UNIT 8: METHODS OF COLLECTION

UNIT STRUCTURE

8.1 Learning Objectives

8.2 Introduction

8.3 Data Collection: Meaning

8.4 Sources of Data Collection

8.5 Tools for Data Collection

- 8.5.1 Interviewing
- 8.5.2 Questioner
- 8.5.3 Observation
- 8.5.4 Focused Group Discussion

8.6 Concept of Interview Schedule and Interview Guide

8.7 Let Us Sum Up

8.8 Further Reading

8.9 Answer to Check Your Progress

8.10 Model Questions

8.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- acquire knowledge about meaning and importance of data collection in research
- know about primary and secondary data
- discuss the tools of data collection
- explain the concept of interview schedule and interview guide.

8.2 INTRODUCTION

Till now, we have learnt about review of literature, design of research and various sampling techniques used in research. In this unit we will learn about collection of data- an important part of research. Starting from meaning and collection of data, the unit covers about primary and secondary data along with the tools of data collection.

8.3 DATA COLLECTION: MEANING

Any statistical enquiry originates from collection of related facts or information. These facts must be capable of measurement and this measurement may either be qualitative or quantitative. The facts and figures collected for any statistical enquiry, are recorded systematically. There are three concepts that are associated with data and its collection:

- Investigator
- Respondents
- Statistical units

An **Investigator** is a person who conducts the statistical enquiry. He must be a trained person and well-acquainted with statistical analysis. He/she counts or measures the characteristic under study. These measurements are further analysed statistically according to the objectives of the study.

Respondents are the persons from whom the information are collected by the investigator.

The measurements or counts in any statistical investigation are done with some identifiable object or a group of objects. These are known as statistical units. These units have to be defined clearly before collection of data.

Thus, data collection is a process. It consists not only of counting or enumerating or measurement of information but recording of the same also. It must be a systematic process. In fact, the entire structure of the statistical analysis for any enquiry is based upon systematic collection of data.

8.4 SOURCES OF DATA COLLECTION

In preceding section, we have learned about the meaning of data collection. Having discussed the preliminaries we have come to the conclusion that any statistical enquiry must ensure original

collection of data. That is, it should be correct and proper. Otherwise valid and useful conclusions cannot be done even after adopting sophisticated techniques of analysis and presentation of results. Data may be obtained from the following two sources:

- The investigator or the organizer may collect data first hand. This type of data is termed as Primary Data.
- The investigator may use the data from other published sources. This type of data is termed as Secondary Data.

It is clear that primary data are originally collected by an investigator or agency for the first time for any statistical investigation and used by them in the statistical analysis. Nature of primary data depends on the objectives and scope of the enquiry. Various variables as well as statistical units for them is defined prior to the collection of data.

On the other hand, secondary data have already been collected and processed by some agency or person, and taken over from here and used by any other agency for their statistical work. It is termed as secondary data as far as second agency is concerned. They may or may not be published. In the use of these data there exists no any problem regarding definitions of various terms and units used. It is because these type of data have already been collected under a given set of definitions of various terms and units used.

It is important to know that the distinction between primary and secondary data is a matter of degree or relativity only. The same set of data may be secondary in the hands of one and primary in the hands of others. A data set is primary to the source who collects and processes them for the first time. The same set becomes secondary for all other sources who later use it.

To clear the distinction, let us consider the data on fertility and mortality rates for India. Office of Registrar General of India, New Delhi publishes reports on various demographic rates such as birth rate and death rates etc. Similarly, United Nations Organisation

(U.N.O.) publishes U.N. Statistical Abstract wherein data on various aspect are taken from individual countries. Therefore, so far as data on birth rate and death are concerned, the office of the Registrar General of India is the primary source, while U.N.O. is the secondary source.

In the following a comparision is made between primary and secondary data:

Table 8.1: Primary and Secondary Data: A Comparison

Sl. No	Criteria	Primary Data	Secondary Data
1	Data Collection method	Data collected by researcher	Data collected by others
2	Originality	Original or unique information	Not original or unique information
3	Adjustment	Doesn't need adjustment is focused	Need adjustment to suit actual aim
4	Sources	Surveys, Observations, Experiments	Internal records, Govt. published data etc
5	Methods	Observation, Experiment, Interview	Desk -research method, Searching online etc.
6	Reliability	More reliable	Less reliable
7	Time consumed	More time consuming	Less time consuming
8	Need of investigators	Needs team of trained investigators	Does not need team of investigators
9	Cost effectiveness	Costly	Economical
10	Collected when	Secondary data is inadequate	Before primary data is collected
11	Capability	More capable to solve a problem	Less capable to solve a problem
12	Suitability	More suitable to achieve objective	May or may not be suitable
13	Bias	Possibility of bias exist	Somewhat safe from bias
14	Collected by	Researcher or his agents	Persons other than who collects primary data
15	Precaution to use	Not necessary	Quite necessary



CHECK YOUR PROGRESS

Q 1: State whether the following statements are true (T) or false (F).

(a) In collection of data, the statistical units should be stable. (T/F)

.....

.....

(b) Primary data is collected from a secondary source. (T/F)

.....

.....

Q 2: What are the main concepts associated with collection of data? (Write in about 60 words)

.....

.....

8.5 TOOLS OF DATA COLLECTION

In Statisics, there are different tools of data collection.But in this section we will study only those tools which are considered to be important from the aspects of Social Work.

8.5.1 Interviewing

Meaning and Definitions: In simple terms, interview is the verbal conversation between two people aiming at the collection of relevant information for the purpose of research. It is a technique where the interviewer asks questions to elicit information from the respondent. That is why it is a very personal and direct form of research.

McNamara defines interview as a particularly useful method of getting the story behind a participant's experiences. According to him through this method the interviewer can pursue in-depth information around the topic.

According to Good & Hatt (1952), “Interview is a close face to face conversation or dialogue between the investigator or interviewer and the interviewee.”

For P.V. Young (1966), Interview is “a scientific method through which a person enters into the life of a stranger”.

Based on the above definitions we can say that Interview is a verbal communication or two way conversation through which the researcher obtains information about a problematic situation.

Objectives of Interview:

Following are the major objectives of a Research Interview

- To collect information about an existing problem in society
- To create a source of true, original and firsthand knowledge
- To identify the true feelings of the respondent regarding a particular situation
- To provide the researcher with a rich hypothesis
- To aid the researcher in observing and understanding the situation quickly
- To gain new knowledge and additional information about a research area
- To help the researcher to draw quantitative facts and conclusions

Categories of Interview: There are different types of interviews in terms of structure, roll of the interviewer, number of respondents, and types of research. Some of those are personal interview, self- administered interview, focus group interview, standardized interview, depth interview, interview over telephone or computer etc. all these types of interview can be brought under three broad categories-

- **Structured Interview**

  Structured interview is a technique of interview generally applied in quantitative research. It is most commonly seen in the form of questionnaires in survey research. In simple terms structured interview is a method of collecting data for a statistical survey. A set of questions both close-ended and open-ended are asked to the respondent and recorded accordingly. For this purpose the researcher prepares an interview schedule in which all respondents are asked the same questions with the same wording and in the same sequence.

- **Unstructured Interview**

  Unstructured interview is an interview process generally used in qualitative research methods. In an unstructured interview, questions are not pre-arranged. Generally the interviewer engages the respondent with open-ended questions as and when required. It is completely opposite to structured interview and is a very flexible technique of Interaction. This kind of interview is also known as discovery interview or guided conversation.

- **Semi Structured Interview**

  This is the most common technique among the interview processes. The procedure of a semi-structured interview depends on the instinct of the researcher. In such an interview, the interviewer may prepare a list of questions but does not necessarily ask them all or in a particular order. Sometimes the researcher only addresses some general topics, hence we can say that semi-structured interview depends on the responses of the respondent.

8.5.2 Questionnaire

Preparing a good questionnaire can be quite a task. It is a time-consuming and complex process. A good quality questionnaire is a must for good quality data, and so the preparation of a questionnaire must always keep in mind the needs of the research concerned. A Questionnaire can be structured or semi-structured depending on the needs of the research. Structured or closed questionnaires usually consist of pre-set options from which respondents need to choose their answer. Semi-structured or open-ended questionnaires consist of fixed questions but leaves space for the respondents to answer in their own words. Semi-structured questionnaire may result in subjective answers for some questions, and hence does not fully qualify as quantitative data. Structured or closed questionnaire results in more objective answers, which help the researcher to quantify the data and analyze it.

A good questionnaire is expected to have simple, short and crisp questions that are easily understandable by the respondents. The more comprehensible the questionnaire is, the more possibility that the data gained from it is objective and rational.

A good questionnaire should be comprehensible by the respondents. It may emphasize on the significant or important areas of the concerned research and yet the questions should be framed in a way so that the respondents do not hesitate to put forward their answers or opinions. The responsibility of a good questionnaire towards research is that the respondents should realize the importance of the questions asked and not brush them off as unimportant or insignificant.

A good questionnaire should consist of logical questions looking for logical answers. The questions should be set in

such a way that the researcher is able to put across her point in the easiest possible language using minimum sentences. Unnecessary length of questions tends to confuse the respondents and run the risk of misinterpretation.

A good questionnaire should avoid questions that may be embarrassing or annoying to the respondents. Although questionnaire as a data collection tool does not leave much ground for the respondents to express themselves subjectively and also in most cases guarantee anonymity, in the ones that involve face to face interaction with the respondents, the questions should be framed with more sensitivity and care so as to not make the respondent uncomfortable. This is important for getting data which is not coloured by uneasiness.

Above all, a good questionnaire should be attractive. It should be designed in such a way that it draws the interest of the respondents. A skilful designing is thus the most important thing to do while preparing a research questionnaire, since no matter how important the questionnaires are, failure to engage the respondents would directly mean failure of the questionnaire to collect adequate data.

8.5.3 Observation

Meaning : In social science research, observation directly refers to a data collection method where the data is collected through observing. According to Marshall and Rossman (1989), observation is the “systematic description of events, behaviors, and artifacts in the social setting chosen for study”. It presents a “written photograph” of the situation under study by allowing the researcher to use her five senses to describe the existing situations (Erlandson et al 1993). Observation can be both structured and unstructured. A structured observation uses pre-defined schedules and specific variables. Unstructured

observation however adopts a more open approach towards the object of study or the field from which the data is to be collected.

There is a varied range of ways in which observation as a method has proved to be useful to researchers. Observation lets the researcher understand his field beyond theories and pre-assumptions. There are a number of facts that the interviewees are unable to reveal during interviews and discussions. Further, there are certain other information on which they tend to overemphasize. Observation helps the researcher to check the validity of what has been told by the interviewees, and also to look for what probably have been left unsaid, for example, feelings, sentiments or practices which could not be verbally expressed by the interviewees, or who is interacting with whom, or how much time is being spent in various activities, what are the practices that have been actually going on, and so on (Schmuk 1997).

Observation method demands prolonged engagement of the researcher in the chosen social setting or situation. It requires the researcher to observe and record data multiple times from the same social setting. Recording each observation and cross-checking the data recorded is important for validation of the data collected and for deriving at a logical and objective understanding of the field.

Observation of the interviewees in their own comfortable space or natural setting thus provides the researcher with a clearer and more empathetic picture of the field. Studying behaviors that occur naturally within natural contexts seldom leaves room for manipulation of data on the part of the researcher. It is also most likely that the researcher would encounter the field in its rawest setting. It is because of these

qualities that observation is considered as an effective tool of data collection while conducting qualitative research.

Types of Observation : Observation is primarily of two types- participant observation and non-participant observation.

- **Participant Observation :** Participant observation is a widely used qualitative method of data collection. Bronislaw Malinowski (1884-1942) coined the term participant observation in his ethnographic account of courtship, marriage and family life among the natives of Trobriand Islands. M.N. Srinivas writes that in participant observation, the fieldworker spends several months/years in a society different from one's own, uses local language to talk to the indigenes, participates in their daily, seasonal and other activities, and as far as possible gathers information during the course of such participation and direct observation. By establishing rapport with the people studied, the researcher collects information about every aspect of the culture that is being studied, thus getting a holistic picture of the same, and finally using that knowledge to make sense of how culture itself functions.Accrding to Demunck and Sobo (1998), participant observation is the primary method used by sociologists and anthropologists doing fieldwork. Fieldwork involves "active looking, improving memory, informal interviewing, writing detailed field notes, and most importantly, patience. Thus, participant observation is the process that enables researchers to learn about the activities of the people they are studying in the natural setting, through observing and participating in those activities (Dewalt and Dewalt 2001).

  Bernard (1994) defines participant observation as the process of establishing rapport within a community and learning to act in such a way that the community begins to feel at ease with the researcher and starts acting naturally; the researcher

should then remove herself from the setting or community to immerse in the data so that she is able to understand what is going on and hence is able to write about it. The researcher must include more than just observation in the process of being a participant observer. For example, as a participant observer the researcher is expected to take part in natural conversations, in ceremonies and other daily activities if the need arises. Taking part in such activities enables the researcher to gain firsthand knowledge about ceremonies of the community which one otherwise might not have had access to.

Participant observation is thus characterized by having an open, non-judgemental attitude in which the researcher is interested in learning more about others, is aware of the diverseness of culture that she might encounter, and most importantly, is open to making mistakes which obviously can be corrected through repeated observations. Being a good observer, a patient listener and learner, being open to the unexpected and being reflexive about oneself is what constitutes a good participant observer (Dewalt and Dewalt 1998).

- **Non-Participant Observation :** Non-participant observation is a research technique in which the researcher takes up the role of a mere observer. Unlike participant observation, here the researcher observes the subjects of her study with the latter's knowledge, but does not take active part in their activities. J. Patrick Williams (2008) cites a number of reasons as to why the researcher might choose non-participant observation over participant observation. First, a limited access to a particular group. For example, conducting participant observation in a riot-affected area is dangerous and often not feasible due to the spontaneity of events that might disturb the research in multiple ways. In such a situation, non-

participant observation is carried out to collect data on that issue, which may include scrutinizing available videos published by news agencies or insiders, or observing films and other audio/visual records available at that time on that issue.

Second, the researcher could be in a position where he/she takes up a particular topic for study but cannot possibly be a participant observer in the same. For example, the researcher wants to study how parents behave with their children in a public place, or what may be the experience of an athlete in sports. However, the researcher might be neither an athlete nor a parent. In that case, the researcher can be present in the same setting as that of her subjects of study and observe them without directly interacting with them. Third, the researcher might be more interested in deriving patterns of social behavior in a certain social setting instead of delving deep into the subjectivity that occurs in extensive personal interactions. An example of this could be the popular social media these days. Social media today provides enough anonymity to study social patterns or mass psychology that might be formed out of online public interactions, without the researcher necessarily having to take active part in those interactions. Thus, in a non-participant observation, the researcher does not actively take part in the activities of the group being studied. It means passive participation in an event where the observer is a mere spectator to what is happening in the particular situation. This also means that obtaining thick data may become extremely difficult as the researcher is not able to penetrate deep into the situation concerned without actually participating in it.

8.5.4 Focus Group Discussion

Meaning : In simple terms, a focus group is a small group of six to ten people who are led through an open discussion by a skilled moderator. The main goal of the moderator is to carry forward an open and spontaneous discussion among the participants so that all the participants can express maximum number of ideas and opinions.

A focus group discussion is a good way to gather people from similar experiences and backgrounds and discuss a specific topic of interest. These kind of interviews were originally called “Focussed interviews” or “ Group-depth interviews”. This technique of interviewing was developed after World War II. Since then, it has become an important method of data collection in social science.

According to Kruger, “Focus group discussion is a carefully planned discussion design to obtain perception on a defined area of interest in a permissive, non-threatening environment.”

Focus group discussions are defined as semi structured group discussions, which yield qualitative data on the community level by facilitating interaction between participants.

Characteristics of Focus Group:

- A focus group discussion usually consists of six to ten participants.
- The moderator engages the participants in small talks and gradually builds rapport amongst them.
- The moderator does not join the conversation between the participants.
- In a focus group interview, the participants should talk no more than about an hour.
- Recording the focus group discussion is a must.

- In a focus group discussion small issues should be discussed and bigger issues must be avoided.
- The moderator should always maintain impartiality particularly in facial expression and verbal comments.
- In a focus group discussion the participants are always explained the purpose and outcome of the discussion.
- Focus group discussion is a voluntary process where conversation flows among the participants.
- This type of discussion is relaxed, comfortable and often enjoyable for participants as they share their ideas and perception.

Steps for Conducting a Focus Group Discussion: Generally there are four distinct step used to conduct focus group discussion. These are-

- Question designing
- Selection and preparation of participants
- Organizing the interview
- Analyzing the data

These are discussed below-

- **Question Designing :** The first and foremost task of the researcher or moderator in a focus group discussion is to design the questions to be addressed during the discussion.
- **Selection and preparation of participants :** Ideally in a focus group discussion the participants do not know each other but should have a common background and must be comfortable with one another. So we can say that in order to get maximum information out of a focus group, the researcher should select a homogenous group of participants.
- **Organizing the Interview :** An ideal focus group is conducted by a moderator with the help of an assistant moderator. Generally the researcher plays the role of the moderator and carries forward the different discussion in the focus group. The assistant

moderator takes notes and records the entire focus group discussion.

- **Analyzing the data :** Analyzing the data gathered is the most important and critical job for the moderator. At first all information gathered from the participants must be boiled down to select the ones useful and essential for the research. The researcher then may use different methods to analyse these information, such as-labelling them under different columns, identifying categories and sub-categories of issues or summarizing different view points on the same issue. Once these findings are systematically organized they are ready for presentation.

Significance of Focus Group Discussion: Focus group discussions are often used as a tool of research so that the researcher can familiarise himself or herself with the situation at hand. The most important objective of a focus group is to find solution to different problems through flows of voluntary discussion. Group attitudes and beliefs can also be investigated through such a discussion. This is difficult to achieve through any other technique of data collection.

Generally a researcher employs the use of focus group discussion in the beginning of studying the problem. This helps him/her to generate ideas at the exploratory stage of the study. Moreover focus groups are very useful when the researcher aims to investigate complex ideas and behaviours. The open-ended questioning of a focus group helps the researcher to arrive at the depth of the problem. At the same time the different opinions of the participants can help the researcher to explore differences and commonalities about the situation.

Focus group discussion can help the researcher to develop the framework for conducting further research. His understanding of quantitative data will also depend on the

analysis of the results of these discussions. However the researcher must always be aware that the participants might not be entirely honest and truthful about their opinions. At the same time misinterpretation and biased findings can be very damaging for further research. Therefore the results of focus group discussion can be most fruitful only when the researcher is experienced, intuitive and does not take the opinions of the participants for granted.



CHECK YOUR PROGRESS

Q.3: What is an interview? (Answer in about 50 words)

.....

.....

Q.4: Mention two objectives of an interview.
(Answer in about 20 words)

.....

.....

Q. 5 : Fill in the blanks :

(a) Observation is a _____ method in which data is collected through _____.

(b) There are primarily _____ types of observation.

(c) The term participant observation was coined by _____.

(d) Participant observation involves _____ participation of the researcher in an activity.

(e) Participant observation provides _____ knowledge about the people that are being studied.

Q6 : What is a focus group? (Answer in about 50 words)

.....

.....

8.6 CONCEPT OF INTERVIEW SCHEDULE AND INTERVIEW GUIDE

Interview Schedule

Preparing a good interview schedule is no less than an art. It requires proper skill, prior experience and knowledge about the topic of research. K.D. Gangrade defines schedule as a form filled in during a personal interview in which both interviewer and the respondent are present and it contains a set of questions which are asked and filled in by an interviewer in a face to face situation. A schedule works as a tool in which all the respondents are usually supposed to respond to the same questions asked in the same language. This is done in order to derive at a uniformity regarding the answers received. The primary requirement of schedule is that the questions to be asked have to be written down before the interview starts and not during the interview. However, schedule provides flexibility to the researcher to reframe questions during the interview, if the respondents find difficulty in understanding the same.

