Ambo University Waliso Campus

School of Law and Governance

Department of Governance and Development Studies

Quantitative Research Methods (GaDS 1032) Teaching Material

For GaDS II Year Students

Course Credit Hour: 3chrs/4 ECTS

Instructor: Gete N. (MA)

May, 2020

Chapter One: Introduction

✪ Unit Introduction

Dear learner, this introductory unit of your module will introduce you to an overview of the course on 'Quantitative Research'. It aims at introducing you to overview of what quantitative research, its rationales and the difference it has with qualitative research. The unit consists of definitions of quantitative research; rationale behind using quantitative research, its advantage and disadvantages, and types of quantitative research

Unit objectives:

Dear learner, by the end of this section you should be able to:

- ✓ Understand the concept of quantitative research;
- ✓ Explain rationale of why we use quantitative research its rationale;
- ✓ Identify the differences between quantitative and qualitative researches;
- ✓ Identify advantages as well as limitations of quantitative research; and
- ✓ Differentiate among types of quantitative research.

✹ Unit pre-test questions

1. What is quantitative research?
2. Are quantitative and qualitative researches the same?
3. What are the rationales for using quantitative research?
4. What are advantages and disadvantages of quantitative research?
5. What types of quantitative research?

1.1 What is Quantitative Research?

✪ Section overview

Research can be divided in to qualitative and quantitative. Under this section you will learn about quantitative research in detail and its difference with qualitative research.

Section objectives

Dear learner, by the end of this section you should be able to:

- ✓ Define quantitative research;
- ✓ Identify the differences between quantitative and qualitative research; and
- ✓ Determine when to use quantitative or qualitative research.

Quantitative research is the systematic empirical investigation of observable phenomena via statistical, mathematical, or computational techniques. The objective of quantitative research is to develop and employ mathematical models, theories, and hypotheses pertaining to phenomena. The process of measurement is central to quantitative research because it provides the fundamental connection between empirical observation and mathematical expression of quantitative relationships.

? What do you know about quantitative data?

Quantitative data is any data that is in numerical form such as statistics, percentages, etc. The researcher analyses the data with the help of statistics and hopes the numbers will yield an unbiased result that can be generalized to some larger population. Qualitative research, on the other hand, inquires deeply into specific experiences, with the intention of describing and exploring meaning through text, narrative, or visual-based data, by developing themes exclusive to that set of participants.

? In which disciplines may quantitative research be used?

Quantitative research is widely used in psychology, economics, demography, sociology, marketing, community health, health & human development, gender studies, and political science; and less frequently in anthropology and history. Research in mathematical sciences, such as physics, is also "quantitative" by definition, though this use of the term differs in context.

? What are the differences between quantitative and qualitative researches?

Quantitative VS Qualitative Research

Quantitative	**Qualitative**
Quantitative information or data is based on quantities obtained using a quantifiable measurement process-	Qualitative information records qualities that are descriptive, subjective or difficult to measure
Produces “numerical data” or information that can be converted into numbers	Generates “textual data” (non-numerical)
Focuses on the measurement of quantity or amount and quantifying it	Focuses on describing a situation, phenomenon, problem or event rather than quantifying it
Surveys and experiment and document review are data collection tools.	In depth interview, participant observation or an in-depth analysis of individual case are data collection tools
Structured research instruments are used for data gathering	Data gathering instruments may not be structured
Data is available in the form of number	
Less attention is given to behaviour, attitudes and motivation of respondents	Aims at to discovering the underlying motives of human behavior.
Large sample size	Small sample size
Requires statistical procedures or in-depth mathematical analysis for data analysis	Doesn’t require statistical procedures or in-depth mathematical analysis
More objective	More subjective

In quantitative research, the researcher, attempts to describe relationship among variables mathematically. The most significant strength of quantitative research is, the increased ability to generalize quantitative results to the greater population. The purpose of quantitative research is to generate knowledge and create understanding about the social world.

