# 1

# QUANTITATIVE RESEARCH – NATURE, CHARACTERISTICS, SIGNIFICANCE, CRITIQUE

**Unit structure**:

1.0 Objectives

1.1 Introduction

1.2 Nature

1.3 Characteristics

1.4 Significance

1.5 Critique

1.6 Summary

1.7 Questions

1.8 References and Further Readings

## 1.0. OBJECTIVES:

- ➢ To familiarize students with nature of quantitative research
- ➢ To explain its characteristics and significance along with its critiques

## 1.1 INTRODUCTION:

Quantitative research is a systematic approach to collect information via samplingmethods, for instance, questionnaires, online polls and online surveys. It is gathered from both potential and existing subjects and depicted in terms of numerics.

Quantitative research is generally used in fields like political science, gender studies,community health, marketing, sociology, economics, psychology, demography, andeducation. Its objective is to employ mathematical theories in relation to phenomena. The process offers a connection between mathematical expression and empirical observation.

Quantitative research is a method to measure variables, analyze them and reportrelationships amongst the studied variables through a numerical system. Its objective is to understand, analyze, describe and make future predictions or to make suitable changes. It deals in objective, logic, and

numbers and puts its focus on convergent reasoning and detailed and unchanging data.

The data in quantitative research is collected through structured research, and the results are based on a larger size of samples that represents or relects the population. An important fact about this kind of research is that it can be repeated and replicated. The quantitative researcher uses several tools to gather numerical data that is in the form of statistics and numbers and is arranged in non-textual forms like figures, charts, and tables.

Some of the most common quantitative research's data collection methods are described here.

### Cross-sectional studies

Cross-sectional studies are surveys undertaken at one point in time, rather like a photo taken by a camera. If the same or similar survey is repeated, we can get good measures of how society is changing. It is a type of research design in which data is collected from many different subject at a given point in time.

### Longitudinal studies

Longitudinal studies follow the same respondents over an extended period of time. They can employ both qualitative and quantitative research methods, and they follow the same group of people over time.

### Opinion polls

An opinion poll is a form of survey designed to measure the opinions of a target population about an issue, such as support for political parties and views about crime and justice, the economy or the environment.

### Questionnaires

Questionnaires collect data in a standardized way, so that useful summaries can be made about large groups of respondents, such as the proportion of all young people of a given age who are bullied. Usually most questions are 'closed response', where respondents are given a range of options to choose from. Researchers have to be careful that the questions are not 'leading', that the options are comprehensive (they cover every possible answer) and are mutually exclusive, so that only one answer is correct for any respondent.

### Social attitude surveys

Social attitude surveys ask more general questions about beliefs and behaviour, for example, how often people go to church, how much trust they have in the police force, whether they think children need a strict upbringing, how content they are with their life, how often they see other family members, and whether they are in employment.

**Surveys and censuses**

A census is a survey of everyone in the population. Because of the vast number of respondents, they are very expensive to organise. Governments now depend much more on sample surveys and administrative records, for example those created by a stay in hospital or tax returns. Surveys use a questionnaire to investigate respondents in a sample. Samples are chosen in such a way that they can represent a much larger population. A precise calculation can be made of how accurate the information from any sample is likely to be.

## 1.2 NATURE:

Using a deductive approach, quantitative research seeks to establish facts, make predictions, and test hypotheses that have already been stated. A large part of the data analysis of quantitative research is statistical, striving to show that the world can be looked at in terms of one reality; this reality, when isolated in context, can be measured and understood, a perspective known as positivism (Gay & Airasian, 1999).

There are some features of quantitative research that are inherently necessary for this approach. A quantitative researcher must: state both the hypothesis studied and the research procedures that will be implemented prior to conducting the study, maintain control over contextual factors that might interfere with the data collected, use large enough samples of participants to provide statistically meaningful data, and employ data analyses that rely on statistical procedures.

Quantitative Research is a systematic investigation of phenomena by gathering quantifiable data and performing statistical, mathematical, or computational techniques. It collects information from existing and potential respondents using sampling methods, and surveys, or questionnaires; the results of which can be depicted in the form of numerical values (Bhat, 2020). Its purpose is to generate conclusion or to make some inferential conclusion by trying to quantify the problem and understand how prevalent it is by looking for projectable results to a larger population.

Quantitative research is widely used in psychology, economics, demography, marketing, political science, and educational studies. Depending on the nature of the study, a researcher can use any of the following four main types of quantitative research, namely: Descriptive Research, Correlational Research, Causal-Comparative Research or Quasi-Experimental Research, and Experimental Research.

**Check your progress:**

1. Elaborate on the nature of quantitative social research.

## 1.3 CHARACTERISTICS:

The goal in conducting a quantitative research is to determine the relationship of one thing to another within a population. It is descriptive, correlational, quasi-experimental or experimental in nature. Descriptive and correlation researches establish associations between things under study while quasi-experimental and experimental studies establish causality.

As researchers, we should know the characteristics of a quantitative research listed below (Spalding University, 2020):

**1. The data is usually gathered using structured research instruments.**

Before conducting a data gathering, it is a must for a researcher to have his or her research instrument validated by experts. This validation process for the structured research instruments is necessary to ensure the reliability and validity of the results.

**2. The results are based on larger sample sizes that are representative of the population.**

It is necessary for a researcher to correctly determine the number of respondents in the research to be conducted in order to absolutely represent the population.

**3. The research study can usually be replicated or repeated, given its high reliability.**

Between qualitative and quantitative research, the latter is easier to replicate than the former. Since, quantitative research uses a structured research instrument, and deals with numbers and systematic procedure; it is highly replicative in nature. The procedure used in a quantitative research can be repeatedly done to measure the validity of previous results.

**4. Researcher has a clearly defined research question to which objective answers are sought.**

This research questions serve as the backbone of the research. These are the bases of the structured research questions developed and validated.

**5. Data are in the form of numbers and statistics, often arranged in tables, charts, figures, or other non-textual forms.**

Unlike in qualitative research where data and responses are in text form, or non-numerical data, in quantitative research data are definitely numbers which are subject to statistical treatment to interpret and generate conclusion.

**6. Project can be used to generalize concepts more widely, predict future results, or investigate causal relationships.**

As mentioned, since the results of a quantitative research are based on larger sample sizes hence it can be used to make inference to a population.

**7. Researcher uses tools, such as questionnaires or computer software, to collect numerical data.**

Examples of these are questionnaires, survey forms, and alike which are validated prior the conduct of the study.

Now, when reporting the results of a quantitative research conducted, it is necessary to a researcher to remember some tips. A researcher should explain the data collected how they are collected and how they are treated. It is also advisable to include all relevant results to the research questions under study. It is a must also to report all the procedure used in data collection to establish among readers the validity and reliability of results. The trustworthiness of the data is the goal. Statistical treatment must also be appropriate and be explained carefully. When using table or any non-textual presentation, make sure to present it with clarity.

**Check your progress:**

1. What are the characteristic features of quantitative social research?

## 1.4 SIGNIFICANCE:

The importance of quantitative research is that it offers tremendous help in studying samples and populations. It discusses in detail relevant questions, for instance, where didthe data come from, where are the existing gaps in the data, how robust is it and whatwere the exclusions within the data research. It is vital to note the process for their selection and describe the methods and tools that are being used by the researcher to collect the information.

There are four main types of research questions that quantitative research is particularly suited to finding an answer to:

1. The first type of research question is that demanding a quantitative answer. Examples are: 'How many students choose to study education?' or 'How many math teachers do we need and how many have we got in our school district?' That we need to use quantitative research to answer this kind of question is obvious. Qualitative, non-numerical methods will obviously not provide us with the (numerical) answer we want.

2. Numerical change can likewise accurately be studied only by using quantitative methods. Are the numbers of students in our university rising or falling? Is achievement going up or down? We'll need to do a quantitative study to find out.

3. As well as wanting to find out about the state of something or other, we often want to explain phenomena. What factors predict the recruitment of maths teachers? What factors are related to changes in student achievement over time?

4. The final activity for which quantitative research is especially suited is the testing of hypotheses. We might want to explain something – for example, whether there is a relationship between the level of poverty and access to health and medicine.

It discusses in detail relevant questions, for instance, where did the data come from, where are the existing gaps in the data, how robust is it and what were the exclusions within the data research. It is vital to note the process for their selection and describe the methods and tools that are being used by the researcher to collect the information.

The quantitative research identities variables that are being measured, gives a detailed description of the applicable method that is used in obtaining relevant data, notes down important criteria about the fact that the data was already in existence or the researcher gathered himself. Since its mainly about the statistics, qualitative research is highly resourceful in gathering numerical data about any phenomena even in social science.

**Check your progress:**

1. Is quantitative social research significant? Why?

# 1.5 CRITIQUE:

According to Saunders et al. (2009), research methodology serves as the backbone of a research study. Quantitative research's main purpose is the quantification of the data. It allows generalisations of the results by measuring the views and responses of the sample population. Every research methodology consists two broad phases namely planning and execution (Younus 2014). Therefore, it is evident that within these two phases, there likely to have limitations which are beyond our control (Simon 2011).

Following are major criticisms/limitations of quantitative social research:

**1. Improper representation of the target population**

Improper representation of the target population might hinder the researcher for achieving its desired aims and objectives. Despite of applying appropriate sampling plan representation of the subjects is dependent on the probability distribution of observed data. This may lead to miscalculation of probability distribution and lead to falsity in proposition.

**2. Lack of resources for data collection**

Quantitative research methodology usually requires a large sample size. However due to the lack of resources this large-scale research becomes impossible. In many developing countries, interested parties (e.g., government or non-government organisations, public service providers, educational institutions, etc.) may lack knowledge and especially the resources needed to conduct a thorough quantitative research.

**3. Inability to control the environment**

Sometimes researchers face problems to control the environment where the respondents provide answers to the questions in the survey (Baxter 2008). Responses often depend on particular time which again is dependent on the conditions occurring during that particular time frame.

**4. Limited outcomes**

Quantitative research method involves structured questionnaire with close ended questions. It leads to limited outcomes outlined in the research proposal. So, the results cannot always represent the actual occurring, in a generalised form. Also, the respondents have limited options of responses, based on the selection made by the researcher.

**5. Expensive and time consuming**

Quantitative research is difficult, expensive and requires a lot of time to be perform the analysis. This type of research is planned carefully in order to ensure complete randomization and correct designation of **control groups** (Morgan 1980). A large proportion of respondents is appropriate for the representation of the target population. So, as to achieve in-depth responses on an issue, data collection in quantitative research methodology is often too expensive as against qualitative approach.

**6. Difficulty in data analysis**

Quantitative study requires extensive statistical analysis, which can be difficult to perform for researchers from non- statistical backgrounds. Statistical analysis is based on scientific discipline and hence difficult for non-mathematicians to perform.Quantitative research is a lot more complex for social sciences, education, anthropology and psychology. Effective response should depend on the research problem rather than just a simple yes or no response.

**Check your progress:**

1. What are the major criticism of quantitative research?

## 1.6 SUMMARY:

Quantitative research is defined as a systematic investigation of phenomena by gathering quantifiable data and performing statistical, mathematical, or computational techniques. Quantitative research collects information from existing and potential customers using sampling methods and sending out online surveys, online polls, questionnaires, etc., the results of which can be depicted in the form of numerical. After careful understanding of these numbers to predict the future of a product or service and make changes accordingly.

Quantitative outcome research is mostly conducted in the social sciences using the statistical methods used above to collect quantitative data from the research study. In this research method, researchers and statisticians

deploy mathematical frameworks and theories that pertain to the quantity under question.

Quantitative research templates are objective, elaborate, and many times, even investigational. The results achieved from this research method are logical, statistical, and unbiased. Data collection happened using a structured method and conducted on larger samples that represent the entire population.

Quantitative research is often placed in opposition to qualitative research.In many cases, this turns into a 'paradigm war', which is seen to result fromapparently incompatible world views underlying the methods. When youlook closer at researchers' actual beliefs, it appears that the so-called subjectivist(qualitative) versus realist (quantitative) divide is not that clear-cut.

## 1.7 QUESTIONS:

- What is quantitative research in social sciences?
- List the important characteristics of quantitative research.
- Which methods are used for data collection in quantitative research?

## 1.8 REFERENCES AND FURTHER READINGS

- Singh, K. (2007). *Quantitative social research methods*. SAGE Publications India Pvt Ltd https://dx.doi.org/10.4135/9789351507741.

- Black, T. R. 1999. Introduction to Research Design. In: Black, T. R. (eds). *Doing Quantitative Research in the Social Sciences*. London: Sage.
- Burns, R. 2000. Introduction to Research Methods. London: Sage.
- Hughes, C. 2006. *Qualitative and Quantitative Approaches to Social Research*. UK: University of Warwick.
- Punch, K. 1998. Introduction to Social Research: Quantitative and Qualitative Approaches. London: Sage.
- Tashakkori, A. and Teddlie, C. 2010. Integrating Qualitative and Quantitative Approaches to Research. In: Bickman, L. and Rog, D. J. (2nd ed). *The Sage handbook of Applied Social Research Methods*. UK: Sage.
- Tashakkori, A. and Teddlie, C. 2010. Integrating Qualitative and Quantitative Approaches to Research. In: Bickman, L. and Rog, D. J. (2nd ed). *The Sage handbook of Applied Social Research Methods*. UK: Sage.

❖❖❖❖

<div align="right">**2**</div>

# TYPES OF DATA- PRIMARY AND SECONDARY, SMALL AND BIG

**Unit Structure**

## 2.0 OBJECTIVES

- To learn about the sources of Data collection – Primary and Secondary
- To learn about Small and Big Data.

## 2.1 INTRODUCTION

In this chapter, you will learn about Primary and Secondary, Small and Big. The concepts described in this chapter are some of the core concepts which are used in research methodology. In other words, they are the foundation for conducting research specially the concepts of Primary and Secondary data. Till the time you are doing research you will encounter these words again. Let it be your Master's program or Ph.D. program. Before understanding the types of Data let us first look into the meaning of Data.

## 2.2 MEANING OF DATA

According to the Cambridge dictionary the meaning of Data is Information gathered which can be facts or that of number. This information is further studied, analyzed, and utilised to aid in the decision-

<div align="right">9</div>

making. Information can also be in digital form that can be saved and accessed by using a computer[i].

## 2.3 IMPORTANCE OF DATA

Today's several companies work only on just individuals' data and thrive with their businesses on it. There are larger multinational companies too like Mc Kinsey which works on data of consumer market. If a company has to sell a product it needs data – as to what is the current demand, trends, age group, emotion factor through which a product could be sold etc. In the same way, in social science too, data is also important for understanding individuals, groups, societies. This is needed for several reasons like for implementing policies, to take enough measures. For example – During the Covid pandemic it was through data that states could figure out which areas had more covid cases and declare them as red zones and prevent outsiders from entering those areas or creating compulsory social distancing. The Sources of Data Collection has two parts namely Primary Data and Secondary Data. Let us learn about Primary Data first.

## 2.4 MEANING OF PRIMARY DATA -

Primary data is information that is created for the first time by the researcher via his or her own efforts and expertise, especially to understand the research problem in better way. Primary data is also known as raw data or first-hand data. Because primary data collecting is done by the organization or by individual himself or herself. At times this kind of data collection is carried out under certain supervision too. Especially when there is large scale data which is needed to be collected there is a team involved than just one individual. Often primary data is generated through interviews, Questionaires. However, there are other important documents too which are viewed as Primary Data for example - letters, emails, diaries, photographs, and daily schedules. Personal records include things like driver's licences and student ID cards. People may preserve a personal notes of events they have witnessed or participated in, which is another form of primary data. Birth and death certificates, marriage licences, and other important legal records are stored for a long period in public institutions and used for genealogical research (family history research) and other research projects. A property title, a birth or marriage certificate, a social security card, a diploma, or any document having a very long-term and significant value[ii] is also primary data[iii].

- **Interview and Primary Data**

Primary Source of data collection goes hand in hand with the interviews. An interview can be used as a tool to gather fresh, genuine, and sensitive information or insights. Interviews, for instance, are a suitable strategy when discussing themes like violence, conflict zones like communal violence areas, etc. However, sufficient rapport-building must be done prior to the interview. Interview also helps in reducing the gaps between the participants and the researcher.

- **Ethnography and Primary Data**

In Qualitative studies Primary Data is used more often. There are research methods like Ethnography which is often used in the discipline of Anthropology and at times even in Sociology which uses Primary Data to a large extent. Here the researcher visits a village which could be even remote away from modern society, facility and then spends several months and years observing, recording, participating the events of the society and thereafter he/she documents it. In Ethnography kind of primary data collection, the researcher at times even learns the language of the people. This he/she does so that they could think in the local language. The Ethnographer while doing data collection using primary method observe every minute details, every day life and makes notes about it. There are several scholars who have used investigates societies themselves like Malinowski – His famous Works are Magic and Religion, Coming of Age in Samoa, Indian writers like M.N. Srinivas – His Study of Religion among Coorg's, The Remembered Village. Clifford Geertz while studying the Cockfight in Bali, Indonesia.

Journalistic writings like Everyone Loves a Good Drought – by P. Sainath is a example of primary data collection – In this book, Sainath travels to different villages of India and he writes about the lives of the people who are effected by droughts. Every chapter is a story which is of farmer difficulties, farmer's widows, village life. The uniqueness about the book is that it is written in simple language and it connects with the reader. To understand primary data collection P Sainath work is a recommended reading. Feminist view primary data as an important source for documenting the marginalized groups experiences specially females.

- **Colonisers and Primary Data**

The Colonisers sponsored several scholars, researchers, Sociologists, Linguistics to write about the colonized locations. This in turn helped them to understand the societies better in which they ruled. Several Indian texts in local languages as a result was translated to English. There was emphasis laid on the documentation of practices, rituals, food, tribal cultures etc. There still exists a debate on the kind of content – quality or the reality of the content been produced. At times, several writings can be viewed as a researcher documenting a text from that of a hierarchical view point.

## 2.5 MEANING OF SECONDARY DATA

The data that has been compiled and gathered from existing sources is known as secondary data. The secondary sources are frequently the easily accessible information that researchers utilise to produce statistical claims and reports for their investigations. Unpublished materials like PhD Theses and records can also be considered secondary data. The secondary data are beneficial since they are already available online and in libraries. In terms of time and location, secondary sources are likewise unrestricted.

It is not necessary for the researcher to be present when and where the participants are assembled. Consequently, it is also economical.

Examples of secondary data include Census Reports, Annual Reports, Company Financial Statements, Statistic Statements, Department of Government Reports, etc[iv].

Primary data and secondary data are typically contrasted. In contrast to the former, which is information gathered from sources other than the user, the latter is derived directly from first-hand sources through questionnaires, observations, focus groups, or in-depth interviews. Or to put it another way, secondary data is data that has already been obtained for a different purpose. Such material, nevertheless, might be quite helpful for one's research.

A variety of secondary data classifications, including those that try to distinguish between raw and compiled data, are used in literature research. If any processing of the earlier kind has occurred, it has been limited (raw data). In the case of the latter category, there has been some sort of selection or summarization (compiled data).

- **Importance of Secondary Data**

Secondary Data is useful for a number of ways, right from saving time and resources to accessibility. Let us look into some of other ways in which it is helpful.

- **Developing Argument**

Secondary Data is often used/can be used to support the primary data. For example – Imagine you are studying about junk food eating habits among teenagers. In your study you found out that out of 100 sample – nearly 85 percent of the samples like maggi. The same study conducted in some other city by some other scholar also had the same conclusion. So, here you can build up an argument that – The findings of my study and X author study has the same kind of findings. Hence, through this we could generalize the finding and locate the eating habits of teenagers and take enough measures, awareness to reprogram the food habits.

- **Review of Literature**

Secondary Data is widely used to build the Review of literature in the thesis, research papers, reports, any kind of evidence. Let us take the example of a court case Applications in Mobile and Data leakage. You must have seen in movies how in the court lawyer to justify their case, point out the earlier verdict given on a similar issue. For example – The problem (case) would be would discuss in Chennai High Court but the lawyer could be using a point from that of verdict of Kerala High Court. Secondary data thus, presents a narrative about the body of work carried out earlier in the given field, area this is then presented through compilation as Review of literature. Through secondary data it is also shown how the existing gaps in the research topic exists. Hence, even

when a researcher/ student is first starting with his study – the first advice is given by the teacher/supervisor is to read on the topic – i.e., collect all the secondary sources.The following are list of resources from where Secondary Data could be collected-

• The OECD's (Organisation for Economic Co-operation and Development).

• UNICEF has contributed to the change of the data environment for more than 20 years by making strategic investments in data collection. The Multiple Indicator Cluster Surveys (MICS) programme is the central component of this methodology. UNICEF supports governments in conducting household surveys in nations as diverse as Argentina, Bhutan, the Democratic Republic of the Congo, and Iraq through a global programme of methodological research and technical assistance. The results of MICS have been widely used as a basis for policy decisions, programme efforts, and voter persuasion on issues involving children and women around the world. All of the available MICS datasets and conclusions are available at mics.unicef.org.

• NSSO Data- The National Sample Survey is a important component in in India which is used for policy building. It has data on Employment and unemployment, housing conditions, domestic tourism, drinking water, sanitation, land and animal holdings, social consumption, health, domestic tourism expenditure, Labour Force, Construction, Industries, Manufacture, and so on are all covered by the NSSO round.

• Annual Industry Surveys - The Annual Survey of Industries includes the production facilities in the registered sector. It includes comprehensive data on the inputs, outputs, value added, personnel, assets, and other factors of the registered factories.

• Consumer Expenditure in the Home - These surveys track how much money and how many different things households spend each year on. Surveys are carried outboth annually and five years.

• Business Surveys - These studies provide data on input, output, value added, employment, and other variables in unorganised industries like trade, production, and services.

• Surveys of Land and Livestock Holdings are conducted approximately every ten years and provide information on active land and livestock holdings in rural areas[v].

## 2.6 DIFFERENCES BETWEEN PRIMARY AND SECONDARY DATA

The following details emphasise the fundamental distinctions between primary and secondary data:

1. "Primary data" refers to information that the researcher has collected for the first time. Secondary data refers to information that has already been gathered others, organisations.

2. Primary data are up-to-date information, whereas secondary data are older information.

3. Primary data is gathered to address the issue at hand, as opposed to secondary data, which is gathered for a number of reasons than just to address a present situation.

4. It takes a long time to gather primary data. The acquisition of secondary data, however, is rapid and simple.

5. Surveys, observations, experiments, questionnaires, and in-person interviews are a few types of primary data collection techniques. Official documents, websites, books, journal articles, internal records, and other materials are examples of secondary data gathering sources.

6. A lot of time, money, and labour must be put into gathering primary data. Secondary data, on the other hand, is freely available and fairly priced.

7. The primary data is constantly modified to meet the needs of the researcher, and the lead researcher is in responsibility of ensuring that the research is of a high standard. However, the quality of secondary data is outside the researcher's control, and it is not adapted to his requirements.

8. Primary data is available in its unprocessed form, whereas secondary data is a polished version of primary data. Secondary data is created when statistical tools are applied.

**Check Your Progress**

1. Primary data is collected by whom?

2. Do you think primary data and Secondary data are same?

## 2.7 BIG DATA

Big data is a collection of structured, semistructured, and unstructured data that is gathered by businesses and may be mined for information for use in advanced analytics applications like machine learning and predictive modelling[vi].

Big Data is large scale data which is stored across several parallel computers. Today more people are using technology, and there are now more connected devices than ever. Companies that connect the physical world to digital media are always thinking of new ways to draw in customers. This shows how much more data and information will be circulating in the digital world now and in the years to come. Future businesses will have to process, look at, and use a lot more data.

- **Use of Big Data**

Big Data is used in a variety of fields and industries, including geoscience, social media, banking, e-commerce, healthcare, environmental and climate studies, life sciences and drug development, e-library and academic publications, cybersecurity, and governance[vii].

Big Data is helpful when business owners must make important decisions for growth. They hire specialists to extract pertinent data from Big Data Analytics, which could benefit the organisation. To make important decisions and advance, businesses may considerably benefit from the insights provided by a big data professional.

The three V's of big data are variety, velocity, and volume.

Data Volume: There is a vast amount of data.

Versatility: Consists of a wide range of data types.

Velocity: Near-real-time evaluation of a vast amount of streaming data.

## 2.8 SMALL DATA

Small data is a term used to describe a collection of small datasets that can influence current behaviour. anything that can be gathered into an Excel spreadsheet. Small data is only meant to have a small impact over a short time, yet it can still be useful in decision-making.Small Data is the term used to describe the specialised datasets produced after sorting through vast amounts of data. There are a number of issues in a corporation that demand rapid attention. Big Data analysis methods don't need to be used in circumstances like these.

Small volumes of data are gathered to be analysed. It's a sample size that the data scientist selected to address the particular question at hand. Small data allows for control over the data. The prepared data has been ready for study for some time now.It has distinct dataset qualities that are easy to understand and may be used to analyse current events.

## 2.9 DIFFERENCES BETWEEN BIG AND SMALL DATA

**OBJECTIVE:** A single task can be finished by looking at modest amounts of data. On the other hand, the goal of big data increases and directs to unexpected scenarios. We may start out with a single goal, but it will change with time.The cloud's numerous servers are used for sharing a Big data. On the otherhand, small data is type of data which is stored in a single computer. The size of Big Data is measured in terabytes and tetabytes. While that of the small data is that of measured in terms of mg or gb.

**LOCATION:** Small data is typically kept on a local computer or in a database as a single file. On the other hand, big data is scattered among numerous servers located in diverse locations using the cloud.

**STRUCTURE:** Big data may be semi-structured or unstructured across multiple sources, in contrast to organised small data, which is provided in a single table.