Interview Guide

Interview guide is a list of general topics to be covered by the interviewer while conducting the interview. It is a tool used during semi-structured interviews. Interview guide helps the researcher to direct the conversation towards the issues that he/she wants to learn about. Interview guides may be highly scripted or relatively loose but serves the same purpose:

- What question to ask
- How to pose the question
- The sequence of questions
- What to do or say next

So we can say that an interview guide is a planning on the part of the researcher that guides him/her while interacting with the respondents. It offers the interviewer some important techniques to

conduct the interview. Generally an interview guide emphasises on four important points- i) There is a difference between questions asked during interviews and research questions. ii) We start our research with one theory and end it with a different one. iii) Interview is a social activity. iv) The answers of the respondent are not evidences. If a researcher understands these facts then he can make a fruitful use of an interview guide. He/she must always remember that an interview is simply a method of data collection and cannot always answer solutions



8.7 LET US SUM UP

- Data collection is a process. It consists not only of counting or enumerating or measurement of information but recording of the same also.
- An efficient and well -planned statistical enquiry must ensure so many preliminaries before collection of data.
- The objectives of the enquiry should be defined in clear and concrete terms.
- The statistical units on which data collection is done must be clearly defined. They should also be unambiguous, specific, stable, uniform and appropriate for the enquiry.
- Two types of data are used in statistical enquiries: primary and secondary
- Primary data seems better in terms of adjustment ability, originality, reliability, capability and suitability. Moreover, there is no need of precaution in using these data.
- Secondary data seems better in terms of less cost involved in terms of short collection period, money and trained men power required. Also, no biasness is involved here.

- In simple terms, interview is a technique where the interviewer asks questions to elicit information from the respondent. That is why it is a very personal and direct form of research.
- Interview is a very important aspect of research methodology and is the most common method of data collection.
- Interview is a technique where the interviewer asks questions to elicit information from the respondent. That is why it is a very personal and direct form of research.
- A questionnaire can be defined as a research instrument which has a series of pre-set questions that are framed in a certain way to gather objective information from respondents.
- In social science research, observation directly refers to a data collection method where the data is collected through observing.
- In simple terms, a focus group is a small group of six to ten people who are led through an open discussion by a skilled moderator.
- Interview guide is a list of general topics to be covered by the interviewer while conducting the interview. It is a tool used during semi-structured interviews.
- Preparing a good interview schedule is no less than an art. It requires proper skill, prior experience and knowledge about the topic of research.



8.8 FURTHER READING

1) Ahuja, R. (2018). *Research Methods.* Jaipur: Rawat Publications.
2) Kothari, C.R. (2011). *Research Methodology, Methods and Techniques.* New Age International (P) Ltd.
3) Kumar, R. (2011). *Research Methodology.* New Delhi: SAGE Publications India Pvt. Ltd.

8.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: (a) True (b) False

Ans to Q No 2: There are three concepts that are associated with data and its collection:

a) Investigator

b) Respondents

c) Statistical units

An Investigator is a person who conducts the statistical enquiry. He must be a trained person and well-acquainted with statistical analysis. He counts or measures the characteristic under study. These measurements are further analysed statistically according to the objectives of the study.

Respondents are the persons from whom the information are collected by the investigator.

The measurements or counts in any statistical investigation are done with some identifiable object or a group of objects. These are known as statistical units.

Ans to Q No 3 : In simple terms, interview is the verbal conversation between two people aiming at the collection of relevant information for the purpose of research. It is a technique where the interviewer asks questions to elicit information from the respondent.

Ans to Q No 4 :
- To collect information about an existing problem in society.
- To create a source of true, original and firsthand knowledge.

Answer to Q No 5 : a) Data collection, observing b) Two
c) Bronislaw Malinowski d) Active
e) Empirical

Ans to Q No 6 : A focus group is a small group of six to ten people who are led through an open discussion by a skilled moderator.



8.10 MODEL QUESTIONS

A) Short Questions (Answer each question in about 150 words)

Q1: What is meant by data collection?

Q2: Briefly explain the tools of data collection?

Q3: What is meant by observation?

Q4: Distinguish between interview schedule and interview guide.

B) Long Questions

(Answer each question in about 300-500 words)

Q1: Discuss the various sources of data collection.

Q2: Briefly discuss the focus group discussion.

Q3: Do you think non-participant observation can be difficult in terms of collecting data? Briefly explain the reasons.

Q4: Briefly compare the interview and observation as tools of data collection.

*** ****** ***

UNIT 9: PROCESSING OF DATA

UNIT STRUCTURE

9.1 Learning objective
9.2 Introduction
9.3 Editing and Coding
9.4 Classification and Tabulation of Data
9.5 Diagrammatic Reresentation of Data
9.6 Graphical Representation of Data
9.7 Let Us Sum up
9.8 Further Reading
9.9 Answers to Check Your Progress
9.10 Model Questions

9.1 LEARNING OBJECTIVES

After going through this unit, you will be able to -

- know how to deal with raw data collected from field
- understand how to categorize data into major groups
- understand how to organize data and present them in a meaningful manner
- understand the diagramatical and graphical representation of data.

9.2 INTRODUCTION

Social sciences in order to have an understanding of the functioning of the society need to look into the various components of the society. In addition to qualitative data, they have to rely a lot on quantitative information to draw inferences about different aspects. Thus, to understand findings it becomes necessary to have an understanding of the elementary quantitative methods. Social Work makes observations about people and their functioning. But the people themselves are not data. Data are the counts, measurements

and observations made on people. They have to be generated by observing the aspect that is intended to be studied. Once the data are generated, they can take a number of different forms. This is when classification and tabulation of data becomes necessary. No inferences can be made from raw data. They have to be arranged into categories. This part of statistics is called Descriptive Statistics. The objective of descriptive statistics is to provide precise description of a phenomenon. They consist of standardized procedures through which masses of data can be reduced to manageable proportions. The variables may be grouped into categories, which can then be arranged into frequency distribution or better still presented graphically to illustrate the relationship between variables. In the previous unit we have discussed the different tools and techniques of data collection. From that unit we have learned that there are two types of data, namely primary and secondary data and there are different methods of collecting these data. From the discussions in the previous units we are now aware that data is the most significant aspect of research and there are not only different techniques of collecting data but there are also different ways and procedures of classifying, analyzing and representing data. In this unit we will therefore discuss about the classification of data, tabulation of data and diagramatical representation of data graphical representation of data.

9.3 EDITING AND CODING

Editing is the process of making changes to a text of an article, research paper, for publication by correcting, condensing, or otherwise modifying it before printing or before making it open for readers. The similar actions related to editing are- to correct, improve, revise, redraft, condense, rewrite, modify and others. It involves the process of selecting and deciding what needs to be kept in a text and what needs to be modified or deleted.

Coding is the process of assigning a code to something for classification or identification of data. In a research, coding is an important activity to categorize the empirical, primary or secondary data collected for a study. Coding helps to identify and categorize the important information gathered for further analysis.

9.4 CLASSIFICATION AND TABULATION OF DATA

Classification and tabulation are actually not two distinct processes but they more or less go together. Classification may be said to be the first step in tabulation. So before studying the tabulation of data, let us study the concept of classification of data.

CLASSIFICATION

The process of arranging data into categories or groups on the basis of similarities, related facts or common characteristics is called classification. After classification the data becomes comprehensible and the important aspects of the data become apparent. The similarities and differences may also be understood, which makes comparison possible. Another advantage of classification is that it makes the data fit for statistical treatment. Unless and until classification is done, no statistical treatment can be done. From the nature of data, we can also deduce what statistical tools and technique to be used for analyzing the data.

Data can be classified mainly into two types. One is Qualitative Data and the other is Quantitative Data.

Qualitative data is based on attributes which are of descriptive nature. Some examples of qualitative data is gender (male/female), marital status (married/unmarried) etc. In other words, the attributes used for classification cannot be measured, but we can find out whether they are present or absent among the population under study.

Quantitative data is based on some characteristic or parameter that can be measured. For example, the height, weight, age, income,

etc. Such measurable characters are arranged into groups or classes called class intervals.

When classification is done according to qualitative aspect, then it is called qualitative classification and when it is done according to quantitative data or class intervals then it is called quantitative classification. Two elements are very essential in quantitative classification. One is the variable and the other is the frequency. The term variable refers to the characteristic that varies. The variable may be the height of the population which varies from person to person. Frequency is the number of times the variable occurs.

TABULATION

Tabulation is a simple device for summarizing data and presenting them in a meaningful manner. It helps in consolidating the primary data. Tabulation may be defined as the systematic and scientific method of arranging data. A table is generally a systematic method of arranging data in columns and rows. Columns are vertical arrangements whereas rows are horizontal ones. Classified data may be displayed in columns and rows so that their relationship can also be understood. The data after being arranged in a table becomes easy to comprehend and fit for analysis and interpretation.We will discuss more about tabulation in unit 13 of this course.



CHECK YOUR PROGRESS

Q1: Define coding. (Answer in about 20 words)

.....

.....

Q2: What is meant by classification of data?
(Answer in about 20 words)

.....

.....

Q3: What is meant by Tabulation of data?
(Answer in about 20 words)

.....

.....

9.5 DIAGRAMMATIC REPRESENTATION OF DATA

Diagrams is a visual form of representing statistical data.Through diagram basic facts and relationship among statistical variables are highlighted in an effective mannar. It reflects the information that might otherwise be lost amid the details of numerical tabulation.

Advantages of Diagrammatic Presentation

- A diagram is easy to understand. No special knowledge and skill are required.
- Data presented in the form of diagram are attractive and appealing, thereby paving way for easy memorization.
- Diagrams also carries the merit of economy of time as they are easy to understand and easy to draw conclusion from.
- With diagrams, comparisons between two sets of data are made easy.
- Diagrams are universally useful as they can be used at any time and any place.

Disadvantages of Diagrammatic Presentation

- Diagrams provide limited information. In order to get detailed and precise information one needs to go through the original statistical tables.
- Diagrams are useful only when comparisons are desired. A single diagram may not be useful.
- Diagrams are not suitable for further analysis.

9.6 GRAPHICAL REPRESENTATION OF DATA

Graphs provide an illustrative method of presenting data. They help in making a concise summary statement of data. The salient features of the data become more apparent through graphical presentation. In this unit we will study only the basics of graphs and in unit 13 of this course we will study in details about the graphical presentation of data.

Difference Between Diagrams and Graphs

- In the construction of a graph generally graph papers are used, but diagrams are drawn on plain paper.
- Graphs are easy to construct than diagrams.
- Graphs are used to determine certain measures of central tendencies, but these values cannot be determined through diagrams.
- Graphs are less attractive and impressive than diagrams.

Advantages of Graphic Presentation

- Complex data can be presented in simple and understandable way through graph.
- Graphs are used for forecasting, interpolation and extrapolation.
- Comparisons of two or more phenomenon are made easier by graph.

- Graphs are now widely used in various fields such as economics, commerce, advertising, government department, etc.

Disadvantages of Graphic Presentation

- Actual values are not known in graphs.
- Graphical presentation depends on scale. If scales are not accurate, graphs can be misleading.
- Graphs cannot be followed if multiple features of data are to be shown.
- Graphs are not easy to interpret for a layman.



CHECK YOUR PROGRESS

Q4: What is meant by diagramatical representation of data? (Answer in about 20 words)

.....

.....

Q5: Distinguish between diagram and graphs.
(Answer in about 20 words)

.....

.....



9.7 LET US SUM UP

- Generated raw data has to be organized into categories so that inferences can be drawn from the data.
- Editing is the process of making changes to a text of an article, research paper, for publication by correcting, condensing, or otherwise modifying it before printing or before making it open for readers.

- Coding is the process of assigning a code to something for classification or identification of data.
- Classification is the first step in organizing data and is the process of arranging data into categories on the basis of similarities.
- Classification may be qualitative or quantitative depending on the nature of data.
- Tabulation is the simple device of summarizing data into a systematic manner in columns and rows.
- A table consists of table number and title, caption, stub, body and footnotes.
- When data are too large, then separate tables should be made.
- Diagrams is a visual form of representing statistical data.Through diagram basic facts and relationship among statistical variables are highlighted in an effective mannar.
- Graphical presentation is an illustrative method of presenting a concise summary statement.
- The main types of graphical presentation of data are bar graphs, histogram, frequency polygon and frequency curve, ogive and pie diagrams.



9.8 FURTHER READING

1. Kothari C.R. (2004) *Research Methodology Methods and Techniques,* 2nd Edition. New Delhi : New Age International Publisher
2. Choudhury, L. (2000) *An introduction to Statistics.* Volume One. Guwahati : Kitapghar

9.9 ANSWER TO CHECK YOUR PROGRESS

Answer to Q1: Coding is the process of assigning a code to something for classification or identification of data.

Answer to Q2: The process of arranging data into categories or groups on the basis of similarities, related facts or common characteristics is called classification.

Answer to Q3: Tabulation is a simple device for summarizing data and presenting them in a meaningful manner. In simple words tabulation implies systematic arrangement of data into rows and columns.

Answer to Q4: Diagrams is a visual form of representing statistical data.Through diagram basic facts and relationship among statistical variables are highlighted in an effective mannar.

Answer to Q5: 1. In the construction of a graph generally graph papers are used, but diagrams are drawn on plain paper.

2. Graphs are easy to construct than diagrams.

9.10 MODEL QUESTIONS

A) SHORT QUESTIONS: (Answer each question in about 150 words)

Q1: Define editing and coding.

Q2: What is classification of data? Describe the types of classification used in statistics.

Q3: Describe the different parts of the table.

Q4: Distinguish between diagramatical and graphical representation of data.

Q5: Mention the advantages and disadvantages of graphical representation of data.

B) LONG QUESTIONS: (Answer each question in about 300-500 words)

Q1: Discuss some of the rules to be followed while tabulating data.

Q2: What are the different ways by which data can be presented graphically?

Q3: Explain the concept diagram in statistics with its advantages and disadvantages.

*** ***** ***

UNIT 10 : BASIC STATISTICS FOR SOCIAL RESEARCH

UNIT STRUCTURE

10.1 Learning Objectives
10.2 Introduction
10.3 Meaning and Definition of Statistics
10.4 Functions of Statistics
10.5 Importance of Statistics in Social Research
10.6 Frequency Distribution and Tabulation
10.7 Let Us Sum Up
10.8 Further Reading
10.9 Answers to Check Your Progress
10.10 Model Questions

10.1 LEARNING OBJECTIVES

After going through this unit, you will be able to -

- put forward the meaning and definition of statistics
- disucss the main functions of statistics
- describe the different methods of presenting statistical information
- explain the meaning of frequency distribution.

13.2 INTRODUCTION

Dear learners, you must have come across various facts in your daily life involving numbers. Such as Temperature, Price of the commodity, Quantity of items, Share market etc. These facts in statistical terms are called as data. The question arises, what to do with these facts or what type of information these facts carry? The answer to the question is – it depends on person to person that what type of information they need.

Around the world, specially the business world, most of the decisions are made with the information gathered about the companies, market place, the economic and the financial environment. Such information usually comes in the form of data or accompanied by data. Statistics provides the tool through which such data are collected, analyzed, summarized, and presented to facilitate the decision making process. This unit shall introduce to you the field of Statistics and its role in Social Research.

10.3 MEANING AND DEFINITION OF STATISTICS

The word 'Statistics' has been derived either from Latin word Status or Italian word Statistica or German word Statistick. Each carry same meaning 'An Organized Political State'. Some people consider Statistics as Data, Facts or Measurements, simply saying numerical facts while for some Statistics is analysis of figures for prediction and for some it is presentation of facts in graphs.

In this way Statistics exists in the following three forms:

a) As Numerical Facts: In this form it includes numerical facts such as National Income, Export, Import, Production, Employment Statistics etc.
b) As a subject: It exists in Singular as well as plural form.
c) It may also refer to the study of, and research into the theory and principles underlying statistical methods.

According to Tate the three meanings of the word statistics are contained in the statements "You compute Statistics by Statistics from Statistics."

According to Spiegal "Statistics is concerned with scientific method of collecting, organizing, summarizing, presenting and analyzing data as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis".

A. L. Bowley defined Statistics as: "Science of Counting, Science of averages and Science of the measurement of social

organism regarded as a whole".

According to P.H. Karmel "The subject Statistics is concerned with the collection, presentation, description and analysis of data which are measurable in numerical form".

10.4 FUNCTIONS OF STATISTICS

After going through the above definitions, you must have gathered some basic idea of statistics. Now lets discuss some of its major functions:

- Statistics reflects aggregate of facts. A single numerical record or fact is not statistics; in fact it is the aggregate of facts. For example, salary of person for a particular month, marks of a student in some examination etc.
- Statistics expressed facts numerically. In other words it is quantitative in nature. For example, if we say that the performance of BSW students in this year is Average, then it is not statistics. The qualitative characteristics like honesty, beauty, intelligence etc. expressed by giving certain scores as numerical standards can be called as statistics.
- In Statistics data are collected in a systematic manner according to a well prepared plan. The purpose for which the data are to be collected is decided and well defined in advance.
- Statistics reflect multiplicity of causes in its data. Statistical data are influenced by a number of factors. One cause alone can not affect data to a marked extent. For example, facts related to agriculture production are affected by a number of factors like rain, quality of soil, quality of seed, methods etc.
- Statistics are being capable of being placed in relation to each other. Statistical data are collected mostly for the purpose of comparison. The data must be homogenous and uniform in nature. For example, if we want to compare standard of living of some class in India then it must be compared with the class

of the same nature not with the other class or foreign country data.

- Statistics tests hypothesis and laws of other subjects.Nearly every law and theory, frequently used today, were based on deductive method. These laws can be tested for their validity with the help of statistics.
- Statistics estimates for the present and predicts for the future. Statistical methods are not only helpful in estimating the present but also in forecasting the future. Extrapolation and interpolation are the techniques uses for the purpose. 5 year plans and other economic plans like budget etc are the examples under this heading.
- Statistics is a tool of all science indispensable to research and intelligent judgment and has become a recognized discipline in its own right.

10.5 IMPORTANCE OF STATISTICS IN SOCIAL RESEARCH

Social Research is a systematic and scientific process which are generally carried out to acquire knowledge regarding social variable and phenomena.We have already studied about the Social Research in unit 2 of this course.So learner are requested to acquire knowledge from that section before studying this section.Social Research requires investigation and interpretation of social facts, social phenomena and social problems for the welbeing of social life and social conditions.In this regard Statistics play a vital role. In the following, we are briefly discussing the importance of Statistics in Social Research.

- Collection of data: Statistics helps in collecting the raw data from the society related to different social phenomena through the application of various tools and technique. We have already studied about the collection of data in unit 8 of this course.

- Arrangement and presentation of data: After collection of data, Statistics helps to arrange, process and present the data and information systematically to achieve the pre-determined social objective. We will study in details about the presentation of statistical data in unit 13 of this course.
- Analysis of data: Effective arrangement of collected data helps a statistician to move to the next important step which is known as analysis of data. Here, by utilising the necessary statistical tools and technique, data is analysed to throw light on the objective of the study and establish cause and effect relationship between different social variables and problems. We will study in details about the analysis of statistical data in unit 11 and 12 of this course.
- Interpretation of result: Collection, presentation and analysis of data is not the sole purpose of Statistics in Social Research. Statistics helps the social scientist to interpret the result of the study with the help of different statistical tools and technique (which we will study in details in unit 12 of this course), so that the result of the study is generalised and used for the betterment of the greater society.