Some examples of quantitative data:

- ✓ A jug of milk holds one gallon.
- The painting is 14 inches wide and 12 inches long.
- The new baby weighs six pounds and five ounces.
- A bag of broccoli crowns weighs four pounds.
- A coffee mug holds 10 ounces.
- John is six feet tall.
- A tablet weighs 1.5 pounds

1.2 Rationale of Quantitative research

✪ Section Overview

Quantitative research is used to quantify behaviors, opinions, attitudes, and other variables and make generalizations from a larger population. The main goal is to understand the relationship between an independent and dependent variable in a population. Under this section we will learn about the rationale of using quantitative research.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Explain the reasons of using quantitative research; and
- ✓ Determine when to use quantitative research.

? Why we choose quantitative research over qualitative research?

It is more scientific: A large amount of data is gathered and then analyzed statistically. This almost erases bias, and if more researchers ran the analysis on the data, they would always end up with the same numbers at the end of it.

It is less biased/ objective: The research aims for objectivity i.e. without bias, and is separated from the data. Researcher has clearly defined research questions to which objective answers are sought.

It is focused: The design of the study is determined before it begins and research is used to test a theory and ultimately support or reject it.

It deals with larger samples: The results are based on larger sample sizes that are representative of the population. The large sample size is used to gain statistically valid results in customer insight.

It is repeatable: The research study can usually be replicated or repeated, given its high reliability.

It is arranged in simple analytical methods: Received data are in the form of numbers and statistics, often arranged in tables, charts, figures, or other non-textual forms.

It is generalizable: Project can be used to generalize concepts more widely, predict future results, or investigate causal relationships. Findings can be generalized if selection process is well-designed and sample is representative of a study population.

It is relatable: Quantitative research aims to make predictions, establish facts and test hypotheses that have already been stated. It aims to find evidence which supports or does not support an existing hypothesis. It tests and validates already constructed theories about how and why phenomena occur. More structured: Researcher uses tools, such as questionnaires or equipment to collect numerical data.

It is pertinent in later stages of research: Quantitative research is usually recommended in later stages of research because it produces more reliable results.

It is consistent with data: With quantitative research, you may be getting data that is precise, reliable and consistent, quantitative and numerical.

It is more acceptable: It may have higher credibility among many influential people (e.g., administrators, politicians, sponsors, donors)

It is fast: Data collection using quantitative methods is relatively quick (e.g., telephone interviews). Also, data analysis is relatively less time consuming (using statistical software).

It is useful for decision making: Data from quantitative research—such as market size, demographics, and user preferences—provides important information for business decisions.

1.3 Advantages and Limitations of Quantitative Method

✪ Section Overview

Using quantitative research has its own advantage and disadvantage. Thus, a research needs to know this merits and demerits in advance. This section will teach you about them.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Identify the advantages and disadvantages of quantitative method.

? What advantages may quantitative research has?

Advantages:

Collect reliable and accurate data: As data is collected, analyzed and presented in numbers, the results obtained will be extremely reliable. Numbers do not lie. They present an honest picture of the conducted research without inconsistency and is also extremely accurate.

Quick data collection: A quantitative research is carried out with a group of respondents who represent a population. A survey or any other quantitative research method applied to these respondents and the involvement of statistics, conducting and analyzing results is quite straightforward and less time-consuming.

Wider scope of data analysis: Due to the statistics, this research method provides a wide scope of data collection.

Eliminate bias: This research method offers no scope for personal comments or biasing of results. The results achieved are numerical and are thus, fair in most cases. While the results of qualitative research can vary according to the skills of observer, the results of quantitative research are interpreted in an almost similar manner by all experts.

Helps to generalize findings- every finding developed through this method can go beyond the participant group to the overall demographic being looked at with this work

You can perform the research remotely.

N.B Both qualitative and quantitative researches have their own advantages and disadvantages. It is recommended that nobody has to simply attach to one of these methods. Since neither are superior to the other. Even many studies require combining both methods. Generally, the research problem should determine whether the study is carried out by quantitative or qualitative methods.

? What are the limitations quantitative research have?

Limitations

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 organizations, public service providers, educational institutions, etc.) may lack knowledge and especially the resources needed to conduct a careful 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. Responses often depend on particular time which again is dependent on the conditions occurring during that particular time frame.

For example, if data for a study is collected on residents' perception of development works conducted by the municipality, the results presented for a specific year (say, 2009), will be held redundant or of limited value in 2015. Reasons being, either the officials have changed or the development scenario have changed (from too effective to minimal effective or vice versa).

4. **Limited outcomes in a quantitative research**

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 generalized form. Also, the respondents have limited options of responses, based on the selection made by the researcher.