End users frequently prepare little amounts of data for their own unique purposes. Because of this, the person entering the data is aware of how to use it and what to anticipate from it. On the other hand, a group of people who may or may not be end users create big data. The collaboration required to manage the data is therefore somewhat complex.

**LONGEVITY:** Short-term data can be kept for a limited time or until the task is finished. On the other hand, big data requires permanent storage.

**REPRODUCIBILITY:** If a tiny amount of data is mistakenly lost or corrupted, it is feasible to recreate it; nevertheless, enormous amounts of data cannot be replicated. Therefore, it should be carefully inspected and evaluated before any dangerous material is removed.

**RISK:** In tiny data, the risks are very low. On the other hand, big data is risky because it requires a lot of resources including cash, labour, materials, and time.

**INTROSPECTION -** When dealing with modest amounts of data, we are given well-organized, discrete data pieces that are simple to find and have explicit metadata that explains all of the columns. Finding many files in various formats, meanwhile, could be challenging in cases with huge data. Insufficiently recorded data can be difficult to understand.

**ANALYSIS:** Several sorts of analysis can be performed using a variety of data.

On a single system, small amounts of data can be analysed in a single process. Huge data may need to be divided up and examined in stages using various approaches in dispersed scenarios.[viii]

**Check Your Progress**

1. List out few uses of Big Data

2. Discuss two comparison with reference to small and big data

## 2.10 SUMMARY

Primary data is information that is created for the first time by the researcher via his or her own efforts and expertise, especially to understand the research problem in better way. Primary data is also known as raw data or first-hand data. Because primary data collecting is done by the organization or by individual himself or herself. On the other hand secondary data is the data that has been compiled and gathered from existing sources. The secondary sources are frequently the easily accessible information that researchers utilise to produce statistical claims and reports for their investigations. Unpublished materials like PhD Theses and records can also be considered secondary data. The secondary

data are beneficial since they are already available online and in libraries. In terms of time and location, secondary sources are likewise unrestricted. It is not necessary for the researcher to be present when and where the participants are assembled. Consequently, it is also economical.Examples of secondary data include Census Reports, Annual Reports, Company Financial Statements, Statistic Statements, Department of Government Reports, etc. Small data is a term used to describe a collection of small datasets that can influence current behaviour. anything that can be gathered into an Excel spreadsheetThe size of Big Data is measured in terabytes and tetabytes. While that of the small data is that of measured in terms of mg or gb.Big Data is large scale data which is stored across several parallel computers. Today more people are using technology, and there are now more connected devices than ever.

## 2.11 QUESTIONS

1. Distinguish between big and Small Data
2. Write a brief note of Secondary Data and discuss some examples
3. Discuss Primary Data and some methods of collecting them.

## 2.12 REFERENCES

[i]https://dictionary.cambridge.org/dictionary/english/data

[ii]https://researchguides.ben.edu/c.php?g=282050&p=7037030

[iii]Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.

[iv]Krishnaswamy O.R. (2010), Methodology of Research in Social Science, Himalaya Publishing House.

[v]http://microdata.gov.in/nada43/index.php/catalog/central

[vi]https://www.techtarget.com/searchdatamanagement/definition/big-data

[vii]https://www.sciencedirect.com/journal/big-data-research/about/aims-and-scope

[viii]https://medium.com/analytics-vidhya/small-data-vs-big-data-30a38f129074

❖❖❖❖

# THEORETICAL CONSIDERATIONS – POSITIVISM

**Unit structure**:

## 3.0. OBJECTIVES:

➢ To familiarize students with Positivist theoretical paradigms
➢ To understand that 'positivism' is a philosophical paradigm, based upon natural science of reason and logic.

## 3.1 INTRODUCTION:

Positivism is a philosophy of science that rejects metaphysical speculation in favor of systematic observation using the human senses. "Positive" knowledge of the world is based on generalizations from such observations that, given sufficient number and consistency, are regarded as producing laws of how phenomena coexist or occur in sequences (Lewis-Beck et al., 2004).

The doctrine of positivism was formulated by Auguste Comte, the French philosopher, who is also known for being one of the founding fathers of sociology. Positivism is regarded as the scientific understanding of society. Although positivism, as a research paradigm, concerns itself with the social sciences, it draws heavily from the natural science. Positivism, being based on the principles of natural science, argued for the study of the society driven by scientific investigation and knowledge.

However, as noted by Bryant (1985), the terms 'positivism' and 'sociology' are both commonly supposed to have originated with Comte, and in particular his Cours de philosophie positive (6 vols, 1830-42), although true of the second, this is misleading with respect to the first term insofar as Comte wrote not about 'positivism' but about 'the positive philosophy'

and 'the positive method', and Saint-Simon before him had also advocated a positive philosophy.

## 3.2 'POSITIVISM' – THE BACKGROUND:

The philosophy of positivism was to solve the problems of social life, its progress by following the principles of natural science. In *A General View of Positivism* French philosopher Auguste Comte (1798-1857) gives an overview of his social philosophy known as Positivism (Comte, [2009] 1848). The French Revolution had deep impact on the thinking of Auguste Comte, to the extent that he rejected religion professed 'religion of humanity'. Comte strongly felt that the scientific study of society would be the only way to solve its problems and thus named it 'sociology'.

Based on scientific principles, Comte formulated the doctrine or philosophy of 'positivism', mainly through his six-volume work *Course of Positive Philosophy*, over a decade. Following the natural science principles, Comte proposed to look at the society as being governed by its own set of laws, just as the physical sciences. He, thus laid down the foundation of the scientific study of society, to become popular as 'sociology'. The uncertainty and chaos in the social life was to be resolved by the use and application of these scientific principles. Positivism in philosophy came to be associated with epistemologies which make experience the foundation of all knowledge. and also with their complementary ontologies which propose a division between objects which are accessible to observation and objects which are not Bryant (1985).

Comte was also significantly influenced by the early intellectuals. From David Hume and Immanuel Kant Comte derived his conception of positivism—i.e., the theory that theology and metaphysics are earlier imperfect modes of knowledge and that positive knowledge is based on natural phenomena and their properties and relations as verified by the empirical sciences ("The New Encyclopaedia Britannica," 1997). From the beginning, Positivism believed in the reality being accessible through our senses. Human experience plays an important role in acquiring the knowledge of reality. Thus, the validity through experience remains crucial for positivist approach. In positivism, the scientific knowledge has to be verifiable through human experience.

Comte's main contribution to positivist philosophy falls into five parts: (a) rigorous adoption of the scientific method; (b) law of the three states or stages of intellectual development; (c) classification of the sciences; (d) conception of the incomplete philosophy of each of these sciences anterior to sociology; and (e) synthesis of a positivist social philosophy in a unified form (Duignan, 2010). Comte's law of three stages – a theological stage, metaphysical stage and positive stage – outlined the process of human intellectual development in the history of society.

**Check your progress:**

1. What is Positivism?

## 3.3 THE FRENCH TRADITION OF POSITIVISM:

It is well known that Comte not only coined the term 'sociology' but also introduced and systematized the positivist philosophy in the social science. Saint-Simon, in fact, provides a better starting point for an analysis of the French tradition of positivism than Comte for two reasons:

(a) he announced the great nineteenth century project of the construction of a positive science at the start of the century,

(b) he did so in a way that attracted admiration from Marx and Durkheim (Bryant, 1985).

The works of Saint-Simon, Comte and Durkheim display interconnections in their thinking and ideas. Bryant (1985) lists twelve tenets, which indicate the basic characteristics of the French tradition of positivism:

i.   There is but one world, and it has an objective existence.

ii.  The constituents of the world, and the laws which govern their movements, are discoverable through science alone, science being the only form of knowledge. Therefore that which cannot be known scientifically, cannot be known.

iii. Science depends upon the combination of reason and observation.

iv.  Science cannot discover all the constituents of the world, and all the laws which govern them, because human powers of reason and observation are limited. Scientific knowledge will remain forever relative to the level of intellectual development attained and to progress in the social organization of science.

v.   What man seeks to discover about the world is normally suggested by his practical interests and his situation.

vi.  There are laws of historical development whose discovery will enable the past to be explained, the present understood and the future predicted.

vii. There are social laws which govern the interconnections between different institutional and cultural forms.

viii.   Society is a reality sui generis.

ix.  Social order is the natural condition of society.

x.   Moral and political choice should be established exclusively on a scientific basis.

xi.  The subjection of man before the natural laws of history and society precludes evaluation of institutional and cultural forms in any terms other than those of conformity with these laws.

xii. The positive, the constructive, supersedes the negative, the critical. The positive, the relative, also supersedes the theological and the metaphysical, the absolute.

**Check your progress:**

1. What are the characteristics of the French tradition of positivism?

## 3.4 CENTRAL TENETS OF POSITIVISM:

Blaikie (2007), through his meta-analysis of the literature on 'positivism', has presented some brief points to be incorporated as positivism's characteristics or its central tenets. There are:

▪ **Phenomenalism:** this rule asserts the acquisition of scientific knowledge only through human experience. Scientific knowledge is nothing but the perception by the senses, a 'pure experience', without any cognitive interference.

▪ **Nominalism:** Any abstract concepts used in scientific explanation must also be derived fromexperience; metaphysical notions about which it is not possible to make any observations have no legitimate existence except as names or words. Hence, the language used to describe observations must be uncontaminated by any theoretical notions. As 'reality' or 'truth' is believed to be observable through one's senses, any theoretical terms, such as 'God', which is non-observable, should be considered as meaningless.

▪ **Atomism:** The objects of experience, of observation, are regarded as discrete, independentatomic impressions of events, which constitute the ultimate and fundamental elements of the world. In so far as these atomic impressions are formed into generalizations, they do not refer to abstract objects in the world, only regularities among discrete events.

▪ **General Laws:** Scientific theories are regarded as a set of highly general law-like statements;establishing such general laws is the aim of science. These laws summarize observations by specifying simple relations or constant conjunctions between phenomena. Explanation is achieved by subsuming individual cases under appropriate laws. These laws are general in scope, in that they cover a broad range of observations, and are universal in form, in that they apply, without exception, across time and space.

▪ **Value Judgments and Normative Statements:** "Facts" and "values" must be separated asvalues do not have the status of knowledge. Value statements have no empirical content that would make them susceptible to any tests of their validity based on observations.

▪ **Verification:** The truth or falsity of any scientific statement can be settled with reference toan observable state of affairs. Scientific laws are verified by the accumulation of confirming evidence.

▪ **Causation:** There is no causality in nature, only regularities or constant conjunctionsbetween events, such that events of one kind are followed by events of another kind. Therefore, if all we have are regularities between types of events, then explanation is nothing more than locating an event within a wider ranging regularity.

**Check your progress:**

1. What are the central tenets of 'Positivism'?

## 3.5 SUMMARY - 'POSITIVISM' AND SOCIOLOGY:

Positivism in sociology has come to be associated with the very idea of a social science and the quest to make sociology scientific (Bryant, 1985). Positivists look at social reality as scientifically verifiable and measurable. Just as the physical world is governed by the laws, human existence is also governed certain laws that should be discovered, the positivists believe. And therefore, sociology, from a positivist perspective, is also modelled on the laws of the natural sciences, which necessarily include logic and reasoning.

As sociology deals with the scientific study of society, positivism ensures scientific explanations – which are observed, verified and logically stated – of social phenomena, as it completely rejects theological or supernatural assumptions and explanations. As social actors remain important in the social phenomena, however, the positivist philosophy, within sociology, looks at the factual aspects of the phenomena, devoid of any subjectivity and value judgments of the actors or the researchers.

Romm (1991), through a meta-analysis of existing literature, affirms that the positivist theory of science is not the only theory which has become incorporated into the sociological enterprise, but this theory of science, and the research practice which it inspires, have assumed a dominant position within sociology. Thus the positivist pursuit of the sociological understanding of the society entails understanding the causal aspects of human behaviours, in order to understand how society operates.

Positivism in sociology shares a strange relationship. Sometimes, to be positivist means no more than being scientific, although that fails to discriminate between positivism and all the other sociologies that have claims to be scientific in perhaps different ways, such as Marxism, functionalism, structuralism, and so on; and sometimes, positivist sociology is synonymous with statistical analysis, as in many sociological research reports and methods textbooks; yet some other times, to practice positivist sociology is to seek to establish causal explanations, or to search for fundamental laws of human behaviour or historical change, or to insist

upon objective empirical information systematically organized to generate or test hypotheses (Halfpenny, 2014).

Thus, for a positivist sociological investigation, we need necessarily need: an empirical basis of knowledge, where direct observation leads to the theory; the place of deduction and induction in linking 'theory' and 'observation'; constructing a hypothesis; tentative character of scientific statements; demarcation between science and non-science; objectivity; and establishing the causality.

## 3.6 QUESTIONS:

- Elaborate on the positivist research methodology within Sociology.
- Elaborate on 'Positivism' as a classical methodological perspective.
- Elaborate on the French Tradition of Positivism.
- What has been Comte's contribution to the philosophy of positivism?
- How is 'positivism' incorporated into Sociology?

## 3.7 REFERENCES AND FURTHER READINGS

- Anand, S. (1997). *Understanding Religion: Theories and Methodology*. New Delhi: Vision and Venture.

- Blaikie, N. (2007). *Approaches to Social Enquiry*. United Kingdom: Polity Press.
- Brennan, J. F., & Houde, K. A. (2017). Sensationalism and Positivism: The French Tradition *History and Systems of Psychology* (7 ed., pp. 123-136). Cambridge: Cambridge University Press.
- Bryant, C. G. A. (1985). *Positivism in Social Theory and Research*. London: Macmillan Publishers Limited.
- Comte, A. ([2009] 1848). *A General View of Positivism* (J. H. Bridges, Trans.). USA: Cambridge University Press.
- Halfpenny, P. (2014). *Positivism and Sociology: Explaining Social Life*. London: Taylor & Francis.
- Mill, J. S. (2009). *Auguste Comte and Positivism*. New York: Cosimo Classics.

❖❖❖❖

# 4

# WRITING RESEARCH PROPOSAL

**Unit Structure**

## 4.0 OBJECTIVES

1.0 To understand the meaning of Research proposal and its uses.
2.0 To learn proposal writing from certain examples of existing proposals.

## 4.1 INTRODUCTION

In this chapter we are going to learn about research proposal its meaning, the objective behind making of writing research proposal, the different kinds of research proposal used. Research Proposal is, an idea/ problem is being proposed to a guide/ committee/ Institution and there after it is accepted or rejected or some modification is suggested.

## 4.2 MEANING OF RESEARCH PROPOSAL

A research proposal is a thorough description of a proposed study to look into a certain issue[i].A detailed summary of the dissertation or research project is also contained in a proposal. It demonstrates the design and methods of the study[ii].The goal of a research proposal is to persuade readers with the idea that the proposal is a wort hwhile study, topic and researcher has the education and experience required to carry it out. The research proposal generally has to answer the questions like following issues: What the researcher is planning to do, how he/she is planning to accomplish it[iii].

Writing a research proposal is a challenging procedure in today's world because of the research design trends that are constantly changing and the requirement to include scientific breakthroughs into the technique. The

most important phase of the research process is the development of a clear, well-thought-out proposal that will serve as the basis for the research. Writing a study proposal is done in order to apply for grants and obtain approval from a number of bodies, including the ethical committee. The requirements for producing a high-quality research proposal change based on the needs of the funding agency or institution because there is no universal standard way that is followed[iv].The most important thing though is that proposal must be able to persuade the review panel evaluating the study of the validity, viability, practicability, and generalizability of the study's design[v]. The process of conducting research includes multiple stages and a flow of documents. The order and stages of the process that have a real impact on the quality and content of the final report and the research materials created at every level. The four documents mentioned in the research proposal are research summary, research abstract, and the research report which are used to evaluate the appropriateness, quality, and validity of every research/study.

- **Length of Proposal –**

It is preferable to contact the institution/organization where the proposal is being sent. To save time, this step should be completed before writing the proposal. The proposal typically has between 10 and 20 pages. References and a reading list are also included at the end of the research proposal.

A research proposal needs to be organised clearly. It must be broken up into paragraphs with subjects and subtopics. Subject changes may be noted with headings and paragraphs. The discussion's direction is indicated by the headings. A research proposal needs to be written in a distinct manner from other academic writing assignments including essays, fiction, and poetry. The latter makes use of the word's richness. However, research reports should be written at a formal level, needs standard English since they are a formal presentation of a problem[vi].

- **Time required –**

Before preparing a proposal, the topic must be defined. Because doing so would save time and enable a person the researcher to concentrate on a single issue. While writing a proposal, there are several modifications that happen. especially if a guide is present. There are modifications made, new points are added, and proofread. Therefore, preparation must begin on day one; a little writing each day would aid in finishing the work.

- **Important Questions**

Some important questions which an individual needs to ask before writing a proposal to oneself is which according to Maxwell (2005) are –

- What do readers need to know in order to grasp the topic better?
- What aspects of the subject are readers unfamiliar with?
- Prior Things to Think About like What does one want through the research?

- What is the environment, context and who will be the subjects of the study?
- What is the process of Data collection and analysis?
- What will be the process employed to cross check the data?
- What will be the theory used?
- What moral dilemmas will the study bring up?
- What do early findings indicate about the applicability?

It is difficult to understand research proposal without understanding the why it is used.

1. **For admission in Philosophy of Doctorate (Ph.D.)**
   To enroll for Ph.D. firstly an individual needs to have a Master's Degree, then he/she needs to clear any one of the entrance exams likeNational Eligibility Test, State Eligibility test or PET (PhD entrance test).When a student wishes to apply for the program of PhD after Masters Degree the first thing the individual does is finds out a research problem/ topic and then build a research proposal based on that. This proposal the candidate prepares individually or with help of some teacher.

2. **Non-Governmental Organizations**
   Writing proposal or research proposal is not just restricted to higher education but it is alsouse in the development sector like non-Government organisations like for seeking funds. The NGOs make proposals to Government seeking funds for implementing a project, conducting Survey, Conducting Research. Proposals are also submitted to Corporates, National and Multinational Companies, Trusts, International Bodies like WHO, UNESCO, IMF, World Bank etc.Based on the funds earned they further take action and work in villages. However, the focus in this chapter is more towards academic writing.

**Check Your Progress**

1. Why is a research proposal made?

2. Discuss the length of research proposal.

## 4.3 THE PROCESS OF WRITING

Writing itself can be a challenging task for several individuals. The challenge is more with that of technology and access to it. As writing itself is a creative task one needs a certain amount of dedication, consistency in it. There are invisible problems one could encounter too like handling criticism of past writing and face one's own fear. As students often tend to generally write only majorly during exam for exam papers. In a way, we tend to write what we have studied or what we have read. However, research proposal is a step towards independent writing and on a topic which you choose to work upon. Beginning to write a research proposal would take some amount of mental preparation.

The individual who is making the research proposal is often expected to make several drafts and a brief proposal or plan explaining the research project's purpose and methodology before submitting the dissertation. This is an efficient method of getting ready for your study and it will motivate you to consider many of the topics that are mentioned in the next section. The proposal will ask you to demonstrate some knowledge of the literature in your chosen field—for example, by naming several key authors or significant research studies—in addition to outlining your proposed research design and methods, the topic area in which your study is going to be located, and the research questions that you intend to address.

Based on this information, a supervisor who is knowledgeable in your study area of interest or who has experience with your suggested research approach may be assigned to you. The proposal can serve as a valuable starting point for discussions with your supervisor about your research topic, and if it includes a timetable for the project, it can be used to schedule frequent meetings to discuss your progress. By giving you a series of continuing objectives to work toward and forcing you to think about various parts of the entire research process, such as the various stages of your research and their timing, creating a timeline may be very helpful.

You will likely need to address a variety of topics when creating a research proposal like What is the subject of your research, or, alternately, what are the goals of your research? • Why is the subject of your study (or the goals of your study) significant?

What is your research question, or what questions do you have? • What does the body of research indicate about your study's subject, goals, and research question(s)? • How will you go about gathering information pertinent to your research question(s)? In other words, what research techniques do you plan to employ? • Why are the sources and research methodologies you've chosen the best fit for your research question(s)? • What equipment (such as postage, travel expenses, or software) will you need for your research, and how will you pay for it? • What schedule do you have for the various phases of the project? • What difficulties do you foresee in conducting the research (for instance, gaining access to organisations)? • What potential ethical issues could arise from your research? • How will you analyse your data?

Writing a proposal is consequently helpful for getting your research project off the ground and motivating you to establish reasonable goals. The research proposal could make up a modest portion of the overall evaluation of the dissertation or report that results from the project in various higher education institutions. It is vital to keep in mind that, although though the research proposal is a working document and the concepts you provide in it can be polished and developed as your study advances, doing so will cost you valuable time that could be used to finish the dissertation by the deadline.

In conclusion, a research proposal can be crafted to appeal to the readership. Popular research proposals should be able to expand readers' knowledge in some little ways, assist readers in finding the information they need inside the proposal. In addition, be aware of at least some of the preferences of the target audience and accommodate those preferences through thoughtful research design. A writer needs to be aware of and steer clear of a few frequent mistakes. As follows:

Avoid overuse of jargon

-Reduce the Verbosity

- Be aware of the individual bias in writing

 - Be conscious of any inaccuracies in the facts

- Revise to reduce the Grammatical mistakes

- Avoid illogical statements

-Try to have a sequence

- Include proper references

While designing a Proposal's Sections some research hints discussed in the below section can be used while creating the overall format of a proposal.

Writing Techniques and Ethical Issues

• Ideas for additional areas often come to mind while working on one section. Create an outline first, and then quickly jot down something for each part to get thoughts down on paper. The sections can then be improved later. However, if you fail to write or build an outline, the process could become time consuming. Look for ideas from other proposal and make points and review them carefully. Request copies of proposals that your advisor/ supervisor thought that were particularly strong and deserving of consideration from committees.Examine the subjects covered, the sequence in which they are addressed, and the level of information employed to create the proposal.

• Check to see whether your programme or school offers a course on developing proposals or a related subject. Such a class would be beneficial as a support system for your project and as a source of people who can respond to your proposal ideas as they emerge.

• Discuss the preferred proposal format with your adviser over a meal. It's possible that your adviser or graduate committee won't find the material they're looking for in the parts included in published journal papers[vii]

## 4.4 A QUANTITATIVE PROPOSAL'S FORMAT –

This format is followed in several quantitative studies and the outcome is reported in journal publications. The structure typically consists of an introduction, a review of the relevant literature, followed by methods, results, and discussion. To create a dissertation proposal and prepare a quantitative study, take into account the study's goals and boundaries theoretical vantage point, Inquiry-based statements or hypotheses, Writing Techniques and Ethical Issues. One needs to also include an analysis of the literature, Methods, Research methodology Participants, sample, and population Instruments, variables, and materials for data collecting techniques, data analysis, ethical problems that one might encounter while conducting the investigation.

**General Format of Research Proposal**

The proposal's front page includes information like -

- Your Name
- Qualification
- Topic of Proposal
- Seeking for Guidance Under (if known and confirm them)
- Name of University
- Year Month

The topic of the proposal need to focus on the core problem under study. The Introduction part of the proposal needs to explain the problem under study with a broader perspective. The Aims and objectives are also included. Review of literature includes the secondary data available on the topic like that of thesis, books, journal articles, reports any other reference material connected to the topic. The next section focusses on the research methodology which discusses the data, the rationale behind choosing the specific subjects, the limitations, number of subject. The next section includes the time, resources. Include budget (if proposal is being submitted to organizations).

## 4.5 FORMAT FOR A MIXED METHODS PROPOSAL

When using a mixed methods design, the researcher combines methods from both the quantitative and qualitative forms (see Creswell & Plano Clark, 2007). One has to look into the deficits in prior research, as well as one gather information related to both quantitative and qualitative data. Information should be also documented about the target audiencesneeded in the research, the project's goal and justifications for using a mixed-methods study. The proposal should also include the research propositions and questions (quantitative questions or hypotheses, qualitative questions, mixedmethods questions, Literature on the philosophical underpinnings of employing mixed methods research should also be included.

An explanation of mixed-methods analysis,the design style employed and its concepts, problems that may arise when adopting this approach and how they can be used to solve the problem has to explained. References

and appendixes with instruments, protocols, visuals needs to be also included at the end.

## 4.6 QUALITATIVE RESEARCH PROPOSAL –

The following are the steps which is used for Qualitative research proposal.

Introduction

Description of the issue (issue, significance of issue)

The study's goals and boundaries

Theoretical vantage point

Inquiry-based statements or hypotheses

Writing Techniques and Ethical Issues

An analysis of the literature

Methods

Research methodology

Participants, sample, and population

Instruments, variables, and materials for data collecting

Techniques for data analysis

Ethical problems that the investigation might encounter.

initial research or pilot tests

Appendices: Tools, Schedule, and Budget Proposal

- **Ethics and Plagiarism**

The researcher should take extra effort to make sure that moral obligations are honoured. The protection of the participants' rights, including their right to informed consent, the institutional review process, their right to autonomy, their right to privacy, their right to confidentiality, their right to fair treatment, and their right to be free from discomfort and harm, are all considered ethical issues (ethical approval). The researcher must offer sufficient details.

Participants, the research site, and the appropriate authorities must all give their informed consent.

Another important point to remember while building a proposal is also that one has to cite all the sources from which the idea, material, citation, paragraph, lines has been taken. If one doesn't cite the content would be seen by the reader as yours and that is inappropriate. Hence, it is very important to acknowledge others work. This works as a cumulative scholarship where other person would acknowledge your work tomorrow when ever your work is published. One has to also cite previously published work if used in the main text of the proposal. As without that

doing that it could lead to self-plagiarism which is also seen as unethical way.

