farmers, doses of nitrogen application in practice, age of farmers, height, length, and weight are examples of data. Incidentally, all of these examples are that of quantitative data. They can be expressed in terms of numerical and can be measured. The quantitative data can be of two types – discrete and continuous. Discrete data is expressed in terms of integers like number of adopters, tractors in a village or age in terms of completed years while continuous data can be fertilizer application in kg, land holding of farmers or yield of the crop.

Qualitative data, on the other hand, is description of quality to which some number may be assigned which are nominal or categorical in nature. Colour of hair, texture of rice grain, smell of flower, gender, caste, statehood, are some of the examples of qualitative data.

In social research, you will have to collect both qualitative as well as quantitative data to get a complete picture of the actual situation. Quantitative data can be easily analysed using statistical tests and inferences can be drawn but they lack in depth analysis. Qualitative data may be less reliable than quantitative data but they gain in terms of validity and largely help the researcher in getting insight into the research results. Say for example, you know that 75% of farmers are adopters but why remaining 25% are non-adopters can be found out by using case studies. The reason could be cost of seed, lack of knowledge, lack of access to
inputs or fear of failure. One can never come to these results by using quantitative methods only.

5.3 Why should we collect data?

Do you remember the cycle of research process? Well, it is given here for your ready reference. After having identified problem, set objectives, formulated hypothesis, finalized research design, you need to identify variables pertaining to your study and decide ways and means of measuring these variables for the purpose of data collection. The data collected using any of data collection methods will be analysed and generalization / results are obtained. In unit 2, we have discussed issue of selecting variables. What we also need to understand is types of variables, so that it will be easy for you to select or devise tools for measuring these variables and identify appropriate method of data collection.

5.4 What are different types of variables?

Before we discuss about types of variables, please recall what is a variable. Variable is central idea in any research. A variable is a symbol to which numerals or values are assigned. In other words, variable is a concept that acquires varying values. There are two types of concepts: those that refer to a fixed phenomenon and those that vary in quantity, intensity, or amount (e.g. cost of cultivation). In order to vary a variable will have to acquire two or more values. These values can be based on quality or quantity. Once you begin to look for them, you will see variables everywhere. For example gender is a variable; it can take two values: male or female (qualitative variable). Income is a variable; it can take on values from zero to billions of Rupees (quantitative variable). Now, take a pause and spell out other variables that you are able to see or visualize around yourself. Please try to understand that any variable can be constant in a particular situation. Say for example, if in your class of PGDAEM there are no female students, then in this situation gender (which may be variable in some other situation)
will not be considered as a variable but a constant. Majorly, variables can be classified as following.

i. Discrete and continuous variables

ii. Independent and dependent variables

However, there are some other types of variables which we will discuss in this unit. They are moderating variables, intervening variables and extraneous variables.

5.5 What are discrete and continuous variables?

Consider the following ruler, suppose the variable that you have chosen can take values like 1, 2, 3, 4, 5, 6, 7 or 8 only then it is a discrete variable. This kind of variable cannot take values in fraction. Such a variable is also called as categorical variable or classificatory variable, or discontinuous variable. Gender, education, religion and caste are some of the examples of discrete variables. If a discrete variable can hold only two values then it is called dichotomous variable e.g. gender - male, female or adopter, non adopter.

A continuous variable, on the other hand can assume any value on this ruler, even fractions. That means it can be assigned infinite number of values. Income, land holding, age, or a test score are examples of continuous variables.

5.6 What are independent and dependent variables?

One of the classifications of variables is independent and dependent variables. You may recall the cause and effect relationships we discussed in Unit 2. As a researcher, you will usually focus on an effect and then go on to find out its cause(s). The cause variables
that force the effect are called independent variables while the variable that is the effect or outcome of another variable(s) is called dependent variable. In statistic, dependent variable is denoted by ‘Y’ while, independent variable is denoted by ‘X’. In the research vocabulary different labels have been associated with the independent and dependent variables like:

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Dependent variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presumed cause</td>
<td>presumed effect</td>
</tr>
<tr>
<td>Stimulus</td>
<td>Response</td>
</tr>
<tr>
<td>Predicted from ...</td>
<td>Predicted to ...</td>
</tr>
<tr>
<td>Antecedent</td>
<td>Consequence</td>
</tr>
<tr>
<td>Manipulated</td>
<td>Measured outcome</td>
</tr>
<tr>
<td>Predictor</td>
<td>Criterion</td>
</tr>
</tbody>
</table>

5.7 What do you understand by moderating variables, intervening variables and extraneous variables?

A **moderating variable** is one that has a strong effect on the independent variable-dependent variable relationship. That is, the presence of a third variable (the moderating variable) modifies the original relationship between the independent and the dependent variable. Say for example, education as independent variable causes the effect, adoption of technology. Now, income can be a moderating variable in this situation as income will affect both education of farmer and farmer’s affordability to adopt the technology. In this relationship, income is a moderating variable i.e. which moderates the strength of the association between X and Y variables.
An **intervening variable** is the variable that appears in between independent and dependent variables. It intervenes in the causal relationship of cause and effect. The intervening variable acts as a dependent variable with respect to independent variable and acts as an independent variable toward the dependent variable. The example that we will discuss here will make the concept clearer to you. Suppose, type of family, an independent variable, is related with adoption of (labour intensive) technology, the dependent variable. A close analysis will reveal that it is the labour availability that affects the adoption of (labour intensive) technology. That means type of family causes the effect, labour availability and in turn labour availability causes the effect—adoption of technology. Therefore, labor availability is an intervening variable.

In order to explain extraneous variables, let us first take an example of effect of technology demonstration on technology adoption. You may assume that the causal relationship in this case is very simple to study. You have been conducting demonstration as part of your job and you have seen the positive results as well. Take a pause and think if there were other variables that affected technology adoption. Yes, your interaction, your frequency of visit, a TV programme on the same technology, previous adopters in the vicinity and a lot of other variables were affecting the technology adoption in your situation. The effects of these variables are consciously and deliberately removed by using suitable research design or by making them constant. All these unwarranted variables are **extraneous variables**.

EXERCISE for Students

List the variables around you and also classify them as per the types of variables discussed above.
5.8 How to measure the variables?

A variable is measured by assigning numerals to it according to rules. The function of rules is to establish isomorphism between certain numerical structure and the structure of the variable being measured. The numerical structure can be one among the four levels of measurement - nominal, ordinal, interval and ratio. Further, the mathematical or statistical treatment permissible on given set of numerals is dependent on level of measurement obtained.

The concepts of validity and reliability are linked with measurement. As mentioned earlier, validity is concerned with the question whether one is measuring what one intends to measure, whereas, the concept of reliability deals with accuracy and stability of measurement. While measuring a variable, first the extension researcher should be aware of level of measurement so that s(he) uses right mathematical or statistical operations so that non-spurious results can be obtained. Secondly, s(he) should ascertain the validity and reliability of the measurement so that estimates of population parameters are estimated accurately.

5.9 How do we record the data?

Basically, there are two broad methods of collecting data - one, by observing people, process, behavior, events and conditions and second, by interacting/communicating with people regarding various topics including their knowledge, skill, behavior, attitude, intentions and motivation.

Observation is the typical method of data collection in scientific research. If one wishes to understand, explain and predict the existing phenomenon, one can simply go and observe it. But if one is interested in systematic results, then the observations must be carried out keeping following points i) what to observe ii) where and when to observe and iii) how much to infer when recording observation.
The observation method has certain advantages and limitations. The data collected relates to current situation rather than *ex-post facto*. It is independent of respondents’ willingness to response. If you want you can eliminate subjectivity from data by accurately recording the behavior. However, observation is time consuming, expensive and does not allow controlling the effect of extraneous factors.

You can record observation by using carefully defining units of observation, data to be collected, method or recording and condition in which data will be recorded. This is also called structured observation. If observation is made in absence of well defined process, it is called unstructured observation. Further, there can be a variation in method of observation by the researcher remaining as an insider (participant observation) or outsider (non-participant observation) while observing the phenomenon.

The second method of data collection is survey. Traditionally, the method of survey was based on face-to-face interaction. Where, the researcher used to administer a questionnaire or conduct interview using a structured schedule. In case of structured schedule, questions, their wording and their sequence are fixed. Nonscheduled interview is conducted with the help of interview guide. Later, other methods like mail questionnaire and telephone interview came into use. Simultaneously, there was the change from traditional paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI). Now, face-to-face surveys (CAPI), telephone surveys (CATI), and mail surveys (CASI, CSAQ) are increasingly replaced by web surveys.

Whether you use PAPI method or CAI method for data collection, you will be required to frame questions, you should be careful about following points (Ray and Mondal, 2004).

1. Your questions should be related to your objectives/hypothesis.
2. The question should be related to knowledge and background of the respondents.
3. The questions should be simple, unbiased and easy to understand.
4. There should be only one idea in one question.
5. The questions should proceed in a logical sequence, from easy to more difficult ones. This should not create any response bias.
6. The questions should be culturally compatible.
7. Personal and intimate questions, if any, should be left to the end.
8. The questions and probable replies should be short.
9. Terms conveying different meaning to different people (ambiguous) should be avoided.
10. Questions carrying value judgments should not be used.
11. Do not repeat questions/items.
12. The number of questions should not be too many.
13. Questions should be sequenced keeping analysis of data in mind.

You may formulate multiple choice questions, which is based on probable answer to the question. Multiple choice questions make it easy for the researcher to code, enter data and analyze data. A multiple choice question can be open ended or close ended. In case of close ended question, the response is limited to the choice given by the researcher but in case of open ended question, respondents can give their own reply.