10.6 FREQUENCY DISTRIBUTION AND TABULATION

A frequency distribution is a comprehensive way to classify raw data of a quantitative variable. It shows how different values of a variable are distributed in different classes along with their corresponding class frequencies. A frequency distribution table consists of class interval and frequencies. Suppose a class of 50 students obtains marks in economics between 0 and 50. This information may be presented in a frequency distribution table in the following way:

Class interval	0-10	10-20	20-30	30-40	40-50
Frequency	1	10	12	18	9

In the above example, against each class interval different frequencies are plotted. The class intervals or class limits are the two ends of a class the lower value is called the lower limit and the upper value is called the upper limit. For the class interval 10-20, 10 is the lower limit and 20 is the upper limit.

The difference between the upper limit and lower limit is called the length or height of the class. In the above example, the height is (30 – 20)= 10.

The middle value of a class is called mid value or mid-point of the class. It is also known as class mark and is calculated in the following way:

Mid Value = (upper limit + lower limit)/2

Preparation of Frequency Distribution

- **Number of Classes :** It is important to decide as to how many classes we need. A frequency distribution with too many classes or too few classes may not reveal the true set of the data, and it will also be difficult to interpret such frequency distribution.

Before we determine the class interval, we need to find the range of the data, which is the difference between the maximum number and minimum number, i.e., **range = (max-min).**

- **Size of the class**: After we determine the range and number of classes, we need to decide upon the height of the class interval using the formula ***h=range/no. of classes***

Given the range of the variable, number of classes can be easily determined if we know the class intervals. Similarly, class interval can be known if the number of classes is known. So, these decisions are interlinked with one another.

- **Determining the class limit :** The value of the upper class limit of a class is obtained by adding the class interval with the value of the lower limit of that class. For example, in the class limit 30 – 40, 30 is the lower class limit and by adding 10, which is the class interval, to

30, we get 40, i.e., 30+10=40. In order to determine the lower class limit of the first class, one has to choose the minimum value of the frequency.

However, before determining the class limit, one must see that the classes are mutually exclusive. This means that no data value can fall into two different classes. Also, the classes must be all inclusive or exhaustive. This means that all data values must be included.

Let us consider the following example.

57 44 86 75 00 18 45 14 04 64 72 51 69 34 22 83 70
20 57 28 100 56 50 47 10 34 61 64 80 46 22 10 84 50
47 73 42 33 48 65 10 34 66 53 75 90 58 46 38 69

In the above example, 0 is the minimum value so the lower class limit of the first class will be 0. Now adding 10, the upper class limit of the first class will be 10. The same process will be followed for the other classes as well.

Table 10.1

Class interval	Frequency
0 - 10	2
10 – 20	5
20 – 30	4
30 – 40	5
40 - 50	8
50 – 60	8
60 – 70	7
70 – 80	5
80 – 90	4
90 - 100	2

In the above table, upper class limit of the first class = lower class limit of the second class = 10. This is true for other classes as well. Here classification is done by exclusive method, where upper

limit of one class coincides with the lower limit of the next corresponding class. So, 40-50 and 50-60 are two consecutive classes. Now the question is in which of the two classes will 50 fall. To answer this we need to know the exclusive method and the inclusive method.

- **EXCLUSIVE METHOD :** When the classes are so fixed that upper limit of one class is the lower limit of the next class, it is called exclusive method. So, the continuity of the data is maintained. This method is more useful in case of continuous variable. Under this method, the value which is equal to the lower limit of a class is included in that very class. On the other hand, the value which is equal to the upper class limit of a class is not included in that very class. So, in case of class 30-40, 30 is included in the class 30-40, but 40 is included in the class 40-50.

- **INCLUSIVE METHOD :** In this method, both the lower limit and upper limit of a class are included in that very class. So, in this case, the upper limit of one class is not equal to the lower limit of its next class. Consider the following example.

Table 10.2

Class interval (daily wage)	Frequency (no. of persons)
100 – 199	50
200 – 299	30
300 – 399	40
400 – 499	80
500 – 599	120
600 - 699	100
	Total = 420

Here no. of persons earning daily wage between 100 and 199 will be included in the class 100-199. If someone earns a daily wage of 200, that person will be included in the class 200-299.

➢ Conversion of Inclusive Series in Exclusive Series

For converting the inclusive series into exclusive series, following steps may be followed:

- Find the difference between upper limit of a class and lower limit of a class. In the above example, the difference is 1.
- Divide the difference by 2. S, ½= 0.5
- Add 0.5 to upper limit and subtract 0.5 from lower limit of each class.

This kind of adjustment restores continuity of data.

The above example, when converted into exclusive series, will be as follows:

Table 10.3

Class interval (daily wage)	Frequency (no. of persons)
99.5 – 199.5	50
199.5 – 299.5	30
299.5 – 399.5	40
399.5– 499.5	80
499.5– 599.5	120
599.5 - 699.5	100
	Total = 420

Construction of Frequency Distribution Table

Consider the following example of marks obtained by 50 students:

57 44 86 75 00 18 45 14 04 64 72 51 69 34 22 83 70
20 57 28 100 56 50 47 10 34 61 64 80 46 22 10 84 50
47 73 42 33 48 65 10 34 66 53 75 90 58 46 38 69

The above information needs to be classified into a frequency distribution table, as shown below.

Table 10.4

Classes	Observations	Tally marks	Frequency	Mid Value
0 - 10	0,4	II	2	5
10 – 20	18,14,10,10,10	IIII	5	15
20 – 30	22,20,28,22	IIII	4	25
30 – 40	34,34,33,34,38	IIII	5	35
40 - 50	44,45,47,46,47,42,48,46	IIII III	8	45
50 – 60	57,51,57,56,50,50,53,58	IIII III	8	55
60 – 70	64,69,61,64,65,66,69	IIII II	7	65
70 – 80	75,72,70,73,75	IIII	5	75
80 – 90	86,83,80,84,	IIII	4	85
90 - 100	100,90	II	2	95

The above table has five columns – classes, observations, tally marks, frequency and mid value. It is evident from the above table that most of the observations are concentrated in classes 40-50 and 50-60. Their frequencies are 8. Out of 50 observations, 16 observations are in these two classes. That is, 32% of the data lies in between 40 – 60. The other classes are sparsely populated.

Cumulative Frequency Distribution

From table 10.4, one can know as to how many students got marks between 40-50, or 50-60 and so on. But if one wants to know as to how many students got above 90 or less than 40, then that table cannot furnish the information. For that one has to calculate cumulative frequency. Cumulative frequency distribution is a modification of simple frequency distribution and is obtained by successfully adding the frequencies of the values of the classes according to a certain rule. It is written as c.f.

Cumulative frequency distributions are of two types – 'less than' and 'more than'. In a 'less than' frequency distribution, the frequencies of each class are added successively from top to bottom. It indicates the cumulative number of observations less than or equal to the class frequency to which it relates. In a 'more than' cumulative

frequency distribution, the frequencies of each class are added successively from bottom to top. It indicates the cumulative number of observations more than or equal to the class frequency to which it relates. This is shown in the following table:

Table 10.5

Classes	Frequency	Cumulative frequency	
		Less than c.f.	More than c.f.
0 – 10	2	2	2+48=50
10 – 20	5	2+5=7	5+43=48
20 – 30	4	7+4=11	4+39=43
30 – 40	5	11+5=16	5+34=39
40 – 50	8	16+8=24	8+26=34
50 – 60	8	24+8=32	8+18=26
60 – 70	7	32+7=39	7+11=18
70 – 80	5	39+5=44	5+6=11
80 – 90	4	44+4=48	2+4=6
90 – 100	2	48+2=50	2
The above table is		presented is presented	below in a more

simplified way –

Less than		More than	
Marks	No. of students	Marks	No. of students
Less than 10	2	0 and more than 0	2+48=50
Less than 20	2+5=7	More than 10	5+43=48
Less than 30	7+4=11	More than 20	4+39=43
Less than 40	11+5=16	More than 30	5+34=39
Less than 50	16+8=24	More than 40	8+26=34
Less than 60	24+8=32	More than 50	8+18=26
Less than 70	32+7=39	More than 60	7+11=18
Less than 80	39+5=44	More than 70	5+6=11
Less than 90	44+4=48	More than 80	2+4=6
Less than 100	48+2=50	More than 90	2

➢ **Purpose/Use/Necessity of Cumulative Frequency Distribution**

- ❍ Cumulative frequencies are used to know as to how many values of the variables are larger than a particular value or less than a particular value.
- ❍ Cumulative frequencies are used in determining mean, median and mode.
- ❍ Cumulative frequencies are also used while drawing Lorenz curve.
- ❍ We also require cumulative frequencies to represent some given data by ogive.

  We have already studied about tabulation of data in unit 9 of this course.So learners are requested to study that section for the concept of tabulation of data.



10.7 LET US SUM UP

- The word 'Statistics' has been derived either from Latin word Status of Italian word Statistica or German word Statistick. Each carry same meaning 'An Organized Political State'.
- Statistics is a tool of all science indispensable to research and intelligent judgment and has become a recognized discipline in its own right.
- A frequency distribution is a comprehensive way to classify raw data of a quantitative variable.
- A frequency distribution table consists of class interval and frequencies.
- Cumulative frequency distribution is a modification of simple frequency distribution and is obtained by successfully adding the frequencies of the values of the classes according to a certain rule.



CHECK YOUR PROGRESS

Q1 : Define Statistics. (Answer in about 30 words)

.....

.....

Q2 : What is meant by classification of data?
(Answer in about 20 words)

.....

.....

Q3 : What is meant by frequency distribution?
(Answer in about 20 words)

.....

.....

Q4 : State the meaning of cumulative frequency distribution?
(Answer in about 30 words)

.....

.....



10.8 FURTHER READING

1) Gupta, S.C. (2009). *'Fundamentals of Statistics,* Himalayan Publishing House. Mumbai : Sixth Edition Reprint.
2) Gupta, S.P. (2010). *'Statistical Methods',* New Delhi : Sultan Chand & Sons.



10.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1 : Statistics is concerned with scientific method of collecting, organizing, summarizing, presenting and analyzing data as well as drawing valid conclusions and making reasonable decisions on the basis of such analysis.

Ans to Q No 2 : The technique of arranging data in different heomogenous groups is known as classification.

Ans to Q No 3 : A frequency distribution is a comprehensive way to classify raw data of a quantitative variable.

Ans to Q No 4 : Cumulative frequency distribution is a modification of simple frequency distribution and is obtained by successfully adding the frequencies of the values of the classes accending to a certain rule.



10.10 MODEL QUESTIONS

A) Short Questions (Answer each question in about 150 words)

Q1 : Define classification.

Q2 : What is meant by frequency distribution?

Q3 : Distinguish between frequency distribution and cumulative frequency distribution.

B) Long Questions (Answer each question in about 300-500 words)

Q1 : State the functions of Statistics.

Q2 : Discuss the steps involved in the preparation of frequency distribution table.

Q3 : Discuss the importance of Statistics in Social Research.

*** ***** ***

UNIT 11: APPLICATIONS OF STATISTICS

UNIT STRUCTURE

11.1 Learning Objectives

11.2 Introduction

11.3 Measures of Central Tendency / Averages

11.4 Measures of Variability / Dispersion

11.5 Let Us Sum Up

11.6 Further Reading

11.7 Answers to Check Your Progress

11.8 Model Questions

3.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- explain the meaning of measures of central tendency or averages
- describe the various types of averages along with their advantages, limitations and uses
- explain the meaning of dispersion or variation
- describe the various methods of obtaining dispersion for various types of distribution.

3.2 INTRODUCTION

In the earlier two units, we have discussed the classification, tabulation, diagrammatic and graphic representations of data that has considerably advanced our knowledge of the significance of statistical data. We need a numerical expression which summaries the characteristics of the whole set of data under study. ‘Measures of Central Location’ or ‘Measures of Central Tendency’ popularly known as ‘Averages’ serve this purpose. Apart from it in this unit, we shall discuss about the degree to which numerical data tend to spread about an average value. Dispersion measures the extent to which the items vary from central value.

3.3 MEASURES OF CENTRAL TENDENCY / AVERAGES

Meaning

An average of a distribution (i.e. a distribution of the values of a variable like height, weight, income etc.) is a representative value of that distribution. This representative value usually lies at the central part of distribution. ‘‘A measure of central tendency or an average of a certain distribution is nothing but a representative value of that distribution which enables us to comprehend in a single effort the significance of the whole.’’



Extreme Value : An extreme value is either very small or very large values in a probability distribution

Characteristics of a Good Average

An average will be termed as a good if it possesses the following characteristics:

i) It should be easy to understand and calculate.

ii) It should be rigidly defined which means that the definition should be so clear that its interpretation does not differ from person to person.

iii) The average should be based on all the values of the variable.

iv) It should be capable of further algebric treatment.

v) The average should not be unduly affected by extreme values.

Different Types of Measures of Central Tendency

The following three types of measures of central tendency or averages are in use.

(a) Mean, (b) Median, and (c) Mode.

a) MEAN: There are three types of mean, namely,

(i) Arithmetic Mean (A.M.), (ii) Geometric Mean (G.M.) and

(iii) Harmonic Mean (H.M.)

So far as mean is concerned, we shall discuss arithmetic mean only since this is the most popular technique among the different types of mean.

The A.M. of a variable x is generally denoted by $\bar{x}$ and is

defined as the sum of the values of divided by the total number of values of x.

Thus, if x_1, x_2, x_3, ..., x_n be n values of x then A.M ($\bar{x}$) will be given by,

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum x}{n}$$

In case of ungrouped frequency distribution table:

x :	x_1	x_2	x_3	...	x_n
f :	f_1	f_2	f_3	...	f_n

the A.M. ($\bar{x}$) will be given by,

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + f_3x_3 + ... + f_nx_n}{f_1 + f_2 + f_3 + ... + f_n}$$

$$= \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$

$$= \frac{\sum fx}{\sum f} = \frac{\sum fx}{N}$$

where, $N = f_1 + f_2 + f_3 + ... + f_n$

Arithmetic Mean (A.M.) of grouped frequency distributions: In case of grouped frequency distribution A.M. ($\bar{x}$) is defined as

$$\bar{x} = \frac{\sum fx}{N}$$

where, $N = f_1 + f_2 + f_3 + ... + f_n = \sum f$ and x is the mid value of a class.

Example 1: Find A.M. of the following frequency distribution:

x	1	2	3	4	5	6	7	8	9
y	7	11	16	17	26	31	11	1	1

Solution: First of all we shall prepare the following frequency table:

x	f	fx
1	7	7
2	11	22
3	16	48
4	17	68
5	26	130
6	31	186
7	11	77
8	1	8
9	1	9
	N = 121	Sfx = 555

$$\text{A.M. } (\bar{x}) = \frac{\sum fx}{N} = \frac{555}{121} = 4.59 \text{ (Approx.)}$$

Example 2: Determine mean of the following distribution:

Daily wages (in Rs.):	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of workers:	6	5	8	15	7	6	3

Solution: Since there are three methods of obtaining mean, namely arithmetic mean (A.M.) method, geometric mean (G.M.) method and harmonic mean (H.M.) method, we shall apply A.M. method to find mean of the given distribution. (We shall see subsequently that A.M. is the most popularly used measure of central tendency.)

To find mean by applying arithmetic mean technique we form the following table:

Wages	Mid value ($\bar{x}$) = $\frac{l_1 + l_2}{2}$	No. of workers (f)	fx
0-10	5	6	30
10-20	15	5	75
20-30	25	8	200
30-40	35	15	525
40-50	45	7	315
50-60	55	6	330
60-70	65	3	195
		N = 50	Sfx = 1670

$$\text{Mean } (\bar{x}) = \frac{\sum fx}{N} = \frac{1670}{N} = 33.4$$

i.e.; The required mean (i.e. average) wage = **Rs. 33.40**

Note : l_1 and l_2 imply respectively the lower limit and the upper limit of a class interval.

ALTERNATIVE METHODS OF FINDING A.M.:

a) Assumed Mean Method or Short-Cut Method: In this method A.M. $\bar{x}$ is calculated by using the following formula:

$$\bar{x} = A + \frac{\sum fd}{N}$$

where A = Assumed mean of x, d = x – A, N = total frequency. This formula is used when the values of 'd' do not have a common factor.

b) Step-deviation Method: If the values of (d = x – A) in the above formula have one or more common factors then this method is applied. The formula is:

$$\bar{x} = A + \frac{\sum fd'}{N} \times h$$

where $d' = \frac{d}{h}$, h = Highest common factor of 'd'.

In case of grouped frequency distributions with equal class intervals, h is equal to the length of the class intervals.

A note on assumed mean: Any value of a variable x can be considered to be the assumed mean of x. However, the assumed mean should be taken from the central part of the values

Example 3: Determine arithmetic mean of the following distribution:

Height (in cm)	130-134	135-139	140-144	145-149
Frequency	5	15	28	24
Height (in cm)	150-154	15-159	160-164	
Frequency	17	10	1	

Solution: We denote height by the variable x and we take the assumed mean A of x to be 147.

Class Interval	Mid value (x)	Frequency (f)	d = x – A, A = 147	$d' = \frac{d}{h}$, h = 5	fd'
130-134	132	5	–15	–3	–15
135-139	137	15	–10	–2	–30
140-144	142	28	–5	–1	–28
145-149	147	24	0	0	0
150-154	152	17	5	1	17
155-159	157	10	10	2	20
160-164	162	1	15	3	3
		N = 100			$\sum fd' = -33$

$$\text{Now, A.M. } (\bar{x}) = A + \frac{\sum fd'}{N} \times h$$

$$= 147 + \frac{-33}{100} \times 5$$

$$= 147 + \frac{-33}{20}$$

$$= 147 - 1.65$$

$$= 145.35$$

i.e. the required A.M. = **145.35 cm**

b) MEDIAN : The median of a series or distribution in ascending or descending order is that observation of the distribution which divides the distribution into two equal parts. Thus there are equal number of observations on the right and on the left of the median value.

In order to determine the median of an individual series, first of all we have to observe whether the values (observations) are in a define order or not i.e. whether the values are in ascending or in descending order or not. If the values are not in a define order, then these values are to be arranged either in ascending or in descending order. If there are odd number of values in the series, then the $\left(\frac{n+1}{2}\right)$th value from the beginning (and also from the end) will be the median. If the number of values is even then the arithmetic mean of the $\frac{n}{2}$th value and the $\left(\frac{n}{2}+1\right)$th value will be the median.

Example 4: Determine median for the following series:

i) 77, 73, 72, 70, 75, 79, 78

ii) 94, 33, 86, 68, 32, 80, 48, 70

Solution:

i) Arranging the values of the series in ascending order, we get: 70, 72, 73, 75, 78, 79

No. of terms in the series = 7 = An odd number

The required median = $\frac{7+1}{2}$th term i.e; 4th term = 75

ii) Arranging the data (values or observations) in ascending order, we get: 32, 33, 48, 68, 70, 80, 86, 94

No. of term in the series = 8 = An even number

Now, $\frac{n}{2}$th term = $\frac{8}{2}$th term = 4th term and $\left(\frac{n}{2}+1\right)$th term = 5th term

\ The required median = $\frac{68+70}{2}$ = 69.

MEDIAN OF A GROUPED FREQUENCY DISTRIBUTION: In case of grouped frequency distributions, one may consider the $\frac{n}{2}$th observation as the median if N is even. When N is odd, $\left(\frac{n}{2}+1\right)$th observation will be the median.

Median (Me) = $L+\frac{\frac{N}{2}-f_c}{f}\times 1$

Where L = Lower class limit (lower class boundary) in case of exclusive (inclusive) classification.

f = Frequency i.e. simple frequency of the median class

f_c = Cumulative frequency of the class preceeding the median class.