**Check Your Progress**

1. Explain ethics while writing research proposal

2. Discuss plagiarism practice in research proposal

## 4.7 SUMMARY

We began the chapter by understanding the meaning of research proposal. A research proposal is a thorough description of a proposed study to look into a certain issue.A detailed summary of the dissertation or research project is also contained in a proposal. It demonstrates the design and methods of the study

The goal of a research proposal is to persuade readers with the idea that the proposal is a worthwhile study, topic and researcher has the education and experience required to carry it out. The research proposal generally has to answer the questions like following issues: What the researcher is planning to do, how he/she is planning to accomplish it. A research proposal is submitted to Universities, Departments for higher education generally. It is also used when applying a research project to an organization. The Research Proposal is also used by Non-Governmental Organisations.

The topic of the proposal needs to focus on the core problem under study. The Introduction part of the proposal needs to explain the problem under study with a broader perspective. The Aims and objectives are also included. Review of literature includes the secondary data available on the topic like that of thesis, books, journal articles, reports any other reference material connected to the topic. The next section focusses on the research methodology which discusses the data, the rationale behind choosing the specific subjects, the limitations, number of subjects. The next section includes the time, resources. Include budget (if proposal is being submitted to organizations). Ethics, consent should be discussed in the proposal too. Citation of the used material has to be made while writing the proposal. While writing the research proposal the language has to be formal. There should not be any unnecessary jargons used. The written text should be based on facts and presented in a logic and sequential order. The chapter also discusses the mixed method proposal, quantitative and qualitative proposal too which is similar in the core points but has little bit of variation amongst each other.

## 4.8 QUESTIONS

1. Discuss the meaning of Research Proposal and write about Mixed Methods Research proposal.
2. Explain the Writing process involved with research proposal
3. Discuss quantitative and qualitative research proposal format.

## 4.9 REFERENCES

[i]Traenkel JR, Wallen NE. How to design and evaluate research in education. On line learning center with power web. Available at www.highered.mcgraw-hill.com/sites/0072981369/student_view0/chapter24/key_terms.html

[ii]Walliman, N. (2006). Writing a research proposal. In *Social research methods* (pp. 186-192). SAGE Publications, Ltd, https://dx.doi.org/10.4135/9781849209939

[iii]Wong P. How to write research proposal. International network on Personal meaning. Available at www.meaning.ca/archives

Al-Riyami A. (2008). How to prepare a Research Proposal. *Oman medical journal*, *23*(2), 66–69.

[iv]Sudheesh, K., Duggappa, D. R., & Nethra, S. S. (2016). How to write a research proposal?. *Indian journal of anaesthesia, 60*(9), 631–634. https://doi.org/10.4103/0019-5049.190617

[v]Saunderlin G. Writing a research proposal: The critical first step for successful clinical research. *Gastroenterol Nurs.* 1994;17:48–56.

[vi]Krishnaswami, Rangantham (2019) Methodology of Research in Social Sciences, Himalaya Publishing House.

[vii]John W. Creswell (2019)Research design: Qualitative, quantitative, and mixed methods approaches. *Sage Publications.*

❖❖❖❖

# 5

# MAIN STEPS IN QUANTITATIVE RESEARCH

**Unit Structure**

5.0 Objectives

5.1 Introduction

5.2 Inductive and Deductive Approach

5.3 Main Steps in Quantitative Research

5.4 Summary

5.5 Questions

5.6 References

## 5.0 OBJECTIVES

1.0 To understand the different steps involved in Quantitative Research
2.0 To learn about the ethics involved in these steps

## 5.1 INTRODUCTION

Modern Research in the market is increasingly being conducted as a team, with each team member offering their specialised knowledge to the project. Therefore, research activities have changed and has even become highly competitive. Organizations, business and industry cooperate and operate by applying mathematical and scientific techniques to solve problems in theorganisation issues and through research studies. Corporate bodies are also adopting research operations more and more to increase productivity, management science research (OR) and methodologies to improve quality and cut expenses. They use quantitative research to address issues with planning, strategy, anddistribution of resources, facility design, inventory management, employee schedules, and distribution techniques. However, as students you will be conducting independent research many a times, guided by a teacher.

Quantitative research focuses on collecting numerical data and using it to understand a specific event or generalise it across groups of individuals.Quantitative methods place an emphasis on precise measurements and the statistical, mathematical, or numerical analysis of data gathered through surveys, polls, and other types of research, as well as the manipulation of statistical data that has already been obtained using computing methods[i].

In this chapter, you will be learning about the important steps which is used in Quantitative Research. Learning these steps would give you a

guideline, base, direction to undertake an independent study. This chapter would act as a foundation chapter for the research methodology as the topics dealt here would be spread across whole of this paper. Some topics would be even common in this paper and you will also find it in the next semester syllabus, i.e., in qualitative research too. Some steps are even common if you are undertaking any research project or going for Higher studies like PhD.

## 5.2 INDUCTIVE AND DEDUCTIVE APPROACH

Before understanding the main steps in detail, we need to first understand, inductive and deductive approach. In Quantitative research many a times it is the deductive approach used. While in Qualitative Research it is the inductive approach used.
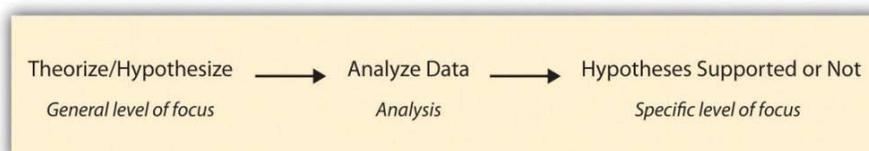
### Inductive Approach

When a researcher takes an inductive technique, they start by gathering information that is pertinent to their study topic. Here the researcher will stop collecting data once a significant volume has been gathered in order to step back and gain a bird's eye view of their data. The theory is used in order to account for the trends, the researcher also tries to find the patterns in the data. In order to proceed from a specific set of experiences to a more general set of propositions about experiences, researchers use an inductive approach. To put it another way, the researcher progress from facts to theories, or from the particular to the general.

| Gather Data | → | Look for Patterns | → | Develop Theory |
|---|---|---|---|---|
| *Specific level of focus* | | *Analysis* | | *General level of focus* |

### Deductive Approach

A convincing social theory will serve as the starting point for deductive researchers, who use the theory first. In other words, they employ the same procedures as inductive research but will carry out the processes backward, going from broad to more precise levels. Scientific study is most frequently connected with the deductive research approach. The researcher analyses existing theories of the phenomena and then try to examine, investigate what others have done, and then tests hypotheses that result from those theories[ii].

| Theorize/Hypothesize | → | Analyze Data | → | Hypotheses Supported or Not |
|---|---|---|---|---|
| *General level of focus* | | *Analysis* | | *Specific level of focus* |

**Check Your Progress**

1. What is Inductive approach

2. What is deductive approach

- **Steps involved in Scientific Research**

Before getting into the details of Quantitative Research steps let us first look into the Scientific Research steps.The steps of the scientific process are:

• Choosing a study topic

• Review of pertinent literature to evaluate prior work in thechosen subject for study;

• Compilation of statistics and facts already known about the subject of the study;

• Creation of a well-thought-out hypothesis;

• Validating the idea in accordance with a careful study plan;

• Method and organisation of the data acquired for analysis;

• Reaching judgments and inferences;

• Generalization, provided the data support it;

• Assembling and presenting the research's findings;

• Publication in peer review journals

## 5.3 MAIN STEPS IN QUANTITATIVE RESEARCH

### 1. Theory

As seen earlier Quantitative Research uses Deductive approach. Hence, the existing theory is used to guide the research. Merriam dictionary defines a theory as a <u>plausible</u> or scientifically acceptable general principle or body of principles offered to explain phenomena[iii]. According to Byrman, the word "theory" has many various meanings, its most common usage is to refer to an explanation for an observable pattern, such as why poor people fall more sick than rich people or why job alienation differs by technology. There are several theories like Grand theories, Middle range theories, Micro theories which exist in social science. Grand theories are those theories which are often applicable universally and is generalizable to a large extent. For example – Marxian theory of Class Struggle, concepts like Alienation. These theories and concepts are applicable in every part of the world as still the industries exists and the capitalists and working class are too present, in addition the conflicts continue too remain in every society. Another example is that of Functionalism. Middle range theories are

like Robert Merton's work, the micro theories are like ethnomethodology, symbolic interactionism, phenomenology.

One can find theories in existing study, literature hence one has to look into the studies conducted similar to one's own research. A researcher can use one or more theories in a given study. The following are some points to remember while reading-
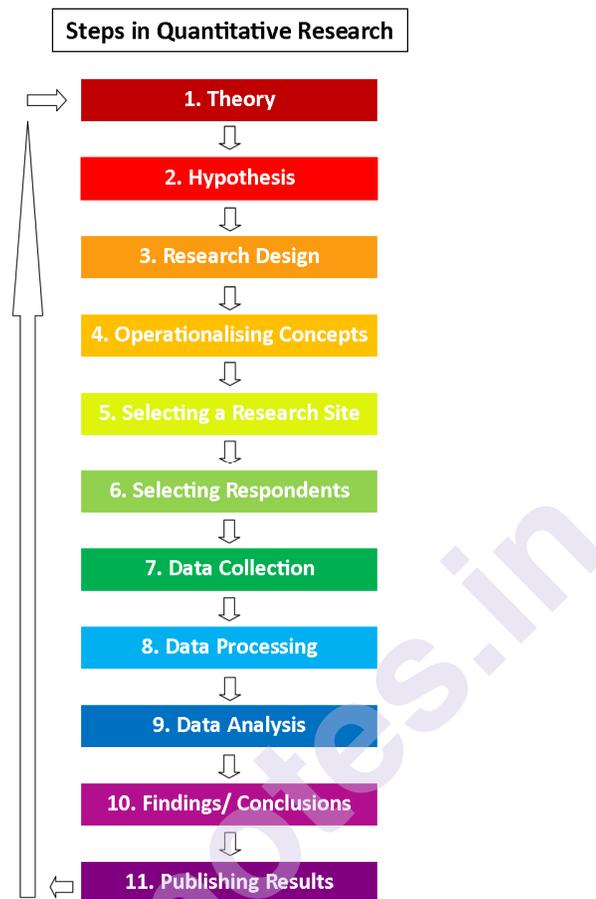
- Read a lot of reputable, pertinent reviews.
- Recognize the study's variables
- Create a list of synonyms or a replacement word
- You can add variables to a Venn diagram while reading.
- Keep citations compiled or located with abstracts
- Read abstracts and remove any publications whichare not useful.
- Determine whether you need to focus your review more or dig deeper.
- Systematically keep track of your pertinent readings
- Read and made annotations on each pertinent item
- Organize and sort your annotations.
- Created a draft of your proposed theories
- Write with intention
- You can use the literature to support your claims.
- Use the proper diction and tone.

A clear understanding of literature on the topic and theories helps to figure out the gaps in the research and that could turn out to be a research problem for the study. Hence formulating the research problem is also one of the main steps and first step of doing quantitative research. A clear research problem helps the investigator to understand what is the problem and what he/she wants to find out about the problem. Hence to do that, literature survey could be carried out through visiting libraries, studying existing research papers on the same topic, referring to articles on internet, reading several thesis on similar topic, discussing with people working on the same area.

2. **Hypothesis** –

According to Collins Dictionary - A hypothesis is an idea which is suggested as a possible explanation for a particular situation or condition, but which has not yet been proved to be correct[iv].In quantitative investigations, a hypothesis is typically put forth before the study and tested later on. Having a predetermined hypothesis helps researchers focus their investigation and interpret the findings in light of a pre-existing theoretical framework. A hypothesis is developed based onexistent knowledge which is a tool for advancing knowledge through testing it throughout a research investigation and haveknowledge about a specific topic. In the event that a certain hypothesis is found to be false during a study, still theoutcome is considered a knowledge breakthrough since we are aware that within the context of the investigation,The listed factors

have a relationship or not. This could also benefit other scholars who might be studying similar studies in the future[v].

**Steps in Quantitative Research**

1. Theory

2. Hypothesis

3. Research Design

4. Operationalising Concepts

5. Selecting a Research Site

6. Selecting Respondents

7. Data Collection

8. Data Processing

9. Data Analysis

10. Findings/ Conclusions

11. Publishing Results

### 3. Research Design

Research Design is the blue print of the study. A lot of decision-making goes into research design. It gives your researcha framework and direction. Following the selection of your topic, one has to make a decision regarding the methodology which will be used in research. It entails developing a strategycovering all phases, from the creation of hypotheses to the data analysis here the research design helps. According to Kerlingerresearch design is a strategy, framework, and blueprint for an investigationas well as to find solutions to issues or challenges in study. It comprises an outline of the different steps the researcher will take, starting with drafting the hypotheses and their operational implications all the way through to the data analysis at the end. According to Thyer, a research design is a precise strategy or blueprint for a studyoperationalizing variable in a study so they may be quantified, to choosinga study sample, gathering information to use as a foundation for testing theories,and finally, reviewing the outcomes. There are different types of Research design like descriptive, experimental, exploratory etc.

A research question is an inquiry that makes clear what the researcher specifically wants to know. A research question forces the researcher to be more specific about what is to be examined. A research purpose might be stated as a statement, such as "I want to find out whether (or why). A research question cannot exist without a question mark at the end of it. In other words it must be a question.

A list of different kinds of research questions has been offered by Denscombe (2010). This list first appeared in an older edition that White has updated (2009). Denscombe suggests the following categories of study questions:

1. Making predictions (does x occur under conditions a and b?).

2. Defining the causes and effects of an event (is y a result of x or an effect of x?).

3. Assessing a phenomenon (does it demonstrate the benefits that are asserted to exist?).

4. Describe the phenomenon (what does it look like or take on?).

5. Establishing positive habits (how can we do y better?).

6. Empowerment (how can we make those whose lives we study better?)[vi].

## 4. Concepts
A concept is an abstract word that unites various meanings. Concepts occasionally appear in the field. A researcher should choose the main concepts from the problem studied. The concepts used in the particular hence has to be clearly defined. As it carries different meaning outside the context too.

## 5. Selecting the Research Site
A research site is the location in which the research is going to carried out. This takes time and effort too. A pilot study helps to understand whether the chosen site is helpful for the study or not. While choosing a research site one has to remember bothtime and money.

## 6. Selection of Respondents
Selection of the respondents has to be based on the research problem, topic. There should be proper representation of the sample. Rapport building helps in choosing the sample easily. The respondents need to have proper representation of age group, class, location, caste, gender.

## 7. Data Collection
Data collection can be carried out at different stages or at one time too. Tools like Survey even help in carrying out studies through online.

However, the response rate of online survey might differ specially if the questions are lengthy. However, large scale studies like Census involves door to door collection and it takes years to complete and also requires large amount of field investigators. Reliability and Validity are very important while conducting any research. Stable, reliable, and dependable research methods, tools, data, or outcomes are characterised by reliability. The crucial quality of the entities, processes, or tools are used to measure the various aspects is validity. Objectivityand subjectivity are traits that have an impact on the research findings in hence one has to be cautious about it. Subjective personal bias may spoil research findings. Hence one has to be very careful[vii].

While collecting the sample one's own bias should be avoided completely. This would help in gaining maximum precision and accuracy of the data generated.

Sampling methods like random/ probability/mixed/ non random sampling could also be used.

### 8. Data Processing
- Data management –Data can be stored in proper software. However, there are other steps like routinely logging in and screening your data, entering the data into a programme, and lastly "cleaning" your data.

- Understanding variable types - Different data kinds call for discrete treatment, so it's crucial to look into variables by their measurement scales and both their causes and effects (dependent or independent) (nominal, ordinal, interval, and ratio).

- One also has to run descriptive statistics to identify the key characteristics of a data collection. These statistics include measures of central tendency (mean, mode, and median), dispersion (range, quartiles, variance, and standard deviation), and distribution (skewness and kurtosis).

- There needs to be proper inferential statistics in order to evaluate the researchers' capacity to make inferences that go beyond the data at hand. A sample representing the population, differences between two or more groups, changes through time, or a relationship between two or more variables also helps.

- Make sure you use the appropriate statistical test - this depends on understanding the nature of your variables, their scale of measurement, their distribution shape, and the types of questions you wish to pose.

- Be on the lookout for statistical significance. This is typically represented by a "p-value," which determines the likelihood that your findings are not just a coincidence. Researchers can be more certain that results are real the lower the p-value[viii].

## 9. Data Analysis

In order to determine if one variable significantly affects another, the researcher use a variety of statistical approaches to search for significant correlations between variables.The simplest technique is to arrange the relationship between variables into graphs, pie charts, and bar charts. These tools are essential for communicating the findings of one's quantitative data analysis to others and provide an immediate, "intuitive" visual impression of whether there is a significant relationship. Even codes are used to make the analysis simpler. Data Analysis takes more time at times even more than data collection. As writing is an art which is shapes with time.

## 10. Findings

The researcher must evaluate the analysis' findings in light of the data analysis. The results will become apparent at this point, including if the hypothesis was validated. What consequences do the findings have for the theoretical concepts that served as the research's foundation?This has to be kept in mind. Often Scientists frequently collaborate with one another and discuss their discoveries. As a result, hypotheses are frequently updated, improved upon, or replaced. As a result, knowledge doubles every five to ten years in some areas of science. The proof that science is a useful tool for discovering new things can be seen through research. As all around us, knowledge is expanding[ix]. The interpretation of the results is also an important thing which has to be carried out while reporting the findings.

In some disciplines and among some universities certain research also provide recommendations and suggestions and limitations of the present study and there is further scope in which the study could be carried out while reporting the present finding. The findings also highlight theuninvestigated areas which other research could carry out as an extension of the present research. After all research is a collective effort and scholarship.

## 11. Publishing Results

Once the study is completed, the data could be summarized and published as journal articles, books. At times, the research is also presented in conference. It is also quoted and cited by other scholars. If the paper is bringing about a social change or solving a problem, even policy makers make use of such research. The researcher also can visit and share the findings with the subjects upon whom the research was conducted. This is also part of knowledge sharing process which would help develop trust and even the community to understand their own lives from a third

person's perspective. However, this has to be carried out with certain degree of caution specially in topics like gender, caste, violence etc.

- **Ethics and Legal requirements**

Laws apply to everyone, even researchers. It is forbidden for researchers and research subjects to do anything that is forbidden for the general population ill health or harm. However, researchers who: (a) desire to explore illegal activities; or (b) come across criminal activities while conducting their investigations face a more typical legal conundrum. The courts can demand your data and files in most nations, you are required to disclose illegal activity. Researcher assurances of secrecy may not be upheld in court, according to legal precedence. You do not have the same rights as a lawyer, doctor, or priest as a researcher in several countries hence, one has to be careful.

- **Moral responsibilities**

Some moral factors one has to take into account when conducting research:

One has to be conscious in prioritising the needs of respondents or participants.

There needs to be Equity while studying and asking some demographic groups to participate in research while leaving other groups off-limits.

Honesty is the need of the hour and researchers need be forthright and truthful as well as that results in the specifics of the research process be made transparent.

The professional code, the discipline, and the institution will all have different ethical standards for research conduct, they need to be consulted when in doubt.

A participant can only offer "informed consent" to participate in a research study if they fully understand the request for their engagement, including the time commitment, type of activity, topics that will be covered, and the potential physical and emotional hazards involved. Participants must be competent, autonomous, willing to participate willingly, aware of their right to withdraw, not duped, not pressured, and not induced in order to give their informed consent.One has to make sure there is no physical, emotional, or psychological harm suffered by respondents.

Protecting the identity of persons who provide research data; all identifying information belongs to the researcher alone; ensuring confidentiality and, if necessary, anonymity. Beyond confidentiality, anonymity refers to safeguards against identification, even by the researcher. This has to be done specially in sensitive topics.

**Check Your Progress**

1. Explain hypothesis in few lines

2. Discuss measures to take while selection of respondents

## 5.4 SUMMARY

Quantitative research focuses on collecting numerical data and it is used to understand a specific event or generalise it across groups of individuals.Quantitative methods place emphasis on precise measurements and the statistical, mathematical, or numerical analysis of data gathered through surveys, polls, and other types of research, as well as the manipulation of statistical data that has already been obtained using computing methods[x]. Further, in the chapter we also learnt about the main steps associated with Quantitative Research. The main steps are Theory – Hypothesis, Research Design, Operational Concepts, Selection of Research Site, Selecting Respondents, Data Collection, Data Processing, Data Analysis, Findings/ Conclusion, Publishing Results. We also learnt about Inductive and Deductive approach used in research. Inductive approach is often used in Qualitative research while that of Deductive approach is used in Quantitative research. In Inductive approach Data is gathered then patterns are formed and thereafter theories are used. On the other hand, in the Deductive approach the theories are used first and through that the research is further guided and conducted and hypothesis is tested accordingly in the research. One of the important steps in the present times while doing Quantitative or Qualitative research is that of remembering the ethics while conducting research. There are even Research Ethical committees who go through the proposal before the research begins.

## 5.5 QUESTIONS

1. Discuss the main steps involved in Quantitative research
2. Explain the ethics and legal requirements while conducting quantitative research
3. Discuss the inductive and deductive approach in quantitative research.

## 5.6 REFERENCES

[i]Babbie, Earl R. *The Practice of Social Research*. 12th ed. Belmont, CA: Wadsworth Cengage, 2010; Muijs, Daniel. *Doing Quantitative Research in Education with SPSS*. 2nd edition. London: SAGE Publications, 2010.

[ii]https://scientificinquiryinsocialwork.pressbooks.com/chapter/6-3-inductive-and-deductive-reasoning/

[iii]https://www.merriam-webster.com/dictionary/theory

[iv]https://www.collinsdictionary.com/dictionary/english/hypothesis

[v]http://epgp.inflibnet.ac.in/epgpdata/uploads/epgp_content/social_work_e
ducation/05._research_methodology_and_statistics/08._hypothesis_and_re
search_questions/et/6056_et_et.pdf

[vi]Bryman, A. (2016). *Social research methods*. Oxford university press.

[vii]https://egyankosh.ac.in/bitstream/123456789/11204/1/Unit-1.pdf

[viii]https://study.sagepub.com/oleary3e/student-resources/analysing-
data/steps-in-quantitative-analysis

[ix]L Mitchell, M., & M Jolley, J. (2010). *Research design explained*.

Creswell, J. W., & Creswell, J. (2003). *Research design* (pp. 155-179).
Thousand Oaks, CA: Sage publications.

[x]Babbie, Earl R. *The Practice of Social Research*. 12th ed. Belmont,
CA: Wadsworth Cengage, 2010; Muijs, Daniel. *Doing Quantitative
Research in Education with SPSS*. 2nd edition. London: SAGE
Publications, 2010.

❖ ❖ ❖ ❖

# 6

# WRITING RESEARCH REPORT

**Unit Structure**

## 6.0  OBJECTIVES

- To learn the meaning of Research Report
- To understand the process involved in making a research report.

## 6.1 INTRODUCTION

In this chapter, we will learn about what is research report and what are the steps involved in writing a good research report. This chapter would be an important study material in terms of career view point. For example - After your graduation if you are interested in working with a Non-Governmental Organisation or with a Human resource management – you are often expected to write reports. This could be about a field visit, about a project. After studying a particular company, problem, field or even after completing a project. Hence, learning about this topic becomes very important. A good report can help you get funding too. Companies hires people or even outsources the work to freelancers too to write reports.

In terms of companies before introduction of any introducing a new service, product, or feature, research is essential. Due to the daily influx of new competitors and speed of production research is needed to find out the gap in the products, markets. Even for a firm to remain relevant in competitive market it has to be updated with new products that meet client requests. They also have to make the appropriate decisions must be made at the right time. Hence, these corporates also need people who conduct

research and produce reports. Even PhD, MPhil students are expected to submit six month or every month progress report to show their progress.

The four core documents in any research are that of the research proposal, research summary, research abstract, and the research report—all these are used to evaluate the appropriateness, quality, and validity of every research/study. Among all these the primary document used to assess the contribution of the research is the research report. The process of conducting research includes multiple stages and a flow of documents. The order and stages of the process have a real impact on the quality of the final report and the research materials created at every level.

## 6.2 MEANING OF REPORT WRITING

A research report is an official, formal declaration that includes facts, provides a record of findings, and/or could be the outcome of a survey or study (Booth 1991). The Oxford English Dictionary defines a report as a declaration of the findings of an inquiry or of any subject requiring specific information.

Report writing can be done for a variety of reasons, including: presenting findings; maintaining records of the information/data gathered; and

• In order to record an organization's accomplishments and shortcomings

• To document the development of a project or research in writing.

Research reports are written statistics that are created by statisticians or researchers after information has been collected through organised research, generally through surveys or other qualitative approaches.

Reports typically include a wide range of subjects, but they are primarily concerned with disseminating knowledge on a single subject and a relatively specific target audience. The main goal of research reports is to offer crucial information about a study that marketers can take into account when developing new tactics. The most efficient method of communicating certain incidences, facts, and other information to the decision-makers is through writing research reports. The best research papers contain material that is incredibly accurate and have a clear purpose and conclusion. These reports must be presented in a clear, organised manner if they are to effectively convey information.

A research report is a trustworthy source to retell specifics about a research project and is frequently regarded as a true testament to all the work put in to gather research specifics.

One can find different sections in research like the Overview, Introduction and Background, Applied Techniques, Findings from the analysis, Discussion.

While many of the components and portions of report writing are generic, it is distinguished by certain themes that are unique to it. Reports are written using factual material supported by statistics and findings. The

authors' own biases or sentiments should not have any bearing on the material, which means the research report has to be objective in nature.

Report writing is an art which has to be cultivated. It has to be tailored according to audience needs. For example – If the report is submitted to an academic institution it has to bedesigned differently focussing more on theory, methodology, literature application of theories and the findings and following the format given by the institution. On the other hand, while writing a research report for business project the length can vary from small to big. It often focusses on the study area, field details, key methodology used, findings of the topic. Research Report connected to Social Policies or an investigation related to schemes, its effects, benefits also at times includes the recommendations at the end of the report. As here researchers are trying to solve the problem too than just studying it.