The questionnaire you have developed may have some inherent problem. In order to remove them, you should pre-test your questionnaire with 2-5 per cent of your sample. To avoid testing effect, the individuals who responded in pre-testing should not be included in the final sample selected for study.

Non-Interactive Media, Human Interactive Media and Electronic Interactive Media

We will discuss different methods of data collection on the basis of its interactive ness (interaction between researcher and respondent). The traditional questionnaire, either self-administered or mailed, does not allow a dialogue or exchange of information
for immediate feedback. Therefore, questionnaires printed on paper are non-interactive method of data collection. If you have ever received such a questionnaire then you must be remembering the difficulty you faced in understanding it as well as filling the questionnaire.

When they think of interviewing, most people envision face-to-face dialogue or a conversation on telephone. Therefore, structured interview, unstructured interview and telephone interview are based on human interactive media having a dialogue or exchange of information for immediate feedback.

Researcher can reach a large number of respondents using electronic interactive media and to personalize individual messages, and to interact with respondents using digital technology. In this case, respondents are not passive audience members but they are actively involved in a two-way communication. The internet is a prominent example of the new electronic interactive media.

Face-to-face Interview

A face-to-face interview is a two way conversation initiated by an interviewer to obtain information from a respondent. As an interviewer, you will have to first establish a rapport with the respondent. Then you will explain your purpose and assure the respondent that the information provided by him/her will be kept confidential and used only for research purpose. The consequences of the interview are usually insignificant for the respondent. The respondent is asked to provide information and has little hope of receiving any immediate or direct benefit from this cooperation. Therefore, it is one of the difficult methods of data collection. Interviews may take place in the farm, in a home, in an office, in a shopping mall, or in other settings.

The advantages of interview are opportunity for feedback, probing complex questions, length of interview can be large, high completion rate, high participation rate, researcher can observe the non-verbal behavior and illiterates can participate in study. The disadvantages of interview can be high cost, scarcity of highly trained interviewers,
lack of anonymity of respondent, difficult in recall, interviewer can influence respondent, interviewer bias may affect interview, opportunity to consult, less standardized wording, limitations in respondents' availability and accessibility and difficulty in visiting some neighborhoods.

**Intercept Interviews in Malls and Other High-Traffic Areas**

Interviews conducted in shopping malls are referred to as mall intercept interviews. Interviewers generally stop and attempt to question shoppers at a central point within the mall or at the entrance. These are low cost. No travel is required to the respondent's home instead the respondent comes to the interviewer, and thus many interviews can be conducted quickly. The incidence of refusal is high. The sampling in this case is not random; therefore, you should be careful in generalizing the findings.

In mail questionnaire data is collected through structured questions which the respondent is supposed to fill up. It is impersonal method of data collection. Its major advantages are: lower cost, reduction in bias, greater anonymity and accessibility while disadvantages are: lower response, no opportunity for probing and little control over respondent. These disadvantages can be overcome in case of structured schedule based face-to-face interview.

Besides primary source of data there are a number of secondary sources of data such as official records, archives etc.

**Choosing a communication media**

Suitable instrument of data collection should be used based on the accuracy of data required, resources available including time. Once the researcher has determined that surveying is the appropriate data collection approach, various means may be used to secure information from individual. A researcher can conduct a survey by interview, telephone, mail, computer, or a combination of these media. An alternative to survey methods is case study which is discussed below in detail.
Case study methods

Rather than using large samples and following a rigid protocol to examine a limited number of variables, **case study methods** involve an in-depth, longitudinal examination of a single instance or event: a *case*. They provide a systematic way of looking at events, collecting data, analyzing information, and reporting the results. As a result the researcher may gain a sharpened understanding of why the instance happened as it did, and what might become important to look at more extensively in future research. Case studies lend themselves to both generating and testing hypotheses (Flyvbjerg, 2006).

Case selection

When selecting a case for a detailed study, researchers often use **information-oriented sampling**, as opposed to **random sampling** (Flyvbjerg, 2006). This is because the typical or average case is often not the richest in information. Extreme or atypical cases often reveal more information because they activate more basic mechanisms and more actors in the situation studied. In addition, from both an understanding-oriented and an action-oriented perspective, it is often more important to clarify the deeper causes behind a given problem and its consequences. It is not necessary to describe the symptoms of the problems and how frequently they occur. Random samples emphasizing representativeness will seldom be able to produce this kind of insight; it is more appropriate to select some cases chosen for their validity.

Steps in Case Study Research

- Problem Identification
- Design Phase
- Prepare to collect the data
- Field Data Collection
- Data Analysis
- Report Preparation
Problem Identification

The first step in case study research is to establish a firm research focus to which you can refer over the course of study of a complex phenomenon or object. The research object in a case study is often a program, an entity, a person, or a group of people. Each object is likely to be intricately connected to political, social, historical, and personal issues, providing wide ranging possibilities for questions and adding complexity to the case study. The researcher investigates the object of the case study in depth using a variety of data gathering methods to produce evidence that leads to understanding of the case and answers the research questions.

Case study research generally answers one or more questions which begin with "how" or "why." The questions are targeted to a limited number of events or conditions and their inter-relationships. To assist in targeting and formulating the questions, researchers conduct a literature review. Careful definition of the questions at the start pinpoints where to look for evidence and helps determine the methods of analysis to be used in the study.

Design Phase

During the design phase of case study research, the researcher determines what approaches to use in selecting single or multiple real-life cases to examine in depth and which instruments and data gathering approaches to use. When using multiple cases, each case is treated as a single case. Each case’s conclusions can then be used as information contributing to the whole study, but each case remains a single case. Exemplary case studies carefully select cases and carefully examine the choices available from among many research tools available in order to increase the validity of the study.

Selection of case

As a researcher, you must determine whether to study cases in terms of the following:
- Unique
- Typical
- Objective
- Geographic region
- Size of parameters
- Variety of parameters
- Purpose of study

**Data collection process**

A key strength of the case study method involves using multiple sources and techniques in the data gathering process. The researcher determines in advance what evidence to gather and what analysis techniques to use with the data to answer the research questions. Data gathered is normally largely qualitative, but it may also be quantitative. Tools to collect data can include surveys, interviews, documentation review, observation, and even the collection of physical artifacts.

The researcher must use the designated data gathering tools systematically and properly in collecting the evidence. Throughout the design phase, researchers must ensure that the study is well constructed to ensure construct validity, internal validity, external validity, and reliability.

**Prepare to Collect the Data**

Because case study research generates a large amount of data from multiple sources, systematic organization of the data is important to prevent the researcher from becoming overwhelmed by the amount of data. To prevent the researcher from losing sight of the original research purpose and questions, advance preparation assists in handling large amounts of data in a documented and systematic fashion. Researchers prepare databases to assist with categorizing, sorting, storing, and retrieving data for analysis.
Conduct a pilot study in advance of moving into the field in order to remove obvious barriers and problem. The investigator training program covers the basic concepts of the study, terminology, processes, and methods, and teaches investigators how to properly apply the techniques being used in the study. The program also trains investigators to understand how the gathering of data using multiple techniques strengthens the study by providing opportunities for triangulation during the analysis phase of the study. The program covers protocols for case study research, including time deadlines, formats for narrative reporting and field notes, guidelines for collection of documents, and guidelines for field procedures to be used. Investigators need to be good listeners who can hear exactly the words being used by those interviewed. Good investigators review documents looking for facts, but also read between the lines and pursue collaborative evidence elsewhere when that seems appropriate. Investigators need to understand the purpose of the study and grasp the issues and must be open to contrary findings. Investigators must also be aware that they are going into the world of real human beings who may be threatened or unsure of what the case study will bring.

After investigators are trained, the final advance preparation step is to select a pilot site and conduct a pilot test using each data gathering method so that problematic areas can be uncovered and corrected. Researchers need to anticipate key problems and events, identify key people, prepare letters of introduction, establish rules for confidentiality, and actively seek opportunities to revisit and revise the research design in order to address and add to the original set of research questions.

Field Data Collection

The researcher must collect and store multiple sources of evidence comprehensively and systematically, in formats that can be referenced and sorted so that converging lines of inquiry and patterns can be uncovered. Researchers carefully observe the object of the case study and identify causal factors associated with the observed phenomenon. Renegotiation of arrangements with the objects of the study or addition of questions to
interviews may be necessary as the study progresses. Case study research is flexible, but when changes are made, they are documented systematically.

Field notes record feelings and intuitive hunches, pose questions, and document the work in progress. They record testimonies, stories, and illustrations which can be used in later reports. They assist in determining whether or not the inquiry needs to be reformulated or redefined based on what is being observed. Field notes should be kept separate from the data being collected and stored for analysis.

Maintaining the relationship between the issue and the evidence is mandatory. The researcher may enter some data into a database and physically store other data, but the researcher documents, classifies, and cross-references all evidence so that it can be efficiently recalled for sorting and examination over the course of the study.

Data Analysis

The researcher examines raw data using many interpretations in order to find linkages between the research object and the outcomes with reference to the original research questions. Throughout the evaluation and analysis process, the researcher remains open to new opportunities and insights. The case study method, with its use of multiple data collection methods and analysis techniques, provides researchers with opportunities to triangulate data in order to strengthen the research findings and conclusions.

The tactics used in analysis force researchers to move beyond initial impressions to improve the likelihood of accurate and reliable findings. Exemplary case studies will deliberately sort the data in many different ways to expose or create new insights and will deliberately look for conflicting data to disconfirm the analysis. Focused, short, repeat interviews may be necessary to gather additional data to verify key observations or check a fact.
Researchers use the quantitative data that has been collected to corroborate and support the qualitative data which is most useful for understanding the rationale or theory underlying relationships. When the multiple observations converge, confidence in the findings increases. Conflicting perceptions, on the other hand, cause the researchers to pry more deeply.