N = Total frequency

I = length of the median class

Note: This formula for obtaining median of a grouped frequency distribution holds only when the distribution is in ascending order. If the distribution is in descending order then it is to be first of all arranged in ascending order in order to apply the above formula.

Example 5: Determine median for the following distribution:

Weekly wages (Rs.)	50-55	55-60	60-65	65-70	70-75
No. of workers	6	10	22	30	16
Weekly wages (Rs.)	75-80	80-85			
No. of workers	12	15			

Solution:

Weekly wages (Rs.)	No. of workers (f)	Cumulative frequency (fc)
50-55	6	6
55-60	10	16
60-65	22	38
65-70	30	68
70-75	16	84
75-80	12	96
80-85	15	111
N = 111		

Here N = 111, which is odd.

Now, $\frac{N=1}{2}$ th term = $\frac{111+1}{2}$ th term = 56th term. From the cumulative frequency table we find that the 56th term lies in the class 65 - 70. 65 - 70 is the median class.

Now, median = $L + \frac{\frac{N}{2} - fc}{f} \times 1$

Here L = 65, f = 30, fc = 38, N = 111, I = 5

Median = $65 + \frac{\frac{111}{2} - 38}{30} \times 5$

$= 65 + \frac{55.5 - 38}{30} \times 5$

$= 65 + \frac{17.5}{6}$

= 65 + 2.92

= 67.92

i.e. the required median = Rs. 67.92

c) MODE: The mode of a distribution is that observation of the distribution whose frequency is the maximum. It is to be noted that mode is not unique which means that a distribution may have more than one modes.

Clearly an individual or single series (distribution) does not have a mode. In many cases of ungrouped frequency distributions, mode can be detected simply by observation. Mode of a grouped frequency distribution is obtained by using the following formula.

$$\text{Mode } (M_o) = L + \frac{f_1 = f_0}{2f_1 - f_0 - f_2} \times I$$

Where L = Lower limit/lower boundary of the model class

f_o = Frequency of the class preceeding the model class

f_1 = Frequency of the model class

f_2 = Frequency of the class succeeding the model class

And I = Length of the model class

- **Note:** i) The class (specified by a class interval) whose frequency is the maximum is called the model class.
  ii) The above formula is used when all the classes are of equal length.

Example 6 : Determine mode for the following distribution:

Marks	1-5	6-10	11-15	16-20	21-25	26-30
No. of students	7	10	16	32	24	18
Marks	31-35	36-40	41-45			
No. of students	10	5	1			

Solution: Since the frequency of the class 16-20 is the maximum, hence this class is the model class. The class intervals of the given distribution are as per the inclusive method of classification and hence in determining mode we must take the lower boundary of the model class.

Now, Mode $= L + \frac{f_1 = f_0}{2f_1 - f_0 - f_2} \times I$

$= 15.5 + \frac{32 - 16}{2 \times 32 - 16 - 24} \times 5$

$= 15.5 + \frac{16}{64 - 40} \times 5$

$= 15.5 + \frac{16}{24} \times 5$

= 15.5 + 3.33 = **18.33 marks**

Example 7: Determine mode/modes, if any, of the following series:

i) 3, 4, 5, 2, 3, 4, 1, 6, 4;

ii) 7, 9, 11, 7, 6,5, 9, 13;

iii) 3, 5, 6, 7, 9, 12, 3, 6, 5, 9, 12, 7

Solution:

i) The number 4 is repeated the maximum of number times (thrice). Hence 4 is the mode of the given series.

ii) We find that the frequency of 7 and 9 is the maximum each being equal to 2. Hence the two modes of the series are 7 and 9.

iii) Since the frequency of each observation is the same (being 2 in each case), hence the given series has no mode.

RELATIONSHIP AMONG MEAN, MEDIAN AND MODE: Prof. Kari Pearson has established an empirical (experimental) relationship among mean, median and mode. For any distribution the following relationship known as empirical relationship approximately holds:

Mean – Mode = 3 (Mean – Median)



CHECK YOUR PROGRESS

Q.1: What is meant by measures of central tendency or averages?

.....

.....

Q.2: Find mean and median from the following distribution:

Age (year):	20	19	18	17	16	15	14	13	12	11
No. of students:	1	2	4	8	11	10	7	4	2	1

Q.3: Determine mean, median and mode from the following data:

Weekly wages (Rs.):	15	16	17	18	19	20
No. of workers:	6	12	23	30	9	1

Q.4: Determine mean, median and mode from the following statistical distribution.

Class interval:	6.5-7.5	7.5-8.5	8.5-9.5	9.5-10.5
Frequency:	5	12	25	48

Class interval:	10.5-11.5	11.5-12.5	12.5-13.5
Frequency:	32	6	1

11.4 MEASURES OF VARIABILITY / DISPERSION

Meaning

Scatteredness of data around an average is termed as dispersion or variation. To quite Spiegel, "The degree to which numerical data tend to spread about an average value is called variation or dispersion".

Objectives of Studying Dispersion

The objectives of studying Dispersion are as follows:

i) To study the reliability of average.

ii) To control the variation.

iii) To make the comparison among series of observations with regards to variations.

iv) To make further statistical analysis.

Essentials of a Good Measure of Dispersion

i) A measure of dispersion should be based on all the observations of a series.

ii) It should be rigidly defined.

iii) It should be easy to understand and calculate.

iv) It should not extremely affected by extreme values.

v) It should not be affected much by fluctuations of sampling.

vi) It should be usable for further statistical analysis.

Different Measures of Dispersion

The following are the various measures of dispersion: (a) Range (b) Interquartile Range and Quartile Deviation (c) Mean Deviation (d) Standard Deviation and Variance (e) Coefficient of variation. While measures (a), (b), (c) and (d) are called absoluate measures, measure (e) is called a relative measure of dispersion. An absolute measure is that which possesses the same unit of the distribution for which the measure is obtained. On the other hand a relative measure is a ratio or percentage and as such it is a pure number. It is to be mentioned that all measures of central tendency are absolute measures. So far as disperson is concerned, we shall discuss only Standard Deviation and Quartile Deviation, since these are the most popular technique among the different types of mean.

a) Quartile Deviation (Q.D.): The interquartile range of a distribution is the difference between the third quartile Q_3 and the first quartile Q_1 of the distribution. Again, half the interquartile range of a distribution is called the quartile deviation (Q.D.) of the distribution. Thus,

Interquartile range = $Q_3 - Q_1$, and Q.D. = $\frac{Q_3 - Q_1}{2}$

Formulas to Calculate the Quartile Deviation:

For Individual Servies:

Inrerquartile Range = $Q_3 - Q_1$

Q.D. = $\frac{Q_3 - Q_1}{2}$

Where Q_1 = size of $\frac{n+1^{th}}{4}$ item

n = number of observations

Q_3 = size of $\left(\frac{3(n+1)}{4}\right)^{th}$ item

n = number of observations

For Continuous Series:

Interquartile Rang = $Q_3 - Q_1$

Q.D. = $\frac{Q_3 - Q_1}{2}$

Here we first locate the size of $\frac{N^{th}}{4}$ item in c.f column and the value of Q_1 will lie in the corresponding class interval. (i.e. Q_1 class)

Where $Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$

l_1 = lower limit of class interval.

N = Sum of frequencies.

c.f. = Cumulative frequencies of the class preceding the Q_1 class.

i = Class interval.

First, we locate the size of $\frac{3N^{th}}{4}$ item in c.f column and the value of Q_3 will lie in the corresponding class interval.

$Q_3 = l_1 + \frac{3\left(\frac{N}{4}\right) - c.f.}{f} \times i$

Where l_1 = lower limit of class interval.

N = Sum of frequencies.

c.f. = Cumulative frequencies of the class preceding the

Q_3 class.

i = Class interval.

Example 8: Calculate Q.D. of the following observation:

20 25 29 30 35 39 41 48 51 60 and 70

Solution:

For Q. D., we need to calculate values of Q_3 and Q_1.

Q_1 is the size of $\frac{N+1^{th}}{4}$ value

Here N being 11. Q_1 is the size of 3rd value.

As the values are already arranged, in ascending order, it can be seen that Q_1, the 3rd value is 29.

Similarly. Q_3 is size of $\frac{3(N+1)^{th}}{4}$ value : i.e. 9th value which is 51

Here Q_3= 51

Q.D. = $\frac{Q_3 - Q_1}{2} = \frac{51-29}{2} = 11$

Example 9: For the following distribution of marks scored by a class of 40 students. Calculate the Q.D.

Class interval	No of students (f)
0–10	5
10–20	8
20–40	16
40–60	7
60–90	4
	40

Solution: For Q.D., first calculate cumulative frequencies as follows:

(Class Interval)	(Frequency)	(Cumulative Frequency)
(c.i.)	(f.)	(c.f.)
0–10	5	05
10–20	8	13
20–40	16	29
40–60	7	36
60–90	4	40
N = 40		

Q_1 is the size of $\frac{N}{4}^{th}$ value in a continuous series. Thus, it is the size of the 10th value. The class containing the 10th value is 10-20. Here Q_1 lies in class 10-20. Now, to calculate the exact value of Q_1, the following formula is used:

$$Q_1 = L + \frac{\frac{N}{4} - c.f.}{f} \times i$$

Where L = 10 (Lower limit of Quartile class)

c.f. = 5 (value of c.f. for the class preceding the Quartile class)

i = 10 (interval of the Quartile class) and

f = 8 (frequency of the Quartile class)

Thus, $Q_1 = 10 + \frac{10-5}{8} \times 10 = 16.25$

Similarly, Q_3 is the size of $\frac{3N}{4}^{th}$ value i.e., 30th value which lies in class 40 – 60. Now using the formula for Q_3 its value can be calculated as follows:

$$Q_1 = L + \frac{\frac{3N}{4} - c.f.}{f} \times i$$

$$= 40 + \frac{30-29}{7} \times 20$$

= 42.87

$$\therefore \text{Q.D.} = \frac{42.87 - 16.25}{2}$$

= 13.31

.

b) Standard Deviation (S.D.): The positive square root of the arithmetic mean of the squares of the deviations of the values of a variable from its arithmetic mean is called the standard deviation of that variable. The symbol σ (a Greek letter pronounced as 'sigma') is used to denote standard deviation.

If a variable x takes n values x_1, x_2, ..., ..., ..., x_n and if $\bar{x}$ be the arithmetic mean of these values, then

$$\sigma = \sqrt{\frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \ldots + f_n(x_n - \bar{x})^2}{n}}$$

$$\text{i.e. } \sigma = \sqrt{\frac{\sum(x_1 - \bar{x})^2}{n}}, \quad \bar{x} = \frac{\sum x}{n}$$

Again, in case of frequency distribution:

x	x_1	x_2	.....	x_n
f	f_1	f_2	.....	f_n

Standard deviation will be:

$$\sigma = \sqrt{\frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \ldots + f_n(x_n - \bar{x})^2}{N}}, \quad N = \Sigma f$$

$$\text{i.e. } \sigma = \sqrt{\frac{\sum_{i=1}^{n} f_i(x_i - \bar{x})^2}{N}}, \quad \bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{N}$$

Instead of the symbol σ we may use the symbol σ_x to clearly signify the standard deviation of the variable x.

Alternative methods of obtaining standard deviation:

i) Assumed Mean Method: In case of individual series we have the following formula:

$$\sigma = \sqrt{\frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2}, \; d = x - A$$

And A = Assumed mean of x.

In case of frequency distribution,

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

ii) **Step deviation method:** If the deviations from assumed mean A denoted by d have some common factors then these deviations are divided by the highest common factor h (say) and the step deviations denoted by d' are obtained. Thus $d' = d/h$, where h is the H.C.F. of the deviations d. In this method, we have the following formula:

For individual series, $\sigma = \sqrt{\frac{\sum f'^2}{n} - \left(\frac{\sum f'}{n}\right)^2} \times h$

For frequency distribution $\sigma = \sqrt{\frac{\sum fd'^2}{n} - \left(\frac{\sum fd'}{n}\right)^2} \times h$

$d' = \frac{d}{h} = \frac{x - A}{h}$, A = Assumed mean of the variable x whose values are the various values of the distribution.

Example 10: Find standard deviation of the following observations: 8,10, 12, 14, 16, 18, 20, 22, 24, 26

Solution: We shall solve this problem by Direct Method as well as by Step-Deviation Method.

Direct Method:

Value x	Deviation from the mean $x = \bar{x}$, ($\bar{x} = 17$)	$(x-\bar{x})^2$
8	–9	81
10	–7	49
12	–5	25
14	–3	9
16	–1	1
18	1	1
20	3	9
22	5	25
24	7	49
26	9	81
$\Sigma x = 170$		$\Sigma(x-\bar{x})^2 = 330$

$$\bar{x} = \frac{\sum x}{n} = \frac{170}{10} = 17$$

$$\sigma = \sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{n}}$$

$$= \sqrt{\frac{330}{10}}$$

$$= \sqrt{33}$$

$$= 5.74$$

Note: When the mean $\bar{x}$ is an integer the formula $\sigma = \sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{n}}$ should be used to find S.D. When $\bar{x}$ is not an integer then the $\sigma = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$ or $\sigma = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$ formula should be used. Because when $\bar{x}$ is not an integer determination of the values of $(x - \bar{x})^2$ becomes vary time consuming.

Step-Deviation Method: Let the given values be the values of the variable x and let the assumed mean of x be 18.

Calculation of S.D.:

Value	d = x – 18	$d' = \frac{d}{2}$	d'^2
8	–10	–5	25
10	–8	–4	16
12	–6	–3	9
14	–4	–2	4
16	–2	–1	1
18	0	0	0
20	2	1	1
22	4	2	4
24	6	3	9
26	8	4	16
Total		$\Sigma d' = -5$	$\Sigma d'^2 = 85$

$$\because \sigma = \sqrt{\frac{\Sigma fd'^2}{n} - \left(\frac{\Sigma fd'}{n}\right)^2} \times h$$

$$= \sqrt{\frac{85}{10} - \left(\frac{-5}{10}\right)^2} \times 2$$

$$= \sqrt{8.50 - 0.25} \times 2$$

$$= \sqrt{8.25} \times 2$$

$$= 2.87 \times 2$$

$$= 5.74$$

Example 11: Calculate mean and standard deviation from the following distribution:

Age (year):	20-25	25-30	30-35	35-40	40-45	45-50
No. of Persons:	170	110	80	45	40	35

Solution:

Age (year)	Mid value (x)	No. of persons (f)	d = x – 32.5	$d' = \frac{d}{2}$	fd'	fd'^2
20-25	22.5	170	–10	–2	–340	680
25-30	22.7	110	–5	–1	–110	110
30-35	32.5	80	0	0	0	0
35-40	37.5	45	5	1	45	45
40-45	42.5	49	10	2	80	160
45-50	47.5	35	15	3	105	315
		N = 480		$\Sigma fd'$	$\Sigma fd'^2$	
				= –220	=1310	

Here, assumed mean A = 32.5

Now, A.M.$(\bar{x})$ = A + $\frac{\sum fd'}{N} \times h$

= 32.5 + $\frac{-220}{480} \times 5$

= 32.5 – 2.99

= 29.51

The required A.M. = 29.51 years.

S.D. (σ) = $\sqrt{\frac{\sum fd'^2}{N} - \left(\frac{\sum fd'}{N}\right)^2} \times h$

= $\sqrt{\frac{1310}{480} - \left(\frac{220}{480}\right)^2} \times 5$

= $\sqrt{2.7292 - 0.2101} \times 5$

= $\sqrt{2.5191} \times 5$

= 1.5871 × 5

= 7.936

∴ The required S.D. = 7.936 years.

Now, A.M. $(\bar{x})$ = $A + \frac{\sum fd'}{N} \times h$

$$= 64.7 + \frac{27}{75} \times 10$$

$$= 64.5 + 3.6$$

$$= 68.10$$

$\therefore$ The required A.M. = Rs. 68.10



CHECK YOUR PROGRESS

Q.5: Determine standard deviation for the following distribution:

Age (year):	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of Persons:	6	42	86	122	78	30	5

Q.6: Below are given yearly profits of a few small companies. Determine arithmetic mean and standard deviation:

Profit (in thousand Rs.):	20-30	30-40	40-50	50-60	60-70
No. of Companies:	30	58	62	85	112
Profit (in thousand Rs.):	70-80	80-90	90-100		
No. of Companies:	112	57	26		



11.5 LET US SUM UP

- A measure of central tendency or an average of a certain distribution is nothing but a representative value of that distribution which enables us to comprehend in a single effort the significance of the whole.
- There are three types of measures of central tendency or averages, they are Mean, Median, and Mode.
- The Arithmetic .Mean. of a variable x is generally denoted by $\bar{x}$ and is defined as the sum of the values of divided by the total number of values of x.

- The median of a series or distribution in ascending or descending order is that observation of the distribution which divides the distribution into two equal parts.
- The mode of a distribution is that observation of the distribution whose frequency is the maximum.
- Scatteredness of data around an average is termed as dispersion or variation.
- The measures of dispersion are Range, Interquartile Range and Quartile Deviation, Mean Deviation, Standard Deviation and Variance and Coefficient of variation.
- The interquartile range of a distribution is the difference between the third quartile Q_3 and the first quartile Q_1 of the distribution.
- The positive square root of the arithmetic mean of the squares of the deviations of the values of a variable from its arithmetic mean is called the standard deviation of that variable.



11.6 FURTHER READING

1) Agarwal, D. R. (2006). *Business Statistics*. Delhi: Vrinda Publications.
2) Gupta S. C. (1994). *Fundamentals of Statistics*. New Delhi: Himalayan Publishing House.
3) Rajagopalan, S. P. & Sattanthan R. (2009). *Business Statistics and Operations Research*. New Delhi: Tata McGraw-Hill
4) Sharma, J. K. (2007). *Business Statistics*. New Delhi: Pearson Education Ltd.
5) Verma, A. P. (2007). *Business Statistics*. Guwahat: Asian Books Private Limited.

11.7 ANSWERS TO CHECK YOUR PROGRESS

Ans. to Q. No. 1: A measure of central tendency or an average of a certain distribution is nothing but a representative value of that distribution which enables us to comprehend in a single effort the significance of the whole.

Ans. to Q. No. 2: Mean (A.M.) = 15.54 years; Median = 16 years

Ans. to Q. No. 3: Mean (A.M.) = Rs. 17.33; Median = Rs. 17; Mode = Rs. 18

Ans. to Q. No. 4: Mean (A.M.) = 9.98; Median = 9.97; Mode = 10.17

Ans. to Q. No. 5: S.D. = 119

Ans. to Q. No. 6: A.M. = Rs. 63.32, S.D. = Rs. 18.69;

11.8 MODEL QUESTIONS

A) Short Question (Answer each question in about 150 words)

Q1 : What is meant by A.M?

Q2 : What is meant by dispersion?

Q3 : Write down the empirical relationship among mean, median and mode.

Q4 : Define standard deviation. Why is standard deviation most widely used as a measure of dispersion?

B) Long Question (Answer each question in about 300-500 words)

Q.1: Define median of a distribution. Discuss its advantages and limitations.

Q.2: Define mode. What are its advantages and limitations?

Q.3: What are the various measures of dispersion?

*** ***** ***

UNIT 12: MEASURES OF ASSOCIATION

UNIT STRUCTURE

12.1 Learning Objectives
12.2 Introduction
12.3 Correlation Coefficient
12.4 The Chi-Square (χ^2) Distribution
12.5 t-Distribution
12.6 Analysis of Variance and F- Distribution
12.7 Let Us Sum Up
12.8 Further Reading
12.9 Answer to Check Your Progress
12.10 Model Questions

12.1 LEARNING OBJECTIVES

After going through this unit you will be able to

- explain the concept of correlation coefficient
- introduce the Chi-square distribution for the comparison of more than two means or proportions
- use the Chi-square distribution for testing the hypothesis about a single population variance
- use t-test in small samples.