Depending on the goal of the research, the financing or sponsoring organisation, the report has to be altered. The field workreports can be divided into a number of different categories. Reports, such as those on budgeting and other organisational tasks, might be very brief and concise, and it can be published for informational purposes. Another kind of report writing that is frequently used in universities for project documentation are case studies and analyses. An informal format may be used for a report intended for an organization's internal audience. If a report is discussing topics like absenteeism, work plans, or processes, an informal conversational tone may be appropriate. In a report which is written for an employee policy, a manual, or a task report, might be casual, semi format language but can still have a defined framework. The third is a formal report that has a specific format and organisation as well as sections for research, analysis, and also draws inferences.

## 6.3 TYPES OF REPORT

### 1.    Technical Report
This is a detailed report on the entire study process and its findings. It is mostly targeting the academic community. For example, the scientists and researchers. The technical report's focus and even the language are both primarily technical in nature. Like in every other report, the problem is explained, the method(s) and techniques employed are described, along with the results and recommendations.

### 2.    Popular Report
In this kind of report there are more headlines, italicised text, images, and concise phrases in this report. To catch the reader's attention, there is white space and extra margin. This report is built like a journalistic writing. The style encourages the reader to have quick reading and comprehension of the topic and finding.

### 3.    Interim Report
This report is essentially a work in progress. It describes how much work has already been done and what remains. If the report is a funded project,

then the Interim Report is used. These aids the sponsors in maintaining their enthusiasm and potentially continuing the funding. Additionally, it displays the analysis and results up to this point.

**4.    Summary Report –**

This kind of report is written for the general public or lay readership. There are no jargons or technical terms in the language. It is straightforward and explains everything with many images and graphs. The reports often only have two or three pages (Krishnaswami, Ranganatham, 2019).

**Check Your Progress**

1. List out two types of Report.

2. Explain your understanding of Reference

## 6.4 FUNCTIONS OF RESEARCH REPORT

1. The research report provides a way for organising the presentation of the topic under investigation, the approaches and methods utilised for gathering and analysing data, the conclusions, and the suggestions.

2. The research report provides as a foundational resource for creating research projects in the same or related fields in the future.

3. The research report is used to assess the calibre of the finished study job.

4. The research report serves as a tool for assessing the researcher's capacity and expertise to do research

5. The research report offers a factual foundation for developing policies and strategies related to the topic under study.

6. It offers organised knowledge about the themes and problems analysed.

## 6.5 FORMAT OF RESEARCH REPORT WRITING

The specifics of a research report may alter depending on the goal of the study, but the fundamental elements will always be present. The market researcher's method of conducting research has an impact on how reports are written as well. The following are the top seven elements of an effective research report:

• **Research Report Summary:**The first thing a research report consists is a summary/ Abstract. In a business research report – the title used is executive summary instead of summary. The research report summary consists of few paragraphs. Itshould incorporate the complete purpose as well as an outline of the research. Under the report summary, each of the research's numerous components is briefly discussed. It ought to be captivating enough to include all the essential aspects of the report. when PhD students write an abstract in the beginning pages of the thesis. This

abstract explains the objectives, methods used, and overview of the thesis and findings. This helps the examiner/ readers to the get the core idea of the thesis.Unlike when we read the Newspaper title and the first paragraph we get an idea of the content of the article in the same way, the research report summary acts.

### •Research Introduction

Every report has a main objective that the author is seeking to accomplish. These objectives are listed out in the Introduction section of the project. In the Introduction the researcher can discuss goals pertaining to this objective in the introductory part and formulate a thesis that will be used to try and provide a thorough response. What is the goal's present situation? should be addressed in this paragraph. Provide such information in the research report's opening section. Did the researcher successfully complete the goal after the research was completed, or are they still a work in progress? Here the discussion theories used in the research is also made. This is done to produce an interconnection between existing studies and that of the current study. It also creates credibility and shows the research gap of the existing knowledge.. However, theories are not a must in the works which are presented for business reports. Here, the limitations of the study are also discussed in detail. The introduction part of the report also by using different examples and arguments, prove the problem's existence; arouse interest in your work by highlighting its relevance and significance; name the key problems; and give context for your proposed solution.

### • Methodology

The most significant information is located in this portion of the report. Readers can learn more about the subject while also evaluating the quality of the content given. As a result, this section must be extremely informative and cover every facet of the research in depth. Information must be presented chronologically in accordance with its value and priority. As a result, this section must be extremely informative and cover every facet of the research in depth. Information must be presented chronologically in accordance with its value and priority.If a researcher used an established technique to obtain information, they should cite their sources. This section explains what was done during the study, where it was conducted and what was the procedure followed, tools, techniques used, the number of people participated or studied.

Who are the subjects, in (a)? The study's subjects are described in the subsection Age, gender, and other pertinent social or demographic factors are used to describe the subjects. (b) The number of subjects. Mention the overall participant count as well as the split by experimental condition. If any subjects chose not to finish the study, please list their number and the reason. (c) How are the subjects chosen? Describe how the volunteers were chosen for the experiment and how they were divided into groups. Was a randomization methodology employed, for instance, or was a

different approach required? Report any promises or payments made to the subjects (Kabir, 2016).

**Research Results-**

In this section the results are discussed. A summary of the findings as well as the methods used in order to accomplish the objective. Usually, the report's discussion section is where the description follows data analysis.

**Research Discussion -**

Here the findings are covered in great length in this section, which also includes a comparison of papers that might be in the same field. In the discussion area, any irregularity found during research is also discussed in detail. The researcher also needs to make associating with existing literature, studies and explain when they write study reports about how the findings will be applicable in the real world.

- **References for the research and its conclusion-**

Here the researcher has to summarize the research findings and include all authors, articles, and other content sources consulted.This final part should follow APA formatting guidelines. The term references should be at included as subheading at the end of the project; it should not be underlined or included in quotation marks. References cite works that are openly accessible. The page number can be included, especially if it's an article. Reference lists of works cited in your text are required, and the details of it has to be given at the end. = Reference lists must be accurate and comprehensive because they are meant to be used by the reader. The following general subsections make up a reference: Author, Year of Publication, Title, Publication Place. If the work is taken from Edited Book that that has to be mentioned too.

## 6.6 TIPS FOR WRITING REPORTS -

A well written report helps the readers to understand the content very easily. The followings are some pointers through which one could write better research reports:

Research Reporting writing becomes easier if one is organized. For drafting research reports, the arrangement of the survey questions may not be the best or most efficient. The goal is to begin with a wide issue, narrow it down, and then concentrate on a conclusion or support that the study should provide evidence for.

Starting a report takes time and effort and motivation too. Hence, taking slow steps towards the goal of writing helps. Begin with the title and introduction. While choosing a format, keep the target audience in mind and make sure it is simple, comprehensible, and relevant. If the report would be viewed by policy decision-makers then write it accordingly. Be consistent in the language you use, the annexes' numbering, and other areas.

Follow the institutions guidelines for the delivery of research findings, and ensure that the project has covered the objectives of the study. Before submitting a researcher needs to read the entire proposal and confirm that the information, they provide supports the goals that were outlined at the outset. If a researcher speculates, they are directly challenging their own research.

Obtain all the data you can on the study, subject. Speak with other researchers who have studied the similar problems and familiar with the terminology used in the field. When terminology is misused, readers of research reports may become less interested in continuing to read aloud while writing.

Read the report loudly and check if there is any error. If the you an improper sound while reading the report, for instance, if you misread some words, the reader will undoubtedly experience the same thing. If the researcher can't express an idea in a single sentence, then it is too long and needs to be changed so that everyone can understand it. Check your spelling and punctuation. Good practises unquestionably aid in comprehension of the report. Utilize present tense, verbs. Consider utilising the present tense to sound more immediate with your results. Avoid using jargons, slangs.

Explain mainly about the significant findings. Do not mention any data that are not very important. Make an effort to stick to the survey's questions. The graphs need to be easy to grasp by themselves. Give the graphs a title, including the indicators, the size of the sample, and the appropriate phrasing to prevent readers from misreading them. Every portion of the report should be written by the researcher with correctness in both language and details.

Use proper titles which suits the content of the topic, especially for segmentation studies. Use namesthat add life to study.Write a effective conclusion. The research report's conclusion is the hardest to write, but it also presents a fantastic opportunity to shine. Write a thorough synopsis. It can be helpful to begin the conclusion with a specific statement, followed by a summary of the most significant findings from the study, and finally a discussion of any ramifications. It's challenging for writers to spot their own errors. However, they are in charge of what is displayed. Before sending the final document, be sure it has received the approval of the supervisor, teacher in charge of the report.

If there is an earlier report already submitted then we have to inform the reader/ sponsor about what are the new things being included or show the continuation of it too.  However, one has to include some background before describing the content of the present proposal.

1. List out two important points to remember while writing Research Report

2. What does Research Results means in Research Report?

## 6.7 SUMMARY

Report writing is a laborious task since it entails compiling and presenting all the information gathered during field research in accordance with a predetermined research strategy. Writing a report requires a very structured style, which takes time. We also saw in the chapter that there are different types of reports like Interim Reports – which discusses the work in progress, Technical report in which we uses technical information, jargon and caters the scientists and researchers than general public. Conventions have been established to create a standard format suitable for readers and/or audiences. The format and convention of written reports often emphasis on the method used to obtain the data necessary to produce the report.In conclusion, we can conclude that a research report is an authoritative, narrative document that summarises the findings of a research project. It offers incredibly detailed information to a very specialised audience. Its communication style is unpersuasive. The subject being delivered comes first, followed by presentation. It is an easy-to-read, precise style of communication.A well-written research report provides an organised way to convey the examined issue, the data collection and analysis procedures, findings, conclusions, and suggestions.The intention of the research report is to express the core ideas.

## 6.8 QUESTIONS

1.    Discuss the meaning of Research Report
2.    Discuss the different tips while writing research report.
3.    Write the different types of Research Report and the functions of research report
4.    Discuss the format for Research Report

## 6.9 REFERENCES

1. Bryman, A. (1988). Quantity and Quality in Social Research. London: Routledge

2. Bryman, A. (2008). Social Research Methods. Oxford University Press

3. Matt, H., Weinstein, M., Foard N. (2006) A Short Introduction to Social Research. New Delhi: Vistaar Publications

4. Goode, W., Hatt, P. (1981). Methods in Social Research. McGraw-Hill Book Company

5. Somekh, B., Lewin, C. (ed) (2005). Research Methods in the Social Sciences. New Delhi: Vistaar Publications

6. Krishnaswami O.R., Rangantham M, (2019), Methodology of Research in Social Sciences.

7. Datar A, Nicosia N. Junk Food in Schools and Childhood Obesity. J Policy Anal Manage. 2012 Spring;31(2):312-337. doi: 10.1002/pam.21602. PMID: 23729952; PMCID: PMC3667628. This article can be used as a sample for research report. As some points are relating.

8. Kabir, Syed Muhammad. (2016). WRITING RESEARCH REPORT.In book: Basic Guidelines for Research: An Introductory Approach for All Disciplines (pp.500-518)Publisher: Book Zone Publication, Chittagong-4203, Bangladesh

❖❖❖❖

# ASPECTS OF QUANTITATIVE RESEARCH

# 7

## SURVEY METHOD

**Unit Structure :**

## 7.0 OBJECTIVES :

- To understand the term survey as a means for collection of data or information.

- To study survey as a process by which quantitative facts are collected.

- To familarise the students with the organisation of survey is a important form of data collection.

## 7.1 INTRODUCTION :

The word survey has been derived from the word 'Sur' and 'Veeir' which mean 'over' and 'see' respectively. Literally survey means something taken from a high place but it has come to be used as a term with specific purpose of its own. In other words the term Survey is used for method of investigation by direct observation of a phenomena or getting information theory interview or questionnaire thus the term survey is not applicable to direct contact but also to the other means say for collection of data or information.

## 7.2 MEANING AND DEFINITION :

The Survey is defined as fact finding study dealing chiefly with working community.

According to mark Abrams" a social survey is a process by which quantitative facts are collected about the social aspect of community composition and activities from the above definition.

It may be drawn the characteristic feature of a social survey. It is a story of immediate and burning problem of the social and constructive programme of social research for remoxing the social evils the scope of its limited and localized geographically. And it may from the basis of for further social research on the matter. In brief, the social survey is designed the investigate some course relationship of the some aspect of human life. A social survey is to perceive with the aims given below.

## 7.3 AIMS OF SOCIAL SURVEY :

1) As we know that the survey will become a basis for further investigation. Supply of information to needs one is it main objective. The needy one may be an institution or individual thus survey is utilitarian in nature and meant of provide information regarding the practical problems of the society. A researcher of himself may carry out surveys but there are some agencies that carry survey for other.

2) Survey describe the phenomena to a social scientists a survey may have descriptive as a way of studying social condition, relationship and behaviour for example survey communities. Socio economic survey describes the living condition of people of a geographical area. The description would be accurate as well as complete if we became to face with it. Further the purpose of the social survey is to get information and it is not meant to test an hypothesis but simply describe the things therefore survey may be started without any hypothesis the description of data. So collected may serve as a basis for hypothesis later on.

3) Alongside of the description, the social survey explains the determinants for the state affairs may be so specific and purposive. Social survey has intensive usage and is widely used in a number of disciplines. In social sciences it can be used for variety of purpose availability of nature of the source of information is the main and source of undertaking a survey. Broadly the subject matter of social surveys are divided into

1) Demographic Features

2) Social conditions

3) Opinion and attitudes

**Demographic features** come under the purview of organisation and

working of family regarding such information as household composition, marital status are family planning programmes etc. the social conditions to which people are subjective this includes occupation, incomes, a housing amenities many of social action as conditioned by the social condition i.e. Social environment in which people live. So social condition are helpful to get the following information and there upon to establish the casual relationship to study the family relation. One would require information on the location the extent of relationship with elders etc. a survey of deliquenets would in complete unless. Following knowledge of their homes and family environment is available.

**Social activities** such as expenditure pattern, radio listening newspaper reading social mobility information for example to know the expenditure putter of a group of family house. It requires to the survey or following knowledge on expenditure habit say, expenditure towards family, clothing, education, cigarettes, cinema and other.

**Opinion and attitude**, includes information regarding opinion and attitudes of the people toward various factor and the motives and the expenditure of them this information may be necessary as the basis of nature of question that may safety be asked for electing rich response. For eg. The opinion or attitude person toward social economic political, incident survey.

## 7.4 TYPES OF SURVEY :

a)  General or Specific Survey

b)  Regular and Adhoe Survey

c)  Preliminary and final survey and

d)  Census and sample survey

The general survey concerns with collecting general information about population institution or phenomena with out any specific object or hypothesis. The types of survey are mostly taken by the government for providing regular data on may socio economic problems. A good example for this surveys is census of population once in a decade. Any it is termed as specific survey information collected through the specific survey is general of very little outside the problem under study. For example, marketing survey on a particular issue, say people, attitude about the use of television.

Regular survey is survey conducted at regular intervals. To this permanent machinery for collecting information is to be set up. Many of such surveys are mainly economic surveys.  For example, a study of family budget can be cited as  example.  Another example is the rural credit survey of the Reserve Bank of India which collects data regarding the rural conditions annually.

Adhoc Survey is undertaken for all it may be conducted in phases if the area of investigation a large. Because of completion of survey in phases due to its largeness, it never initials as regular or repetitive survey. A preliminary survey is one which is to be undertaken well in advance to get the hand knowledge of the universe to be surveyed. Further, it helps the survey or to get acquaint himself with the nature of the problem so that he can be able to get the rich response. After conducting the preliminary survey only, the construction of schedule or questionnaire may be taken into picture. And it also guides the way of planning as well as organizing The preliminary survey may also be called as A pilot studying which is conducted well before taking a detailed study about the main work. After the pilot study has been completed from the final survey is made.

Every unit is the universe is to be contacted and data collected from it called census survey while a few units of the universe covered it refers to sample survey. Planning of a social survey : the quality of survey results to be considerable degree on the preparation made before the survey is conducted so planning a survey is of paramount importance in social sciences. The researcher has to give careful and conductive though out the planning of a survey. Planning of social research programme must for its successful execution part on remarked that only be carefully planning the survey from start to finish can reliance be place upon results and in many cases will the finding ever rich the publication stage therefore it is pertinent to think twice properly at the planning of a social survey with regards to the following points. Scope of the survey : next to the purpose of survey, scope of it is important step regarding to the type of information, subject matter, geographical area. For instance, an enquiry may reliable to India or a particular style or an industrial level town or a particular industry in particular location.

## 7.5 METHOD OF STUDY :

The method of study : in selecting a particular method for a survey. The researcher should take the consideration of its suitability and has own knowledge of it.

Unit of data collection and the unit or units of universe must be clearly defined for the purpose of an enquiry before taking the task of collection information. However, defining the units is not as simple as it appear to be for eg. to study the size of a mill we have different oriteria of measuring the size of a small such as capital employed number of employees production etc. for efficient collection of information the unit of data collection should have the characteristic which are as under.

a) the purpose of enquiry being studied to the unit.

b) the unit should be specific.

c) the unit should be stable in character.

d) the unit should be uniform though out the report.

*Sources of Data :*

After the purpose, scope and methodology have been defined clearly and properly the investigator has to plan about the sources of data the sources of data may be either primarily or secondary. Either the two devices or only one device to be used in the investigator may be depended upon the objective and purpose of the survey. Quite often in social sciences investigation both may be used mode of data collection of social phenomena is very complex and influence by a number of variables. Therefore, it is essential to decide hand what type of information is to be collected adopting the questionnaire or schedule.

*Organisation :*

Organisation of survey is very much important from the point of getting quantitative as well as qualitative data. A number of field worker may be required and training must be given to them if necessary. Arrangement with regard to checking and supervision should be made so as to avoid giving false information. Editing and coding if needed have to be undertaken. Following it classification and analysis of data has to be carried out. All this organisation work should be pre planned.

*Report :*

After the data have been classified and analysed the next step in a survey is to be drafting the report to whom it meant determines the style of presenting the report.

## 7.6 ADVANTAGES :

The following advantages have with the survey method in compensation to other methods they are :

1) The researcher comes in contact with the people whom he wants study. And with this he can observe things personally with relaties and there upon his inferences are not based upon any other theory or dogma till upon the fact of the life.

2) The survey method leads is greater objectivity many field workeruse in the survey remove possible faios and collect correctinformation.

3) Survey may very well lead to the introduction of new theory for example poverty was regarded as the course of crime for fairly long time till increasing crime in advanced countries have falsified this theory.

4) Survey method enables to have fill knowledge of social institution. The actual experience with the situations amounts none to any amount of investigation.

## 7.7 DISADVANTAGES :

1) For carrying the survey field workers are to be recruited and training is to be given to them. All this need for money.

2) It is a prolonged and time consuming process.

3) The reliability of the data collected through survey is not always without doubt, honestey and efficiency of the field-operation of the respondents the suitability of schedule etc. Paramount for collecting valuable data. But all these requirement are very seldom and the data collected may obtain invalid data.

4) Survey is conducted on sample basis and therefore it is subject of the sampling error.

## 7.8 SUMMARY :

A social survey in its broder sense, has a reference to a first hand investigation analysis and co-ordination of economic, sociological and other related aspects of a selected community group. A survey may be undertaken with the primary purpose of a selected community or group. A survey may be undertaken with the primary purpose of formulating a programme for amelioration of the conditions of life and work of a community or a group, implying some frame in the mind of the survey or as to what the conditions ideally ought to be. The purpose of a social survey may also be used to provide scientifically gathered facts or materials affording some empirical basis for the social theorist to set up their conclusions.

## 7.9 CHECK YOUR PROGRESS :

1) What do you mean by Survey?

2) Discuss aims of Survey Method.

3) Explain various types of survey method.

4) Highlight advantages of survey.

5) Analyse disadvantages of survey.

## 7.10 REFERENCES :

1) Wilkinson and Bhandenkar : Methodology and Techniques of Social Research, Himalaya Publishing House, Mumbai 1977.

2) Denzine N. K. and Lincoln, Y. S. (Eds) Handbook of Qualitative Research, London Sage Publications 1994.

3) Bryman A; Quantity and Quality in Social Science Research,London

Routledge 1988.

4) Dillon W. R. and Goldstein M. : Multivariate Analysis Methodsand Applications, New York, John Wiley and Sons 1984.

## 7.12 QUESTIONS :

1) Discuss survey method and highlight advantages anddisadvantages of survey.

2) What do you mean by survey method? Explain advantages and disadvantages of survey method.

❖❖❖❖

# 8

# QUESTIONNAIRE

**Unit Structure**

8.0   Objectives

8.1   Introduction and Meaning

8.2   Objectives of Questionnaire

8.3   Advantages of Questionnaire

8.4   Disadvantages of Questionnaire

8.5   Techniques of constructing a Questionnaire

8.6   Summary

8.7   Questions

8.8   References

## 8.0 OBJECTIVES

1) To make students aware about the tealines , uses and importance of the Questionnaire method as a tool of datacollection.

2) To Tamilaise students to understand Questionnaire method as scientific method to collect reliable method for collection of data.

## 8.1 INTRODUCTION AND MEANING

It is defined as "A list of questions given to a human for them to answer. It secures standardized result that can be tabulated and tested statistically. (Bogandus – Sociology) "A questionnaire refers to a device for securing answers to questions by using a form which the respondent fills in himself. (Goode Half- Methods of social Research) Questionnaire is a set of predetined questions. Generally it is mailed to the respondents for collecting data. It is employed when the area of study is wide and the subjects are widely dispersed. In this method the researcher does not collect the data by himself. He relies on the information provided by the respondents.

## 8.2  OBJECTIVES AND TYPES OF QUESTIONNAIRE

There are two basic objectives of Questionnaire :

1)   To collect information from the respondents who are scattered in wide area.

2)   To achieve success in collecting reliable and dependable information.

60

**Types of Questionnaire :**

a) **Structured Questionnaire :-**

This questionnaire is named before the study is started and it is not

possible to change it offer beginning the study.

b) **Non Structured Questionnaire :-**

This kind of Questionnaire is used more like a guide. It consists of definite subject matter areas, the coverage of which is required during the process of data collection. It is generally used in the technique of interview in which case it is called an interview schedule.

Questionnaire is also divided on the basis of nature of the questions that it contains. They are closed, open, pictorial and mixed. Closed questionnaire usually contains itemized answers to the questions being asked various alternatives to the real answer are also given. The respondent is only to select the answer and put it down.

Open Questionnaire is just the reverse of the closed questionnaire. It is used in the cases where new facts are to be found out. The respondent is given the liberty to express his views freely.

Pictorial questionnaire is similar to closed types of questionnaire. Generally it is use for children of small age group on illiterates on person with lower. I.Q. Along with the questions, pictures showing the meanings of those questions one given. The respondent is required to give out the answers on the basis of pictures.

Mixed questionnaire is neither completely closed non open.

It consist of both the type of questions. Since it is combination of the types of the questions it is popular in social research.

## 8.3 ADVANTAGES OF QUESTIONNAIRE:-

Questionnaires are very cost effective when compared to face interviews. This is especially true for studies involving large sample size and large geographic areas. Written questionnaire become even more cost effective as the number of research questionnaire increases.

Questionnaires are easy to analyze. Data entry and tabulation for nearly all surveys can be easily done with many computer software packages.

Questionnaires are familiar to most people. Nearly everyone has had some experience completing questionnaires and they generally do not make people apprehensive.

Questionnaires reduce bias. The researcher's own opinion will not influence the respondent to answer questions in a certain manner. There are no verbal on visual clues to influences the respondent.

Questionnaires are less intrusive than telephone on face-to- face surveys. When a respondent receives a questionnaire in the mail, he is free to complete the questionnaire on his own time-table. Unlike other research methods, the respondent is not interrupted by the research instrument.\

**Check your profress :**

1) Discuss the meaning of questionnaire.

2) What are advantages of Questions.

## 8.4  DISADVANTAGES OF QUESTIONNAIRE :-

One major disadvantage of written questionnaire is the possibility of low response rates. Low response is the curse of statistical analysis. It can dramatically lower own confidence in the results. Response rates very widely from one questionnaire to another (10%-10%), however, well-designed studies consistently produce high response rates.

Another disadvantage if questionnaire is the inability to probe responses, Questionnaires are structured instruments. They allow little flexibility to the respondent with respect to response format.

By allowing frequent space for comments, the researcher can partially overcome this disadvantage. Comments are among the most helpful of all the information on the questionnaire, and they usually provide insightful information that would have otherwise been lost.

Nearly ninety percent of all communication is visual. Gestures and other visual cues are not available with written questionnaires. The lack of personal contact will have different effects depending on the type of information being requested. A questionnaire requesting factual information will probably not beaffected by the lack of personal contact. A questionnaire probing sensitive issues on attitudes may be severely affected.

Finally, questionnaires are simply not suited for some people. For example, a written survey to a group of poorly educatedpeople might not work because of needing skill problems. People are turned off by written questionnaires because of misuse.

Check your progress :

What are the disadvanteges of questionare.

## 8.5  TECHNIQUE OF CONSTRUCTING AQUESTIONNAIRE

In case, the study is to be conducted through questionnaire method, the questionnaire has to be drawn up in a scientific manner. The framer of questionnaire should keep certain things in view will constructing this device. In this method the respondent gives his answers from a distance. The language and the wordings of the questions should be stimulating to the respondents to give replies. The psychology of the respondent should

be kept in mind and the questionnaire should be framed keeping in view the factors that are likely to encourage him to give correct answers.

**1)** *Number of questions :-*

A questionnaire should be contain a large number of questions. If there are two many questionnaire, generally respondents loose interest and start giving irrelevant answers. Thus a questionnaire should neither be too long no to brief.