**Report Preparation**

Exemplary case studies report the data in a way that transforms a complex issue into one that can be understood, allowing the reader to question and examine the study and reach an understanding independent of the researcher. The goal of the written report is to portray a complex problem in a way that conveys a vicarious experience to the reader. Case studies present data in very publicly accessible ways and may lead the reader to apply the experience in his or her own real-life situation. Researchers pay particular attention to displaying sufficient evidence to gain the reader’s confidence that all avenues have been explored, clearly communicating the boundaries of the case, and giving special attention to conflicting propositions.

Techniques for composing the report can include handling each case as a separate chapter or treating the case as a chronological recounting. The researcher uses representative audience groups to review and comment on the draft document. Based on the comments, the researcher rewrites and makes revisions.

**Case writing**

Case development and writing should be an ongoing process for any institution using the case method. Its importance arises from the fact that recent cases not only provide an element of interest among programme participants, but also bring to the class the latest situations being faced by decision-makers.
5.10 Let us sum up

Data is a series of disconnected facts and observations. After proper analysis of data do we get some meaningful information. Pattern of information gives rise to knowledge.

Qualitative data relates to attribute like gender, color, caste, religion etc. while quantitative data are expressed in terms of numerals and can be measured. Height, Weight, distance, volume, number of adopters and yield of a crop are some of the examples of quantitative data.

There are four levels of measurements – nominal, ordinal, interval and ratio. Qualitative data are mostly nominal in nature or at the most it can be ordinal e.g. shades of a color, whereas, quantitative data are mostly interval or ratio in nature.

A variable is a symbol to which numerals are assigned. Mainly, there are two types of variables – discrete & continuous, and independent & dependent. Data regarding variables can be collected using various methods like observation, survey (using questionnaire or interview method) and case study.

Steps in Case Study Research

- Problem Identification
- Design Phase
- Prepare to collect the data
- Field Data Collection
- Data Analysis
- Report Preparation
Unit 6

Analysis of Data and Interpretation

Structure

6.0 Objectives
6.1 How do I process the collected data?
6.2 What do you mean by level of measurement?
6.3 What is reliability and validity of measurement?
6.4 How do I analyse my data?
6.5 Notes on Univariate Analysis
6.6 Notes on Bivariate Analysis
6.7 Notes on Multivariate Analysis
6.8 What is Statistical Inference - Test of Hypothesis and Estimation?
6.9 Let us sum up

6.0 Objectives

By the end of the unit you will be able to understand

- Levels of Measurement
- Use of Statistical Tests
- Parametric and non-Parametric Tests
- Interpretation of Analyzed Results

6.1 How do I process the collected data?

Data processing links data collection with data analysis whereby observations are transformed into codes that can be used for quantitative analysis. At first numerous individual observations are converted into smaller number of categories. This system of conversion is called coding scheme which links with the theory and the problem under study. The coding schemes are developed in such a way that the categories are
exhaustive and mutually exclusive. By ‘exhaustive’ we mean that no observation is left out of the coding scheme, while ‘mutually exclusive’ means no single observation can be categorized into more than one category. Further, the coded data are processed with computers. The data are fed in what is called spreadsheet. Spreadsheet is array of rows and columns. Finally, you can use batch processing or interactive processing to process the data according to the objectives of the study.

Let us take part of a questionnaire for example.

Part of a questionnaire for example

<table>
<thead>
<tr>
<th>A. Personal Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Name .................. Gender – M/F Age .... Years Caste ........</td>
</tr>
<tr>
<td>2. Contact Address with Telephone Nos.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>3. Education: Illiterate</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>12th standard Pass</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

In the above example, we will take variables like name, gender, age, caste and education. Name is text so we will enter in the spreadsheet as it is. A serial number is assigned to each respondent as its unique identification number. Next is gender which you know is a dichotomous variable and it can be male or female. For the purpose of coding, we can assign 2 for male and 1 for female (avoid scoring as ‘0’, since it will cause
problem in further analysis). In case of age, we can take as it is (45, 30, 25, 37, 22, ....). In case of education, you can score it as follows.

<table>
<thead>
<tr>
<th>Education</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illiterate</td>
<td>1</td>
</tr>
<tr>
<td>Literate</td>
<td>2</td>
</tr>
<tr>
<td>Matric Pass</td>
<td>3</td>
</tr>
<tr>
<td>12th standard Pass</td>
<td>4</td>
</tr>
<tr>
<td>ITI/Diploma</td>
<td>5</td>
</tr>
<tr>
<td>Graduate</td>
<td>6</td>
</tr>
<tr>
<td>Post graduate</td>
<td>7</td>
</tr>
<tr>
<td>Above Post graduate</td>
<td>8</td>
</tr>
</tbody>
</table>

Based on above coding scheme, the responses of individual respondents are coded and data is transferred to a master table. You can see the below table as an example of master table which contains variable fields up to 58 and the respondents are 100 in number.

<table>
<thead>
<tr>
<th>Sr. no. of respondent</th>
<th>Name of respondent</th>
<th>Gender</th>
<th>Age</th>
<th>Education</th>
<th>Var_57</th>
<th>Var_58</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ram Kumar Singh</td>
<td>2</td>
<td>55</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Sita Pandey</td>
<td>1</td>
<td>34</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>Tejpal Yadav</td>
<td>2</td>
<td>39</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Y.P. Singh</td>
<td>2</td>
<td>25</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>A. Devi</td>
<td>1</td>
<td>29</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100.</td>
<td>ABC</td>
<td>1</td>
<td>30</td>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data from the master sheet is then transferred to any spreadsheet. A spreadsheet is both a sheet of paper designed to capture and show data in rows and columns. Using the original meaning as a metaphor, a computer spreadsheet is a programme having rows and columns for data entry and further processing. In case of
Before we discuss about analysis of data, we need to understand the level of measurement as it is crucial for applying any statistical analysis on the data.

### 6.2 What do you mean by level of measurement?

First of all, let us discuss measurement per se. Measurement is the process or the result of determining the value of a physical quantity or socio-psychological trait, such as a length, time, temperature, attitude, IQ etc., to a unit of measurement, such as the meter, second, degree Celsius, attitude score or IQ score. The science of measurement is called metrology.
Stanley Smith Stevens in 1946 gave the level of measurement which is also known as theory of scale of measurement. According to him, there are four levels of measurement – nominal, ordinal, interval and ratio. Before discussing these levels, we will discuss three basic postulates related to measurement. These three postulates are necessary in order to make it possible to equate objects, to rank order them and to add them. These postulates are:

1. Either (a=b) or (a≠b), but not both
2. If (a=b) and (b=c) then (a=c)
3. If [(a>b) and (b>c)] then (a>c)

The first two postulates are necessary to classify the object on the basis of the fact that whether one object is equal to another object or not. You will understand better if we consider the example of gender where a person can be a male or a female. If there are 100 persons, we can check if person 1 is equal to person 2. If number 1 is a female and number 2 is equal to number 1, then the second person has to be a female too. Likewise, we can compare and classify a group of 100 persons. The result may show us that there are 55 males and 45 females in the group. This level of measurement is called nominal measurement. **Nominal** means based on the nomenclature or name. So we assign names to different characteristics (male or female) of a criterion (gender). Please note that in case of nominal measurement you cannot say if male>female or female>male.

Second level of measurement is ordinal measurement. As the name suggests, **ordinal** means something that can be put in an order. If a measurement satisfies all three postulates mentioned earlier, then its level is that of ordinal measurement. Let us take the example of education which was discussed earlier while discussing coding. We can always say a graduate is more educated than a 12th standard pass person. Similarly, a literate is least educated than all other categories (literate, metric pass, 12th pass,
graduate, post graduate and so on). Again, you should note that you cannot say if the difference between Post Graduate and Graduate is equal to the difference between 12th standard pass and 10th standard pass [numerically, \((7 - 6) \neq (4 - 3)\)]. Graphically, this kind of scale can be represented as follows.

Virtual position of numerals assigned to an ordinal level variable (e.g. Education)

![Ordinal Scale Diagram]

Actual position of level of education on the ordinal scale

Next level of measurement is interval or equal-interval measurement. **Interval** scales possess the characteristics of nominal and ordinal scales. Besides, it has equal interval which in terms of postulate can be mathematically and graphically expressed as

4. \((a - b) = (b - c)\)

Virtual position of numerals assigned to an interval level variable (e.g. I.Q.)

![Interval Scale Diagram]

Actual position of level of I.Q. on the interval scale

You will be able to understand better, if we take an example of Intelligence Quotient (IQ) score. We can say the difference between 110 and 100 is equal to 120 and 110 i.e. a
difference of 10. Similarly, take an example of temperature in degree Celsius or height above sea level. In all the cases, the zero is only arbitrary and not absolute. Arbitrary zero means it has been just fixed at some point. Like, 0°C Celsius is actually 273°C Kelvin. Similarly, the height of Mount Everest from sea level (also arbitrarily fixed zero) is only 8.85 km but if you consider the centre of earth as zero (which is absolute zero) then the height of Mount Everest will be 6,381 km.

The level of measurement which is based on the absolute zero is called ratio scale. It is so called because it fulfills following postulate.

5. \((a/b) = k\), where \(k\) is some constant

The meaning of above equation is that you can find out the actual value of the ratio of two measurements. Say for example, height of a child and her father may be \((3\text{ft} / 6\text{ft}) = 0.5\) which is possible because height is in ratio scale and zero height is absolute zero. A summary of all the level of measurement is given in the table below.