12.2 INTRODUCTION

Hypothesis testing begins with an assumption commonly termed as “hypothesis” that a researcher makes about an unknown population parameter. Hypothesis testing is a well-defined procedure which enables us to decide objectively whether to accept or reject the hypothesis based on the information obtained from the sample.In hypothesis testing there are two types of test viz. parametric tests and non-parametric test. In the parametric test, tests are based on

the assumption that the samples are drawn from a normally distributed population or, more appropriately, that the sample means are normally distributed. Such tests are known as parametric tests as the testing procedure involves the assumption about the type of population parameters. On the other hand, in some cases, the researcher may not be sure of whether the population distribution is normal. The statistical tests that do not require prior knowledge about the population are termed as non-parametric tests. Chi-square test is one of the important non-parametric methods.

A non-parametric method, also known as distribution free test satisfies at least one of the following criteria:

(i) The method does not take into consideration any population parameter.

(ii) It is applied only on categorical data that are non-numerical and frequency counts of categories for one or more variables.

(iii) It does not depend on the form of the underlying population distribution, in particular, the requirement of normality.

In this unit we will focus on χ^2 (Chi-square) test, t-distribution and F-distribution .

12.3 CORRELATION COEFFICIENT

In the simplest term, correlation means relation between two variables (note that ‘co’ simply means two things together). In case of bivariate distributions, many a times, we may be interested in looking at how the two variables are related among themselves. Obviously, when two variables are related, change in any one variable is responded by the other variable showing some changes in it. Correlation analysis tries to measure the 'extent' of such responses among the variables. Thus, correlation is an analysis of co-variation between two (or even more) variables. Two variables are said to be correlated if the change in any one of them results in a corresponding change in the other.

Correlation may be classified as Positive correlation, Negative correlation and Zero Correlation.

When increase (or decrease) in one variable leads to a corresponding increase (or decrease) in the other variable then it is called positive correlation. Negative correlation is, on the other hand, is a situation when increase (or decrease) in one variable is accompanied by a decrease (or increase) in the other variable.In case, variables are found as uncorrelated i.e., change in one of the variables fail to produce any corresponding change in the other, and then it is called zero correlation.

Scatter plots are particularly useful to identify whether the variables are correlated and in case they are then what is the type of such relation. It, however, can't quantify or quantitatively measure the 'amount', 'extent', 'magnitude' or 'degree' of relation that exists between the variables. Karl Pearson introduced a simple technique for measuring the magnitude of such correlation. His measure is called Pearson's Product-Moment Correlation Coefficient or simply Pearson' Correlation Coefficient and usually denoted by 'r' (standing for 'relation') which is given as:



Karl Pearson (1857-1936) was an English mathematician and biostatistician

$$r_{x,y} = \frac{Cov(x,y)}{\sigma_x.\sigma_y}$$

Where $Cov(x,y) = \frac{1}{n}\Sigma(x-\bar{x})(y-\bar{y})$

$$= \frac{1}{n}\Sigma xy - \bar{x}\,\bar{y} \quad \text{(after simplification)}$$

$$\sigma_x = \sqrt{\frac{1}{n}\Sigma(x-\bar{x})^2}$$

$$= \sqrt{\frac{1}{n}\Sigma x^2 - \bar{x}^2} \quad \text{(after simplification)}$$

$$\sigma_y = \sqrt{\frac{1}{n}\Sigma(y-\bar{y})^2}$$

$$= \sqrt{\frac{1}{n}\Sigma y^2 - \bar{y}^2} \qquad \text{(after simplification)}$$

Thus, the simplified form of the coefficient of correlation is given as:

$$r_{xy} = \frac{\frac{1}{n}\Sigma xy - \bar{x}\,\bar{y}}{\sqrt{\frac{1}{n}\Sigma x^2 - \bar{x}^2}\sqrt{\frac{1}{n}\Sigma y^2 - \bar{y}^2}}$$

$$= \frac{\frac{1}{n}\Sigma xy - \frac{\Sigma x}{n}.\frac{\Sigma y}{n}}{\sqrt{\frac{1}{n}\Sigma x^2 - \left(\frac{\Sigma x}{n}\right)^2}\sqrt{\frac{1}{n}\Sigma y^2 - \left(\frac{\Sigma y}{n}\right)^2}}$$

$$= \frac{\frac{1}{n}\left[\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}\right]}{\sqrt{\frac{1}{n}\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right].\frac{1}{n}\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}}$$

$$= \frac{\frac{1}{n}\left[\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}\right]}{\sqrt{\left(\frac{1}{n}\right)^2\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}}$$

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

Example 1: Find Karl Pearsons correlation coefficient from the following information :

X :	2	3	5	5	6	8
Y :	9	8	8	6	5	3

Solution : Let us prepare the following table :

x	y	xy	x^2	y^2
2	9	18	4	81
3	8	24	9	64
5	8	40	25	64
5	6	30	25	36
6	5	30	36	25
8	3	24	64	9
$\Sigma x = 29$	$\Sigma y = 39$	$\Sigma xy = 166$	$\Sigma x^2 = 163$	$\Sigma y^2 = 279$

Now,

$$r_{xy} = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{n}}{\sqrt{\left[\Sigma x^2 - \frac{(\Sigma x)^2}{n}\right]}\sqrt{\left[\Sigma y^2 - \frac{(\Sigma y)^2}{n}\right]}}$$

$$= \frac{166 - \frac{29 \times 39}{6}}{\sqrt{\left[163 - \frac{29^2}{6}\right]}\sqrt{\left[279 - \frac{39^2}{6}\right]}}$$

= 0.93 (approx) (after simplification)

12.4 THE CHI-SQUARE (χ^2) DISTRIBUTION

The symbol χ stands for the Greek letter 'chi'. This test was first developed by Karl Pearson in testing of hypothesis in 1900. The sampling distribution of χ^2 is called χ^2 distribution. This distribution has only one parameter called the degrees of freedom (df) and the shape of the curve depends on the degrees of freedom. The distribution is skewed to the right. Being a square quantity, χ^2 distributi3on can never be negative. In fact, χ^2 is a continuous probability distribution with range 0 to ∞ (Figure 12.1)

The probability density function of distribution is given by

$$f(\chi^2) = C(\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}$$

where ν is is the degrees of freedom, C is a constant depending upon the degrees of freedom, and $e = 2.71828.$

χ^2 test statistic can be defined as below:

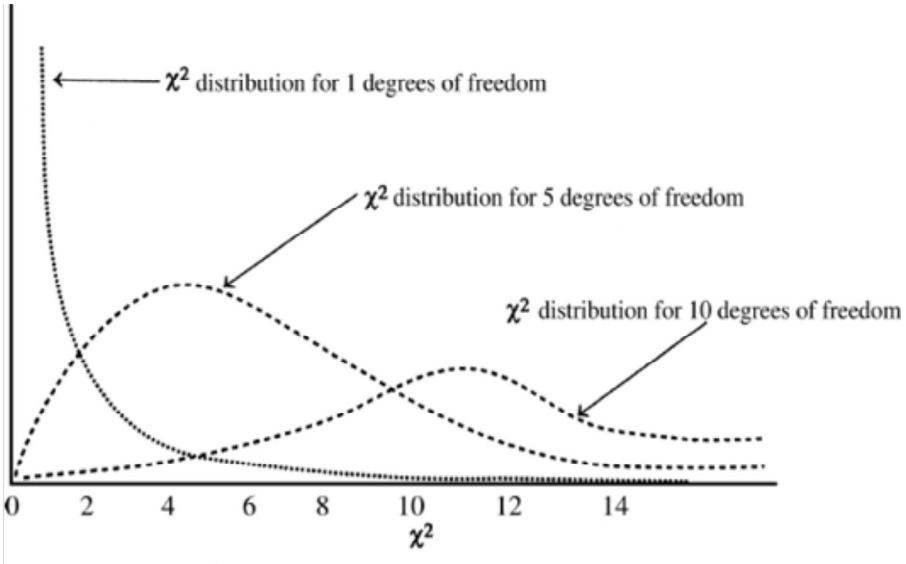
$$\chi^2 = \sum \frac{(O-E)^2}{E} \text{ with } df = k - 1 - c$$

where O is the observed frequency, E the expected frequency, k the number of categories, and c the number of parameters being estimated from the sample data.

The computed value of χ^2 is compared with the critical value of χ^2 at a specified level of significance and degrees of freedom. The decision to accept or reject the null hypothesis is given by

If $\chi^2_{cal} < \chi^2_{critical}$ then we may accept the null hypothesis, otherwise we reject the null hypothesis.

Figure 12.1: χ^2 Distribution with 1, 5 and 10 Degrees of Freedom



Properties of Chi-square test:

The following properties are useful while using χ^2 test statistic to analyze its sampling distribution:

(i) The density function of χ^2 does not involve any parameter and the shape of the curve depends upon the degrees of freedom only. The distribution becomes more and more symmetrical as the degrees of freedom increase and tends to normal distribution for high value of degrees of freedom ($\nu \geq 30$).

(ii) Mean = Degrees of freedom = ν, mode= $\chi^2 - 2$ and variance= 2ν

(iii) The values of χ^2 lie between 0 and ∞.

(iv) If $x_1, x_2, ..., x_n$ be a random sample drawn from a normal population with mean μ and variance σ^2, then

(a) $\sum_{i=1}^{n} (x_i - \mu)^2 / \sigma^2$ follows χ^2 -distribution with n degrees of freedom.

(b) $\sum_{i=1}^{n} (x_i - \bar{x})^2 / \sigma^2$ follows χ^2 -distribution with degrees of freedom $n-1$, $\bar{x}$ being the sample mean.

(v) For large degrees of freedom (ν), the standard variate

$$Z = \frac{\chi^2 - E(\chi^2)}{S.E.(\chi^2)} = \frac{\chi^2 - \nu}{\sqrt{2\nu}}$$

may approximately be considered as a standard normal variate.

Conditions for applications of Chi-square test:

The following conditions need to be satisfied before applying χ^2 as a test statistic for hypothesis testing:

(i) In a contingency table, an expected frequency of less than 5 in a cell is less than the frequency required to apply the test. In such cases we apply the technique of 'pooling' which consists

in adding the frequencies which are less than 5 with the preceding or succeeding frequency, so that the resulting sum is 5 or more and adjust for the degrees of freedom accordingly.

(ii) The sample should consist of at least 50 observations and should be drawn randomly from the population.

(iii) The data should be expressed in original units, rather than in percentage or ratio form.

(iv) The constraints on the cell frequencies, if any, should be linear such as $\sum O = \sum E = N$, the total frequency.

12.5 t-DISTRIBUTION

We have already discussed that in case of large samples (i.e. $n \geq 30$,) the sampling distribution of the statistic is approximately normal. But in many decision making situations, due to time, money and other constraints, sometimes a researcher may have to take small sample of size less than 30 (i.e. $n < 30$,). In case of small samples also, the researcher can apply standard normal test for a specified population mean or equality of two population means provided the sample is drawn from a normal population where S.D. σ is known. But if σ remains unknown then standard normal test fails and in all such cases the sampling distribution turns out to be the student's 't' distribution. Sir William Gosset of Ireland in early 1900, under his pen name 'Student', developed a method for testing of hypothesis known as 't-test'. It is to be noted that the concept of degrees of freedom is associated with t-test. The graph of the 't' distribution approaches the normal curve as the degrees of freedom increase.

If $x_1, x_2, \ldots\ldots\ldots\ldots, x_n$ be a random sample of size drawn from a normal population with mean μ and variance σ^2, then the Student's 't' statistic is defined by

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}},$$

where $\bar{x}$ = sample mean

and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$, sample mean square

is an unbiased estimator of σ^2 . If s^2 be the sample variance given by

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

Then $ns^2 = (n-1)S^2$

Thus we have, $\frac{S}{\sqrt{n}} = \frac{s}{\sqrt{n-1}}$

$$\therefore \quad t = \frac{\bar{x} - \mu}{S/\sqrt{n}} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

Here, 't' follows t-distribution with (n-1) d.f.

Characteristics of t-Distribution

The characteristics of t distribution are given below:

(1) It is an exact distribution.

(2) The distribution is unimodal with

Mean=Median=Mode=0

(3) The t-distribution is less peaked than normal distribution at the centre and higher at both the ends.

(4) The shape of the -distribution depends on depends on the sample size . As increases, it approaches the normal distribution.

(5) The limits of t is $-\infty < t < \infty$.

(6) The critical values of at level of significance vary with the degrees of freedom.

Applications of t-Distribution

The -distribution has a number of applications in Statistics some of which are given below

(i) It is used to test the significance of single mean, population standard deviation being unknown.

(ii) It is used to test the significant difference between two population means under the assumption that population standard deviations being equal but unknown.

(iii) It is used to test the significance of an observed sample correlation coefficient.

Assumptions for Student's -test: The following assumptions are made in Student's -test:

(i) The parent population from which the sample is drawn is normal.

(ii) The drawn sample is random.

(iii) The population standard deviation σ remains unknown.

12.6 F- DISTRIBUTION AND ANALYSIS OF VARIANCE

In this section, we will discuss only the concept of F-Distribution and analysis of variance

F-Distribution: F-variate is defined as the ratio of two independent chi-square variates divided by their corresponding degrees of freedom. If χ_1^2 and χ_2^2 are two independent chi-square variates with n_1 and n_2 degrees of freedom respectively then F is defined as

$$F= \frac{\chi_1^2 / n_1}{\chi_2^2 / n_2}$$

It has (n_1 and n_2) degrees of freedom.

Analysis of Variance: Analysis of Variance (ANOVA) is a technique of partitioning the total variance of an experiment or trial into component variances responsible for contributing towards total variance. If any gap exists between total variance and sum of component variances, it is attributed to experimental/ random error

and so is true for degrees of freedom. ANOVA utilises F-test. Here, each of the component variance is tested against error variance and conclusion is drawn.



CHECK YOUR PROGRESS

Q 1. What is meant by correlation?
(Answer in about 20 words)

.....

.....

Q 2. State whether the following statements are true or false:

(i) The critical value of a statistic is the value which cuts off the region for the rejection of null hypothesis.

(ii) The χ^2 -test requires that the sampling distribution should be normally distributed.

(iii) In order to reject the null hypothesis $\chi^2_{cal} > \chi^2_{table}$

differ significantly.

Q 3. Mention any two characteristics of t-distribution.

(Answer in about 20 words)

.....

.....

Q 4. What is meant by F- variate? (Answer in about 20 words)

.....

.....



12.7 LET US SUM UP

- Statistical tests that do not require any prior information about the population are termed as non-parametric tests. This unit focuses on only χ^2 (Chi-square) distribution and the related χ^2 test. χ^2 distribution

is the family of curves with each distribution being defined by the degree of freedom associated to it.

- χ^2 test can be used for a variety of purposes. χ^2 test provides a platform that can be used to ascertain whether the theoretical probability distribution coincides with the empirical sample distribution. This is commonly known as χ^2 goodness of fit test.
- χ^2 test can be used to test the independence of independence of two variables. χ^2 test of independence makes use of the contingency table to determine the independence of two variables.
- Sir William Gosset of Ireland in early 1900, under his pen name 'Student', developed a method for testing of hypothesis known as 't-test'.
- The concept of degrees of freedom is associated with t-test. The graph of the 't' distribution approaches the normal curve as the degrees of freedom increase.
- F-variate is defined as the ratio of two independent chi-square variates divided by their corresponding degrees of freedom.
- Analysis of Variance (ANOVA) is a technique of partitioning the total variance of an experiment or trial into component variances responsible for contributing towards total variance. If any gap exists between total variance and sum of component variances, it is attributed to experimental/ random error and so is true for degree of freedom.



12.8 FURTHER READING

1) Sharma, J.K. (2007). *Business Statistics.* New Delhi. Pearson Education Ltd.
2) Hazarika, P.L. (2016). *Essential Statistics For Economics And Business Studies.* New Delhi. Akansha Publishing House.
3) Lind, D.A., Marshal, W.G., Wathen, S.A. (2009) *Statistical Techniques in Business and Economics.* New Delhi. Tata McGraw Hill Education Private Limited.

12.9 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: Correlation is an analysis of co-variation between two (or even more) variables. Two variables are said to be correlated if the change in any one of them results in a corresponding change in the other.

Ans to Q No 2: (i) T, (ii) F, (iii) T.

Ans to Q No 3: (1) It is an exact distribution.

(2) The distribution is unimodal with

Mean=Median=Mode=0

Ans to Q No 4: F-variate is defined as the ratio of two independent chi-square variates divided by their corresponding degrees of freedom.

12.10 MODEL QUESTIONS

Q 1: Define coefficient of correlation.

Q 2: Define Chi-square distribution. Discuss the important properties of the distribution.

Q 3: Under what situations Chi-square test can be applied?

Q 4: Write down the applications of Chi-square test.

Q 5: State the meaning and characteristics of t-distribution.Explain the applications of t-test.

*** ****** ***

Unit 13 : INTERPRETATION AND PRESENTATION OF STATISTICAL DATA

UNIT STRUCTURE

13.1 Learning Objectives

13.2 Introduction

13.3 Tabulation of Data

- 13.3.1 Parts of Table
- 13.3.2 General Rules for Making Tables
- 13.3.3 Types of Table

13.4 Diagrammatic Representation of Data

- 13.4.1 General Rules of Drawing Diagrams
- 13.4.2 Types of Diagram

13.5 Graphical Representation of Data

- 13.5.1 General Rules for Constructing Graphs
- 13.5.2 Types of Graph

13.6 Analysis of Findings, Limitations and Scope for Further Research

13.7 Let us Sum Up

13.8 Further Reading

13.9 Answers to Check Your Progress

13.10 Model Questions

13.1 LEARNING OBJECTIVES

After going through this unit, you will be able to -

- know the general rules of drawing tables,diagrams and graphs
- understand the types of drawing tables,diagrams and graphs
- understand how to interprete data in a meaningful manner.
- understand how to illustrate relationship between variables through graphical representation.

13.2 INTRODUCTION

Presentation of data means exhibition of data in a clear and attractive manner so that these can be easily understood and analysed. Statistical data may be presented in the forms of diagrams and graphs. Diagrammatic presentation is the most popular and commonly used form of representing statistical data. It presents data in an attractive and appealing way. In case of graphical presentation, statistical data are presented in the form of lines and curves in graph paper.In this unit, we will discuss in details about the interpretation of statistical data.

13.3 TABULATION OF DATA

In unit 9 of this course, we have already studied about the classification and tabulation of data and we learnt that they are actually not two distinct processes but they more or less go together. In this unit, we will study about the parts of table, rules of making table and types of table.

13.3.1 Parts of Table

A table comprises of different parts which are as follows:

- Number and title – Each table should be numbered. The number is accompanied by a suitable title. The title is a short description of the contents of the table. The title should be clear, brief and written in a form different from the contents of the table.
- Caption – The heading of a column is called the caption. The caption tells us what the column represents. A column may also have subheads and more importantly it should indicate the unit of measurement of the data contained in the column.

- Stub – The contents of the rows are indicated from the stub. The stub is the first column in a table and tells us what the data contained in the rows represent. Column is the first horizontal row and stub is the first column in the table.
- Body – The numerical information contained in the columns and rows of the table is called the body of the table. This is the most important and vital part of the table.
- Footnotes – Any explanation regarding any part of the table which is necessary for understanding the table is added as footnotes. The footnotes are placed just below the body of the table, outside the margin of the table. The footnote is added either, to point out exceptions, clarify something about the table or cite source of secondary data. If there is more than one footnote necessary, then they should be numbered or simply explained by adding a star.