For example, answer to a question– "Does your manager motivates you to take up challenges"; can be yes/no/can't say or Strongly Agree to strongly disagree.

But to know what are the strategies applied by the manager to motivate the employee or on what parameters the employee does not feel motivated (if responded no), the researcher has to ask broader questions which somewhat has limited scope in close-ended questionnaires.

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. 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. Quantitative research is a lot more complex for social sciences, education, anthropology and psychology.

7. **Requirement of extra resources to analyze the results**

The requirements for the successful statistical confirmation of result are very tough in a quantitative research. Hypothesis is proven with few experiments due to which there is ambiguity in the results. Results are retested and refined several times for an unambiguous conclusion. So, it requires extra time, investment and resources to refine the results.

1.4 Types of Quantitative Research

✪ Section Overview

Dear learner quantitative research in not only of one type, there are different types of quantitative research. In this section detail of them will be presented.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Identify different types of quantitative research; and
- ✓ Determine the situation in which the different quantitative research types are used

? What are four main types of quantitative research designs?

1. Descriptive Research

- ➢ More focused on the 'what' of the subject matter rather than the 'why'.i.e.
- ➢ It aims to describe the current status of a variable or phenomenon.
- ➢ It describes circumstances.
- ➢ It can be used to define respondent characteristics, organize comparisons, measure data trends, validate existing conditions.
- ➢ Data collection is mostly by observation and the researcher does not begin with a hypothesis but, creates one after the data is collected.
- ➢ Albeit very useful, this method cannot draw conclusions from received data and cannot determine cause and effect.

2. Correlational Research

- ➢ Correlational research is a non-experimental research method, where the researcher measures two variables, and studies the statistical relationship i.e. the correlation between variables.

- The researcher ultimately assesses that relationship without influence from any peripheral variable.

Example: louder the jingle of an ice cream truck is, the closer it is to us

- The most prominent feature of correlational research is that the two variables are measured – neither is manipulated.
- A correlation has direction and can be either positive or negative.
- It can also differ in the degree or strength of the relationship.

3. Experimental Research

- Often referred to as 'true experimentation', this type of research method uses a scientific method to establish cause-effect relationship among a group of variables.
- It is commonly defined as a type of research where the scientist actively influences something to observe the consequences.
- It is a systematic and scientific approach to research in which the researcher manipulates one or more variables, and controls/randomizes any change in other variables.
- Experimental research is commonly used in sciences such as sociology and psychology, physics, chemistry, biology and medicine and so on.

4. Quasi-experimental Research

- ✓ The prefix quasi means "resembling".
- ✓ Quasi-experimental research resembles experimental research but is not a true experimental research.
- ✓ It is often referred to as 'Causal-Comparative'.
- ✓ Quasi-experimental involves 'comparison.'
- ✓ The study of two or more groups is done without focusing on their relationship.

? Which type of quantitative reseach shows the statistical relationship?

Activity 2

1. Discuss about quantitative research and its difference with qualitative research
2. Discuss the rationale of quantitative research
3. Discuss the merits and demerits of quantitative research
4. Discuss the types of quantitative research

Chapter Two: Sampling in Quantitative Research

✪ Unit Introduction

Dear learner, this unit will introduce you to an sampling in quantitative research. Sampling is the very important thing in research. It is the process of selecting the representative the population under the study. The unit consists: basic concepts in sampling, sampling process, merits and demerits of sampling, sampling techniques and sample size determination.

Unit objectives:

Dear learner, by the end of this section you should be able to:

- ✓ Understand what is sampling;
- ✓ Understand basic concepts in sampling;
- ✓ Discuss the sampling process;
- ✓ Explain the advantages and disadvantages of sampling;
- ✓ Identify the sampling techniques; and
- ✓ Determine the sample for a given population.

✹ Unit pre-test questions

1. What is sampling?
2. What is sampling process?
3. What advantage and disadvantage does sampling have?
4. What sampling techniques are there?
5. How can we determine sample for some population?

2.1 Basic concepts of sampling

✪ Section Overview

This section will introduce you with some basic concepts that are common with sampling. Understand these concepts are very useful to understand sampling well.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Understand the basic concepts in sampling; and
- ✓ Determine how to use sampling related concepts.