<table>
<thead>
<tr>
<th>Scale Level</th>
<th>Scale of Measurement</th>
<th>Scale Qualities</th>
<th>Statistical analysis</th>
<th>Example(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Ratio</td>
<td>Magnitude</td>
<td>All statistics permitted for interval scales plus the following: geometric mean, harmonic mean, coefficient of variation, logarithms</td>
<td>Age, Height, Weight, Temperature in Kelvin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Equal Intervals</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Absolute Zero</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Interval</td>
<td>Magnitude</td>
<td>mean, standard deviation, correlation-r, regression, analysis of variance, Factor analysis</td>
<td>Temperature in degree Celsius or Fahrenheit, IQ</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Equal Intervals</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Ordinal</td>
<td>Magnitude</td>
<td>Median and mode, percentile, rank order correlation, non-parametric analysis of variance,</td>
<td>Likert Scale, Anything rank ordered</td>
</tr>
<tr>
<td></td>
<td></td>
<td>None</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Nominal</td>
<td>None</td>
<td>mode, chi-squared</td>
<td>Names, Lists of words</td>
</tr>
</tbody>
</table>

### 6.3 What is reliability and validity of measurement?

When we construct some measuring instruments like scale or test, we have to ensure that the scale is reliable and valid. Reliability means accuracy and consistency of
the measuring instrument. You should verify if the instrument measures the variable accurately and is it consistent in measuring. On the other hand, validity means truthfulness of the measuring instrument. That means you have to verify if the measuring instrument measures what it intends to measure. It is more important that a measuring instrument is valid than reliable. If an instrument is not measuring what is should be measuring, then even if it is measuring accurately and consistently, it is meaningless.

The most important two methods of estimating reliability are test-retest and split half (internal consistency) methods. In case of test-retest, you have to administer the measuring instrument on a group and get data and again after 15 days or so you re-administer the test. If you get higher correlation between these two sets of scores then the instrument can be considered as reliable. While in case of split half, the instrument is administered only once and then correlation is ascertained between scores of odd items and even items. Higher correlation value indicates higher reliability of the measuring instrument.

The common methods followed to ascertain validity are content validity, construct validity, concurrent validity and predictive validity. When the contents of the items individually and collectively are relevant to the test, then it represents content validity. It is also called Judge’s validity since experts rate the validity of the items. Construct validity is the extent to which the scale measures a theoretical trait or construct. E.g. intelligence and IQ test. In case of concurrent validity, the test is correlated with some earlier defined criterion scores, whereas, predictive validity may be defined as the degree to which a measure predicts a second future measure.

6.4 How do I analyze my data?

Data analysis is based on the research design determined for the study. The most basic of them is descriptive analysis. Descriptive analysis provides description of the encoded data using frequency distribution and percentage summarizing the results in
tabular form. Often it is useful to obtain some average value that is representative of the distribution. The three commonly used measures of central tendencies are mean, median and mode. However, measures of central tendencies are misleading without the information about the measures of dispersion. The commonly used measures of dispersion are range, standard deviation and variance. CV or coefficient of variance is yet another useful statistics that combines the properties of central tendency and dispersion enabling comparison among different statistics of different samples. All these methods can be termed as univariate analysis.

Next is the bivariate analysis focused on the nature of relationships between two variables. A relationship between two variables can be assessed by special measures of relationship that reflect the relative utility of using one variable to predict another. The measures of relationship usually correspond to the level of measurement of the variables. Nominal relations are assessed by lambda or by Goodman and Kruskal’s Tau. Either gamma or Kendall’s Tau is used to calculate relationship between ordinal variables. In case of interval or ratio variables Pearson’s $r$ is used to calculate relationship.

Although bivariate analysis provide measures of relationship, in reality X and Y do not exist without being influenced by other variables. Therefore, control and prediction based on bivariate analysis is difficult to achieve. Whereas, multivariate analysis provides you tools for control, interpretation and prediction. Statistical control acts as substitute for experimental control and is accomplished through partial correlation. This explanation is to provide you bird’s eye view of statistical tools available for data analysis. Brief notes are given below so that you can apply statistical analysis to your data.

<table>
<thead>
<tr>
<th>Category</th>
<th>Frequency (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$f_1$</td>
</tr>
<tr>
<td>II</td>
<td>$f_2$</td>
</tr>
<tr>
<td>III</td>
<td>$f_3$</td>
</tr>
<tr>
<td>IV</td>
<td>$f_4$</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>F</strong></td>
</tr>
</tbody>
</table>
6.5 Notes on Univariate Analysis

Frequency Distribution

After the data has been coded and prepared for analysis, the first task is to construct frequency distribution to examine the pattern of data collected for dependent and independent variables. Frequency distribution is constructed by listing the categories within a variable and counting number of responses given in each category by the respondents. A general format of univariate frequency distribution table is given on the right.

In case of nominal variables, categories may be presented in any order but in case of ordinal data, categories should be arranged in ascending or descending order. The interval variables are usually continuous, the classification into categories is quite arbitrary. Many of the extension researcher use Mean + S.D. to classify interval and ratio variable. Naturally, they end up having maximum concentration of frequency, say around 68%, in the middle category and low frequency concentration at extreme categories. Researchers should refrain from such classification unless it is a requirement of the research.

Percentage Analysis

The frequencies in itself do not give a meaningful result unless it is compared with the frequencies of other categories. The simplest method is to convert the frequencies out of hundred. That is the frequencies are converted into frequency per cent (hundred). The percentages thus obtained can also be compared with across samples. The percentage of males as compared to females in Haryana is more than that in Kerala.

Measures of Central Tendency

When only a short summary is required, all the categories of the data cannot be presented. Under such a

\[
\bar{X} = \frac{\sum X_i}{N}
\]

Where
- \(\bar{X}\) = The arithmetic mean
- \(\sum X_i\) = The sum of total observations
- \(N\) = The number of observations
situation, measure of the central tendency, the central value around which observations tend to cluster, that is characteristic of a sample becomes very useful. Say for example, average income or per capita income of different countries can be easily compared rather than a table of different categories for each country. The commonly used measures of central tendency are Mode, Median and Mean (Arithmetic mean).

**Mean (Arithmetic Mean)**
The arithmetic mean is most frequently used measure of central tendency. It is easy to calculate. It is suitable for variables measured on an interval scale. It is amenable to further statistical operations. Mean is usually represented by $\bar{X}$.

**Median**
The median is another commonly used measure of central tendency. It is a positional measure that divides the distribution into two parts. Median is defined as the observation located halfway between smallest and largest observation. Unlike mean, median is not based on all the observations rather on number of observation.

**Mode**
The mode is the category or the observation that appears most frequently in the distribution. It is mostly used in case of nominal variables.

**Measures of Dispersion**
Any distribution of observation cannot be described completely without the mention about its distribution. The measures of dispersion around central value are range, mean deviation, quartile deviation, variance, standard deviation and coefficient of variation. We shall be discussing only commonly used measures of dispersion.

$$Md = L_i + \left( \frac{N/2 - CF}{f_i} \right) W_i$$

Where
- $M_d$ = Median
- $L_i$ = The lower limit of median class
- $CF$ = The cumulative frequency of class preceding median class
- $f_i$ = The frequency of median class
- $W_i$ = The width of median class
Range
The range is measured as the difference between the lowest and highest values. It is simplest of all measures of dispersion.

Variance and Standard Deviation
The variance of a distribution is the average of squares of the distances from the values drawn from the mean of the distribution. In other words, arithmetic mean is subtracted from each observation; the differences are squared, added and divided by total numbers of observations. The square root of variance is standard deviation.

Coefficient of Variation
Coefficient of Variation or CV is commonly used by general researchers. In social researches, it is seldom used. In fact, the standard deviation in absolute terms cannot be compared. A standard deviation of 1.5 will convey different meaning in relation to a mean of 10 than to a mean of 20. Mathematically coefficient of variation is the ratio of standard deviation and mean.

Univariate Distribution
After explaining descriptive statistics, one may be interested to describe the distribution of the responses. Statistically, one of the most important distributions is normal distribution. It is symmetrical and bell shaped. In case of normal distribution, the mode, median and mean coincide at the center of the distribution and is asymptotic.
6.6 Notes on Bivariate Analysis

After summarizing and describing the pattern of distribution of single variables, the next task before the extension researcher is to examine the relationship between variables.

Nominal Measures of Relationship

In order to measure relationship between two variables, the following measures could be of use which can be carried out with the help of SPSS.

Coefficient of Contingency: A measure of association based on chi-square. The value ranges between zero and 1, with zero indicating no association between the row and column variables and values close to 1 indicating a high degree of association between the variables. The maximum value possible depends on the number of rows and columns in a table.

Phi and Cramer’s V: Phi is a chi-square based measure of association that involves dividing the chi-square statistic by the sample size and taking the square root of the result. Cramer's V is a measure of association based on chi-square.

Lambda: A measure of association which reflects the proportional reduction in error when values of the independent variable are used to predict values of the dependent variable. A value of 1 means that the independent variable perfectly predicts the dependent variable. A value of 0 means that the independent variable is of no help in predicting the dependent variable.

Goodman and Krushkal’s tau: Goodman and Krushkal’s tau measures the association between nominal variables and can be applied in situations when lambda is inappropriate. That is when the marginal totals of the dependent variable are extremely uneven.

Ordinal Measures of Relationship

In case of ordinal variables, following measures of relationship could be used.
**Gamma:** A symmetric measure of association between two ordinal variables that ranges between negative 1 and 1. Values close to an absolute value of 1 indicate a strong relationship between the two variables. Values close to zero indicate little or no relationship. For 2-way tables, zero-order gammas are displayed. For 3-way to n-way tables, conditional gammas are displayed.