13.3.2 General Rules for making Tables

While tabulating the data certain precautions should be taken, which are as follows:

- If the data are too large then separate tables should be made. Each table should portray a particular set of data. This would make the data more comprehensible.
- If any statistical calculation like total, average, percentage, etc. is to be shown, then they should be placed together after the data.
- The caption, stub and body of the table should follow a systematic order. Then only the data becomes readable and allows important items to be emphasized. The data should be arranged alphabetically, chronological, or by following some accepted conventional arrangement depending on the nature of the data.

- Zero should be used to indicate zero quantity and not non availability of information. If data for a particular box is not available then a dash (-) can be used.

The objective of tabulation is to portray clearly and understandably the collected data. Three things that need to be kept in mind while making a table are that the data should be clearly represented; secondly, it should contain only those information which are important for understanding a particular aspect of the total data; Thirdly it has to be assured that the contents of the table are accurately represented.

13.3.3 Types of Table

There are mainly two types of tables. One is the simple type where only one characteristic is shown and the other is the Complex tables in which information about two or more inter-related characteristics is shown together. This type of table is commonly used as they facilitate a comparison and consideration of all related facts. A simple table is also known as a one way table. When two characteristics are shown, it is known as a two way table. When four or more characteristics are simultaneously shown then such a type of tabulation is called a manifold tabulation. Tabulation helps in consolidating and presenting raw data. Tables or frequency distributions are not the end in itself. They are commonly a preface to a more intensive analysis. Tabulation can be said to be the last step in the collection and compilation of data. It leads to analysis and interpretation of data.



CHECK YOUR PROGRESS

Q1: What is meant by manifold Tabulation? (Answer in about 20 words)

.....

.....

Q2: What are Footnotes? (Answer in about 20 words)

.....

.....

13.4 DIAGRAMMATIC REPRESENTATION OF DATA

In unit 9 of this course we have already studied about the basics of diagramatic representation of data. In this unit, we will study about the general rules of drawing a diagram and types of diagram.

13.4.1 General Rules of Drawing Diagrams

- **Title:** First and foremost it is very essential that a suitable title is given to the diagram. The title should be written either on top or just below the diagram. It should be very clear precise, self-explanatory and convincing.
- **Size:** It is very important to decide the size of the diagram. It should be neither to small nor too big.
- **Scale:** The scale should be proper and accurate so that no mistake creeps in and the diagram can properly depict the data.
- **Actual Drawing:** The diagram should be neatly and clearly drawn. The different parts of the diagram should be drawn in such a way that each part is easily distinguishable.
- **Use of colours and signs:** Each part of the diagram should be separated with the use of appropriate types of lines or shades or colours.

- **Footnotes and Source Notes:** this is a very important part of the diagram. Source notes and footnotes are given at the bottom of the diagram. Footnotes are given in order to clarify some points related to the diagram and source notes are given to indicate the source from which data have been given.

13.4.2 Types of Diagram

There are different types of diagrams. But here we will discuss only bar diagram and pie diagram.

Bar Diagram

There are three types of bar diagram – simple, multiple and sub-divided bar diagram

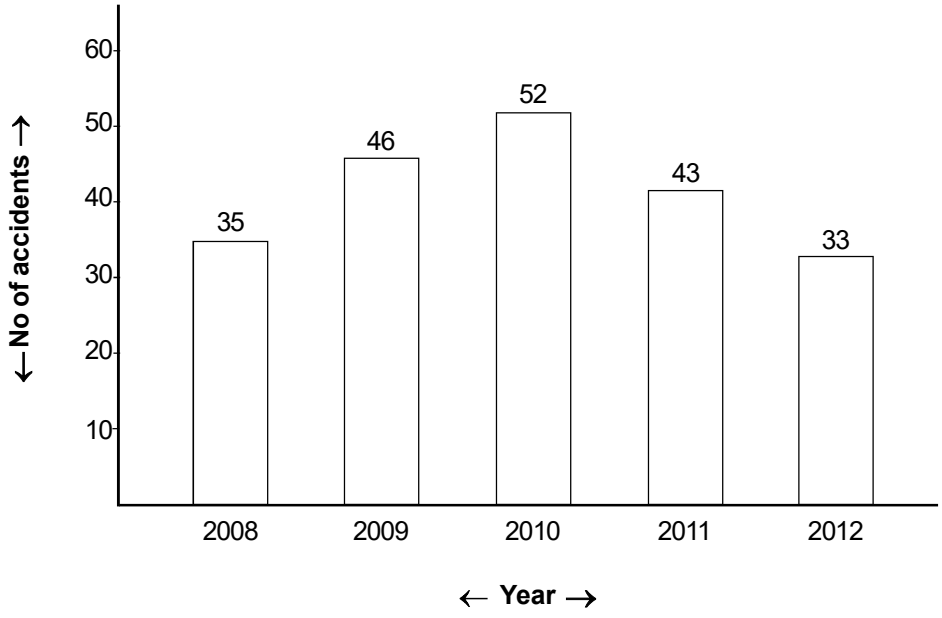
Simple Bar Diagram :

A simple bar diagram takes the form of single bar to represent values of a single variable.

Example 1: Represent the following data with the help of a suitable diagram.

Year	2008	2009	2010	2011	2012
No. of Accidents	35	46	52	43	33

Figure 13.1



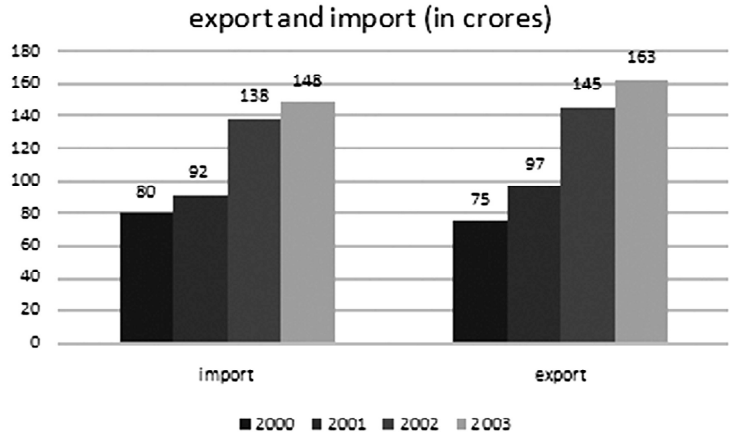
Multiple Bar Diagram:

Multiple Bar Diagram is drawn to represent two or more sets of interrelated data.

Example 2 : Draw a multiple bar diagram t represent the following data.

Year	2000	2001	2002	2003
Import(in crores)	80	92	138	148
Export(in crores)	75	97	145	163

Figure 13.2: Export and Import (in Crores)



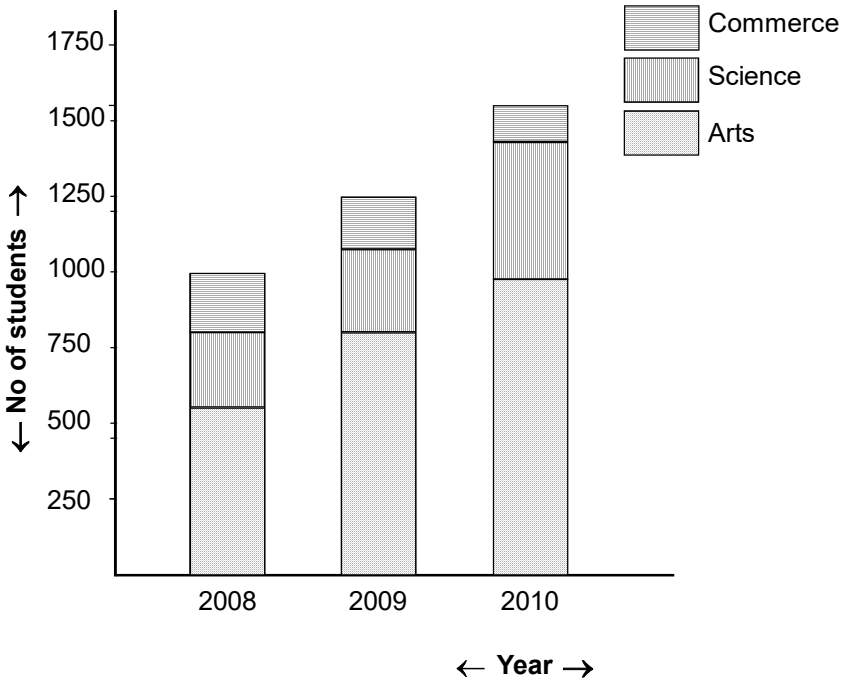
Sub-Divided Bar Diagram:

The diagram which represents the total values and parts of a set of a data in one diagram is known as sub-divided bar diagram.

Example 3: Construct a sub-divided bar diagram to represent no. of students admitted in science, arts and commerce stream of a college in the last three years.

Year	Arts	Science	Commerce	Total
2008	650	250	100	1000
2009	800	300	150	1250
2010	900	400	250	1550

Figure 13.3



Pie-Diagram

Just as sub-divided bars are used to represent the total magnitude and its various components, the circle may be divided into various sections or segments, viz, sectors representing certain proportions or percentage of the various component parts of the total. Such a sub-divided circle diagram is known as an angular or pie-diagram as the various segments resembles slices cut from a pie.

STEPS TO CONSTRUCT A PIE DIAGRAM

- Using a compass, construct a circle of an appropriate size.
- Then express each of the component values as a percentage of respective totals.
- Then the percentage values are covered into corresponding degrees in the circle. One circle contains 360 and percentage values of all the items are equal to 100, therefore one percentage value is represented by 360/100 = 3.6. Therefore to get size of angles each percentage value is multiplies with 3.6.

- Take a radius as base line to draw the angle represented by the first component. The new line will then become the base for second component's angular representation. This process is repeated till all the components are represented.
- Then different component parts are distinguished from one another by using different shades, dotting, colours, etc.

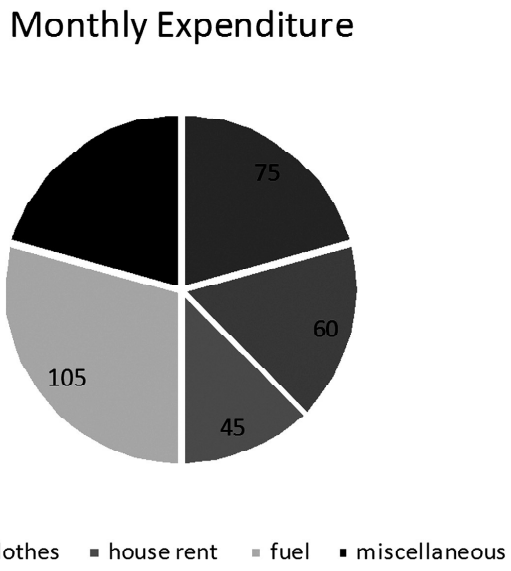
EXAMPLE 4: Expenditure of a family in different heads in a year:

HEADS OF EXPENDITURE	EXPENDITURE (in Rs.)
Food	50
Clothes	40
House Rent	30
Fuel	70
Miscellaneous	50

In order to present the above information in pie diagram, the various components are converted into degrees of the circle as shown below –

HEADS OF EXPENDITURE	EXPENDITURE	PERCENTAGE VALUE (in Rs.)	PROPORTIONATE ANGLE
Food	50	50/240x100=20.83	20.83x3.6=7.4.98=75
Clothes	40	40/240x100=16.66	16.66x3.6=59.976=60
House Rent	30	30/240x100=12.5	12.5x3.6=45
Fuel	70	70/240x100=29.16	29.16x3.6=104.976=105
Miscellaneous	50	50/240x100=20.83	20.83x3.6=74.988=75
Total	240		360

Figure 13.4



Disadvantages of Diagramatic Presentation

- Diagrams provide limited information. In order to get detailed and precise information one needs to go through the original statistical tables.
- Diagrams are useful only when comparisons are desired. A single diagram may nt be useful.
- Diagrams are not suitable for further analysis.

13.5 GRAPHICAL REPRESENTATION OF DATA

Another useful method of presenting data is graphic presentation. In this method statistical data are shown in the form of line or curve in the graph paper.In unit 9 of this course, we have studied about the basics of graphical presentation. In this section we will study the rules of constructing the graphs and types of graphs.

13.5.1 General Rules for Constructing Graphs

- Every graph should have a heading or title.
- Graphs are prepared on graph paper. Considering the size and nature of data a suitable point on the graph is fixed as the

point of origin. Horizontal and vertical lines intersecting through this point are shown as bold lines which are known as axis. Horizontal line from left to right is known as Horizontal axis or X-axis. On the other hand, the line going upward r downward is known as vertical axis or Y-axis. Intersected by X-axis and Y-axis a graph paper is divided into four parts and each part is called quadrant. Point of origin in the graph is indicated by the word "O". To the right and above the point of origin, positive values of a statistical series are indicated. Therefore, in the first quadrant, both X and Y are +X and +Y respectively. In the second quadrant X is –X while Y is –Y respectively; in the third quadrant both X and Y are –X and –y respectively; in the fourth quadrant while +X, Y is +Y (Choudhury, et al., 2013).

- A suitable scale should be chosen such that the whole data can be accommodated.
- The scale n the Y-axis should begin from zero.
- The curve drawn must be distinct and the curve must not be overcrowded with curves.
- An index is to given to indicate the scale and meaning of different curves.
- Sometimes the actual figures are also indicated in the graph.
- The source of information must be given at end of the graph.

13.5.2 Types of Graph

Although there are many types of graph, here we will discuss only two types – Histogram and Ogive.

Histogram or Column Diagram

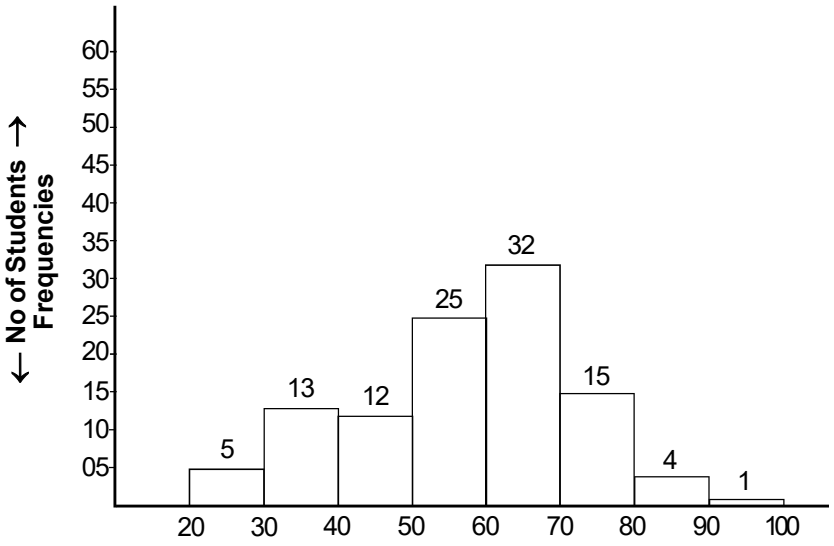
Histograms are the most commonly used forms of graphs. They are a series of rectangles that are drawn adjacent to each other without any gaps between them. They are drawn on the basis of number of class intervals. The shapes of class

interval depend on whether the class intervals are equal or unequal.

When class intervals are equal: Marks obtained by 100 students in Economics are given below. Construct a Histogram.

Marks (Classes)	No. of Students
20-30	5
30-40	13
40-50	12
50-60	25
60-70	32
70-80	15
80-90	4
90-100	1

Figure 13.5



When class intervals are unequal: Let us consider the following data.

Age	25-29	30-34	35-39	40-44
No. of persons	50	30	40	25

In order to prepare a histogram with unequal class intervals, we have to first make the class intervals equal. For this, we delete 0.5 from the lower limit and 0.5 in the upper limit. This is shown in the following table.

Classes (age)	No. of persons (f)	Class limit	Class length (h)	Frequency density = f/h
25-29	50	24.4-29.5	5	10
30-34	30	29.5-34.5	5	6
35-39	40	34.5-39.5	5	8
40-44	25	39.5-44.5	5	5

Difference Between Histogram and Bar Diagram

- A histogram is a two dimensional diagram where both the width and length are important. On the other hand, bar diagram is one dimensional diagram in which only length matters while width is arbitrary.
- In histogram, the bars are adjacent to each other whereas in bar diagram proper spacing is given between different bars.
- In histogram the class frequencies are represented by the area of rectangles while in a bar diagram they are represented by the heights of the corresponding bars.
- Histogram is drawn only for continuous variable while bar diagram is drawn both for continuous and discrete variable.

Cumulative Frequency Curve or Ogive

Ogive is a graphical representation of cumulative frequency distribution of continuous variable. Cumulative frequencies are plotted against the class boundaries. Cumulative frequencies are placed in the Y-axis and class boundaries are placed in the X-axis. As there are two types of cumulative frequencies — less than and more than – so there are two types of ogives – less than and more than.

LESS THAN OGIVE : In order to construct less than ogive, less than cumulative frequencies are plotted against the upper class boundaries of the respective classes. Points so obtained are joined by a smooth free hand curve to get less than ogive. Less than ogive is an increasing curve, slopping upwards from left to right and has the shape f an elongated S.

MORE THAN OGIVE : In order to construct less than ogive, more than cumulative frequencies are plotted against the upper class boundaries of the respective classes. Points so obtained are joined by a smooth free hand curve to get more than ogive. More than ogive is a decreasing curve, slopping downwards from left to right and has the shape of an elongated S, upside down. Point of intersection of the less tan and more than ogives give the median. The foot of the perpendicular from their point of intersection on the X-axis gives the value of the median.

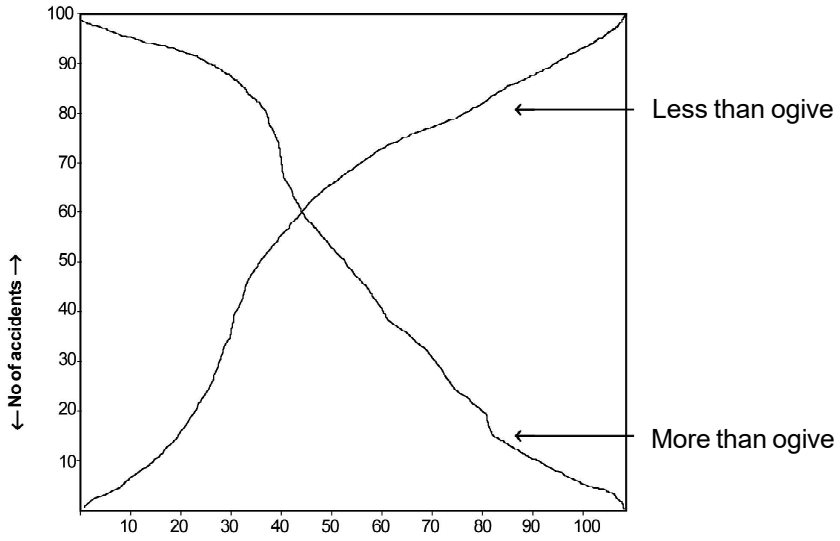
Example: Construct less than and more than ogive with the following data.

Marks	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90
No. of Students	5	13	20	22	17	9	10	4

In order to construct the ogive, we first work out the following table.

Less than ogive Marks	Students	More than ogive Marks	Students
<20	5	>20	100
<30	18	>30	95
<40	38	>40	82
<50	60	>50	62
<60	77	>60	40
<70	86	>70	23
<80	96	>80	14
<90	100	>90	4

Figure 13.6



13.6 ANALYSIS OF FINDINGS, LIMITATIONS AND SCOPE FOR FURTHER RESEARCH

ANALYSIS OF FINDINGS: In order to fulfil the predetermined objective and to test the hypothesis, a researcher is required to analyse and examine the collected data by using the appropriate statistical tools and techniques. Although it is not a routine process, but it requires careful, logical and critical examination of the results obtained from the analysis of data. There is always an element of subjectivity depending on the limitations of sample chosen and statistical tool utilised. So as a researcher our aim and objective is to reduce such subjectivity to a minimum level while interpreting the results. In the light of interpretations of the result, a researcher have to formulate the conclusions and generalisations.Here, critical and logical thinking is required in summarizing the findings of the study and compare the findings with the hypothesis formulated in the beginning of the study. A researcher cannot draw any conclusions which are inconsistent among themselves or with external facts or known laws of nature. The conclusion may be subject to further examination. Sugessions should also be put to apply the research findings in practical field.