**2)** *Questions should be unambiguous, clean and simple :-*

Double barreled questions should not be used. One should not include two or more question in one. (e.g. Does your department have a special recruitment policy for racial minorities and women). Such a question typically leads to hesitation and indecision on the part of the respondent. Some wounds are themselves vague and ambiguous. Themes such as social integration for example, many not well known to the respondents. The meaning of some wounds may be known only to highly educated respondents. Slang on colloquial phrases may be known only to one group, on may have different meanings to different groups. Such differences can present a real communication problem if the group of respondents is not homogeneous. After resolving to avoid ambiguous wordings is appropriate. This vary often depends upon the educational level of the respondents. Many researchers feel that they should phrase their questions in the respondents everyday slang so as to maximize rapport between respondent and researcher. This is perhaps on those matters for which there is no right or wrong choice. Moreover the questions should refer to concrete and specific matters like age on sex are specific but opinion questions are especially difficult. The respondent often does not have an option because he or she has never thought about the topic. He or she is concerned about appearing stupid and must be measured that there is no right or wrong answer.

**3) Leading questions :-**

Questions should be carefully structured in order to minimize the probability of biasing the respondents answer by leading him or her and thus artificially increasing the probability of a particular response. The researchers task is to avoid leading questions as for as possible or to use neutral wordings instead.

**4) Technical and special words should be clearly explained :-**

While using technical jargon the researcher needs to provide an adequate explanation to all the words so as to enable the respondent to understand it in a connect way.

**5) Personal questions :-**

Very personal questions should be avoided . Respondents are generally unwilling to write down such information. Personal questions should be

asked only if completely required.

Besides these considerations certain technical considerations should be kept in mind like

a) Quality of paper used.

b) Response category format.

c) Mailing facilities etc.

d) Layout of a questionnaire

When technique of questionnaire is used in scientific surveysand when the sample size is big ; it is always advisable to attach a <u>Covering letter</u> to the questionnaire. A good covering letter includes the following.

1) A brief introduction of a researcher/researchers and basic information about the institution involved in the process ofresearch.

2) Statement regarding the purpose of survey undertaken.

3) Enclose a self-addressed envelope for the respondents convenience in returning the questionnaire.

4) Assume the respondent that the information he gives will be kept confidential.

5) Promise the respondent that he will bet solicited after he fills up questionnaire.

6) If the respondent is interested, promise a copy of the results of the survey to him.

7) If possible and if required affer social incentives to there spondents on the return of the questionnaire.

## 8.6  SUMMARY :-

The use of questionnaire depends on devoting the right balance of efforts to the planning stage, rather than rushing too early administering the questionnaire. Therefore the researcher should have a clear plan of action in mind and costs, production, organization, time limit and permission should be takencare in the beginning while designing of questionnaire, the characteristics of a good questionnaire should be kept in mind.

## 8.7  QUESTIONS :-

1) Explain questionnaire, Discuss advantages and disadvantages of questionnaire?

2) Write a detail notes on questionnaire method.

3) Discuss meaning of questionnaire.

4) Analyze objectives of questionnaire.

5) What are the advantages of questionnaire ?

6) Explain disadvantages of questionnaire

7) What are the techniques is conduct questionnaire explain.

## 8.8 REFERENCES :-

1) Anthony Gliders, Positirsm and sociology, Cambridge, Gower1987.

2) Baily Kenneth, Methods in social research, The freepren 1978

3) Ker linger F.N., Foundations is behavioral Research, HiltRinehart and Winston Inc 1964.

4) Kumar Ranjit, Research Methodology Ed. 2, Pearson Education,2006.

❖❖❖❖

# SAMPLING TECHNIQUES

**Unit Structure**

## 9.0 OBJECTIVES

➢    To know the principal and objective and concept of sampling

➢    To know the advantages and disadvantages of sampling

➢    To know various types of sampling and its application in statistics

## 9.1 INTRODUCTION:

Sampling is the process of drawing sample from the population. A sample is a portion, selected from the universe or population in statistics.

In statistics 'population" means "all individual things, event or observation that a particular study wants to cover for e.g. the population of a study dealing with the views of college students studying in various colleges in

the city. The sample is considered to be the representative of the universe. If the sample has been properly selected and its size is appropriate, whatever holds good for the sample also holds good for the universe.

## 9.2 OBJECTIVES OF SAMPLING:

i. Selecting the sample of adequate size

ii. Collecting the information

iii. Making Inferences about the population

## 9.3 MERITS OF SAMPLING:

**1.** A sample survey is time saving and less expensive as compared to census survey.

**2.** A sample survey requires small administrative organization because;

  **i.** The field of survey is small.

  **ii.** Number of staff needed is small

  **iii.** The volume of information to be collected and processed is small.

**3.** The result obtained from the sample method is accurate and more reliable then the census method. In this method the investigation comes to know about the statistical error.

**4.** Since the coverage is limited detailed information can be obtained.

**5.** If the population is very large, hypothetical sampling is the only method of studying the population characteristic.

## 9.4 DEMERITS OF SAMPLING:

1. This method is not suitable where high degree of accuracy is required.

2. In the absence of expert investigations, the result obtained from this method cannot be realied on.

3. This method is not suitable when there is heterogeneity.

4. If care is not taken in the selection of sample the conclusion drive from them will be mis-leading.

## 9.5 CLASSIFICATION OF SAMPLING METHOD:

Sampling method can be classified into two broad categories,

**a)**    **Probability sampling :-**

In probability, selection of each respondent is known.

**b)**    **Non-probability sampling:-**

In non-probability sampling the probability of selection is not known.

## 9.6 TYPES OF PROBABILITY SAMPLING:

Hess (1985) in her historical analysis of sampling provides many good examples of different types of probability which are as follows.

### 9.6.1 Random Sampling:

The best known form of probability sampling is random sampling. In a random sampling each person in the universe has an equal probability or chance of being selected regardless of similarities or differences among them as long as they are members of same universe.

All that is required to conduct a random sample is to select a person without showing bias for any personal characteristics. Accuracy of the random sample depends on the accuracy sample frame. If same person are listed for more than once, then they will have greater probability of being selected. If other persons are omitted from the list they will not be selected at all. In either of the cases the sampling will not be random.

Random sampling is used to obtained a sample i.e. most likely to be representative of the population. Random sample has a technical meaning in statistics. It does not mean hap-hazards or unplanned, when items are not selected by choice but by chance. It is called random sampling. This method is also known as 'chance selection.'

The following methods are used for obtaining a random sampling:

- Lottery method.

- Table of random numbers.

- Arrangement of all number in same order and every 5th, 10th, 100th or nth unit is selected.

One important property of simple random sample is that the larger the size of sample the more likely is that its mean will be close to the population.

**Merits of Random Sampling:**

i.     Since the selection of the item in the sampling depends entirely upon chance there is no possibility of personal bias affecting the result.

ii.    Compared to judgment a random sample represents the universe in a better manner.

iii.   We can calculate the margin of errors because the sampling error follows the principle of chance.

**Demerits of Random Sampling:**

i. Simple random sampling requires a complete catalogue universe from which a sample is drawn. However, it is very difficult to have up-to-date list of all the items of the population to be sampled.

ii. In a field survey the cases are selected by random tend to be too widely dispersed geographically. Therefore, it is costing in terms of time and cost of collecting data.

iii. The size of the sample requires ensuring satisfied reliable results in usually larger under stratified sampling then random sampling.

### 9.6.2 Stratified Sampling:

In this method the population is divided into different groups or classes called a strata and a sample is drawn from each stratum at random. The purpose of stratification is to increase the efficiency of sampling by dividing a heterogeneous population in such a way that there is great homogeneity with each strata and marked difference between different strata.

A stratified sample is controlled so that it reflect exactly some known characteristics of the population. In a stratified sample everything is not left to the chance. For.eg: In a public opinion poll it is important that the sampling reflects the ratio of population who are Hindus, Muslims, Christians and others.

Stratified sampling of a group or ranked ordered such as professors associated professor assistant professors. Stratified sampling consist of listing all professors together in a homogeneous group, then all professor and assistant professors after that is done a random sampling is drown within each group the procedure not only avoids the biases but also saves time and money.

Stratified sampling is not limited to stratification of only one variable. One can stratify on two more variable simultaneously.

### 9.6.3 Multistage/Cluster/Area Sampling:

Cluster sampling is sometimes called as area sampling cluster sampling samples among clusters. It is mainly concerned with the particular geographical area or a particular aspect of population blocks.

Under this method the random selection is made at primary, intermediate and final units from a given population. At first the 1st stage unit are sampled by some suitable methods such as simple of 2nd stage unit is selected from 1st stage unit. Again by some or different form the method used from the 1st stage unit. Further stage may be added be added as required. For.eg:- as we take e.g. of 10,000 students from Bombay university we take colleges as 1st units, then draw departments at the 2nd stage and choose students at the 3rd and the final stage.

### Merits:

i. When the area of inquiring is wide this method is used.

ii. It brings flexibility in sampling which is lacking in other methods.

iii. The obvious advantage of cluster sampling is great saving in time and money.

**Demerits:**

a. This method is very complicated and less accurate than other method of sample selection by a single stage process.

b. It is not a single sample but two or more, with a possibility of sampling error in each.

c. This means that the investigator must be concerned about sample size and accuracy not once but at every stage.

## 9.7 TYPES OF NON-PROBABILITY SAMPLING:

Non-probability sampling cannot claim to be representative since the probability of person to be chosen is not known. This greatly limits the investigators ability to generalize his/her findings beyond the specific sample studies.

The obvious advantage of non-probability sampling is that it is much less complicated, much less expensive and may be done on a spur of the movement basis to take advantage of available response without the statistic of the probability sample. A non-probability may prove perfectly adequately if the researcher has no desire to generalize his/her findings beyond the sample or if the study is nearly a trial run for a larger study.

### 9.7.1 Convenience sampling/accidental sampling:

In convenience or accidental sampling the investigator chooses the closet person as respondent, this method is also known as 'hit' or 'miss' or 'chunk.' A 'chunk' is a part of the population which is selected by convenience. The investigator selectes certain items from the domain as per his/her convenience that is the method by which a tourist studies the country of his visit. He comes across certain people and things have transaction with them and then try to generalize about the entire population in his travelogue.

What is lost in sampling accuracy is saved in time and money. If a person is to submit a report from labor management relation in textile industries, he is following convenient sampling. It is suitable for making pilot studies. However, the results rarely represent the population. They give biased unsatisfactory results.

### 9.7.2 Quota sampling:

It is type of judgment sampling. In a quota sampling, quotas are set according to some specific characteristic. For.eg:- sex, occupation,

education, age, etc. i.e. so many in each occupation, so many in each age and so on. Each interviewer is asked to interview a certain number of people who constitute his quota. Within the quota the selection of sample items depends on personal judgment. For.eg:- in a radio listening survey the interviewer may be asked survey the interviewer 500 people living in a certain area and that out of every 100 person interview 60 are to be housewives, 25 farmers and 15 children within this quota the interviewer is free to select people.

The cost per person's interview may be relatively small for quota sample but there are numerous opportunities for biases which may invalidate the results. Because of the risk of personal prejudice and bias the quota sample is not used in public opinion studies.

The success of this method depends upon the integrity and professional competence of investigations. It provides satisfactory results if the interviewers are carefully trained and if they follow their instruction closely.

### 9.7.3 Purposive or judgment sampling:

In this method investigation has complete freedom in choosing his sampling according to his wishes and desire. To choose or to leave the item for the purpose of study depends entirely on the wishes of the investigation and he chooses items which he thinks are the most representative of the universe. For.eg:- if a sample of 10 students is to be selected from the class of 60, the investigation would select 10 students who in his opinion are representative of the class.

Use of judgment sampling is justified where a small number of sampling units are there in universe or when we want to study some unknown traits (quality) of population or in solving everyday business problems. Executive or public officials are often pressed for time. Judgment sampling is then the only practical method to arrive at solution to the urgent problems.

The disadvantage of this method is that although this simple, it is not scientific because the sample units may be biased by the personal prejudices of the investigation. Since an element of subjectiveness is possible. This method cannot be recommended for general use.

The success of this method depends upon the excellence in judgment. If the individual making the decision is knowledgeable about the population and has good judgment only then the resulting sample may be representative.

## 9.8 Questions

1) outline sampling & explain its classification.

2) What do you understand by the concept orf sampling. Explain probability/ non-probability sampling in details.

## 9.9 REFERENCE AND FURTHER READINGS

➢ Best, J., Kahn, J. (2008) Research in Education (10th ed.). Prentice Hall. Pearson Education

➢ Bryman, A. (1988). Quantity and Quality in Social Research. London : Routledge

➢ Bryman, A. (2008). Social Research Methods. Oxford University Press

➢ Goode, W., Hatt, P. (1981). Methods in Social Research. McGraw-Hill Book Company

➢ Matt, H., Weinstein, M., Foard N. (2006) A Short Introduction to Social Research. New Delhi: Vistaar Publications

➢ Somekh, B., Lewin, C. (ed) (2005). Research Methods in the Social Sciences. New Delhi: Vistaar

❖❖❖❖

<div align="right">

# 10A

</div>

# MEASURES OF CENTRAL TENDENCY

**Unit Structure:**

10A.0   Objectives

10A.1   Introduction & Meaning

10A.2   Types of Average

10A.3   Summary

10A.4   Check your progress

10A.5   References

10A.6   Questions

## 10A.0 OBJECTIVES

- To acquaint students with statistical terms.

- To familiarize students with Mean, Median & Mode for presenting data.

## 10A.1 INTRODUCTION & MEANING

One of the objectives of the analysis of data is to get one single value which can describe characteristics of the entire mass of the data, which can be considered as representative of the entire distribution. A value satisfying this criterion is a **central value** or an "**average**"

In practice, the word **"average"** is used with different meanings. For instance, an average student, average height of boys, average Hindi film, average actor, average income, etc. In some cases, we use the term **"average"** to denote a mediocre type e.g. average student, average actor, average film, etc. In some other cases by the expression "average" we mean "typical" or "usual" e.g. average Indian, average housewife etc. In statistatical terms the average refers to a value of obtained  by a  specific process like average height or average income.

In Statistics, the average is representative or typical value of the data. It usually lies somewhere near the centre of the group and that is why the averages are termed as measures of Central Tendency or Central Value. It depicts the main characteristics of the data. Large volumes of data cannot be easily understood or remembered. So a single value, summarizing prominent features of the data is needed.

If two or more sets of data are to be compared, it is not possible to compare each and every item. So, we require  one figure, representing the

entire data in condensed form. For example, average salaries of employees of two companies of sametype can be compared. Suppose these are Rs. 2,500 and Rs. 2,150. The employees of the second company can demand a raise in salary based on these results, or suppose, average marks at the terminal examination of students of two divisions of F.Y.B.Com. are 65.2 and 45.8 respectively. Then, some arrangement of special coaching can be made for students of second division. Thus, averages can facilitate inter-comparison of different characteristics.

While drawing conclusions, care has to be taken to study the number of forces affecting the data. For instance, in the previous example of students of two divisions, the divisions might have been formed according to marks at H.S.C. examination and the first

division may have students with higher percentage, which explains the average marks of 65.2 at terminal examination. Another points to be noted is the same type of measure must be used to compare two or more sets of data.

**Requisites of a Good Average:**

1. It should be easy to understand and easy to calculate.

2. It should be based on all observations.

3. It should be capable of further algebraic treatment.

4. It should not be affected by extreme values.

5. It should not be affected much by sampling fluctuations.

## 10A.2 TYPES OF AVERAGES

The averages can be classified into two groups - **mathematical averages and positional averages.**

The **mathematical averages are** based on all observations and they are calculated using mathematical formulae. The averages are:

(i) Arithmetic Mean

(ii) Geometric Mean

(iii) Harmonic Mean

However, we will study only the first two averages.

The **Positional averages** are based on only some of the observations and are located at a specific place in the sets. They are also called **"measures of location."** They are:

(i) Median

(ii) Mode

Let us study these measures in detail.

**Arithmetic Mean:**

The most popular measure of central tendency is Arithmetic mean. It is the arithmetic mean, which is referred to by a common man as an "average".

In simple terms, the mean is computed by dividing sum of all observations by the number of observations.

**For ungrouped data** (where frequency distribution is not formed)

If there are n observations in the set and these represent different values $X_1, X_2, X_3, \ldots \ldots X_n$ of the variate x, then the arithmetic mean, denoted by $\overline{X}$, is calculated as follows.

$$\overline{X} = \frac{Sum\ of\ the\ Observations}{Number\ of\ the\ Observations}$$

$$= \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{n}$$

If we denoted the sum $X_1 + X_2 + X_3 + \ldots\ldots + X_n\ as \sum X_1$

$$\boxed{\overline{X} = \frac{\sum x}{n}}$$

**Note:** We may choose any constant as the value of C.

| Daily Wages in Rs.(Class-Interval | Number of Workers f | ClassMarks x | u | fu |
|---|---|---|---|---|
| 50 - 55 | 10 | 52.5 | 0 | 0 |
| 55 - 60 | 22 | 57.5 | 1 | 22 |
| 60 - 65 | 30 | 62.5 | 2 | 60 |
| 65 - 70 | 20 | 67.5 | 3 | 60 |
| 70 - 80 | 12 | 75.0 | 4.5 | 54 |
| 80 - 100 | 6 | 90.0 | 7.5 | 45 |
| Total | 100 | | | 241 |

Now, $N = \sum f = 100, \sum fu = 241.0, A = 52.5\ and\ C = 5$

The formula is

$$\overline{X} = A + C \times \left[ \frac{\sum fu}{N} \right]$$

**75**

$$= 52.5 + 5 \times \left[ \frac{241.0}{100} \right]$$

$$= 52.5 + 12.05 = 64.55 = Rs.64.55$$

So, the average daily wages are Rs. 64.55

**Merits of Arithmetic Mean:**

(i) It is easy to understand and easy to calculate.

(ii) It is rigidly defined, so that a unique answer is obtained.

(iii) It is based on all the observations, which can be seen from the

formula $\overline{X} = \dfrac{\sum x}{n}$

(iv) Further mathematical treatment is possible in case of arithmetic mean. For instance, mean for the combined group can be calculated knowing means of individual groups.

(v) If number of items and their average are known, the sum of the values of these items can be directly obtained.

(vi) As sum of the deviations of the values from the arithmetic mean is zero, it balances the values on either side of it. So, it is better representative than any other average.

(vii) It is less affected by sampling fluctuations and so it has a sampling stability.

**Demerits of Arithmetic Mean:**

(i) As its computation requires all values, if some values are not known, it can not be calculated.

(ii) It is a value which may not be present in the data. That is, there may not be even one item whose value coincides with value of the arithmetic mean.

(iii) Sometimes, it may give absurd results like the average number of students per class is 50.4.

(iv) It is affected by extreme values, i.e. those, which are either too large or too small.

e.g. the mean of 50, 75, 65, 57, 48 is 59.

If we consider the last observation as 480 instead of 48, i.e. the mean of 50, 75, 65, 57 and 480 is 145.5.

(v) In case of open end class intervals, the arithmetic mean can not be computed, unless some assumption about size of class intervals is made.

**MEDIAN:**

Median is an important measure of **location**. When the raw or ungrouped data are arranged in ascending or descending order, the **middle observation** or the arithmetic mean of two middle observations is the **median**. The median can be obtained without any formula as follows.

i) For Ungrouped Data :

If $x_1, x_2, \ldots \ldots x_n$ are n observations, arranged in order, then median is defined as Median $= \left( \dfrac{n+1}{2} \right)$th observation if n is an odd number = arithmetic mean of $\left( \dfrac{n}{2} \right)$ the observation and $\left( \dfrac{n}{2} + 1 \right)$ the observation if n is even.

**Illustration 1:**

Find median for the following set of observations 53, 42, 30, 55, 75,

50, 32, 39, 62.

**Solution:**

First arrange the values in ascending order as 30, 32, 39, 42, 50, 53, 55, 62, 75.

Now, n = no. of observations = 9

So, the middle observation is the fifth observation. Hence, median = 50.

**Illustration 2:**

Find median for the following data containing ten observations. 93, 32, 47, 55, 78, 65, 95, 100, 86, 70.

**Solution:**

First arrange the data in descending order as 100, 95, 93, 86, 78, 70, 65, 55, 47, 32. Now as the number of observations is ten, an even number, there is no single middle observation. But the pair (78, 70) can be considered as the middle pair.

So, median = average (arithmetic mean) of the pair $\dfrac{78 - 70}{20} = 74$

Hence, median is 74.

**(i) For Grouped Data:**

Consider the case of discrete variate.

Median is defined as the value of the variable, for which cumulativefrequency exceed $\frac{n}{2}$ where N represents the total number of observations.

**Illustration 3:**

Find median for the following data representing the ages in years of children.

| Age in Years | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|
| No. of Children | 14 | 20 | 40 | 54 | 40 | 18 | 7 | 7 |

**Solution:**

We prepare the table of cumulative frequencies "less than" type.

| Age<br>x | No. of Children<br>f | Cumulative Frequency<br>(less than) Cf |
|---|---|---|
| 3 | 14 | 14 |
| 4 | 20 | 34 |
| 5 | 40 | 74 |
| 6 | 54 | 128 |
| 7 | 40 | 168 |
| 8 | 18 | 186 |
| 9 | 7 | 193 |
| 10 | 7 | 200 |
| Total | 200 | |

Here, $N = \sum f = 200 \therefore \frac{N}{2} = \frac{200}{2} = 100$

By comparing 100, with the cumulative frequencies, we get 128 as the first cumulative frequency exceeding 100 i.e. $\frac{N}{2}$. The value of x, corresponding to the cumulative frequency 128 is 6, which gives the median value.

So, the median is 6.

**Continuous Variate Case:**

Now, consider the case when the variate is continuous expressed with the help of class intervals.

Firstly, the median class is located as the one for which the cumulative frequency exceeds $\frac{N}{2}$.

The following procedure is followed.

Let $l_1$ = lower class limit of the median class

$l_2$ = upper class limit of the median class

f = frequency of the median class

cf = cumulative frequency of the premedian class

Now, median is calculated using the formula,

$$Median = l_1 + \frac{l_2 - l_1\left[\dfrac{N}{2} - cf\right]}{f}$$

Now consider the following example.

**Illustration 4:**

Calculate the median for the following data:

| Monthly Income in Rs. | No. of Families |
|---|---|
| 1500-1700 | 70 |
| 1700-1900 | 100 |
| 1900-2100 | 120 |
| 2100-2300 | 150 |
| 2300-2500 | 100 |
| 2500 - 2700 | 60 |

**Solution:**

We prepare the table of cumulative frequencies to locate the median class.

| Monthly Income in Rs. | No. of Families | Cumulative Frequency |
|---|---|---|
| 1500-1700 | 70 | 70 |
| 1700-1900 | 100 | 170 |
| 1900-2100 | 120 | 290 |
| 2100-2300 | 150 | 440 |
| 2300-2500 | 100 | 540 |
| 2500-2700 | 60 | 600 |
| Total | 600 | |

Now, $N = \sum f = 600,$ an even number $\dfrac{N}{2} = 300$

By comparing $\dfrac{N}{2}$, that is, 300, with the cumulative frequencies, find the first cumulative frequency exceeding 300. It is 440, so the corresponding class interval 2100-2300 is the median class.

Here, $l1 = lower\ class\ limit = 2100$

$l2 = upper\ class\ limit = 2300$

$f = frequency = 150$

$cf = cumulative\ frequency\ of\ the\ previous = 290$

Substituting these values in the formula,

$$\text{Median} = l_1 + \frac{l_2 - 1\left[\dfrac{N}{2} - cf\right]}{f}$$

$$= 2100 + \frac{2300 - 2100}{150} \quad \frac{300 - 290}{}$$

$$= 2100 + \frac{200}{150} \quad \frac{10}{} = 2100 + \frac{2000}{150}$$

$$= 2100 + 13.3333 = 2113.3333$$

So, the median income is Rs. 2113.33

**Illustration 5:**

Find the median for the following data:

| Class Interval | 5-9 | 10-14 | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 |
|---|---|---|---|---|---|---|---|
| Frequency | 8 | 18 | 27 | 21 | 10 | 8 | 7 |

**Solution:**

First we take the class intervals exclusive type, by adding $\dfrac{10-9}{2} = 0.5$

to the upper class limits and subtracting 0.5 from the lower class limits. So, the new classes with class boundaries are 4.5-9.5, 9.5-14.5 and so on. The following table is prepared to find cumulative frequencies and the median class.

| New Class Interval | Frequency | Cumulative Frequency |
|---|---|---|
| 4.5 - 9.5 | 8 | 8 |
| 9.5 - 14.5 | 18 | 26 |
| 14.5 - 19.5 | 27 | 53 |
| 19.5 - 24.5 | 21 | 74 |
| 24.5 - 29.5 | 10 | 84 |
| 29.5 - 34.5 | 8 | 92 |
| 34.5 - 39.5 | 7 | 99 |
| Total | 99 | |

Now, N=99 and $\dfrac{N}{2} = \dfrac{99}{2} = 49.5$

After comparing 49.5 with the cumulative requencies, as 53 exceeds 49.5, the median class is 14.5-19.5.

$l1 = 14.5, l2 = 19.5, f = 27, cf = 26$

Using the formula,

$$\text{Median} = l_1 + \dfrac{l_2 - l_1 \left[ \dfrac{N}{2} - cf \right]}{f}$$

$$= 14.5 + \dfrac{19.5 - 14.5 \quad 49.5 - 26}{27}$$

$$= 14.5 + \dfrac{5 \quad 23.5}{27} = 14.5 + \dfrac{117.5}{27}$$

$$= 14.5 + 4.35 = 18.85$$

Hence, median is 18.85

**Merits of Median:**

(i) It is easily understood and the calculations are also simple. In some cases, it can be obtained by mere inspection.

(ii) It is not affected by extreme values.