**Somers':** A measure of association between two ordinal variables that ranges from -1 to 1. Values close to an absolute value of 1 indicate a strong relationship between the two variables, and values close to 0 indicate little or no relationship between the variables. Somers' d is an asymmetric extension of gamma that differs only in the inclusion of the number of pairs not tied on the independent variable. A symmetric version of this statistic is also calculated.

Kendall’s tau-b: A nonparametric measure of association for ordinal or ranked variables that take ties into account. The sign of the coefficient indicates the direction of the relationship, and its absolute value indicates the strength, with larger absolute values indicating stronger relationships. Possible values range from -1 to 1, but a value of -1 or +1 can only be obtained from square tables.

Kendall’s tau-c: A nonparametric measure of association for ordinal variables that ignores ties. The sign of the coefficient indicates the direction of the relationship, and its absolute value indicates the strength, with larger absolute values indicating stronger relationships. Possible values range from -1 to +1, but a value of -1 or +1 can only be obtained from square tables.
Interval Measures of Relationship

When the variables being analyzed are at least interval, we can be more precise in describing nature and form of the relationship.

Pearson’s Product Moment Correlation Coefficient (r): It is a measure of linear association between two variables measured at least by interval scale. Values of the correlation coefficient range from -1 to 1. The sign of the coefficient indicates the direction of the relationship, and its absolute value indicates the strength, with larger absolute values indicating stronger relationships. The graphical representation on various kinds of relationships between dependent variable, Y and independent variable, X is presented.

Linear Regression: The method of specifying the nature of a relationship between two variables is referred to as regression analysis. Linear Regression estimates the coefficients of the linear equation, involving independent variable, that best predict the value of the dependent variable.

6.7 Notes on Multivariate Analysis

Establishment of bivariate relationship is the first step in data analysis. An observed correlation between two or more variables does not allow the researcher to make causal interpretations. A bivariate relationship may be due to chance, or it may be due to presence of a third variable, the intervening variable. Further, a phenomenon may be due to a number of independent variables. In any case introduction of more than one independent variable helps in clarifying and elaborating
the real relationships. Partial correlation and multiple linear regression are two measures of relationship that are employed to achieve the purpose of explaining complex relationship between more than two variables.

**Partial Correlation**

In partial correlation we study the effect of the independent variable on a dependent variable by excluding the effect of other independent variables/factors. Thus, if adoption (a dependent variable) is being affected by age, education and income, we can study the relationship between adoption and education by excluding the effects of age and income. This is a study of partial correlation between adoption and education. In case of simple correlation, the effect of age and income is not excluded but simply ignored.

**Multiple Correlation**

In multiple correlation we study effects of ALL the independent variables simultaneously on a dependent variable. Considering the same example cited above effects of age, education and income on adoption at the same time would be called a multiple correlation study.

**Multiple Regression Analysis**

Multiple regression is similar to linear regression, whereby we estimate the coefficients of regression equation. In case of linear regression equation two variables are involved, whereas there are more than two variables involved in multiple regression equation. Two or more independent/predictor variables are regressed on dependent/criterion variable. The measure of prediction is depicted by \( R^2 \), coefficient of determination.

The combination of two or more items or indicators yields a composite measure referred to as index.
Index Construction and Scaling Methods

When the extension researcher develops tools for data collection s(he) comes across the problem of measuring variables maintaining higher validity and reliability. Some of the variables, particularly nominal variables are easy to measure but the interval and ratio variables are not so easy to measure in social sciences. Like measuring Sex or caste is easy but in case of empowerment special measuring instruments are required. Indexes and scales are such measuring instruments which the extension researchers develop during the course of study.

Index Construction

Before index construction it is essential to answer these questions: What does one attempt to measure and how is the measurement going to be used? The various forms of the index are proportion, percent or ratio. A proportion is the frequency of observation in a category divided by total number of observation, mathematically f1/N. The proportion ranges from zero to one. A proportion becomes per cent when multiplied by 100, that is, f1/N X 100. A per cent varies from zero to 100. The ratio on the other hand is relative amount of one frequency of a category to another, for example male female ratio, ratio of input and output, ratio of small farmers to marginal farmers etc. Another common method of constructing indexes is by computing aggregate values. Depending on the purpose of index, these aggregates could be simple or weighted. Weight helps in giving relative influence of each indicator of an index. For example, The SES scale uses simple aggregate of various indicators like education, income, social participation etc. While relative effect of these indicators may be taken by future researchers if they assume that all indicators does not contribute equally to SES.

Scaling Methods

In social sciences, the commonly used scaling methods are Thurstone’s technique of equal-appearing intervals, Likert’s summated rating technique and Gutman’s scaling technique, and paired comparison. These techniques are dealt in detail in third section. But in case of multidimensional variables such as SES, empowerment and leadership,
There is a problem of multicollinearity. In case of multidimensional variables. A method developed by a group of Indian researchers called MKJBD method of multidimensional scale construction overcomes the problem of multicollinearity.

6.8 What is Statistical Inference - Test of Hypothesis and Estimation?

The process of selecting and using sample statistic (mean, standard deviation etc.) from a sample in order to draw conclusion about the population, is known as statistical inference. It is concerned with using probability concept to deal with uncertainty in decision making. Statistical inference deals with two types of problems:

(A) Hypothesis testing, and
(B) Estimation.

Hypothesis Testing

Test begins with assumptions is referred to as hypothesis testing. Hypothesis testing is made about population parameters and it includes the test of significance.

Procedure for Hypothesis Testing

(1) Setting up a hypothesis:

First and foremost thing in hypothesis testing is to set up a hypothesis about the population parameter. The samples are collected for sample statistics (mean, standard error, etc.) and the information is used to see how close is the hypothesized population parameter with the samples collected. To test the validity of hypothesis, the difference between populations mean (hypothesized value) with that of sample is compared. If the difference is small, then the hypothesized value for the mean is correct. If the difference is large hypothesized value for the mean is incorrect. Approach to hypothesis testing is not to construct a single hypothesis about population parameter, but to set up two different hypotheses (Null hypothesis and Alternative hypothesis). If one hypothesis is accepted, the other one is rejected and vice versa.

The null hypothesis is very useful tool in testing the significance (procedure for distinguishing whether the difference connotes any real difference or is merely sampling
fluctuations) of difference. In null hypothesis we consider that there is no real difference in the sample and that the difference, if at all exists, is due to sampling fluctuations. In alternative hypothesis, we consider real difference in the sample. For example, if we want to find out whether a particular insecticide is effective in controlling a particular plant disease, the postulated null hypothesis is that the insecticide is not effective in controlling the disease. The rejection of the null hypothesis indicates that the differences has statistical significance and the acceptance of the null hypothesis indicates that there is no difference or whatever differences is there, it is purely due to chance. The null and alternative hypotheses are distinguished by using two different symbols ($H_0$, denoting null hypothesis and $H_a$, denoting alternative hypothesis).

(2) **Setting up suitable significance level:**
After postulating the hypothesis, the next step is to test the null hypothesis symbolically designated as $H_0$ against the alternative hypothesis, designated as $H_a$ at a certain level of significance i.e. the researcher rejects or retains the null hypothesis depending on the specified level of significance. Generally, the significance level is expressed in percentage (5% or 1%).

(3) **Setting a test criterion:**
The next step in testing the hypothesis is to construct test criterion by selecting an appropriate probability distribution which can be properly applied. The most commonly used probability distributions in testing are $t$, $F$ and $\chi^2$ etc.

(4) **Computations:**
The fourth step involves using statistical tests (mean, standard error etc.). A random sample is necessary for this purpose to design a statistical test.

(5) **Decision making:**
Finally, it is necessary to make statistical conclusion to decide either to reject or to accept the null hypothesis. The decision to reject the null hypothesis indicates the acceptance of alternative hypothesis. The decision depends on the computed value of the test criterion.
If the observed statistic is less than the value of 5% probability level, then the difference between the sample statistics and the hypothetical parameter is considered to be significant and the null hypothesis is rejected (false) i.e. the sample observations are not consistent with the null hypothesis. Rejection of null hypothesis automatically leads to acceptance of alternative hypothesis.

While testing the hypothesis, four possibilities are involved.

(a) Test is rejected when the hypothesis is true. (Type I error)
(b) Test is accepted when the hypothesis is false. (Type II error)
(c) Test is accepted when the hypothesis is true and
(d) Test is rejected when the hypothesis is false.

The above four possibilities are stated in a tabular form as:

<table>
<thead>
<tr>
<th></th>
<th>Accept $H_0$</th>
<th>Reject $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_0$ is true</td>
<td>Correct decision</td>
<td>Type I error</td>
</tr>
<tr>
<td>$H_0$ is false</td>
<td>Type II error</td>
<td>Correct decision</td>
</tr>
</tbody>
</table>

In statistical hypothesis testing, Type I error is committed by rejecting the null hypothesis when it is true. It is more dangerous to accept Type II error (the false hypothesis) than to reject Type I error (the true hypothesis). The probability of committing a Type I error is denoted by (level of significance), and can be written as:

$$ = \text{Probability (Type I error)}$$

$$ = \text{Probability (Rejecting $H_0$, when it is true)}$$

Alternatively, the probability of committing Type II error is denoted by which can be written as:

$$ = \text{Probability (Type II error)}$$

$$ = \text{Probability (Accepting $H_0$, when it is false)}$$

In testing hypothesis, the main motto is to reduce both types of errors (Type I and Type II). But it is not possible to reduce both these errors simultaneously due to fixed sample
size. i.e. with the decrease of one type of error, the other type of error increases. Hence, to deal with this type of situation, it is necessary to decide the appropriate level of significance. In most of the statistical tests, the level of significance is generally set at 5% (P=0.05). This means the probability of accepting a true hypothesis is 95%.