LIMITATIONS OF THE STUDY : Limitations of the study are those conditions that are beyond the control of the researcher while conducting a research. These conditions might impact the results of the study. For example, due to certain unavoidable circumstances, the researcher might not be able to have access to certain samples, resources or places. The limitations of the study have to be acknowledged by the researcher.

SCOPE FOR FURTHER : A researcher is required to provide some suggestions for addressing the problems of further research associated with the current study. Scope for further research includes the following:

1. Building upon findings of your research
2. Addressing limitations of your research
3. Constructing the same research in a new context, location and/or culture
4. Re-assessing and expanding theory, framework or model you have addressed in your research.



CHECK YOUR PROGRESS

Q3 : What is meant by limitations of the research study? (Answer in about 20 words)

.....

.....

Q4 : What is histogram? (Answer in about 20 words)

.....

.....

Q5 : What is meant by ogive? (Answer in about 20 words)

.....

.....

Q6 : What are less than ogive and more than ogive? (Answer in about 20 words)

.....

.....



13.7 LET US SUM UP

- Any explanation regarding any part of the table which is necessary for understanding the table is added as footnotes.
- If the data are too large then separate tables should be made. Each table should portray a particular set of data. This would make the data more comprehensible.
- Data may be presented in the form of diagrams or graphs.
- Diagrammatic presentation can take two forms – bar diagram and pie-diagram.
- First and foremost it is very essential that a suitable title is given to the diagram. The title should be written either on top or just below the diagram.
- Bar diagrams are of three types – simple, multiple and sub-divided.
- In graphic presentation of data statistical data are shown in the form of line or curve in the graph paper.
- Graphs may be presented in the form of histogram and ogive. Ogives are of two types – more than ogive and less than ogive
- In order to fulfil the predetermined objective and to test the hypothesis, a researcher is required to analyse and examine the collected data by using the appropriate statistical tools and techniques analysis.
- Limitations of the study are those conditions that are beyond the control of the researcher while conducting a research. These conditions might impact the results of the study.
- A researcher is required to provide some suggestions for addressing the problems of further research associated with the current study.



13.8 FURTHER READING

1) Gupta, S.C. (2009). *'Fundamentals of Statistics,* Mumbai : Himalayan Publishing House, Sixth Edition Reprint.

2) Gupta, S.P. (2010). *'Statistical Methods',* New Delhi. : Sultan Chand & Sons.



13.9 ANSWERS TO CHECK YOUR PROGRESS

Answer to Q 1: When four or more characteristics are simultaneously shown on a table, such a type of tabulation is called a manifold tabulation.

Answer to Q 2: Footnotes can be described as the explanations regarding any part of table which are placed at the bottom of the table i.e below the body and outside the margin of the table.

Ans to Q No 3 : Limitations of the study are those conditions that are beyond the control of the researcher while conducting a research.

Ans to Q No 4 : They are series of rectangles that are drawn adjacent to each other without any gaps between them.

Ans to Q No 5 : Ogive is a graphical presentation of c.f. distribution of continuous variable.

Ans to Q No 6 : Less than ogive is an increasing curve, slopping up wards from left to right; and more then ogive is a decreasing curve, slopping downwards from left to right.

13.10 MODEL QUESTIONS

A) Short Question (Answer each question in about 150 words)

Q1 : What is meant by scope of further research?

Q2 : Define limitations of the study.

Q3 : What are the different types of graph?

Q4 : Briefly explain the importance of graphical presentation of data.

B) Long Question (Answer each question in about 300-500 words)

Q1 : Explain the importance of analysis of findings of research.

Q2 : Discuss the general rules of drawing diagrams.

Q3 : What is a pie diagram? What are the general rules for drawing a pie diagram?

Q4 : Discuss the general rules for drawing graphs.

Q5 : Explain how the method of drawing a histogram with equal class interval different from the method of drawing a histogram with unequal class interval.

*** ***** ***

UNIT 14: QUALITIES OF RESEARCHER

UNIT STRUCTURE

14.1 Learning Objectives
14.2 Introduction
14.3 Maintaining Objectivity
14.4 The Unbiased Approach
14.5 Curiosity to Know the Unknown
14.6 Desire to know the Cause and Effect Relationship of Social Problems
14.7 Desire to discover New and Test Old Scientific Procedures
14.8 Patience and an Alert Mind
14.9 Courage and Consciousness
14.10 Let Us Sum Up
14.11 Further Reading
14.12 Answers to Check Your Progress
14.13 Model Questions

14.1 LEARNING OBJECTIVES

After going through this unit, you will be able to-

- understand how to proceed with your research work
- identify the qualities important for a researcher
- adopt the correct approach towards objective research.

14.2 INTRODUCTION

The quality of a research work determines its validity and importance in the field of knowledge. The approach adopted by the researcher while doing the research determines the quality of the research. Hence, it is very important for the researcher to:

- Maintain Objectivity

- Have an unbiased approach
- Be curious to know more and more about his field of study
- Be desirous to know the Cause and Effect Relationship of Social Problems
- Be desirous to discover new and test old scientific procedures
- Have patience and an alert mind
- Be courageous and conscious

In the subsequent sections, we will discuss each of these qualities to have a deeper and proper insight into the path of proper research.

14.3 MAINTAINING OBJECTIVITY

The Merriam-Webster Dictionary defines 'objective' as expressing or dealing with facts or conditions as perceived without distortion by personal feelings, prejudices, or interpretations. Being objective in research means to eliminate all kinds of bias. The main intention of the researcher is to discover the truth behind the research undertaken. The principal aim of research is to obtain accurate and reliable results. In order to achieve this, research must be carried out with total integrity. There should not be any manipulation or bias or any subjective effort to find a desired outcome. In other words, an objective research can be interpreted as a research that is impartial and unbiased.

Maintaining objectivity is a difficult task in social research as it involves human beings as the principal subject matter. For example, if a researcher undertakes a research work about a social issue, he/she might be influenced by cultural, societal, political, economic, or any other bias. Such biases will definitely affect the objectivity in the approach and the true result might be lost. The objective of the research will go in vain.

However, the researcher must not spare any effort towards adopting an unbiased and impartial approach while addressing the

research problem. The approach of the researcher should be such that the findings of the study depend on the nature of what was studied. Factors like personal values, beliefs, biases, or ideologies should not be reflected in the research.

14.4 THE UNBIASED APPROACH

An unbiased research is defined as "A sample drawn and recorded by a method which is free from bias. This implies not only freedom from bias in the method of selection, e.g. random sampling, but freedom from any bias of procedure, e.g. wrong definition, non-response, design of questions, interviewer bias, etc. An unbiased sample in these respects should be distinguished from unbiased estimating processes which may be employed upon the data." (https://stats.oecd.org/glossary/detail.asp?ID=3878)



Unbiased : Showing no prejudice for or against something

When a researcher is biased, the research process is hampered and the researcher cannot have an unprejudiced approach to the problem. Moreover, the researcher must be careful at all the stages of research as bias can occur at any phase. It might occur during the process of research design, or data collection and even during the analysis of the data. If it occurs then the objectivity of the research is also lost.

Let us discuss different types of research bias:

Quantitative Research

- Design Bias: It happens when the preferences and prejudices of the researcher influence the research design, survey questions and research method.
- Selection or Participant Bias: It happens when the researcher deliberately exclude some part of the population/ sample to arrive at a desired result. In quantitative research, this type of bias occurs when the researcher selects a specific research population ignoring others that might have alternative experiences.

Qualitative Research

- Bias from the Moderator: The moderator or interviewer can be biased in the process of data collection. Various factors like the background of the researcher, his ideological leanings, expression, tone, and faith might impact his approach and his relation with the research participants.
- Biased Questions: Data collection in qualitative research is based on questions. The researcher must be careful while framing and presenting the questions. The questions form the primary basis of collection of information in the research. Biased questions might lead to invalid and unrealizable research findings.
- Biased Reporting: In qualitative research, biased reporting happens when factors like personal beliefs, ideologies, cultural and social perceptions, attitude, etc. impacts and alters the results of the research. In short, biased reporting means that the researcher ignores the actual responses of the respondents and he or she is governed by personal beliefs while analyzing the research data.

14.5 CURIOSITY TO KNOW THE UNKNOWN

The universe provides infinite possibilities for research. Curiosity is often regarded as one of the main habits (or qualities) of successful people. Inquisitiveness about the unknown leads to a continuous process of learning. Curiosity fuels the desire to discover the unknown and the untrodden paths that open new dimensions and enriches human knowledge. All the scientific inventions, discoveries and knowledge about different cultures and societies are the results of the curious mind of a researcher. Curiosity drives the researcher to work on the new and unknown that leads to creation of a new thing or to find out the undiscovered. Inquisitiveness helps the researcher to add to human knowledge and also to develop things

and ideas to positively impact human life and society. Curiosity motivates the researcher to think out of the box and pave new paths towards knowledge.

A true researcher is the one who would shun the oft repeated and conventional well tried and proven paths. Instead he or she would be motivated to discover or invent something new with his creativity and imagination. Curiosity and inquisitiveness would be the driving force behind all the efforts made by the researcher in his research.

14.6 DESIRE TO KNOW THE CAUSE AND EFFECT RELATIONSHIP OF SOCIAL PROBLEMS

"Cause and effect" is a relationship between events or things, where one is the result of the other or others. This is a combination of action and reaction. Something happens (a cause) that leads to an effect. In short, it is "relationship between two phenomena in which one phenomenon is the reason behind the other." (Salkind)

For example, eating too much food and no physical exercise makes a person gain weight. In this case eating too much food and no physical exercise is the cause and weight gain is the effect.

It is a human tendency to look for the reason why a particular thing happens. This is the study of cause and effect. Therefore, a proper understanding of the cause and effect relationship is required by a researcher for better perception of the concerned field of research.

Elwell's Glossary of Sociology (undated) defines causation as:

A 'cause and effect' relationship exists wherever a change in one variable (the independent variable) induces change in another (the dependent variable). Causal factors in sociology include individual motivation as well as many external influences on human behavior that often go unrecognized.

In case of social problems the cause and effect relationship is very complex in nature. In many occasions an event may be both the

cause and effect of another event. For this reason, it is very important that the researcher should be able to perceive the situation and the mechanism of the problem he is dealing with. The researcher must be able to comprehend and associate the findings or results with the correct uses. His approach to the problem must be objective and flexible so that he can accommodate all the complexities that arise in his study to arrive at a valid and correct conclusion. The researcher must also have problem solving abilities in order to associate and consider the cause and effect relationships with the social issues and problems the study deals with. This will result in a successful accomplishment of the project. In other words, the ability to understand the cause and effect in relation to the social issues will enable the researcher to anticipate both the positive and negative possibilities.

14.7 DESIRE TO DISCOVER NEW AND TEST OLD SCIENTIFIC PROCEDURES

The desire to discover new and test old scientific procedures is one of the primary requirements to gain useful and fundamental knowledge. In fact the aim social science research is to discover new facts and verify and test old facts. The purpose of research is the application of scientific procedures to discover answers to the problems or questions.

There have been significant activities in the field of social research. Scientists have been making efforts in formulating and devising new methods and techniques while doing research in various fields of study. In addition to that, they are also testing the existing scientific procedures. This helps in revalidating or negating the authenticity of the procedures which in turn will help in making the old procedures more precise and acceptable. Social scientists, nowadays, use quantitative or statistical tools to make research results more definite, objective, and precise. The whole process can be summed up by the definition of social research forwarded by P. V. Young.

According to Young, social research is “a scientific undertaking are means of logical and systematized techniques seeks to:

(i) discover new facts or verify and test old facts;

(ii) analyze the sequences, inter-relationships and casual explanations which were derived within an appropriate theoretical frame of references; and

(iii) develop new scientific tools; concepts and theories which would facilitate reliable and valid study of human behavior.

14.8 PATIENCE AND AN ALERT MIND

Patience is a virtue that a researcher must possess. Research involves a significant expanse of time. A true researcher will continue his or her research when or where they would accomplish their project. It might take years to achieve the breakthrough they aimed at. If the researcher becomes impatient and losses the passion and persistence in the long run, then the whole endeavour would result in futility. Patience, however, does not mean slowing down but it is actually working steadily without getting frustrated. Research, whether scientific or social, requires the scholar to be calm and composed and continue his or her work by abjuring recklessness and disappointments. The researcher must also be ready to face disappointments, delays by external factors, and other challenges that are beyond his control. The researcher must possess the ability and determination to face the challenges in his or her life and must be ready to face failures (which may be numerous). But he/she must not lose hope. The most important quality the researcher must have is to stay focused on his research work.

Another important quality that a researcher must possess is to have an alert mind. Some of the synonyms of ALERTNESS are awareness, readiness, vigilance, watchfulness, sharpness, attentiveness, and wakefulness. The synonyms clearly suggest the

meaning of staying alert. Observation is a very important factor of research. The researcher must be cautious and alert throughout his research so that no significant aspect of the study should go unnoticed. Alertness in research begins with the selection of the field of study and continues till the results are accomplished. The researcher must be aware of all the factors and aspects associated with his field of study. This will enable him or her to avoid any pitfalls. He or she must also be alert enough to keep track of all the developments and happenings that are deeply or even remotely linked to his study.

For example, during data collection, the respondents might not respond to the queries properly. They might provide frivolous and incorrect responses that would not fulfil the object of the research. The researcher or interviewer must be alert enough to find out the real situation and perhaps reformulate the questions in order to get valid responses.

14.9 COURAGE AND CONSCIOUSNESS

One of the basic requirements of research is courage. Research involves adding something new to the existing knowledge. Hence, it is always a challenge. Sometimes research possess a challenge to the existing system or body of beliefs which is deeply ingrained in the society. In such a scenario, the researcher must not be afraid of unpleasant consequences of his/her findings. He or she must have the courage to record and deliver the truth. The researcher must not waver from his research and follow all the stipulated procedures with the aim to conclude his study, even though, the results may bring social disapproval or unpopularity. There had been many instances when researchers had to suffer a lot of disgrace and humiliation when their findings came in conflict with the established belief systems and dogmas.

The researcher must also be alert and conscious of all the factors associated with his research. His or her main aim is to discover the truth and nothing else.



CHECK YOUR PROGRESS

Q1: Why maintaining objectivity is a difficult task in social research? (Answer in about 20 words)

.....

.....

Q2: What happen, if a researcher is biased ?
(Answer in about 20 words)

.....

.....

Q3: Why is curiosity important for a researcher?
(Answer in about 60 words)

.....

.....

Q4: What is meant by cause and effect relationship?
(Answer in about 60 words)

.....

.....

Q5: State the importance of Patience in research work.
(Answer in about 60 words)

.....

.....



14.10 LET US SUM UP

- Research is an inseparable part of human knowledge. From the discovery of the power and uses of fire, the invention of wheel to the present age of internet, human civilization is based on research. Therefore, the importance of research cannot be denied.

- Research involves a number of qualities that a scholar must possess apart from being aware of the technical aspects associated with the field of study.
- Maintaining objectivity in research is a primary requirement. Research must be undertaken in such a manner that it is free from all sorts of distortions caused by personal feelings, ideals, beliefs or prejudices.
- A biased approach will diminish significantly the authenticity and validity of the research work. Biased approach in research impacts all the aspects of research including the findings, thereby, denying the actual spirit of academic research.
- Inquisitiveness or curiosity is driving force behind man's quest for knowledge. All discoveries and inventions in all fields of human knowledge is the result of human curiosity to know the unknown.
- The basis of the development of human civilization is the desire of the human beings to unravel all the mysteries they are aware of.
- Research, in a way, is the study of the cause and effect relationship. Social research involves certain complexities while understanding the cause and effect relationships the researcher must be able to correctly identify and understand the cause and effect in order to anticipate all kind of possibilities.
- Research involves the discovery of new scientific methods and procedures. These help to arrive at more effective and authentic conclusions.
- Research also involves the testing of the old scientific procedures in order to revalidate the existing knowledge and belief or in some cases denying them as they were proved to be invalid.
- Patience and alertness are the most important virtues of a researcher. Research, usually, is a time consuming process.

The researcher must be patient and he must also work steadily in order to accomplish his goal. An alert mind would help the researcher keep track of all the developments and aspects concerned with his research. These qualities will enable a researcher go ahead with his research in an objective and steady manner.

- Research needs courage. Many times research poses a challenge to existing beliefs and dogmas. History has many examples of researchers be punished for challenging these beliefs. But, the primary aim of research is to unravel the truth. Courage with a conscious approach enables a research in his quest for truth and contribute to human advancement.
- As the civilization progresses, many more new dimensions will be added to the process of research. Scientific and social advancements are the results of research. Research will help civilization to progress and civilization will provide more scope to the quest for knowledge.



14.11 FURTHER READING

1) Salkind, N. J. (2010). *Encyclopedia of research design.* Thousand Oaks, Calif: SAGE Publications.
2) Kothari, C.R. (2004). *Research methodology : Methods and techniques* (2nd revised edition). New Delhi: New Age International (P) Limited
3) https://www.editage.com/insights/7-biases-to-avoid-in-qualitative-research
4) https://theconversation.com/how-to-find-the-knowns-and-unknowns-in-any-research-26338
5) https://open.lib.umn.edu/socialproblems/chapter/1-4-doing-research-on-social-problems/

14.12. ANSWER TO CHECK YOUR PROGRESS

Answer to Q1: Maintaining objectivity is a difficult task in social research as it involves human beings as the principal subject matter.

Answer to Q2: When a researcher is biased, the research process is hampered and the researcher cannot have an unprejudiced approach to the problem.

Answer to Q3: Curiosity fuels the desire to discover the unknown and the untrodden paths that open new dimensions and enriches human knowledge. All the scientific inventions, discoveries and knowledge about different cultures and societies are the results of the curious mind of a researcher.

Answer to Q4: "Cause and effect" is a relationship between events or things, where one is the result of the other or others. This is a combination of action and reaction. Something happens (a cause) that leads to an effect. In short, it is "relationship between two phenomena in which one phenomenon is the reason behind the other."

Answer to Q5: Patience is a virtue that a researcher must possess. Research involves a significant expanse of time. A true researcher will continue his or her research when or where they would accomplish their project. It might take years to achieve the breakthrough they aimed at. If the researcher becomes impatient and losses the passion and persistence in the long run, then the whole endeavour would result in futility.

14.13 MODEL QUESTIONS

A) SHORT QUESTIONS: (Answer each question in about 150 words)

Q1: Why maintaining objectivity is necessary in research?

Q2: What is meant by unbaised research?

Q3: Describe the different types of research bias?

Q4: What is meant by cause and effect relationship?

B) LONG QUESTIONS:

(Answer each question in about 300-500 words)

Q1: Discuss the importance of patience and courage in research.

Q2: Briefly explain the good qualities of a researcher.

Q3: The important qualitiy that a researcher must possess is to have an alert mind. Explain.