(iii) It is a value which exists in the data in many cases.

(iv) For attributes, median can be calculated.

(v) If some extreme values are not known and the total number of observations is known, median can be obtained.

(vi) When the distribution of the data is not symmetrical, median is an appropriate average.

(vii) Median can be located graphically with the help of ogives.

(viii) The sum of the absolute deviations of the values from the median is minimum.

**Demerits of Median:**

(i)    It is not based on all observations so it may not be a good representative of the data in some situations.

(ii)   It is affected by sampling fluctuations.

(iii)  The median is not capable of further mathematical treatment.

(iv)   Its calculation requires prior arrangement of the data in ascending or descending order.

(v)    For continuous variate case, the formula is obtained on the assumption of uniform distribution of frequencies over the class intervals. This assumption may not be true.

**MODE:**

The **mode** is defined as the value of a variable which occurs most frequently. It is a value which is repeated maximum number of times or with highest frequency. So, mode is considered as the most typical average. Graphically, it is the value on x-axis corresponding to the peak of the frequency curve.

If the data are ungrouped, mode can be obtained from inspection as the value with the maximum frequency. If we want to calculate the most common height for a group of students or the most common size of ready made shirts we have to consider the mode as the average. In marked surveys, to know consumers' Preference, mode is considered as the most suitable average.

For ungrouped data, for small sets mode can be found by inspection. But for grouped data, mode is calculated with the help of interpolation formula. If a distribution has two or more values of maximum frequency, then the distribution is known as bimodal, trimodal or multimodal.

**Ungrouped Data:**

Mode is determined by observing the given set of values and then locating the one which is repeated maximum number of times.

**Illustration: 7**

Following are the marks of 15 students in a certain test. Find the model marks.

18, 22, 25, 42, 39, 35, 25, 33, 34, 25, 29, 37, 35, 25 and 40

**Solution:**

As the marks 25 are repeated maximum number of times, that is 4 times out of 15 observation. The modal marks are 25.

Now consider,

**Grouped Data - Discrete Variate Case**

Here mode can be obtained as the value of the variable with the maximum frequency.

**Illustration 8:**

Find the mode for the following data, representing size of ready made pants.

| Size of pants in cms | 60 | 65 | 70 | 75 | 80 | 85 | 90 |
|---|---|---|---|---|---|---|---|
| No. of Pants | 11 | 15 | 25 | 40 | 20 | 15 | 10 |

**Solution:**

Here the variable x is the size in cms and the frequency f is the number of pants. As the size 75 cms has the maximum frequency of 40, it is the modal size.

So, the modal size of pants is 75 cms.

Let us consider,

**Grouped Data - Continuous Variate Case**

If the distribution has only a single maximum frequency, the mode can be calculated as follows:

First of all, modal class is located as the class interval with the maximum frequency.

Let $l_1 = lower\ class\ \lim it\ of\ the\ \mod al\ class$

$f_1 = frequency\ of\ the\ \mod al\ class$

$f_0 = frequency\ of\ the\ pre-\mod al\ class$

$f_o = frequency\ of\ the\ post-\mod al\ class$

Then mode is given by

$$\boxed{Mode = l_1 + \frac{l_2 - l_1}{f_1 - f_0 + f_1 - f_2}\ \ f_1 - f_0}$$

But $f_1 - f_0 + f_1 - f_2 = f_1 - f_0 - f_2$

So mode can also be calculated as,

$$Mode = l_1 + \frac{l_2 - l_1}{2f_1 - f_0 - f_2}\, f_1 - f_0$$

**Illustration 9:**

Find the mode for the following data:

| Income in Rs. | 200-400 | 400-600 | 600-800 | 800-1000 | 1000-1200 |
|---|---|---|---|---|---|
| No. of Persons | 16 | 34 | 60 | 37 | 13 |

**Solution :**

Here the variable is the income in Rs. and the frequency is the no. of persons. The maximum frequency is 60, so that the modal class is 600-800, corresponding to the maximum frequency 60.

$l_1$ = lower class limit of the modal class = 600

$l_2$ = upper class limit of the modal class = 800

$f_1$ = frequency of the modal class = 60

$f_0$ = frequency of the pre-modal class = 34

$f_2$ = frequency of the post-modal class = 37

$$Mode = l_1 + \frac{l_2 + l_1}{2f_1 - f_0 - f_2}\, f_1 - f_9$$

$$= 600 + \frac{800 - 600}{120 - 34 - 37}\, 60 - 34$$

$$= 600 + \frac{200}{49}\, 26$$

$$= 600 + 106.1225 = 706.1224 = Rs.706.12$$

So the modal income is Rs. 706.12

**Merits of Mode:**

(i) It is easy to understand and easy to calculate.

(ii) By definition, it is the most typical or representative value.

(iii) Mode is not affected by values which are too large or too small.

(iv) It is an appropriate average in qualitative data.

(v) It can be obtained graphically from a histogram.

(vi) It can be calculated in open end class intervals or in those cases where the neighbourhood of point of concentration is known.

**Demerits**

(i)   It is not rigidly defined. A distribution may be bimodal and multimodal.

(ii)  It is not based on all observations.

(iii) It is affected by sampling fluctuations.

(iv)  It is not capable of further mathematical treatment.

## 10A.3 SUMMARY

The definitions of arithmetic, mean, geometric mean, median and mode are different in the sense that, means are mathematical averages and the other two are positional averages**.**

The arithmetic mean and geometric mean are based on all observations while median and mode are not. This fact can be a disadvantage to mean, because if only one item is not known, then mean can not be obtained.

As calculations of means require all observations, they are affected by extreme values. But median and mode are not affected by extreme values.

All the measures, except geometric mean, are easy to understand and easy to calculate. Median and mode can be found by inspection, in ungrouped data. But for grouped data, all these measures require some calculations.

In case of open end class intervals, arithmetic mean and geometric mean can not be calculated, but median and mode can be computed for open end classes as well. The arithmetic mean and geometric mean are capable of further algebraic treatment, but no such treatment is possible in case of median and mode. Usually, as arithmetic mean satisfies most of the requisites of a good average, it is most widely used.

## 10A.4 CHECK YOUR PROGRESS

1)  Explain Mean.

2)  Discuss Median.

3)  Highlight Mode and its demerits.

## 10.5 REFERENCES

1)  Borwankar P.V. Research Methodology Seth Publisher 1995.

## 10A.6 QUESTIONS

1)  Discuss measures of central tendency

2)  Give the merits and demerits of median and mode.

❖❖❖❖

# MEASURES OF DISPERSION

**Unit Structure:**

## 10B.0 OBJECTIVES

- To introduce students with quantitative method of datacollection.

- To enable students to apply statistical terms in socialresearch.

## 10B.1 INTRODUCTION AND MEANING

We have studied various measures of central tendency such as mean, median, mode in the previous chapter. But they are not adequate to describe the distribution. For instance, consider the following sets of observations:

Set A: 35, 37, 34, 38 and 46 with mean 38

Set B: 10, 90, 45, 12 and 33 with mean 38

Set C: 38, 38, 38, 38, and 38 with mean 38

All the sets have the same mean 38, but if the values in the sets are observed carefully, it can be seen that in set C, the average 38 completely represents the distribution; in set A, only one value is represented by the average and in set B, average 38 represents none of the values. Also, the variation of the items is nil in set C and is maximum in set B.

Thus, it is quite clear that in addition to averages, some additional information about the variation of items is required, to know the extent to which the values vary from one another andfrom central value.

A measure of spread of scatter of the data is called a measure of variation or dispersion.

The measures of dispersion can give us idea about reliability of the averages. When the dispersion is less, the average is more reliable so that it is a better estimate of the population average, and if, the dispersion is more, the average is not a good representative of the data.

The measures of dispersion determine the extent of variation in the data, by which, some steps can be taken to control the variability. For instance, in factories quality control techniques can be applied to control the variation.

The measures of dispersion can be used to compare two or more distributions. The one with less dispersion is more consistent or homogeneous and the one with more dispersion is less consistent.

The study of dispersion leads to further advanced techniques in analysis such as Statistical Quality Control, Cost Control, Inventory Control, etc.

**Requisites of a Good Measure of Dispersion:**

(i) It should be easy to understand and easy to calculate.

(ii) It should be rigidly defined.

(iii) It should be based on all the observations.

(iv) It should be capable of further algebraic treatment.

(v) It should not be affected much by extreme values.

(vi) It should have sampling stability.

There are two types of measures of dispersion.

(a) Absolute Measures giving actual extent of scatter of the data and

(b) Relative Measures expressed as pure numbers, independent of the unit of measurement.

Corresponding, to each absolute measure of dispersion, a relative measure can be defined which can be used to compare two or more distributions. Now, let us study these measures.

## 10B.2 RANGE

It is defined as the difference between the maximum and minimum values. It is an absolute measure.

$$\boxed{Range = Maximum - Minimum}$$

If we consider the set A,B,C, mentioned earlier.

Range for Set A = 46-34 = 12

Range for Set B = 90 -10 = 80

Range for Set C = 38-38 = 0

So, for the sets A,B,C, even though the means are the same, theranges are quite different.

**Illustration 1:**

The following are the prices in Rs. of different brands of televisionsets. Find the range of prices.

Rs. 17850, 16990, 17500, 19850, 16650, 19300

**Solution:**

Here, Range = Maximum Price - Minimum Price

$$= 19850 - 6650 = 3200$$

So, the range of prices is Rs. 3200.

**2) Coefficient of Range :**

It is a relative measure of dispersion and is defined as,

$$Coefficient\ of\ Range = \frac{Maximum - Minimum}{Maximum + Minimum}$$

Thus, it is a ratio of the difference and the sum of maximum and minimum values. Being a ratio, it is independent of the unit in which the original variable is measured.

**Illustration 2:**

Find the coefficient of range for the following data relating to pricesof shares of "ABC" company during a week.

Rs. 25.75, 23.10, 23.50, 29.25, 24.50

**Solution:**

The maximum price = Rs. 29.25

The minimum price = Rs. 23.10

$$Coefficinet\ of\ Range = \frac{Maximum - Minimum}{Maximum + Minimum}$$

$$= \frac{29.25 - 23.10}{29.25 + 23.10} = \frac{6.15}{52.35}$$

$$= 0.1175$$

The range and coefficient of range are used to measure the variation in prices of commodities, shares, temperatures, rainfall, etc. over a time period. The range is used to measure variations which are highly sensitive like gold and silver prices. In day-to-day life, question like "Kow many answer papers are assessed in a week? Or How many marks are expected in a particular paper? Or How much profit does a shopkeeper make per day? Are always answered in the form between, two extreme value.

But statistically it is represented by single number which is interval between maximum all minimum value called as range.

**Merits of Range:**

(i)    It is easy to understand and easy to compute.

(ii)   It is rigidly defined.

(iii)  For small sample size, range is a good measure.

(iv)   It is affected by extreme values and its value changes from sample to sample.

## 10B.3 QUARTILES

Quartiles are not measures of central tendency but they represent three values dividing the entire distribution of data into four equal parts. As the calculations are similar to that for median, we consider Quartiles along with median. The quartiles are called first quartile $Q_1$, second quartile $Q_2$ and the third quartile $Q_3$. Quartiles occupy specific positions in the distribution of data. if we arrange all N observations in order, then the quartiles $Q_1$, $Q^2$ and $Q_3$ are three points such that they divide the distribution into four equal parts, each consisting of $\dfrac{N}{4}$ observations.
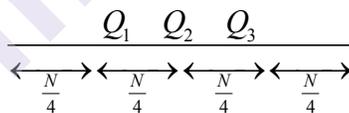


The figure show four parts, containing $\dfrac{N}{4}$ observations and three values $Q_1$, $Q_2$ and $Q_3$.

The number of observations less than $Q_1$ is $\dfrac{N}{4}$

The number of observations less than $Q_2$ is $\dfrac{N}{2}$

The number of observations less than $Q_3$ is $\dfrac{N}{4}$

For continuous distribution, using interpolation formula:

**(i)  Consider the First Quartile Q₁**

The first quartile class is locate as the one for which cumulative frequency exceeds $\dfrac{N}{4}$. After locating the class interval,

$l_1$ = lower class limit of the first quartile class,

 $l_2$ = upper class limit of the first quartile class,

f = frequency of the first quartile class,

cf = cumulative frequency of the class preceding the first quartile class, and and $Q_1$, is calculated using the formula

$$Q_1 = l_1 + \frac{l_2 - l_1 \left[ \dfrac{N}{4} - cf \right]}{f}$$

**(ii) The Second Quartile**  $Q_2$ coincides with the median.

$$Q_2 = l_1 + \frac{l_2 - l_1 \left[ \dfrac{N}{4} cf \right]}{f}$$

**(iii) Consider the Third Quartile Q₃**

The required class interval is obtained by comparing the cumulative frequency with $\dfrac{3N}{4}$ and then finding the class for which cumulative frequency exceeds $\dfrac{3N}{4}$

$l_1$ = lower class limit of the third quartile class,

$l_2$ = upper class limit of the third quartile class,

f = frequency of the third quartile class,

cf = cumulative frequency of the class preceding the third quartileclass.

$Q_3$ is calculated as

$$Q_3 = l_1 + \frac{l_2 - l_1 \left[ \dfrac{3N}{4} - cf \right]}{f}$$

**Note:** There are 25% of the total observation which lie below $Q_1$. 50% of the observation are below $Q_2$ and 75% are below $Q_3$. So, there are

50% observations in between $Q_1$ and $Q_2$, and so Q1 and Q3 are the limits within which middle 50% of the observations lie.

**Illustration 6:**

Find the three quartiles for the following data:

| Daily Wages in Rs. | 10-15 | 15-20 | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 | 45-50 | 50-55 |
|---|---|---|---|---|---|---|---|---|---|
| No. of Workers | 12 | 28 | 36 | 50 | 25 | 18 | 16 | 10 | 5 |

**Solution:**
We prepare the table for cumulative frequencies.

| Daily Wages in Rs. | No. of Workers | Cumulative Frequency less than |
|---|---|---|
| 10-15 | 12 | 12 |
| 15-20 | 28 | 40 |
| 20-25 | 36 | 76 |
| 25-30 | 50 | 126 |
| 30-35 | 25 | 151 |
| 35-40 | 18 | 169 |
| 40-45 | 16 | 185 |
| 45-50 | 10 | 195 |
| 50-55 | 5 | 200 |
| Total | 200 = N | |

For $Q_1$, consider $\dfrac{N}{4} = \dfrac{200}{4} = 50$

From the table as 76 is greater than 50, the required class interval is 20-25.

Now, $l_1 = 20, l_2 = 25, f = 36, cf = 40$

$$Q_1 = l_1 + \frac{l_2 - l_1 \left[ \dfrac{N}{4} - cf \right]}{f}$$

$$= 20 + \frac{25 - 20 \quad 50 - 40}{36}$$

$$= 20 + \frac{5 \quad 10}{36}$$

$$= 20 + 1.3889 = 21.39$$

So, the first quartile is Rs. 21.39.

Now, for $Q_2$, consider $\dfrac{N}{2} = \dfrac{200}{2} = 100$

By comparing 100, with the cumulative frequencies, as 126 exceeds 100, the required class is 25-30.

Now, $l_1 = 25, l_2 = 30, f = 50, cf = 76$

$$Q_2 = l_1 + \dfrac{l_2 - l_1 \left[ \dfrac{N}{2} - cf \right]}{f}$$

$$= 25 + \dfrac{30 - 25}{\phantom{5}} \dfrac{100 - 76}{50}$$

$$= 25 + \dfrac{5}{\phantom{5}} \dfrac{24}{50} = 25 + \dfrac{120}{50}$$

$$= 27.4$$

So, the second quartile is Rs. 27.4.

Now, for $Q_3$, consider $\dfrac{3N}{4} = \dfrac{3 \times 200}{4} = 150$

By comparing 150, with the cumulative frequencies, as 151 exceeds 150, the required class is 30-35.

Now, $l_1 = 30, l_2 = 35, f = 25, cf = 126$

$$Q_3 = l_1 + \dfrac{l_2 - l_1 \left[ \dfrac{3N}{4} - cf \right]}{f}$$

$$= 30 + \dfrac{35 - 30}{\phantom{5}} \dfrac{150 - 126}{25}$$

$$= 30 + \dfrac{5}{\phantom{5}} \dfrac{24}{25} = 30 + \dfrac{120}{25}$$

$$= 30 + 4.8 = 34.8$$

So, the third quartile is Rs. 34.8.

**Limitation:**

(i)     It can not be calculated for open end classes.

(ii)    It does not take into account the deviations of individuals items from a measure of central tendency.

(iii)   It is not based on all the observations.

### 3) Semi-inter-quartile-range or quartile deviation:

It is defined as follows:

$$Semi-\text{int}er-qurtile\ Range = \frac{Q_3 - Q_1}{2}$$

The semi-inter-quartile range considers only the middle 50% of the observations and it ignores the first and the last quarter. It is an absolute measure. The quartile deviation also measures the average amount by which the two quartiles $Q_1$ $_3$ differ from median. median.

### 4) Coefficient of quartile deviation:

It is a relative measure and is defined as

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

As, it is a ratio, it is a pure number, so that it can be used to compare two or more distributions.

### Illustration 3:

Find the quartile deviation for the following data of pocket allowances (Rs.) of 15 student. Also find the coefficient of quartile deviation.

Set A (Rs.): 34, 45, 53, 42, 39, 35, 40, 51, 57, 52, 47, 62, 55, 50, 63

### Solution:

Arrange the sets in ascending order of magnitude as

Set A: 34, 35, 39, 40, 42, 45, 47, 50, 51, 52, 53, 55, 57, 62, 63

Here, n = the total no. of observation = 15

$$Q_1\ is\ \left(\frac{n=1}{4}\right)^{th}\ observation\ = 4^{th}\ observation$$

$$\therefore Q_1 = Rs.40$$

$$Q_3\ is\ 3\left(\frac{n=1}{4}\right)^{th}\ observation = 12^{th}\ observation$$

$$\therefore Q_3 = Rs.55$$

Quartile Deviation $= \dfrac{Q_3 - Q_1}{2} = \dfrac{52-40}{2} = 7.5$

Coefficient of Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1} = \dfrac{55 - 40}{55 + 40}$

$$= \dfrac{15}{95} = 0.1579$$

Now consider the quartile deviation for grouped data.

**Illustration 4:**

Calculate quartile deviation for the following distribution of ages of 800persons. Also find the coefficient of quartile deviation.

| Age in Years | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 | 45-50 | 50-55 | 55-60 |
|---|---|---|---|---|---|---|---|---|
| No. of Persons | 50 | 70 | 100 | 180 | 10 | 150 | 70 | 60 |

**Solution:**

As it is a continuous distribution, first prepare the following table to obtain cumulative frequencies and to locate the required class intervals.

| Age in Years | No. of Persons | Cumulative Frequency (Less than) |
|---|---|---|
| 20-25 | 50 | 50 |
| 25-30 | 70 | 120 |
| 30-35 | 100 | 220 |
| 35-40 | 180 | 400 |
| 40-45 | 150 | 550 |
| 45-50 | 120 | 670 |
| 50-55 | 70 | 740 |
| 55-60 | 60 | 800 |
| **Total** | **800** | |

Here, $N = \sum f = 800$

a) For $Q_1$, consider $\dfrac{N}{4} = 200$. As 220 is the first cumulative frequency greater than 200, the required class for $Q_1$ is 30-35.

Now $l1 = 30, l2 = 35, f = 100, cf =$ cumulative freq. of previous class = 120.

$$Q1 = l_1 + \dfrac{l_2 - l_1 \left[ \dfrac{N}{2} - cf \right]}{f}$$

$$= 30 + \frac{35 - 30}{100} \frac{200 - 120}{}$$

$$= 30 + \frac{50}{100} \frac{80}{} = 30 + \frac{400}{100}$$

$$= 34 \ yrs.$$

b) For $Q_3$, consider $\frac{3N}{4} = 600$. As 670, is forst cumulative frequency exceeding 600, the required class interval for $Q_3$ is 45-50.

Now, $l_1 = 45, l_2 = 50, f = 120, cf = 550$

$$Q3 = l1 + \frac{l2 - l1\left[\dfrac{3N}{4} - cf\right]}{f}$$

$$= 45 + \frac{(50 - 45)(600 - 550)}{120}$$

$$= 45 + \frac{5}{120} \frac{50}{} = 45 + \frac{250}{120}$$

$$= 45 + 2.08 = 47.08 \ years$$

Quartile Deviation $= \dfrac{Q_3 - Q_1}{2}$

$$= \frac{47.08 - 34}{2} = \frac{13.08}{2}$$

$$= 6.54$$

Coefficient of Quartile Deviation $= \dfrac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{47.08 - 34}{47.08 + 34} = \frac{13.08}{81.08}$$

$$= 0.1613$$

So, the quartile deviation is 6.54 years and coefficient of quartile deviation is 0.1613.

**Merits of Quartile Deviation:**

(i)    It is simple to understand and easy to calculate.

(ii)   As it is based on middle 50% of the observations, it is not affected by extreme values. So, it is useful in erratic data.

(iii)  It can be calculated for open end classes.

(iv) The quartile deviation can be obtained for qualitative datawhich can not be measured but can be ranked.

**Limitations:**

(i) It is not based on all observations.

(ii) It is not capable of further mathematical treatment.

(iii) It is affected by sampling fluctuations.

(iv) It does not consider the observations in the first and lastquarter.

However, it gives a rough idea about the scatter of the data and it is a better average than range.

## 10B.4 MEAN DEVIATION

The range or quartile deviation do not take into account, the deviations from the central value. The mean deviation considers these differences in absolute values and averages these differences.

Thus, mean deviation, in which is an absolute measure is defined as the arithmetic mean of absolute values of deviations of all the observations taken from the mean, median or mode. Mean deviation from median is minimum.

**For Ungrouped Data**

a) M.D. from mean $= \dfrac{\sum |x - \bar{x}|}{n}$

M.D. from median $= \dfrac{\sum |x - Median|}{n}$

M.D. from mode $= \dfrac{\sum |x - \text{mod}\,e|}{n}$

Where n represents total number of observations.

For Grouped Data, $N = \sum f$

a) M.D. from mean $= \dfrac{\sum f |x - \bar{x}|}{N}$

b) M.D. from median $= \dfrac{\sum f |x - Medial|}{N}$

M.D. from mode $= \dfrac{\sum f |x - Mode|}{N}$

**2) Coefficient of Mean Deviation:**

It is a ratio of the mean deviation and the measure from which the deviations are considered.

Being a relative measure of dispersion, it is a pure number, independent of the unit of measurement of the variable. Hence, it can be used to compare two or more sets of data. It is defined as follows:

a) Coefficient of M.D. from mean $= \dfrac{M.D.\,fromMean}{Mean}$

b) Coefficient of M.D. from median $= \dfrac{M.D.\,fromMedian}{Median}$

c) Coefficient of M.D. from mode $= \dfrac{M.D.\,fromMode}{Mode}$

**Illustration 1:**

Find the mean deviation from mode and the correspondingcoefficient of mean deviation for the following data:

| Income in Rs. | 800-1000 | 1000-1200 | 1200-1400 | 1400-1600 | 1600-1800 |
|---|---|---|---|---|---|
| No. of Persons | 16 | 34 | 60 | 37 | 13 |

**Solution:**

First calculate mode as follows:

Here the modal class is 1200-1400 as the corresponding frequency 60 is the maximum frequency.

Now, $l_1 = 1200, l_2 = 1400, f_1 = 60, f_2 = 37, f_0 = 34$

Mode $= l_1 + \dfrac{l_2 - l_1}{2f_1 - f_0 - f_2} \dfrac{f_1 - f_0}{}$

$$= 1200 + \dfrac{1400 - 1200}{120 - 34 - 37} \dfrac{60 - 34}{}$$

$$= 1200 + \dfrac{200}{49} \dfrac{26}{} = 1200 + \dfrac{5200}{49}$$

$$= 1200 + 10.6.12 = 1306.12$$

So that modal income is Rs. 1306.12

Now, prepare the following table to calculate $\sum f |x - mod\,e|$.

| Income in Rs. | No.of persons | X | $\lvert x - Mode \rvert$ | $f\lvert x - mode \rvert$ |
|---|---|---|---|---|
| 800-1000 | 16 | 900 | 406.12 | 6497.92 |
| 1000-1200 | 34 | 1100 | 206.12 | 7008.08 |
| 1200-1400 | 60 | 1300 | 6.12 | 367.20 |
| 1400-1600 | 37 | 1500 | 193.88 | 7173.56 |
| 1600-1800 | 13 | 1700 | 393.88 | 5120.44 |
| Total | 160 | | | 26167.20 |

No, $\sum f\lvert x - Mode \rvert = 26167.2, N = 160$

M.D. from mode $= \dfrac{\sum \lvert x - mode \rvert}{n} = \dfrac{26167.2}{160} = Rs.163.545$

Coefficient of M.D. from mode $= \dfrac{M.D.\,from\,Mode}{Mode} = \dfrac{163.545}{1306.12}$
$= 0.1252$

**Merits of Mean Deviation**

i)   It is rigidly defined.

ii)  It is easy to understand.

iii) It is based on all the observations.

iv)  Its value is minimum when calculated from median.

v)   It is less affected by extreme values.

vi)  As it takes into account deviations from averages it is more scientific than range or quartile deviation.

**Limitations:**

i)   It requires more calculations for continuous variables.

ii)  It ignores the negative signs for deviations and only absolutevalues are considered.

## 10B.5 STANDARD DEVIATION

It is defined as the positive square root of the arithmetic means of the squares of the deviations of the observations from thearithmetic mean. It is denoted by $\sigma$ (sigma). It is an absolute measure.