While testing hypothesis, it generally refers to one-tailed or two-tailed test. In one-tailed test, the hypothesis is rejected if the sample statistic is significantly different from the hypothesized population parameter under study and in case of two-tailed test; the hypothesis is rejected if the sample statistic is significantly higher or lower than the hypothesized population parameter. For example, in testing the null hypothesis that the average income per household in a village is Rs.500/- as against alternative hypothesis that it is not Rs.500, the average income per household may be less or more than the average income of Rs.500. To reject the alternative hypothesis in studying the income is more or less than the average income, two-tailed test is applicable. In case of rejecting the hypothesis based on only whether the average income per household is Rs.500 or not, one-tailed test is applicable. Before going further, it is necessary to have some idea about sample distribution, standard deviation, standard error etc.

**Sampling Distribution**

Sampling is a tool which helps to know the characteristics of the universe (population) by examining only a small part of it. The basic objective of sampling is to draw inference about the population. The values obtained from the study of a sample, such as mean, standard deviation etc., are known as statistics and such values for the population are known as parameters.

**Estimation**

When samples are collected from population, it is desirable to draw inferences from these samples about the population parameter. Statistical estimation or simply estimation is concerned with the method by which population characteristics are estimated from sample. i.e. from the sample, attempt is made to estimate the true but
unknown values of the population parameter. For parameter estimation, two types of estimates are possible. i.e.

a. Point estimates and
b. Interval estimates.

**Point Estimates**

A point estimate is a single value used as an estimate to unknown population parameter. The procedure in point estimation is to select a random sample of size n (say) from a population and then use some preconceived method to arrive from the sample size n at a real number which can be regarded as estimator of a population parameter.

**Interval Estimates**

An interval estimate of a population parameter, on the other hand, is a statement of two values between which the estimated value lies. Hence, an interval estimate is always specified between two values. In statistical sense, interval estimation refers to the estimation of a parameter by a random interval, called Confidence interval, whose end points, say, L (Lower value) and U (Upper value). For example, if the estimated average income of people in a village is Rs.500, it will be point estimate and whereas, if the average income lies between Rs. 500(lower value) and Rs.600 (higher value), it will be interval estimate.

In general, between the above two estimates, point estimation has an advantage over interval estimation, because it provides single value. At the same time it has also a disadvantage that, as it provides a single value, it does not indicate how close the estimator is to the parameter being estimated. Further, in scientific investigation it is not necessary to know the exact value of a population parameter, rather, it is desirable to have some degree of confidence that the value obtained within a certain range. As the interval estimate provides the values within a range, these estimates are generally employed in practice.
Properties of Good Estimator

A good estimator is one which is close to the population parameter being estimated. A good estimator has the following properties.

(a) Unbiased: An estimator is said to be unbiased if its expected value is equal to the population parameter being estimated.

(b) Consistent: As the sample size increases, the estimator is closer and closer to the population parameter. In case of large samples (size), consistency is a desirable property for an estimator. For small samples, consistency is of little importance.

(c) Efficient: It refers to sampling variation. For instance, if two estimators are both unbiased, the one with smaller variation for a given sample size is more efficient than the larger variations.

(d) Sufficient: An estimator is said to be sufficient, if it conveys as much information as is possible about the parameter contained in the sample. The above statement indicates that, if sufficient estimator exists, then it is unnecessary to consider any other estimator.

Several methods are available satisfying above properties of a good estimator. Generally two methods (Least Square and Maximum Likelihood) are in use with following test of significances.

Test of Significance

Different test of significance under various situations may be summed up under the following three heads.

I. Test of significance for attributes

II. Test of significance for large samples and

III. Test of significance for small samples.

I. Test of Significance for Attributes:

In case of attributes, a characteristic under study tells about its presence or absence. For example, a sample was drawn in a village to study number of literates and illiterates. As
this situation involves only two types of attributes, we assume that the sampling distribution is binomial. The appearance of an attribute may be taken as success and its non-appearance as failure. For example, out of 1,000 people, if 100 people are found to be literates and 900 illiterates, then the number of success is 100 and failure is 900. The probability of success or \( p = \frac{100}{1000} = 0.1 \) and the probability of failure or \( q = \frac{900}{1000} = 0.9 \) so that \( p+q = 0.1+0.9 = 1 \).

Various test of significance for attributes can be grouped into the following heads:

i. Test for number of success,

ii. Test for proportion of success and

iii. Test for difference between proportions.

i) Test for Number of Success: As the sampling distribution is a binomial, its standard error is given as:

\[
\text{S.E. of number of success} = \sqrt{npq}
\]

where \( n = \) sample size,

\( p = \) probability of success in a trial and

\( q = \) probability of failure i.e. \( 1-p \)

Illustration: In a sample of 500 people from a village, 300 are rice eaters and rest wheat eaters. Can it be assumed that both the cereals are equally popular?

Solution: Let us assume that both the cereals are equally popular and the expected frequency for either rice or wheat would be \( \frac{500}{2} = 250 \) each.

The standard error = Difference between observed and expected number of rice eaters = \( 300 - 250 = 50 \)

Hence, Difference/S.E of difference = \( \frac{50}{11.18} = 4.47 \)

Since the difference is more than 2.58 at 1% level of significance, it is not assumed that both the cereals are equally popular i.e. the hypothesis postulated is rejected.
ii) Test for Proportion of Success: Sometimes record on proportion of success is available instead of simply number of success. i.e. \(1/n\)th of the number of success, where \(n\) is the number of record. Standard deviation of proportion (S.E\(_p\)) of success is given by:

\[ S.E_p = \sqrt{\frac{pq}{n}} \]

iii) Test for difference between proportions: If two samples are drawn from the different populations, it may be interesting to know the significant difference between the proportions. The standard error for the difference between the two proportions is given by:

\[ S.E (p_1 - p_2) = \sqrt{pq \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \]

where, \(p_1\) & \(p_2\) = proportions of two different samples with \(n_1\) & \(n_2\) sample sizes

\(p\) = pooled estimate of proportion in the population i.e. the value of \(p\) is obtained as:

\[ p = \frac{n_1p_1+n_2p_2}{n_1+n_2} \]

If \((p_1 - p_2)/S.E\) is less than 1.96 (5% significance level), the difference of proportion is not significant and the difference can be due to random sampling variation.

II. Test of Significance for Large Samples:
Testing the significance of large samples (30 or more) refers to sampling variables such as plant height, panicle weight etc. which takes different values. Hence, it may not be possible to classify under one or two heads as in case of attributes. For dealing with large samples following assumptions are made.

(a) the random sampling distribution of a statistic is approximately normal and

(b) sample value is close to population value so that this value can be made use for calculating standard error of the estimate.
In case of large sample, the standard error of mean is given by:

\[ \text{S.E. of mean} = \frac{\sigma}{\sqrt{n}} \] if standard deviation of population is known.

where, \( \sigma \) is the standard deviation of population.

If the standard deviation of population is not known, then standard deviation of sample is used for calculating the standard error of mean. The significant difference for large samples can be tested using normal probability integral table. If the difference is less than the table value at 5% (or 1%) level, then the difference is due to sampling fluctuations and the null hypothesis is accepted.

**Illustration:** The mean height of students, from random sample of size 100 obtained from a school, is 48 inches. The standard deviation of the distribution of height of the population is 3 inches. Test whether the mean height of the population is 54 inches at 5% level of significance.

Solution: Let us assume the hypothesis that there is no significant difference between the sample mean and the population mean.

Standard error (S.E.) of mean = \( \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{100}} = \frac{3}{10} = 0.3 \)

Therefore,

\[ \text{Difference/S.E of difference} = \frac{54 - 48}{0.3} = \frac{6}{0.3} = 20 \]

Since the difference is greater than the normal probability integral table value, the postulated hypothesis is rejected and hence, the mean height of the population cannot be 54 inches.

The comparison of sample mean with its hypothetical value (population mean) is not a problem of frequent occurrence. A problem more commonly met with agricultural research is the comparison between two sample means. To compare two random samples two-tailed test is used.
If two independent random samples with sizes \( n_1 \) and \( n_2 \) (both \( n_1 \) and \( n_2 \) less than 30) respectively are drawn from the same population, the standard error of the difference between the sample means is:

\[
\text{S.E of difference between two sample means} = S.E. \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}
\]

If two random samples are drawn from two different populations with means, standard deviations and sample sizes are respectively \( x_1, \sigma_1, n_1 \) and \( x_2, \sigma_2, n_2 \) then the S.E of the difference between the two means is:

\[
\text{S.E of difference between two means} = \sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}
\]

Illustration: A man buys 50 electric bulbs of Philips make and 50 bulbs from HMT make. He finds that Philips bulbs gave an average life of 1500 hours with standard deviation 60 hours and HMT bulbs gave an average life of 1512 hours with standard deviation of 80 hours. Test the difference of means between two makes.

Solution: Let us assume the hypothesis that there is no difference in the mean life of two makes.

The standard error of difference of means is:

\[
= \sqrt{\left(\frac{60^2}{50}\right) + \left(\frac{80^2}{50}\right)}
\]

\[
= 14.1 \text{ (approx.)}
\]

The observed difference in mean between two makes = 1512-1500 = 12.
Hence, Difference/S.E of difference = 12/14.1 = 0.85 (approx.)

Since the difference is less than 1.96 SE (at 5% level of significance), it can be concluded that difference of means is not significant and hence two means are same and the null hypothesis is accepted.

III. Test of Significance for Small Samples:
Tests of significance used for large samples are different for small samples (less than 30) for the reason that the assumption stated that (a) the random sampling distribution of a statistic is approximately normal and (b) sample value is close to population value so that this value can be made use for calculating standard error of the estimate. With the small samples, the main interest is not to estimate the population value (unlike in case of large sample) but to test the given hypothesis.