*** ***** ***

UNIT 15 : THE RESEARCH REPORT

UNIT STRUCTURE

15.1 Learning Objectives
15.2 Introduction
15.3 Research Report: Meaning and Significance
15.4 Different Steps in Writing the Report
15.5 Contents of a Research Report
15.6 Precautions in Writing Research Reports
15.7 Research Report Writing
15.8 Concept of Research Proposal and Research Abstract
15.9 Let Us Sum Up
15.10 Further Reading
15.11 Answers to Check Your Progress
15.12 Model Questions

15.1 LEARNING OBJECTIVES

After going through this unit, you will be able to -

- understand the meaning of research report
- know about the contents of the research report
- identify the different steps in writing the research report
- understand the qualities of a good research report

15.2 INTRODUCTION

Research Report, though the last, but is a very important part of research work. Any research work would remain futile if its findings are not documented. In the previous units, we have learned about the meaning of research, steps in conducting research, techniques and methods of data collection, analysis and presenting of data. In this unit we shall learn about and understand the meaning and

importance of research report, contents of research report, steps involved in writing a research report and also qualities of a research report.

15.3 RESEARCH REPORT: MEANING AND SIGNIFICANCE

By now we have learned about the meaning of research, techniques of data collection and also the various methods of analysing data. But Research doesn't simply end with all these. A research work becomes meaningful only when the findings of the research are systematically written/documented and presented. Analysis of data becomes futile and unrewarding if the research report is not written. Report writing is therefore very crucial to research. Though it is the last activity in a research work, yet it is of immense significance. The research work becomes meaningful and can be popularised only when the research findings are systematically assembled and written down with a valid conclusion. Research report writing is thus the culmination of the research investigation. It is a brief statement of most significant facts that are necessary for understanding the generalisations drawn by the researcher. Research report is prepared on the basis of the findings and the arguments drawn from it by the researcher with the intention of transmitting the information gathered to the wider and interested public. Efforts are made by the researcher to convey through his/her research study the entire result of the study in sufficient detail and so arranged as to enable each reader to comprehend the date and to determine for himself/herself the validity of the conclusions. Report writing not only enables mere dissipation of knowledge but also helps in further research in similar field.

15.4 DIFFERENT STEPS IN WRITING THE REPORT

Research reports are the final outcome of a lengthy process of collecting, analysing and chronologically arranging data and then finally transforming data into views and arguments. It therefore takes a long time to assemble the data or the findings and prepare a report. Several steps are involved in the process of preparing a report- a) logical analysis of the subject matter b) preparation of the final outline c) preparation of the rough draft d) rewriting and polishing e) preparation of the final bibliography f) writing the final draft. Let us now try to understand each one of them.

- ***Logical analysis of the subject matter:*** It is the first and a time consuming step in preparing the research report. There are two ways to develop the subject matter or the report: a) Logically and b) Chronologically. In the logical development, the researcher gives all his/her effort in establishing the logical connections and links between the findings of the study. He/she tries to relate one thing with the other by means of analysis. Secondly the researcher also tries to arrange the findings chronologically i.e based on a connection or sequence in time or occurrence. For example if a researcher is conducting research on some historical event, he shall first try to connect various incidences related to the event with one another and establish a logical connection among them. Secondly he/she shall try to arrange the incidences chronologically i.e which incident first took place and what fallowed after that. Arranging the data or findings of the study in the logical and chronological sequence gives clarity to the study and also depicts the analytical ability of the researcher.
- ***Preparation of the final outline:*** The step which follows the arrangement of the research findings in logical and

chronological order is preparing the outline or the structure of the report. The preparation of outline of the report depends upon the judgement of the researcher. But it is important clearly explain the problem with which he/she was concerned, the method he/she applied to gather the findings, the analysis and conclusions and the bases for his conclusion. One the basis of the outline, the researcher would write down the report. Outline is therefore very important as it delineates the themes and sub themes to be covered. They are an aid to the logical organisation of the material and a reminder of the points to be stressed in the report. Without an outline, there is a possibility that the report may get jumbled up; it may not have a logical sequence or important points may be missed by the researcher.

- ***Preparation of the rough draft:*** This follows the logical analysis of the subject and the preparation of the final outline. This is perhaps the most time consuming of all the steps as the researcher now begins to assemble his findings and thoughts from the research study and write down the report. What he/she has done; what methodology was adopted; what were the challenges; what are the findings, his/her views, arguments, generalizations and also suggestions are to be written down systematically. Since every relevant aspect of the research has to be incorporated, it takes much time of the researcher to put his/her thoughts into words
- ***Rewriting and polishing the rough draft:*** This is also one of the most difficult tasks in report writing and takes a longer time. In fact this step requires same time as that the writing of the rough draft. After the rough draft has been prepared, the researcher has to then carefully review it. While reviewing he/she would come across the weaknesses and flaws of the rough draft which shall have to be modified. Lot of rewriting,

modification, and polishing of the rough draft takes place in this step to make the report flawless. Omissions and Errors of the rough draft including mistakes in grammar, spelling, sentence construction so on and so forth are rectified to bring the report to perfection.

- ***Preparing the bibliography:*** After the report has been brushed up; polished and final shape has been given to it, the next task of the researcher is to prepare the bibliography. The bibliography, which is generally appended to the research report, is a list of books in some way pertinent to the research which has been done. The researcher should carefully list the names of the books which he/she has consulted for the study in a proper instructed referencing style. This too is a very crucial stage because if the names of the books from which the researcher has taken ideas and materials from are left behind that he/she might have to face legal consequences of plagiarism.

15.5 CONTENTS OF THE REPORT

By now we know that once the research study is completed, the researcher has to write a report without doing which his/her entire research would be a meaningless effort. Now, in this section we shall learn about the contents of the research report i.e. what does the research report contain. A reader who is reading the report must be communicated about the entire research study so that he/she can judge the adequacy of its methods and thus form an opinion of how seriously the findings are to be taken. Thus the report has to be written systematically for which a proper layout is necessary. The layout of the report means as to what the research report should contain. A comprehensive layout of the research report should contain a) Preliminary pages b) Main Body c) the Reference Section

PRELIMINARY SECTION: The preliminary section will vary according to the type of research one has undertaken. The preliminary section of the report usually comprise of the cover page indicating the Title, purpose of Research, Institutional Affiliation so on and so forth. The next page is usually the certificate page. Certificate page is usually found in dissertations and theses. It may not be required in other type of research reports. It is followed by acknowledgement page wherein the researcher offers gratitude to all those who have helped him/her to successfully complete the study. After this page there is the Content page which lists the details of contents in the report along with page numbers. The list of contents followed by list of tables and illustrations so that the decision-maker or anybody interested in reading the report can easily locate the required information in the report.

MAIN BODY: The main body of the research is the key component of the research report wherein the entire study and its findings are written down in a systematic manner. The main body is divided into the following subsections:

- ***Introduction:*** The purpose of the introduction is to introduce the research project to the readers. In the introduction, which is the opening section of the main body of the report, the researcher must state the reasons behind taking up that particular study.. The report should clearly explain the Background of the study i.e. what the study is about and why the study is important Depending on the particular piece of research, the review of the literature is introduced. This division will show what is known already and how our research will fill a gap in knowledge or replicate earlier work. Further, the Objectives of the study, the Research Questions; Methods of data collection and analysis, and the theoretical Framework also needs to be described in the report. The researcher should give detailed elaboration of each of these so that the readers

can easily comprehend about the study and judge its merit. Further, the scope of the study should be stated and the various limitations, under which the research project was completed, must also be narrated.

- ***Statement of the findings:*** This generally comprises the main body of the report, extending over several chapters.. After introducing the study in the introductory chapter, a detailed description of the findings of the study supported by data's, tables and figures should be given. Here the main task of the researcher is to convert the raw data derived from the study into meaningful statement and arguments. All relevant findings should be incorporated in the report and it should be written down systematically following a logical sequence. The analysis of the findings must also be clearly explained. The research report must state the findings of the study in a lucid, intelligible language.
- ***Implications of the results:*** Towards the end of the main text, the researcher should again put down the results of his researcher clearly and precisely. He/She should illustrate the implications that flow from the results of the study, for the general reader is interested in the implications for understanding the human behaviour. Such implications may have three aspects as stated below:

➢ A statement of the inferences drawn from the present study which may be expected to apply in similar circumstances

➢ The conditions of the present study which may limit the extent of legitimate generalizations of the inferences drawn from the study

➢ The relevant questions that still remain unanswered or new questions raised by the study along with suggestions for the kind of research that would provide answers for them.

- ***Conclusion:*** The report remains incomplete without a conclusion. Hence the researcher should conclude the study by summarising his/her entire study, its key findings also major arguments and generalisations. The researcher should also state the recommendations and suggestions if any in the conclusion. He/she also try to indicate the possibility and usefulness of future research in that particular field.

END MATTER: After completing the main text, the final task of the researcher, as we have discussed above, is to write down the bibliography. The researcher must carefully list down the books, articles, reports, government documents and all other secondary sources which he/she has referred to in the study. Apart from the bibliography, the researcher must also attach questionnaires, sample information, mathematical derivations, and the like ones. Index should invariably be given at the end of the report. The value of index lies in the fact that it works as a guide to the reader for the contents in the report. Photographs relating to the study, if any, should also be incorporated towards the end of the report with appropriate illustrations or captions

15.6 PRECAUTIONS IN WRITING RESEARCH REPORTS

Apart from the contents of the report, there are several other technicalities involved in preparing the research report. There are definite and set rules which should be followed in the actual preparation of the research report or paper. The report must follow the instructions relating to Font size, Font Style, Line Spacing, Alignment, Margin, so on and so forth. Let us now try to understand some of the important technicalities that a researcher should confirm to while preparing the research report.

- **Size and Physical design:** The report should be on unruled, neat and legible paper. Equal margin must be kept on both

sides of the paper. The format relating to font size, font style, line spacing, so on and so forth should be followed as per instruction.

- **Layout:** Keeping in view the objective and nature of the problem, the layout of the report should be thought of and decided and accordingly adopted.
- **Quotations:** Quotations should be placed in quotation marks and double spaced, forming an immediate part of the text. But if a quotation is of a considerable length then it should be single – spaced and indented at least half an inch to the right of the normal text margin.
- **The Footnotes :** Footnotes are placed at the bottom of the page on which the reference or quotation which they identify or supplement ends. Footnotes are customarily separated from the textual material by a space of half an inch and a line about one and a half inches long. Footnotes should be numbered consecutively, usually beginning with 1 in each chapter separately. The number should be put slightly above the line, say at the end of a quotation. At the foot of the page, again, the footnote number should be used to correlate the reference in the text with its corresponding note at the bottom of the page, except in case of statistical tables and other numerical material, where symbols such as the asterisk (*) or the like one may be used to prevent confusion.
- Footnotes are always typed in single space though they are divided from one another by double space.
- Subsections must be clearly highlighted.
- In-text Citations should be properly given.
- Bibliography should be written in alphabetical order and must follow the instructed referencing style like APA etc..



CHECK YOUR PROGRESS

Q1: What is meant by a research report? (Answer in about 30 words)

.....

.....

Q2: What is Bibliography? (Answer in about 30 words)

.....

.....

Q3: True or False:

a) Report writing is the first step in research

b) Acknowledgment page comprise of the lists of contents of the research report.

15.7 RESEARCH REPORT WRITING

The research report is indispensible if the research is to be meaningful. There is a great temptation to rush into writing because this gives a feeling of something accomplished and indeed it is true that certain sections of the report can sometimes be written in draft before the research is complete. But an illegible and confusing report gives a very bad impression about the researcher and his/her research study. Therefore the researcher should very cautiously write the research report incorporating all the relevant findings of his study in a systematic manner. He/she should use simple, uncomplicated language while writing the report. The report should be easily understandable and should represent the knowledge and wisdom of the researcher. Elaborations and illustrations must be provided wherever necessary. To be acknowledged and appreciated as a good research report, a report must have the following basic and essential qualities.

- **Make it easy to follow:** The report must be written in a way that it becomes easily understandable. Long paragraphs

should be avoided and the paragraphs should focus only on one particular point or theme.. Trying to incorporate multiple points, ideas or themes in one particular paragraph may make it confusing and difficult to understand. The findings must be incorporated systematically and the paragraphs must follow a logical sequence.

- **Make it clear:** Clarity in writing is a quality of communication. A good research report must have clarity. The researcher must be able to clearly communicate his ideas and thoughts to the readers. Further accuracy is also an essential quality in research report.
- **Use of Simple Sentences:** Sentences should be simple and short. It is better to use simple and precise sentences rather than long and confusing statements. The sentences should proceed methodically and should be effortlessly comprehensible.
- **Use of Non-technical language:** Avoid the use of technical terms; aim the report at the experience level of the common reader. . If it is necessary to use technical terms, a brief description or explanation should be included in the report.
- **Make it brief:** The report should be long enough to cover the objects of the study. Highlight major points by stressing the big issues and taking them up first.
- **Stress Practical actions:** Use analysis, specific examples, or compare drawn from experiences familiar to the reader. These show that the researcher has taken a realistic point of view.

15.8 CONCEPT OF RESEARCH PROPOSAL AND RESEARCH ABSTRACT

Research Proposal : After extensive review of literature on a proposed area/ issue for study, a research gap is recognized and a

research problem is identified to be studied. Research proposal consists of the overall idea and plan for the study to be conducted. It includes introduction to the problem, rationale/ importance of the study, research questions, objectives of the study and methodology to be used.

Research Abstract : It is a brief summary of a research study conducted that includes introduction to the problem/ research topic, importance of the study, objectives and a brief explanation of the expected outcomes of the study. Research abstracts, within one or two paragraphs gives an overall idea about the study conducted regarding a particular research problem, without discussion of the results in detail



15.9 LET US SUM UP

- A research work becomes meaningful only when the findings of the research are systematically written/documented and presented.
- Research report writing is the culmination of the research investigation. It is a brief statement of most significant facts that are necessary for understanding the generalisations drawn by the researcher.
- Research reports are the final outcome of a lengthy process of collecting, analysing and chronologically arranging data and then finally transforming data into views and arguments.
- Several steps are involved in the process of preparing a report- a) logical analysis of the subject matter b) preparation of the final outline c) preparation of the rough draft d) rewriting and polishing e) preparation of the final bibliography f) writing the final draft
- A comprehensive research report should contain a) Preliminary pages b) Main Body c) the Reference Section

- Apart from the contents of the report, there are several other technicalities involved in preparing the research report. The researcher must follow certain rules and instructions relating to the layout, citation style, and font style so on and so forth while preparing the research report.
- There are several qualities of a good research report. Some of the important qualities are: clarity, accuracy, preciseness, use of simple sentences, so on and so forth.



15.10 FURTHER READINGS

1) Guthrie, Gerard.(2010). *Basic Research Methods: An Entry to Social Science Research*. New Delhi: Sage publication. India.

2) Kothari, C. R. (2004). *Research Methodology – Methods and Techniques*, New Age International (P) Ltd, New Delhi.



15.11 ANSWERS TO CHECK YOUR PROGRESS

Answer to Q No.1: Research report writing is the culmination of the research investigation. It is a brief statement of most significant facts that are necessary for understanding the generalisations drawn by the researcher

Answer to Q No.2: Bibliography is the lists of books, articles, reports, government documents, and all other secondary sources that the researcher has consuted and reffered to in his/her study.

Answer to Q No.3: a) False b) False

15.12 MODEL QUESTIONS

A) Short Questions (Answer each question in about 150 words)

Q1: Briefly explain the different technicalities involved in writing the research report.

Q2 : Discuss the Qualities of a good research report.

B) Long Questions (Answer each question in about 300-500 words)

Q1 : Discuss the different steps involved in writing a research report.

Q2 : Explain the main components of a research report.

*** ***** ***



REFERENCES (FOR ALL UNITS)

1. Babbie, E. 2001. *The practice of social research* (9th ed.). Belmont, CA: Wadsworth.
2. Bachman, R. & Schutt, R. 2007. *The practice of research in criminology and criminal justice* (3rd ed.). Thousand Oaks, CA: Sage.
3. Corley, K. & Gioia, D. 2011. Building theory about theory building: What constitutes a theoretical contribution? *Academy of Management Review*, *36*, 12-32.
4. Cozby, P. 2009. *Methods in behavioral research*. Boston: McGraw Hill Higher Education.
5. C.R. Kothari 2004 Research Methodology Methods and Techniques, 2nd Edition: New Delhi, New Age International Publisher
6. Dubin, R. (1978). *Theory building* (Rev. ed.). NY: Free Press
7. Guthrie Gerard, *Basic Research Methods: An Entry to Social Science Research*. New Delhi: Sage publication. India. 2010.
8. Gupta S.P. 2010. Statistical Methods. New Delhi: Sultan Chand & Sons.
9. Gupta S.P. 2010. Statistical Method. New Delhi: Sultan.
10. Goode, W. J and Hatt, P.K. 2006. Methods in Social Research. Delhi: Surajeet Publication.
11. Gelso, C. (2006). Applying theories to research: The interplay of theory and research in science. In F.T. Leong & J.T. Austin (Eds.), *The psychology research handbook*. Thousand Oaks, CA: Sage. Retreived from http://www.sage-ereference.com/hdbk_psychrsch/Article_n32

12. Hay, J. & Lee, S. (2009). Theory building through qualitative research: Marshalling opportunities to advance cancer screening efforts. *Health Education & Behavior*, *36*, 145-149. doi: 10.1177/10908109340515
13. Heinen, J. (1985). A primer on psychological theory. *The Journal of Psychology*, *119*, 413-421.
14. Kerlinger, R. (1986). *Foundations of behavioral research*. NY: Holt, Rinehart, & Winston.
15. Jahoda, Marie, Morton Deutsch, and Stuart W. Cook. 1951. Research Methods in Social Relations. New York: Dryden.
16. Kothari C.R. *Research Methods and Techniques.* New Delhi: New Age International Publisher. 2004
17. Lincoln, Y., & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences. In N. Denzin &Y. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed.) (pp. 163–188). London: Sage.
18. Lynham, S. (2002). The general method of theory-building research in applied disciplines. *Advances in Developing Human Resources*, *4*, 221-241.
19. Locke, E. (2007). The case for inductive theory building. *Journal of Management*, *33*, 867-890. doi: 10.1177/0149206307307636
20. L. Choudhury 2000 An introduction to Statistics, Volume One: Guwahati Kitapghar
21. Medhi, J. 2006. Statistical Methods : An Introductory text. New Age International (p) Ltd.
22. Morrow, R. (1994). *Critical theory and methodology*. Thousand Oaks, CA: Sage.
23. Quinney, R. (1970). *The social reality of crime*. Boston: Little, Brown.
24. Rynes, S. & Gephart, R. (2004). From the editors: Qualitative research and the academy of management journal. *Academy of Management Journal*, *47*, 454-462.

25. Stinchcombe, A. (1968). *Constructing social theories*. Chicago: University of Chicago Press.
26. Sutton, R. & Staw, B. (1995). What theory is not. *Administrative Science Quarterly*, *40*, 371–384.
27. Smith, J., Bekker, H., & Cheater, F. (2011). Theoretical versus pragmatic design in qualitative research. *Nurse Researcher*, *18*, 39-51.
28. Southern, S. & Devlin, J. (2010). Theory development: A bridge between practice and research. *The Family Journal: Counseling and Therapy for Couples and Familes*, *18*, 84-87.doi: 10.1177/1066480709358422
29. Thurstone, L. L, .and E. J. Chave. 1929. The Measurement of Attitudes Chicago: University of Chicago Press.
30. Torraco, R. (2002). Research methods for theory building in applied disciplines: A comparative analysis. *Advances in Developing Human Resources*, *4*, 355-376.
31. Young, P.V. *Scientific Social Surveys and Research*. New Delhi: Prentice Hall, 2009
32. Young, P. V. 1992. Scientific social surveys and research, New Delhi: Prentice Hall of India.
33. Zikmund, W., Babin, B., Carr, J., & Griffin, M. (2008). *Business research methods* (8th ed.). Mason, OH: Cengage.