**a) For ungrouped Data**

If $X_1$, $X_2$.........,$X_n$ are n observations then $\bar{x} = \dfrac{\sum x}{n}$

and $\sigma = \sqrt{\dfrac{\sum x - \bar{x}^2}{n}}$

The alternative formula

$\sigma = \sqrt{\dfrac{\sum x^2}{n} - \bar{x}^2}$

**b) For Grouped Data**

If $x_1, x_2$.............$x_n$ are district values of a variable with frequesncies $f_1$, $f_2$.............$f_n$ then

$\bar{x} = \dfrac{\sum fx}{N}$, where $N = \sum f$ and $\sigma = \sqrt{\dfrac{\sum fx - \bar{x}^2}{N}}$

The alternative formula is

$\sigma = \sqrt{\dfrac{\sum fx^2}{N} - \bar{x}^2}$

for continuous variable, $x_1, x_2$.........$x_n$ are the class marks.

**Note :** There are two possibilities for value of $\bar{x}$. They are (i) $\bar{x}$ is integer and (ii) is not an integer. Any one of the two formulae can be used in any case. But usually, the first formula is used when $\bar{x}$ is not an integer, to simplify the calculation. Now consider some examples.

**Illustration 1:**

Find standard deviation for the following data:

| Class Interval | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 |
|---|---|---|---|---|---|
| Frequency | 11 | 15 | 25 | 12 | 7 |

**Solution:**

The variable is a continuous variable so x represents the class marks of the class intervals i.e. x values are 5, 15, 25, 35and 45.

Prepare the following table for the product term fx, $fx^2$ to obtain mean and S.D.

| Class Interval | Frequency | Class-mark | fx | fx$^2$ |
|---|---|---|---|---|
| 0-10 | 11 | 5 | 55 | 275 |
| 10-20 | 15 | 15 | 225 | 3375 |
| 20-30 | 25 | 25 | 625 | 15625 |
| 30-40 | 12 | 35 | 420 | 14700 |
| 40-50 | 7 | 45 | 315 | 14175 |
| Total | 70 | | 1640 | 48150 |

$$N = \sum f = 70$$

$$\sum fx = 1640$$

$$\bar{x} = \frac{\sum fx}{N} = \frac{1640}{70} = 23.4286$$

As it is not an integer, we use the following formula for standard deviation $\sigma$.

$$\sigma = \sqrt{\frac{fx^2}{N} - \bar{x}^2}$$

Now, $\sum fx^2 = 48150, \bar{x} = 23.4286, N = 70$

Substituting these values, in the formula

$$\sigma = \sqrt{\frac{\sum fx2}{N} - \bar{x}^2}$$

$$= \sqrt{\frac{48150}{70} - 23.4286^2}$$

$$= \sqrt{687.8571 - 548.8993} = \sqrt{138.9578}$$

$$= 11.7880.$$

So, the standard deviation is 11.788.

**Illustration 2:**

Find standard deviation for the distribution of weights of 90 children.

| Weight in kgs. | 20-25 | 25-30 | 30-35 | 35-40 | 40-45 |
|---|---|---|---|---|---|
| No. of Children | 11 | 15 | 24 | 26 | 14 |

**Solution:**

We prepare the following table to calculate the product terms fx And $fx^2$ .

| Weight in kgs | No. of Children | x | fx | $fx^2$ |
|---|---|---|---|---|
| 20-25 | 11 | 22.5 | 247.5 | 5568.75 |
| 25-30 | 15 | 27.5 | 412.5 | 11343.75 |
| 30-35 | 2 | 32.5 | 780.0 | 25350.00 |
| 35-40 | 26 | 37.5 | 975.0 | 36562.50 |
| 40-45 | 14 | 42.5 | 595.0 | 25287.50 |
| Total | 90 | | | 104112.50 |

$\therefore \sum fx = 3010$ and $N = 90$, $\sum fx^2 = 104112.5$

$$\bar{x} = \frac{\sum fx}{N} = \frac{3010}{90} = 33.4444$$

The standard deviation is given by

$$\sigma = \sqrt{\frac{\sum fx^2}{N} - \bar{x}^2}$$

$$= \sqrt{\frac{104122.5}{90} - 33.4444^2}$$

$$= \sqrt{1156.8055 - 1118.5278} = \sqrt{32.2777}$$

$$= 6.1869$$

Hence, the standard deviation is 6.1869.

It can be observed that if the class marks are expressed as fractions or if the class intervals are more, the calculations become lengthy and tedious.

### Merits and demerits of standard Deviation

Merits of standard Deviation.

1) It is rigidly defined.

2) It is based on all the observations.

3) It is not affected much by sampling fluctuations.

4) It is amicable to further mathematical treatment.

**Demerits:**

1) It is not easy to understand and easy to calculate

2) As it consider the sum of the squares of deviations of items from the mean. The items away from mean gets more weight age than those near the mean. Thus, standard deviation gives more weight age to extreme value.

## 10B.6 CHECK YOUR PROGRESS

1) What do you mean by Range?

2) What is Mean Deviation?

3) Explain Quartile deviation.

## 10B.7 REFERENCES

1) Borwankar P.V.Research -Methodology Seth Publisher 1995.

## 10B.8 QUESTION

1) What is standard deviation? What are it merits and demerits?

2) What is mean deviation? Discuss its as merits and demerits.

❖❖❖❖

# 11

# CORRELATION

**Unit Structure**

## 11.0 OBJECTIVES

- To learn the meaning and Correlation

- To learn about different uses of Correlation

## 11.1 INTRODUCTION

The use of statistics and statistical tools in social research has been since a long time, especially with the use of computers in social science research. Scientists, researchers have been using various software's, tools to make the analysis, documentation, writing, referencing process quicker. In this chapter we are going to learn about one such topic called 'Correlation'. Correlation is not just used in social science research but it is used in several disciplines starting from Finance, stock markets analysis, to Economics, natural science etc. In this chapter, we will be looking into how this topic is relevant and useful in social science and research.

## 11.2 MEANING OF CORRELATION

The Cambridge dictionary defines Correlation as a <u>connection</u> between two or more things, especially when one of them <u>causes</u> or <u>influences</u> the other. For example – Smoking can cause health related issues. Hence, there is a correlation between Smoking and health.

According to Prof. King, Correlation suggests that there are some random linkages between two series or groups of data.Correlation can indicate whether a change in one variable will result in a change in the other or not. The degree associationbetween the two sets of characters or variables can be stated statistically and this is known as the correlation, coefficient.

The study of correlation has many applications in daily life. The French astronomer Bravis was the first to advance the fundamental ideas of the science of correlation analysis.Nevertheless, he was also the first to advance linear correlation theory. Karl Pearson, a statistician, developed the mathematical formula for computing the coefficient of correlation in the year 1896. Galton and Karl Pearson used the coefficient of correlation to examine a variety of biological and genetics-related issues. According to Professor Neiswanger, who explained the significance of correlation in the field of economics, correlation analysis contributes to the understanding of economic behaviour, which helps in locating the critically important variable. It also reveals how connections and suggest to him the paths through which stabilising forces may become effective.

- **Regression and Correlation**

Regression and correlation are two distinct yet complementary approaches. Roughly speaking, correlation is used to assess the strength of a relationship while regression is used for prediction (which does not extend beyond the data utilised in the research).

Both correlation and regression analysis focus on how different variables relate to one another. A measure of the linear link between two variables is the correlation coefficient. The correlation coefficient always has a value between -1 and +1. A correlation value of 1 denotes perfect linear positive correlation between two variables, a correlation coefficient of 1 denotes perfect linear negative correlation between two variables, and a correlation coefficient of 0 denotes no linear link at all between the two variables. The sample correlation coefficient for simple linear regression is the square root of the coefficient of determination, and its value is identical to the sign of $b_1$, the coefficient of determination.There are instances where the x variable has a random covariate to the y variable rather than being fixed or readily selected by the experimenter. Here covariate means Here covariate means an independent variable that can influence the outcome of a given statistical trial but which is not a direct one.

Analysis methods like as regression and correlation cannot be used to demonstrate cause-and-effect relationships. They can only show if or how closely different variables are related to one another. Only the strength of the linear link (straight line) between two variables is measured by the correlation coefficient. Any findings on a cause-and-effect link must be supported by the analyst's judgement. (11)

## 11.2 CORRELATION TYPES

There are three different types of correlations:

Correlations that are positive, zero, and negative:

### (A) POSITIVE CORRELATION:

Positive correlation is a term which is used to describe a relationship where two variables' values move simultaneously and in the same direction.Perfectly positive or slightly positive correlations of are both possible in the positive correlation.

(i)    Perfect Positive Correlation: This relationship exists when both variables rise and decrease in the same proportion.

(ii) Moderately Positive Correlation: In this situation, two variables have a positive correlation, but the changes are not proportional. The coefficient's value ranges from 1 to 0.

### B) NEGATIVE CORRELATION

Negative correlation is used to describe a relationship where one variable increases (or declines) while the other lowers (or increases).

For example - Size and the quantity of fruits or plants, for example, effects the supply/ shortage adverselyin market.

This adverse relationship may also take one of two forms like:

(i) Perfect Negative Correlation: This type of association is incredibly uncommon in biological contexts, such as when a rise in temperature causes a drop or fall in the cell's lipid content.

(ii) Moderately Negative Correlation: In this relationship, the variables are negatively connected, but not very completely. For example, a longer post-harvesting interval reduces seed viability. The coefficient value in this case is also between 0 and -1.

### (C) ZERO CORRELATION

The two values of the variables are said to have zero correlation when there is no correlation between them, i.e., no consistency in the value of the observation.

## 11.3 COEFFICIENT OF CORRELATION

The Coefficient of Correlation is a numerical term that describes the degree of relationship between two variables when there is any direct interaction between them. The degree of closeness of the linear relationship between the two variables is expressed quantitatively by this metric.

The correlation coefficient, often known as Karl Pearson's Coefficient of Correlation and denoted by the letter "r," and is calculated as follows:

$$r_{x.y} = \frac{\Sigma dx \cdot dy}{\sqrt{\Sigma dx^2 \cdot \Sigma dy^2}} \quad \text{or} \quad \frac{\Sigma dx \cdot dy}{n \cdot \delta x \cdot \delta y}$$

where

$dx = x_1 - \bar{x}$ [deviation of x variable]

$dy = y_1 - \bar{y}$ [deviation of y variable]

$\delta x$ = S.D. of x variable

$\delta y$ = S.D. of y variable.

$n$ = total no. of observation.

How to determine the value of "r":

(A) The variables x and y form two series

(b) The means of the two series, x and y, are calculated.

(c) Each observation's deviation is determined using the variables dx and dy.

(d) The squared variances are noted.

(e) Multiplying the deviations of the two variables.

(f) The total data are added together using the method to determine the value of "r."

Correlation coefficient properties include:

1. The correlation coefficient ranges from -1 to +1, or -1 r +1.

2. If r is greater than 1, the correlation is perfect and positive; otherwise, it is somewhat positive.

3. The connection is complete and negative if r = - 1, and substantially negative if it is higher than - 1.

4. There is no connection between the variables if r is equal to 0.

5. Change and scale of origin have no impact on the coefficient of correlation.

**Check Your Progress**

1. What is positive correlation

2. Discuss Zero Correlation

(a) Moderately positive
r < + 1

(b) Perfectly positive
r = + 1

(c) Moderately negative
r > − 1

(d) Perfectly negative
r = − 1

(e) Zero correlation
r = 0
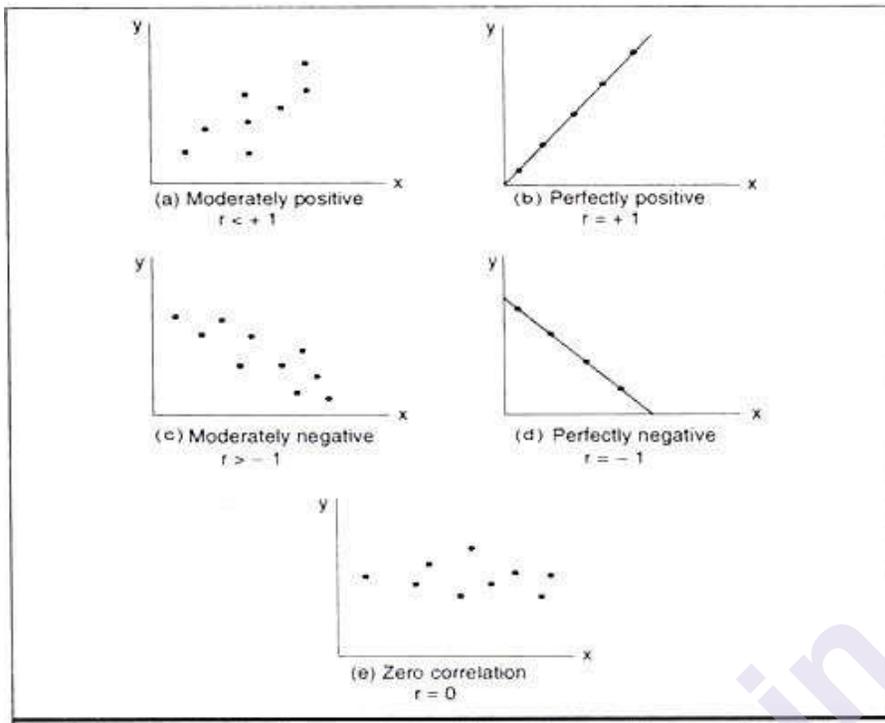
Scatter Diagram

An illustration of a correlation is very much helpful. To make this create a scattergram (also known as a scatterplot, scatter graph, scatter chart, or scatter diagram).The links or creates affiliations between two numerical variables (or co-variables) which are graphically depicted as points (or dots) for each pair of score in a scattergram.The degree and direction of the correlation between the co-variables are shown on a scattergraph.

It doesn't matter which variable is plotted on the x-axis and which is plotted on the y-axis when creating a scattergram.Since paired scores are always involved in correlations, the values of the two variables combined will be utilised to create the diagram. Thereafter,Choose which variable belongs on each axis, then draw a cross where the two values meet.

It should be noted that the coefficient of correlation is one of the statistical measurements that is both utilised and abused the most. It is misused as sometimes people fail to realise that correlation measures are just the strength of a linear relationship, even in the absence of causality. In other words, a cause-and-effect link is not always implied.

When two variables are correlated, they are related quantitatively and can be quantified on either continuous or ordinal scales. Although correlation implies a link between two variables, it does not imply causality. The correlation coefficient can be used to determine how strong a correlation is.

In order to determine the population correlation coefficient for a given sample of data, the correlation coefficient which is statistic is calculated. Correlation coefficients typically range from -1 to +1. High values of one

variable are linked to high values of the other, hence a positive value denotes a positive correlation between variables, whilst a negative value implies a negative association between variables (i.e., high values of one variable are associated with low values of the other). A coefficient of 1 denotes a perfect negative relationship between the variables, while a coefficient of 1 denotes a perfect positive relationship. If the correlation coefficient is 0, the variables are not connected.

If the sample statistic is unlikely to have been taken from a population with a genuine rho of 0, the null hypothesis that the population correlation coefficient rho is 0 is rejected when conducting hypothesis testing. The null hypothesis won't be disproved if the correlation coefficient has a value of 0. The likelihood of rejecting the null hypothesis rises as the sample size increases as the coefficient deviates from 0.

Correlation coefficients can be measured using a variety of methods. The two most well-known statistics are Spearman's rank correlation and Pearson's product-moment correlation, examples of nonparametric statistics.The linear relationship between variables is quantified by the Pearson product-moment correlation coefficient (r) in terms of the variables' actual raw values. When using the Pearson correlation coefficient, a normal distribution and linearity are both taken for granted.

The covariance of two variables X and Y divided by the sum of their respective standard deviations is known as the Pearson correlation coefficient:

One outlier point has the potential to have a significant impact on the correlation coefficient's value. For purposes of interpretation, r2 denotes the percentage of a variable's variance that is "explained" by another variable.

For ordinal variables, or any data that can be ranked, the Spearman rank correlation coefficient (rs) is utilised, which necessitates less assumptions about the distributions of the variables of interest. It is a measure of correlation for which there may be a nonlinear relationship because it assesses the strength of the association between the ranks of the data.

The Pearson correlation coefficient and the Spearman rank correlation both have the same mathematical formula. The quantity of links between data points has an impact on the rank correlation coefficient. The Spearman coefficient can be stated more succinctly as follows if there are no ties in the rankings. Where it represents the rank discrepancy between xi and yi. The Spearman coefficient is unreliable if more than half of the rankings are tied.

For example - An investigation of the consequences of mercury exposure at a thermometer manufacturer is one example of how correlation coefficients are used. The research discovered a strong relationship (r = 0.92) between mercury levels in the air and urine, blood, and hair.

- **Coefficient and Correlation Significance**

It is impossible to overstate the value of correlation in social science research. Even though it may appear simple, establishing links and associations between variables has several benefits for the social science researcher. Below, are few of the importance and significance are briefly describedin social science research on correlation.One of the most used methods is correlation matrices, which was often used by Pearson whether exploratory or explanatory, for examining the concept validity of data in factor analysis.This approach is also used to find factor solutions and is confirmatory (Holgado-Tello P. et al 2011).

Regression uses correlation as a starting point to predict the values of the dependent variablesbased on the established link between the independent variablesas well as the reliant variable too.

The development and evaluation of theoretical models can benefit greatly from correlational research. The nature of bivariate relations may then be established, and theoretical models can be created using this knowledge. Here, the goal is rather than just reporting the bivariate (two variants) correlations, to describe their nature.

In light of this, techniques like factor analysis, path analysis, and structural equationModeling may be utilised (Duncan, 1966).

Correlational research has played and will play a significant role in exploring the nature of the relationships among a collection using quantitative researchand variables. Unrelated variables can be partially removed from future analysis,enabling the researcher to take relevant issues more seriously.

Researchers can study numerous factors at once thanks to more advanced multivariate extensions (Stockwell, 2010). Correlational research can be performed to look at the relationships between the key variables once descriptive research has helped to identify them.For instance, researchers might be curious to know which variables are mosthighly correlated with a specific result, like student achievement. It may then result inexperimental studies where the causal relationships between those important variables canbe investigated under more regulated circumstances even one independent variable in this casecan be altered by the researcher (e.g., instruction method), with other associateda method of controlling the variables (e.g., grade, level of school funding). This results in a determination of the independent variable's impact.

## 11.4 LIMITATIONS OF CORRELATION

Correlation can be perceived incorrectly, much like other statistical analysis features. Even if it may seem as though there is a high correlation between two variables, data from small sample numbers may not be accurate. In contrast, a small sample size could produce uncorrelated results even if the two variables are in fact connected.

When an outlier is present, correlation is frequently distorted. Correlation does not explicitly demonstrate how a specific occurrence or event can affect the correlation coefficient; it merely illustrates how one variable is related to another.

If there is a nonlinear relationship between two variables, correlation may also be misconstrued. Finding two variables that are correlated either positively or negatively is much simpler. A more complex link can still be connected with two variables, though.

## 11.5 STATISTICAL CORRELATION ANALYSIS'S USEFULNESS IN SOCIAL SCIENCE

These kinds of statistical analyses are helpful because they can illuminate how experiences and social traits influence what occurs in a person's life and how various trends or patterns in society, such as unemployment and crime, might be connected. With the help of correlation analysis, we may determine with certainty whether or not there is a connection between two patterns or variables, which enables us to estimate the likelihood of a result in the population under study.

The amount of education and the divorce rate have a significant negative link, according to a recent study on marriage and education. According to data from the National Survey of Family Growth, women's divorce rates for first marriages decline as their level of education rises.

Although there is a substantial association between education and divorce rates, it's crucial to remember that correlation does not imply causation, therefore the decline in divorce among women may not necessarily be caused by this correlation but caused by the amount of education received. (Crossman, 2020).

- **Steps to use Correlation in Excel**

Step 1: Enter your data into an Excel worksheet. Two columns make for the ideal layout. Your x values should go in column A, and your y values should go in column B.

Step 2: Select "Data Analysis" from the "Data" menu.

The next step is to select "Correlation" and then "OK."

Step 4: Fill out the Input field with the position of your x-y variables.

Variety box. Alternatively, you can use your cursor to identify the region where your variables are.

Step 5: To tell Excel how your data is organised, click either the "rows" or "columns" option. You'll typically click "columns" because that's how Excel typically arranges data.

Step 6: If there are column headings, step 6 is to verify the "Labels in first row" box.

Step 7: choose a location on the worksheet for your output by clicking the "Output Range" text box. (9).

Example 1: Calculate the correlation between the two characteristics of five different plants.

| Height of the plant (x) | 5 cm | 9 cm | 13 cm | 17 cm | 21 cm |
|---|---|---|---|---|---|
| Number of leaves per plant (y) | 12 | 20 | 25 | 33 | 35 |

The data is arranged in the following table to calculate 'r':

| x | dx | dx² | y | dy | dy² | dx.dy |
|---|---|---|---|---|---|---|
| 5 | 5 – 13 = – 8 | 64 | 12 | 12 – 25 = – 13 | 169 | 104 |
| 9 | 9 – 13 = – 4 | 16 | 20 | 20 – 25 = – 5 | 25 | 20 |
| 13 | 13 – 13 = 0 | 0 | 25 | 25 – 25 = 0 | 0 | 0 |
| 17 | 17 – 13 = + 4 | 16 | 33 | 33 – 25 = + 8 | 64 | 32 |
| 21 | 21 – 13 = + 8 | 64 | 35 | 35 – 25 = + 10 | 100 | 80 |
| $\Sigma x = 65$ | | $\Sigma dx^2 = 160$ | $\Sigma y = 125$ | | $\Sigma dy^2 = 358$ | $\Sigma dx.dy = 236$ |
| $\bar{x} = \dfrac{65}{5} = 13$ | | | $\bar{y} = \dfrac{125}{5} = 25$ | | | |

Now,

$$r = \frac{\Sigma dx \cdot dy}{\sqrt{\Sigma dx^2 \cdot \Sigma dy^2}} = \frac{236}{\sqrt{160 \times 358}} = \frac{236}{\sqrt{57280}} = \frac{236}{239.33} = 0.986 \,(\text{approx.})$$

Given that the two characteristics, such as the plant's height and the number of leaves, are positively associated and that the value of r = 0.986 is as close to 1 as possible, it may be inferred that the relationship is on the verge of being a perfect positive relationship.

Figure 2:

On the germination of Phaseolus seedlings, the pesticide "Nuvan" is tested. The correlation coefficient should be known.

| Conc. of pesticide (ppm) | 0 | 1 | 2.5 | 5 | 7.5 | 10 | 12.5 | 15 | 20 |
|---|---|---|---|---|---|---|---|---|---|
| % germination | 90 | 81 | 65 | 52 | 39 | 32 | 28 | 17 | 6 |

To compute "r," the information is set up in the following table:

| x | dx | dx² | y | dy | dy² | dx.dy |
|---|---|---|---|---|---|---|
| 0 | – 8.16 | 66.58 | 90 | + 44.45 | 1975.8 | – 362.71 |
| 1 | – 7.16 | 51.26 | 81 | + 35.45 | 1256.7 | – 253.82 |
| 2.5 | – 5.66 | 32.03 | 65 | + 19.45 | 378.3 | – 110.08 |
| 5 | – 3.16 | 9.98 | 52 | + 6.45 | 41.6 | – 20.38 |
| 7.5 | – 0.66 | 0.436 | 39 | – 6.55 | 42.9 | + 4.32 |
| 10 | + 1.84 | 3.38 | 32 | – 13.55 | 183.6 | – 24.93 |
| 12.5 | + 4.34 | 18.83 | 28 | – 17.55 | 308.0 | – 76.16 |
| 15 | + 6.84 | 46.78 | 17 | – 28.55 | 815.1 | – 195.28 |
| 20 | + 11.84 | 140.18 | 06 | – 39.55 | 1564.2 | – 468.27 |
| Σx = 73.5 $\bar{x} = \dfrac{73.5}{9}$ = 8.16 | | Σdx² = 369.46 | Σy = 410 $\bar{y} = \dfrac{410}{9}$ = 45.55 | | Σdy² = 6566.2 | Σdx.dy = – 1507.31 |

The correlation coefficient between the two variables, i.e., pesticide concentration and germination %, is 0.9677, which indicates that they are negatively associated, though not exactly so.

**Check Your Progress**

1. List out the steps of Correlation in Excel.

2. Discuss some limitations of Correlation

## 11.6 SUMMARY

In this chapter we saw the meaning of Correlation i.e. According to Prof. King, Correlation suggests that there are some random linkages between two series or groups of data.The relationship known as correlation can indicate whether a change in one variable will result in a change in the other or not. The degree of this association, known as the correlation, coefficient, between the two sets of characters or variables can be stated statistically.The study of correlation has many applications in daily life. The French astronomer Bravis was the first to advance the fundamental ideas of the science of correlation analysis; nevertheless, he was also the first to advance linear correlation theory. Karl Pearson, a statistician, developed the mathematical formula for computing the coefficient of correlation in 1896. Galton and Karl Pearson used the coefficient of correlation to examine a variety of biological and genetics-related issues. The chapter also discusses about Regression and Correlation, Coefficient and Correlation. The Chapter also discusses the importance of Correlation in Social Sciences.

## 11.7 QUESTIONS

1. Discuss the meaning of Correlation

2. Explain Correlation and Coefficient

3. Discuss Correlation and Regression

4. Explain Correlation and different types.

## 11.8 REFERENCES:

**1.** Bryman, A. (1988). Quantity and Quality in Social Research. London:Routledge

**2.** Bryman, A. (2008). Social Research Methods. Oxford UniversityPress

**3.** Matt, H., Weinstein, M., Foard N. (2006) A Short Introduction to Social Research. New Delhi: VistaarPublications

**4.** Goode, W., Hatt, P. (1981). Methods in Social Research. McGraw-Hill BookCompany

**5.** https://www.investopedia.com/terms/c/correlation.asp

**6.** Crossman, Ashley. (2020, August 26). Correlation Analysis in Research. Retrieved from https://www.thoughtco.com/what-is-correlation-analysis-3026696

**7.** https://www.simplypsychology.org/correlation.html

**8.** Samuel, M., & Okey, L. E. (2015). The relevance and significance of correlation in social science research. *International Journal of Sociology and Anthropology Research*, *1*(3), 22-28.