Assumption of normality:
Unless otherwise stated, small samples require an assumption that the parent population is normal (of course, the assumption of normality is not very much warranted in case of small samples). If the parent population is not normal, the methods of testing the significance for small samples cannot be applied with much confidence. However, some suitable tests were developed in testing the significance of small samples which can be applied with some degree of confidence. Sir William Gosset, and R.A. Fisher developed tests to deal with such small sample problems. Test developed by Sir Gosset under the pseudonym as Student (Student-t test) while, Fisher developed another test known as z-test. These tests are popularly known as t-test and z-test. These tests are based on t-distribution and z-distribution.

Testing the significance of the mean of random sample:
The t-distribution is used when the sample size (n) is 30 or less and if the population standard deviation is unknown, the t-statistic is:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$
where,
\[ \bar{X} = \text{sample mean}, \mu = \text{population mean}, n = \text{sample size} \]
\[ s = \frac{\sum (X_i - \bar{X})^2}{n-1} \]

Illustration: A random sample of size 16 has 53 as mean. The sum of square of the deviations taken from the mean is 135. Can the sample be regarded as taken from the population having 56 as mean?

Solution: Let us assume the hypothesis that there is no difference between the sample and population means. Substituting the appropriate values in t-statistics explained above, we have:
\[ S = \sqrt{\frac{135}{15}} = 3 \text{ and hence; } t = \left[ (53 - 56) \sqrt{16} \right] / 3 \]
\[ = [3 \times 4] / 3 = 4 \]

Since calculated t-value is more than the table t-value, it may be considered as significant and hence, the postulated hypothesis is rejected.

Testing the significance between two independent random samples: In case of two independent random samples of sizes \( n_1 \) and \( n_2 \) the hypothesis testing that the samples come from the same population, the statistic is given as:
\[ t = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1, X_2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \]

and
\[ s_{X_1, X_2} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \]

The connotations are as earlier.
Illustration: Two types of drugs were used on 5 and 7 patients for reducing their weight. The drug A was imported while the drug B was indigenous. The decrease in the weight after using the drugs for six months was as follows:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A (X1):</td>
<td>10</td>
<td>12</td>
<td>13</td>
<td>11</td>
<td>14</td>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>Drug B (X2):</td>
<td>8</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>15</td>
<td>10</td>
<td>9</td>
<td>11</td>
</tr>
</tbody>
</table>

Test: Is there any difference in the efficacy of the drugs?

Solution: Let us assume the hypothesis that there is no difference (significant) in the efficacy of the two drugs. Applying the t-test; we have

\[ \bar{x}_1 = 12; \quad \bar{x}_2 = 11; \quad (n_1-1) S^2_1 = 10 \text{ and } (n_2-1) S^2_2 = 44 \]

Hence, \[ t = \frac{\text{Difference of means}}{\text{Standard Error of difference}} = \frac{1}{2.3} = 0.43 \]

The table t-value at 5% probability level (t_{0.05}) = 2.23 at \( n_1 + n_2 – 2 \) degrees of freedom which is greater than the calculated t-value. Hence the difference in efficacy of two drugs is not significant and the hypothesis is accepted.

**Testing the difference of two dependent or paired samples:**

There are some situations where the condition of independent samples does not hold good. Two samples are said to be dependent when the elements of one sample are related to those in other. In other words, two samples may consist of pair of observations made on the same selected population elements. For instance, in order to find out the efficacy of two drugs; one is made within the country and the other one is made outside the country. When samples are dependent, they comprise of the same number of elements. The paired t-test is based on \((n – 1)\) degrees of freedom and is defined as:

\[ t = \frac{\sum a}{\sqrt{\frac{(n \sum a^2 - \sum a^2)/(n-1)}}} \]
Testing the significance of correlation coefficient:

In testing the significance of correlation coefficient (r) from a bivariate normal population, the postulated hypothesis is that the correlation coefficient of population is zero, i.e. the variables in the population are uncorrelated and the t-test is given by:

\[ t = r \sqrt{\frac{n-2}{1-r^2}} \]

Here, t is based on n – 2 degrees of freedom.

If the calculated value of t exceed at 5% level of significance for n – 2 degrees of freedom, the value of r is significant and the conclusion is that the variables are consistent with the hypothesis of an uncorrelated population.

Z-test for significance of correlation coefficient: For testing whether r differs significantly from zero, the t-test is appropriate. Prof. Fisher developed a method of testing the significance of the correlation coefficient in small sample in which, the correlation coefficient is transformed into Z and hence it is known as z-transformation. The z-transformation is used to test (a) whether an observed value of r differs significantly from some hypothetical value or (ii) whether two sample values of r differ significantly. z-transformation is given as:

\[ z = \frac{1}{2} \log_e \left( \frac{1+r}{1-r} \right) = \frac{1}{2} \left\{ \log_e (1+r) - \log_e (1-r) \right\} \]

This quantity z is nearly normally distributed for all values of n with standard error as:

\[ S_z = \frac{1}{\sqrt{n-3}} \]
Limitations of tests of significance:

i) Tests of significance are to make decisions but not decisions in themselves. In some situations where real differences exist but do not produce evidence that they are statistically significant. Outmost care is to be taken before taking any decisions. Tests of significance should not be used mechanically.

ii) Conclusions are to be given in terms of probabilities but not certainties. When a test shows that a difference is statistically significant, it suggests that the observed difference is real and not due to sampling fluctuations.

iii) Though tests may indicate that the difference has statistical significance, yet they do not tell why the difference exists. However, they do suggest the need for further investigations in order to reach definite answer.

iv) Tests of significance should not be applied when assumptions are not met about the population from which the samples are drawn.

χ² Test and Goodness of Fit

Various tests of significances such as t, F and Z are all based on the assumption that the samples are drawn from normally distributed populations. As the testing procedure requires assumption about the population parameters, these tests are known as parametric tests. There are many situations in which it is not possible to make any assumptions about the population. This limitation has led to the development of a group of alternative techniques known as non-parametric tests. In non-parametric (distribution free) tests, no assumption about the population is required. Chi-square test is one of such non-parametric test in which no assumption is required about the population. The chi-square test is defines as:

\[ \chi^2 = \frac{(O-E)^2}{E} \]

where, O refers to observed and E refers to expected frequencies.

Uses of χ² tests:
This test is one of the most popular statistical inference procedures. It is applicable widely to many problems in practice. The test is applicable only for the comparisons of observed and expected values of absolute frequencies; as such, it should not be used for comparing the observed and expected values either of relative frequencies (proportions).

6.9 Let us sum up

We have learned the coding, scoring and data entry in this unit, followed by level of measurements. Basically, there are four level of measurements – nominal, ordinal, interval and ratio. It is important to know the level of measurement as it is crucial for further data analysis and inferences from the statistical analysis.

If we will get the same measurement repeatedly then the measuring instrument can be said to be reliable. Similarly, if the measuring instrument measures what it intends to measure then it is called a valid instrument.

The univariate analyses are frequency analysis, percentage, measure of central tendency and measure of dispersion. While, bivariate analyses include non-parametric statistics, correlation and regression. Multiple correlation and multiple regression are examples of multivariate analyses.

Statistical inference deals with two types of problems:

(a) Hypothesis testing, and

(b) Estimation.
Unit 7

Report Writing

Structure

7.0 Objectives

7.1 Why should we prepare research report?

7.2 What are the different forms of research report?

7.3 Reports for administrators and policy makers

7.4 A General Format for Preparation of Report

7.5 Let us sum up

7.0 Objectives

By the end of this unit you will be able to follow

- the procedure of preparing a research report
- how to identify the targeted users & sponsor of research, administrators or policy makers, academic community and then preparation of the research report to suit their needs
- an appropriate logical framework for preparing a research report

7.1 Why should we prepare research report?

I would like you to recall the Unit 3 which dealt with Literature Review and Developing Theoretical Orientation. Just imagine if earlier researcher might not have prepared their research report. How could you have done literature review and developed theoretical orientation for your research? Also recall the cycle of research which we have discussed as an ongoing process. The research report is like a baton in a relay race. It is your report that maintains the continuity in research in the area of your study. Report preparation is one of the most important aspects in research.
Your research may have the most brilliant hypothesis, the most carefully designed, scientifically conducted study with the most striking findings; but unless these components are compile together in a meaningful way and communicated to others, it will not have much relevance in the field of research. Hence, after completion of your research project, it is obligatory for you to submit a report to the authority within the stipulated period for appropriate evaluation of the research. Some feel that writing of the report is an unpleasant task, sometimes put off as far as could be managed and actually when time comes for writing, usually sufficient time may not be there for preparation of the report.

Report writing is a process of analyzing, understanding, appreciating and consolidating the findings of the project in a meaningful way and study the phenomenon in a coherent manner. The sole purpose of scientific writing is that it helps to know what has been done in the course of research, why it has been conducted, how it was done, what results were obtained, and what conclusions and recommendations have been arrived at. Since the report is submitted after the completion of the study, it is usually written in the past tense and sometimes present tense may be used for those statements of continuing and general applicability. Above all, the research report should be written in a clear and unambiguous language so that the readers will be able to assess the adequacy and validity of research. The purpose of the report is not communication with oneself but communication with the other readers. The researcher makes an attempt to communicate his own thoughts through a lucid and methodical research report in order to make it meaningful to the reader and For this purpose, research may have to clearly be aware of what the readers/users want to know about the study and how can this information be best presented.