Sampling may be defined as the procedure in which a sample is selected from an individual or a group of people of certain kind for research purpose. In sampling, the population is divided into a number of parts called sampling units. It would normally be impractical to study a whole population, for example when doing a questionnaire survey. Sampling is a method that allows researchers to infer information about a population based on results from a subset of the population, without having to investigate every individual. Reducing the number of individuals in a study reduces the cost and workload, and may make it easier to obtain high

quality information, but this has to be balanced against having a large enough sample size with enough power to detect a true association. Sampling refers to the process of selecting a few (a sample) from a bigger group (the sampling population) to become the basis for estimating or predicting the prevalence of a situation or old come regarding the bigger group.

? What type of sample a researcher should select?

If a sample is to be used, by whatever method it is chosen, it is important that the individuals selected are representative of the whole population. This may involve specifically targeting hard to reach groups. For example, if the electoral roll for a town was used to identify participants, some people, such as the homeless, would not be registered and therefore excluded from the study by default.

It is incumbent on the researcher to clearly define the target population. There are no strict rules to follow, and the researcher must rely on logic and judgment. The population is defined in keeping with the objectives of the study. Sometimes, the entire population will be sufficiently small, and the researcher can include the entire population in the study. This type of research is called a census study because data is gathered on every member of the population. Usually, the population is too large for the researcher to attempt to survey all of its members. A small, but carefully chosen sample can be used to represent the population. The sample reflects the characteristics of the population from which it is drawn.

? What basic terms in sampling you know?

- **Sample**- Is a group of people, objects, or items that are taken from a larger population for measurement. The sample should be representative of the population to ensure that we can generalise the findings from the research sample to the population as a whole.
- **Population Size -**The number of records in the overall data set to be sampled. Before performing a sampling routine, the population size must be determined. Within the sampling function, the population size can be retrieved at design time, generated live at run time, or defined as a fixed value. Although the population size is usually the total number of records in the file, there will be times when you want to sample a subset of the file. To ensure a statistically valid sample, the specified population size must be the same as the number of records sampled.
- **Confidence Level**- Is the probability, expressed in percent, that the selected sample will represent the total population. Most guidelines establish a minimum acceptable confidence level of 90%, 95%, or 98%.
- **Margin of Error**- Represents the amount of error, expressed in percent, that you can tolerate. Lower margins of error require larger sample sizes.
- **Response Distribution**- Allows you to correct for skewness in the sample (if the sample deviates from the normal standard deviation). Use the Response Distribution percentage to account for the skewness in population.
- **Seed-**The statistical sampling routines use a seed value for the pseudo random number generator. This generator produces a series of random numbers from the entered seed. These random numbers are then used to determine each record to include in the

sample. The seed has no effect on the number of records included in the sample, it only affects which records are selected. A single seed will produce the same set of random numbers, so if you want to replicate a sample, use the same seed, population size, and record order. To generate a unique sample, enter a new seed value each time.

- **Sample Size-**The number of records that you want to store in the output file. The number should be a positive integer, greater than zero and less than the population size.

Example for population size and sample size;

Suppose you choose 1000 students among 4 million students to generalize about the result from the 1000 students for the 4 million students. Then

- 4 million students is population
- 1000 is the size of sample

- **Sample design** is a definite plan for obtaining a sample from population. It refers to the technique or the procedure the researcher would adopt in selecting items for the sample. The sample design may as well lay dawn the number of items to be included in the sample i.e. the size of the sample. It has its own sequential steps/ Sampling processes.

2.2 Sampling Processes

✪ Section Overview

It is meant that sampling is the process of selecting representative of the population. Thus, in selection of these representatives the researcher needs to follow some processes and this section introduces you with these processes.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Identify the sampling processes; and
- ✓ Apply sampling processes in sample selection

? What sampling processes do you know?

The followings are sampling are sampling processes:

1. Defining the population: - This is the whole group to which the researcher would want to generalize after selecting some amount of sample. The whole population towards which the researcher directs his/her attention is called as study population usually represented by the letter (N).

2. Specify the sampling frame: - The sampling frame is the physical material from which sample is chosen. This is a document that lists the study population. It may include telephone directory, list of business establishment in a town, list of households and residents in a kebele etc.