**9.** https://www.statisticshowto.com/probability-and-statistics/correlation-analysis/

**10.** A. G. Asuero, A. Sayago& A. G. González (2006) The Correlation Coefficient: An Overview, Critical Reviews in Analytical Chemistry, 36:1, 41-59, DOI: 10.1080/10408340500526766

**11.** https://www.britannica.com/science/statistics/Residual-analysis#ref367510

**12.** Wells, George "Correlation Coefficient ." Encyclopedia of Public Health. . Retrieved June 21, 2022 from Encyclopedia.com: https://www.encyclopedia.com/education/encyclopedias-almanacs-transcripts-and-maps/correlation-coefficient

❖❖❖❖

# INTRODUCTION TO SPSS

**Unit Structure**

## 12.0 OBJECTIVES

* To learn about the SPSS Software its uses, benefits.
* To understand its application.

## 12.1 INTRODUCTION

The computer is the most recent addition to the research process. A computer is a necessary instrument for research, let it be used for academic or commercial purposes. Computers are now used in almost all fields of science, from genetic engineering to astrophysics research. It has paved the way for the World Wide Web, a multinational information platform. Researchers can perform large-scale study with the help of computers.

Since the introduction, computers have always aided in the solution of issues encountered by human beings. Computers have shrunk in size from the size of a room to that which can fit in the palm of a human hand. The computer or a machine today does computations automatically. It performs a wide range of tasks with incredible speed and efficiency.

People nowadays utilise computers in nearly every aspect of their lives. Computers play a critical part in all scientific endeavours. Different tools and software are helping the research process to simplify in the present times. Various software programmes are now helping such as data gathering, analysis, and so on. One such Statistical software is that of SPSS – Statistical Package for Social Sciences.

In this chapter, we will try to learn about SPSS a Statistical package which is used to analyse data. You can expect a basic introduction to the software, its uses, benefits, limitations. You also learn about some steps with pictorial images as to how the software actually works. This chapter, is included in your syllabus as this software SPSS is widely used in research in recent times. One can even earn a livelihood by learning such software's like SPSS, R (name of another Statistical software) etc. Several research firms hire individuals who are excelling in these software's. In future, with the introduction of this chapter, you can read more and try out practically and become an expert if you are interested and even build careers along these lines.

## 12.2 UNDERSTANDING SPSS

SPSS was initially released in 1968 and stands for "Statistical Package for the Social Sciences." Since its acquisition by IBM in 2009, SPSS has been renamed IBM SPSS Statistics, however most users still refer to it as "SPSS." This software is mainly used for analyzing quantitative data.

SPSS, is a powerful and comprehensive data analysis tool/ software, has a plethora of features meant to make the execution of a wide range of statistical studies easier. It was created for data analysis in the social sciences as SPSS stands for Statistical Package for Social Science. It is highly suited to evaluating data from surveys and databases.SPSS is a groundbreaking piece of software mostly used by researchers that allows them to handle vital data in easy and quicker ways. Working with data is a difficult and time-consuming operation, but with the aid of specific strategies, this programme can simply handle and operate data. Different methods are used to examine, and create a distinctive pattern between various data variables. Furthermore, the output may be retrieved using a graphical representation, allowing the user to comprehend the result quickly.

As seen earlier, SPSS is social science analytical software. It is one of the best programs for statistical data analysis and computing. It is widely employed in other fields like as mathematics, health sciences, and even marketing. It aids in providing ad hoc analysis, selecting and performing analyses, hypothesis testing, data management, and report creation. It also provides numerous data analysis methods.

SPSS is an analytical software tool for Windows. It may be used for data entry, analysis, and the development of tables and graphs. SPSS is capable of handling massive volumes of data at the same time. SPSS's objective is

to convert raw data into useful information. SPSS provides a variety of data management capabilities, such as data recording features, a macros programmer on the visual basic editor, and complicated data sets for completing aggregations.

SPSS is a statistical analysis tool that is frequently used in the field of social science, such as market research, surveys, competitor analysis, and others.

It is a fast and adaptable statistical analysis and data management tool. It is a popular statistical tool that can easily execute very complicated data manipulation and analysis. It is intended for both engaged and passive users.

It analyses data for descriptive statistics, numerical result forecasts, and group identification. This program also includes data processing, charting, and direct marketing functions for efficient data management.

- **What SPSS can do?**

Let's look at what SPSS can achieve now that we have a fundamental understanding of how it works. SPSS is ideal for following a normal project workflow.

- Opening data files, either in SPSS's native format or in a variety of different formats;
- Data editing includes sums and means computations over columns and rows of data. SPSS also includes excellent tools for more sophisticated procedures.
- Generating tables and graphs with frequency counts or summary statistics for (groups of) instances and variables
- ANOVA, regression, and factor analysis are examples of inferential statistics.
- Data and output can be saved in a number of file formats.

## 12.3 CHARACTERISTICS OF SPSS

SPSS is a program that allows you to modify and analyze many types of data. These data can originate from a variety of places, including scientific studies, customer databases, Google Analytics, or even from a website's server log files. The uniqueness about SPSS is also that it can open any file type that is typically used for structured data, such as MS Excel or OpenOffice spreadsheets. (.txt or.csv) simple text files, relational databases (SQL); SAS and Stata.

## 12.4 FUNCTIONS OF SPSS

SPSS's Primary Functions. SPSS has four applications to help researchers with their sophisticated data analysis requirements like-

- Program on Statistics - The Statistics module in SPSS offers a variety of fundamental statistical operations, including frequencies, cross-tabulation, and bivariate statistics, to name a few.

- Modeling Software - Researchers may use the Modeler software in SPSS to create and evaluate prediction models using advanced statistical processes.

- Program for Text Analytics in Surveys - The Text Analytics for Surveys tool from SPSS assists survey administrators in gaining valuable insights from open-ended survey responses.

- Designer of Visualization- With SPSS's Visualization Designer application, researchers may easily build a range of graphics from their survey data, such as density charts and radial boxplots.

- Frequencies, cross-tabulation, and descriptive ratio statistics are examples of descriptive statistics.

- Prediction techniques such as cluster analysis and factor analysis are used to identify groupings.

- Data Transformation: This method is used to change the data's format. It unifies the same type of data in one location after altering the data type, making it easier to handle. You may put any type of data into SPSS, and it will adapt its structure according to the system's specifications. This implies that even if you switch operating systems, SPSS will still be able to deal with previous data.

- Linear regression is a type of numerical result prediction.

- Regression Analysis is a technique for determining the relationship between dependent and interdependent variables in a data collection. It also shows how the dependent data might be affected by a change in the value of an interdependent variable. The fundamental goal of regression analysis is to figure out what kind of relationship exists between various variables.

- ANOVA (Analysis of Variance) is a statistical method for comparing events, groups, or processes and determining their differences. It can assist you in determining which strategy is best for completing a task. The feasibility and efficacy of a procedure may be determined by looking at the results.

- MANOVA (Multivariate analysis of variance): This approach compares data from random variables with unknown values. The

MANOVA method may also be used to investigate different sorts of populations and the factors that influence their decisions.

- T-tests: This approach is used to determine the difference between two sample types, and researchers use it to determine the differences in the interests of two groups. This test can also determine if the output is useless or beneficial.

## 12.5 PROGRAM IN STATISTICS:

The SPSS statistics application provides a wide range of fundamental statistical features, including frequencies, cross tabulation, and bivariate statistics, among others.

- Program for modelers- Using modern statistical processes, researchers may create and validate prediction models.

- Program of Text analytics for Surveys- It provides reliable feedback analysis. It results in a vision for the real plan

- Visualization Designer- Researchers discovered this visual designer data and used it to generate a broad range of graphics such as density charts and radial box plots.

## 12.6 FEATURES OF SPSS:

- Data from every Survey Device may be readily exported to SPSS for thorough and detailed analysis.

- Data in SPSS is saved in SAV format. The majority of this information comes from surveys. This simplifies the process of processing, analyzing, and pulling data.

- SPSS has simple access to data with many variable kinds. This variable data is simple to comprehend. SPSS makes it easy for researchers to put up models because the majority of the procedure is automated.

- Once the data is in SPSS, the magic begins. There is no limit to what we can do with this data

- SPSS also includes a unique method for obtaining specific data from large data sets. SPSS features include trend analysis, assumptions, and prediction models.

- SPSS is simple to learn, use, and apply.

- It aids in the acquisition of a data management system and editing tools.

- SPSS provides extensive statistical tools for examining the actual outcome.

• SPSS assists us in improving the clarity of our design, graphing, reporting, and presentation features.

## 12.7 SPSS Statistical Methods:

SPSS supports a variety of statistical procedures, including the following:

• Prediction for a range of data for group identification, including approaches such as cluster analysis, factor analysis, and so on.

• Descriptive statistics, especially SPSS techniques, are important for frequencies, cross tabulation, and descriptive ratio statistics.

• Bivariate statistics, which include procedures such as analysis of variance (ANOVA), means, correlation, and nonparametric tests, among others.

• Prediction of numerical outcomes, such as linear regression.

It is a self-descriptive utility that automatically assumes you want to open an existing file and displays a dialogue box asking which file you wish to open. SPSS's methodology makes navigating the User experience relatively simple.

Aside from statistical data analysis, the SPSS program also has data management tools, such as the ability to choose data, produce derived data, and reshape files, among other things. Data documentation is another feature. Along with the data file, this feature keeps a metadata dictionary.

**Check Your Progress**

1. Discuss Data management tool of SPSS.

2. SPSS is most often used in which kind of research methodology.

## 12.8 The Benefits of Using SPSS

There are several benefits of using SPSS some of them are discussed below -

• Data files from other applications may be imported and exported using the statistical analysis tool. Some of its data processing techniques are very useful, such as its capacity to combine files, regardless of whether they include the same subjects and different variables or separate subjects with the same variables.
• Users are not obligated to work with syntax in SPSS, despite the fact that syntax files can be stored and edited as needed. When syntax files are saved, it greatly aids documentation and also provides insight into how new variables were computed and missing values were handled.
• It provides accurate and timely responses.
• It's interactive and includes informative tables and graphs.

- Many individuals can use it because it supports a wide range of languages.
- It provides good data management
- It is not difficult to get started with the program.
- Both quantitative and qualitative data might be used.
- With SPSS, the odds of making a mistake are slim.
- One of the simplest statistical methods for data analysis
- Users of SPSS can choose the graph type that best fits their data distribution needs.

## 12.9 USES OF SPSS

SPSS use is not just restricted to that of social science research but it has wider usage. SPSS statistics is a widely used statistical analysis tool in the corporate sector. Users may manage and analyze data and show it in visually appealing graphical formats because of its great features and robustness. It has a graphical user interface and a command-line interface, making the program more user-friendly.SPSS simplifies the processing of complicated data. Working with such data is not straightforward, and it is also a time-consuming procedure.

Let's look at four of the most common sectors where SPSS has been used to a large extent -

- **Market Analysis**

Businesses desire actionable information to help them make difficult and effective business choices. Businesses create massive amounts of data, and manually scanning them is not the best approach to examine them. In such cases, SPSS helps immensely.

SPSS is the finest tool for market researchers seeking for a dependable solution to help them comprehend their data, evaluate trends, forecast, plan, and draw conclusions.

SPSS assists market researchers in extracting meaningful insights from consumer data by employing advanced statistical analytics. It is possible to obtain precise information about market trends thanks to its strong survey data analysis technologies.

There is also Psychographic segmentation, preference scaling, predictive analysis, statistical learning, and a slew of other sophisticated techniques including stratified, clustered, and multistage sampling are all available through the application of SPSS.

- **Education**

Every year, educational institutions face the challenge of enrolling and keeping students. Not to mention the fact that they must recruit new pupils each year. SPSS has large use over here.

SPSS software is presently used by more than 80% of all institutions in the United States.

The capacity of SPSS software to focus on trends allows them to predict a student's future achievement. It employs a number of indicators to identify pupils who are at risk.

The university can utilize SPSS software to analyze a wide range of complicated data sets in order to reveal hidden patterns.

- **Healthcare**

To deliver excellent healthcare, we must address a number of concerns. Some of the most serious concerns in healthcare organizations are obsolete patient delivery techniques and unbalanced incentives for caregivers. This is where analytics can literally save millions of life. There are several applications for using SPSS statistical analysis in healthcare delivery.

When it comes to the healthcare industry, patient data is very important. Not only can bad data have disastrous consequences, but it is also timely, sensitive, and rapid.

Healthcare businesses can use SPSS to develop a patient delivery program based on data. It will not only improve patient outcomes but will also save costs.Univariate and multivariate analysis are used for data sets with complicated connections.

- **Retail**

Analytics are frequently used in the retail business for everything from initial stock planning to anticipating future trends. Customers have a lot of options when it comes to retail merchandise now that social media, forums, and review sites have emerged.

Customers make purchasing selections based on internet reviews about the brand. As a result, it is critical that retail firms provide the finest service possible. Fortunately, statistical analysis can save the retail business.

Data generated by retail enterprises must be gathered, evaluated, and translated into meaningful insights. Businesses will provide exceptional customer experiences if they use data successfully with SPSS software.

SPSS analysis enables merchants to better understand their consumers, supply them with the best solutions, and distribute them through the most effective channels.

SPSS analysis may help you understand anything from how various client groups act to why they make particular purchasing decisions.

SPSS statistics will profile clients based on historical expenditure and behavior trends. It will generate consumer preferences and provide a study

of what causes customers to convert from casual visitors to shoppers by utilizing this data.

## 12.10 BEGINNING WITH SPSS

- **SPSS Installation Guide -**

Start – All programs – SPSS Inc – SPSS 16.0, then double-click the icon to launch SPSS.

First and foremost, we must review the minimal system requirements at SPSS Statistics System Requirements.

The selection then identifies the operating system installed on your machine and determines the prerequisites.

Open a browser and navigate to the SPSS website, which will result in the download of the software application. Begin with the free trial version of SPSS.

A Data Editor window with Data View and Variable View would open by default. The first thing when we open SPSS is the Data Editor where we add data, edit, modify, save, we can also define etc. Like in Excel sheet we have a display window. The second thing to understand is the output Viewer where one can see the output the finishing. There are sample packs available where you may practice learning.

- **Variable View**
1. Name: This is a column field that accepts the unique ID. This aids in data sorting. For example, different demographic characteristics such as name, gender, age, and educational degree can be used to sort data. The sole restriction is that special characters are not permitted in this kind.
2. Label: As the name implies, it provides the label. This also allows for the addition of special characters.
3. Type: This is particularly important when inserting various types of data.
4. Width: Character length may be measured.
5. Decimal: When inputting a percentage figure, this type helps us determine how many digits are necessary following the decimal.
6. Value: This assists the user in entering the value.
7. Missing: This allows the user to skip over superfluous data during analysis.
8. Align: As the name implies, alignment aids in left or right alignment. However, in this situation, for example, left align.
9. Measure: This aids in the measurement of data entered into instruments such as ordinal, cardinal, and nominal. (Refer image 1)
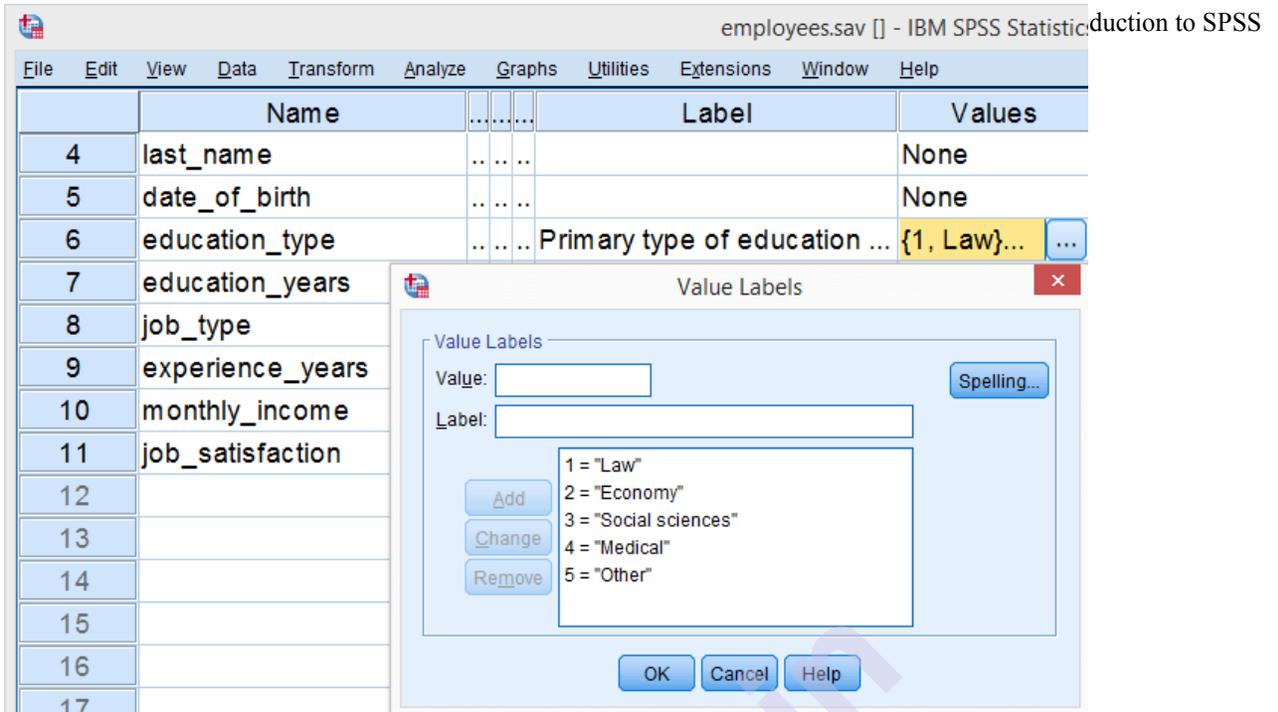
Image 1 - The above given image shows the screenshot of Variable page.

The data must be entered into the "variable view" page. It enables us to adapt the data type as needed for analysis.

In short, to evaluate the data, fill out the various column headers such as Name, Label, Type, Width, Decimals, Values, Missing, Columns, Align, and Measures.

These headers represent the many aspects that assist to characterize the data.

• **Data Viewing -**

The data display is organized into rows and columns. We can operate with SPSS by importing a file or manually entering data.

• **Procedure for importing EXCEL files into that of SPSS**

The first step is to select File.

=> Click to open

=> Choose Data

=> Dialog Box

=> Type files

.xls spreadsheet

After selecting the excel file to be imported for data analysis, we must verify that the "read variable names from the first row of data" option is chosen in the dialogue box.

Finally, press the OK button. SPSS has now imported your file.
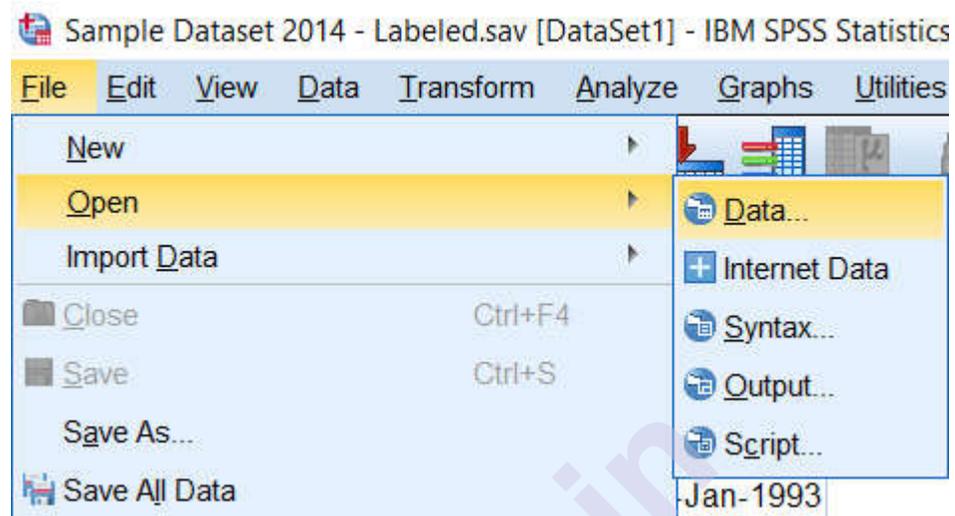


Image2- shows the importing procedure of Excel file into SPSS

- **Analyzing Data**

Once, the Data sets are imported or feeded into the SPSS software the specific commands have to give and depending upon that the data is generated. One can even get a graphical kind of a data like charts, pie diagram etc. The following are some of the images which would help you understand this in pictorial format. At the endnote of the chapter there is also a YouTube video link provided where you could see the video of how to do the data analysis. (Refer image 2)

Depending upon the required output of the series of questions and the derived answers of the subject. The data can be produced in the required form like Pie chart, Bar chart etc. and can be seen in the output viewer.
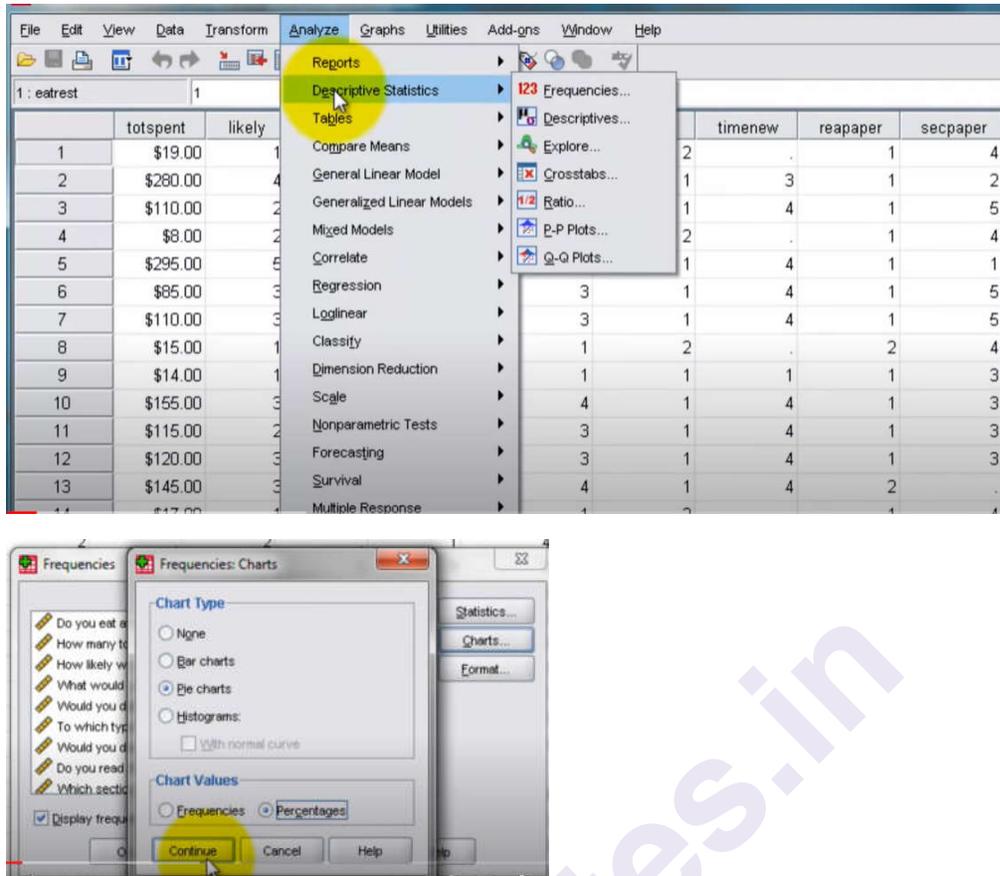
Image 3 – In the above given procedure in the images explains the data analysis procedure.

## 12.11 LIMITATIONS OF SPSS

Unlike anyother other research tool SPSS too has some limitations like -

If researchers gather data using inaccurate or biased procedures, the statistical analysis that results will not provide accurate results. There is no problem if the difference between the sample and the real population is insignificant. However, if the discrepancy is significant, the results will be deceptive.

The SPSS analysis will not fail if researchers do not measure the specific thing they intend to assess. Another problem with utilizing a statistical analysis program like SPSS is that you wind up with easy answers to complicated problems.

**Check Your Progress**

1. List out two uses of SPSS

2. Discuss one limitation of SPSS.

## 12.12 SUMMARY

The use of computers and software's has made the research faster in social sciences than ever before. In this chapter, we learnt about one such software called SPSS a statistical software which can analyze data at a very faster pace, especially large amount of data. Often SPSS is used in quantitative research-based studies. While Excel is useful for data organization, SPSS is better suited for in-depth data analysis. This tool is quite handy for data analysis and visualization. We also saw, the uses of SPSS in several fields like medicine, retails etc. We also discussed about certain limitation like if error is at the data collection point then the output could be different.

## 12.13 QUESTIONS

1. Discuss SPSS uses and its limitations.

2. Discuss the functions of SPSS

## 12.14 REFERENCES

1. https://johnnoels.medium.com/what-is-spss-and-its-importance-in-research-data-analysis-5f109ab90da1

2. https://www.spss-tutorials.com/spss-what-is-it/

3. https://lo.unisa.edu.au/mod/book/view.php?id=646443&chapterid=1 06606https://www.alchemer.com/resources/blog/what-is-spss/

4. https://surveysparrow.com/blog/what-is-spss/
5. https://johnnoels.medium.com/what-is-spss-and-its-importance-in-research-data-analysis-5f109ab90da1

❖❖❖❖