Different procedures have been used to prepare the report depending upon the type of users aimed at by the report. It may start with the statement of the problem,
presentation of the research methodology, results, discussion, conclusions and implications.

There are some basic qualities of good scientific writing, which are accuracy and clarity. It would be desirable to have someone to go through the detailed outline and make a comment in regard to the coverage, sequencing, the points of omission, etc. There are different outlines used for presentation of the report, which may vary with the nature of the study and the particular discipline in which the report has been prepared.

7.2 What are the different forms of research report?

The form of a research report will depend upon the targeted user for whom it is intended. The research report can be prepared for the academic community, the sponsor of research, the policy makers or the general public.

Reports prepared for academic users may take the form of a thesis/dissertation leading to some academic degree or it may take the form of a monograph incorporating in detail the entire process of research or it may take the form of a research paper dealing with different aspects of the study published in different journals. Here the main objective is to indicate the scope, the methodological variations made in undertaking the study rather than the findings *per se*.

The research reports, prepared for the sponsor of research, may want to highlight the things that lead to understand a problem or solve a problem. The purpose of the research report for the sponsoring authority is to inform them various aspects of the problem studied based on the objectives which guided the sponsorship of the study. Therefore, the researchers of this type have specific objectives in terms of fulfilling the requirements of the sponsor in giving answers to the research questions proposed by them in an effort to solve or understand the problems.

As for the research reports published for use the consumption of the general public, it may take the form of summary reports, articles or brochures. Their main concern
would be to know the salient features of the findings without bothering about technicalities. So based on the targeted user, research reports can be classified into:

1. Comprehensive research report
2. Research articles
3. Summary reports.

Each of these may take the form in which it meets the requirements of the targeted user. However, it is felt that attempting to write a multi-purpose report to serve all users becomes difficult because what each one looks for in the report will be uninteresting for those who are not concerned with the details. It is felt that the most useful approach is to choose and write for a single targeted user at a level appropriate to it.

The form, content and style of the research report should be chosen to suit the level of knowledge, experience and interest of the targeted user as well as to anticipate the likely uses for the research findings.

Some of the suggestions for preparing comprehensive research report could be as follows:

1. The report should be an effort to narrate the total research process and experience. The experience of the researcher should be documented in selective and organized way.

2. The objective of the research report is to communicate with the targeted user and hence the level or knowledge, understanding of the users should be kept in mind.

3. The major thrust of the report should be to communicate what actually happened throughout the research process and not simply what we hoped would happen at each stage.

4. Research experience and findings that may not seem to have any relation to the immediate objectives of the study should not be eliminated in a hurry.
What may appear to be irrelevant information may turn out to be important to understand the difficult situation.

5. The report should not be merely selective-narration of our successes but may also contain the limitations along with the reasons for it.

6. It is more appropriate time first to prepare an adequate outline and then to follow it by drafting a report.

7. Since most readers may not go through the entire report, it is necessary to organize the chapters and sections effectively with meaningful titles so that those interested may read the concerned sections as required.

Research articles deal with different facets of a research problem in a more detailed fashion emphasizing the methodology adopted to study those aspects and the conclusions drawn to be related to the specific objectives. A research article may contain following parts.

i. Title of the research article
ii. Name(s) of authors
iii. Abstract
iv. Introduction
v. Selected Literature review
vi. Methodology followed
vii. Results
viii. Conclusions
ix. Acknowledgement
x. References

**Summary Reports:** These are generally meant for the general public. Since summary report may receive wide distribution, they are frequently written in less technical language and may make greater use of better methods of presenting the data. It focuses on major findings, their elucidation in simple and clear language and categorical statements about the implication of the study and recommendations.
7.3 Reports for Administrators and Policy Makers

You may have to prepare reports for administrators and policymakers of researches which may have implication for their activities.

If the research is sponsored by some organization that has assigned to diagnose the situation, evaluate the action programme or advice them about alternative programmes to solve some particular problem situations. The reports should be different as the administrators or the policymakers are already concerned and interested in the problem.

However, in these reports, the technical details concerning design and execution of research are of subsidiary interest to the administrators or policy makers as they are primarily concerned about our diagnosis of the problem and our recommendations of the programmes studied. In case your research is aimed at evaluation of on-going action programme, the report should contain -

i. An assessment or appraisal of the quality and quantity of activity or effort
ii. Describe performance (results of efforts)
iii. Evaluate adequacy of performance with regard to the total need,
iv. Report efficiency (in terms of cost, time and personnel)
v. Specify how and why the programme did not work or worked or what attributes of the programme made it more or less successful, which recipients are more or less affected, under what conditions the programme was more or less successful and the nature of the effects of the programme (unitary or multiple effects and cognitive, attitudinal and behavioural effects)

The report may be divided into two parts; in the first part a brief introduction containing the background of the problem, the broad aspects of the programme, and the terms of reference or objectives of the study may be given. It may be followed by a summary of the findings and suggestions and recommendations (sometimes this part is referred as executive summary)
The second part, it may contain a detailed description of the programme, sources of data, procedures followed, statistical analysis of the data, discussion, conclusions and recommendations. This may be followed by selected references and appendices of technical matters. Language used for preparing the report for the administrators and policy makers should be carefully chosen. As far as possible, the observations made or suggestions offered should have the findings as the base. The understanding of the practical problems of the implementing agencies should also be borne in mind while making the recommendations. By and large, the evaluation report should aid in understanding the problem areas, method of overcoming them and contribute to the constructive development of the programme.

7.4 A General Format for Preparation of Report

Finally the broad sequence of contents of a comprehensive research report would be as follows:

Initial pages: The initial pages contain Title page, Foreword, Acknowledgement, Table of contents, List of Tables, List of Figures, List of Photographs, Table of Abbreviations, Table of Symbols.

**Evaluation of Rice Development Programme of Odisha**

*A Project Report*

**Title page:** The title of the report should be clear, concise and indicative of its contents. In other words, it must state the problem adequately and in clear terms indicating the relationships of the variables studied.

**Foreword:** Generally a foreword of about a page indicating the content of the study, brief statement of the problem, objectives, agencies sponsoring the study etc to be written by a person who conducted the research or by the head of the research organization or someone of importance associated with the type of research that was done.
Table of contents: The purpose of table of contents is to enable the readers to identify the parts of the report and to see the relationship of the parts to one another and indicate the title of each chapter and section, sub-section along with the page number.

List of Tables, List of Figures, List of Photographs, Table of Abbreviations, Table of Symbols: These initial pages are optional and it depends upon you whether you want to specify it or not.

Introduction: It should contain information about the nature of the research problem, its importance in the larger context, the scope and the type of attempts made to answer the research question. It may also contain a brief narration of the way in which the material of the report is organized.

Review of previous studies: It is also necessary to provide a brief review of the existing information in relation to the research study. This may involve an analysis of the various findings related to the type of research in order to bring home the existing knowledge in the area and to indicate the knowledge gap leading to the formulation of the research hypothesis, for the present study. Sometimes this can also form part of the introduction wherever information about previous findings are limited in scope. As otherwise, a separate chapter or section may be given delineating the various aspects on which previous studies have been conducted and the conclusions arrived at, in order to put the research problem under study in the proper perspective.

Methodology: The methodology chapter may contain information about the objectives of the study in clear and precise terms or the hypotheses formulated, the method of formulating the hypotheses, etc., followed by the brief description of the study area, the type of respondents covered (sampling), the nature of sample and the method of selecting the sample given in detail. It should contain information about the type of measuring tools used, the method of developing it, how the variables were identified, the decision about the nature of measuring instruments or the method
followed to collect data, the type of secondary data collected along with their source and justification for using any particular method be given in a more detailed manner.

The method of collecting the data, the type of problems experienced and the limitations of the data may also be indicated in a detailed form to provide the context under which the data were collected.

Following this, the method of treatment of the data, the procedure and statistical techniques adopted for treatment of the data, both qualitative and quantitative should be indicated. The need for using different statistical techniques and the type of inferences that are expected to be drawn should be emphasized in relation to the research question or hypothesis as the case may be.

**Discussion and Interpretation:** While preparing the data, proper tabulations must be made giving information about the type of data contained in each table or the type of measures utilized. The way in which the table has to be read should be implied in the title and explanation of the figures contained may be given as footnotes.

The first task of the research is to draw inferences from the table to indicate the trends of the data and relate these trends to research hypothesis or objectives as the case may be or to the findings of the previous studies. The interpretation of the data can be based on one’s general knowledge of the area, previous findings or the social context in which the study was done. In case the interpretation is repetitive, then it is necessary to combine a series of tables of results in order to derive the type of findings and then attempt an interpretation in the larger context.

One can use graph to predict the trends of the results as it would explicitly bring home the type of information more effectively than the table of contents. The use of table, the statistical techniques relevant to the analysis of the data and for drawing inferences and relating it to other larger findings should constitute a part of the discussion of the findings of the study.
However, following major aspects should be taken into account.

1. A statement of the inferences drawn from the findings of a particular research may be expected to apply in similar situations.

2. As a qualification of these inferences, the researcher should indicate the condition of his study that limit the extent of legitimate generalization. It is also necessary to make an observation on typical results obtained and related to the methodological contribution, if any. The discussion of the implications also should include relevant questions that are still unanswered or new questions that are raised by the correct study.

7.5 Let us sum up

The method of organizing the research findings and experience in a coherent manner constitutes the research report. While writing the report, the researcher should keep in mind the targeted user: the academic community, administrators and policy makers, or the general public. The approach followed to write the report depends on for whom it is intended and for what purpose. The comprehensive research report is generally intended to the academic community for purposes of a university degree. Care need to be taken in the writing style, language used as the purpose is to inform others what one wants based on one’s research findings rather than show one’s prowess in vocabulary.