3. Specify the sampling unit (element): - sampling unit represents the basic unit containing the element of the population to be sampled. It may contain one or several population elements.

4. Specify the sampling method: - refers to the method that you need to follow in selecting the sample from the whole population. For instance, what sampling methods did you use to select only 200 civil servants from a sum total of 20,000.

5. Determine the sample size: - Here the amount to be selected from the total population (sample size) is determined. The sample size is the specific population from which you obtain information.

6. Specify sampling plan: -In this step you need to specify operational procedure by which you communicate each sample unit.

7. Select the sample: - here all office and field activities necessary to select the sample is completed and the sample is selected for the study.

? Can a researcher by pass one of the above sampling processes in sampling?

2.3 Advantages and disadvantages of sampling

✪ Section Overview

Dear learner, in the last section of this chapter you have learnt about sampling and sampling processes. Sampling has its own merit that researchers need to know. This section will introduce you with the advantages and disadvantages sampling has.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Understand merits and demerits of sampling; and
- ✓ Explain how to cope up with the limitation of sampling

? What types of merits sampling has?

2.3.1 Advantages of Sampling

Sampling ensures convenience, collection of intensive and exhaustive data, suitability in limited resources and better rapport. In addition to this, sampling has the following advantages also.

1. Low cost of sampling- If data were to be collected for the entire population, the cost will be quite high. A sample is a small proportion of a population. So, the cost will be lower if data is collected for a sample of population which is a big advantage.

2. Less time consuming in sampling-Use of sampling takes less time also. It consumes less time than census technique. Tabulation, analysis etc., take much less time in the case of a sample than in the case of a population.

3. Scope of sampling is high-The investigator is concerned with the generalization of data. To study a whole population in order to arrive at generalizations would be impractical. Some populations are so large that their characteristics could not be measured. Before the measurement has been completed, the population would have changed. But the process of sampling makes it possible to arrive at generalizations by studying the variables within a relatively small proportion of the population.

4. Accuracy of data is high-Having drawn a sample and computed the desired descriptive statistics, it is possible to determine the stability of the obtained sample value. A sample represents the population from which it is drawn. It permits a high degree of accuracy due to a limited area of operations. Moreover, careful execution of field work is possible. Ultimately, the results of sampling studies turn out to be sufficiently accurate.

5. Organization of convenience-Organizational problems involved in sampling are very few. Since sample is of a small size, vast facilities are not required. Sampling is therefore economical in respect of resources. Study of samples involves less space and equipment.

6. Intensive and exhaustive data-In sample studies, measurements or observations are made of a limited number. So, intensive and exhaustive data are collected.

7. Suitable in limited resources-The resources available within an organization may be limited. Studying the entire universe is not viable. The population can be satisfactorily covered through sampling. Where limited resources exist, use of sampling is an appropriate strategy while conducting marketing research.

8. Better rapport-An effective research study requires a good rapport between the researcher and the respondents. When the population of the study is large, the problem of rapport arises. But manageable samples permit the researcher to establish adequate rapport with the respondents.

2.3.2 Disadvantages of Sampling

? What types of demerits sampling has?

1. Chances of bias

The serious limitation of the sampling method is that it involves biased selection and thereby leads us to draw erroneous conclusions. Bias arises when the method of selection of sample employed is faulty. Relatively small samples properly selected may be much more reliable than large samples poorly selected.

2. Difficulties in selecting a truly representative sample

Difficulties in selecting a truly representative sample produce reliable and accurate results only when they are representative of the whole group. Selection of a truly representative sample is difficult when the phenomena under study are of a complex nature. Selecting good samples is difficult.

3. Inadequate knowledge in the subject

Use of sampling method requires adequate subject specific knowledge in sampling technique. Sampling involves statistical analysis and calculation of probable error. When the researcher lacks specialized knowledge in sampling, he may commit serious mistakes. Consequently, the results of the study will be misleading.

4. Changeability of units

When the units of the population are not in homogeneous, the sampling technique will be unscientific. In sampling, though the number of cases is small, it is not always easy to stick to the, selected cases. The units of sample may be widely dispersed.

Some of the cases of sample may not cooperate with the researcher and some others may be inaccessible. Because of these problems, all the cases may not be taken up. The selected cases may have to be replaced by other cases. Changeability of units stands in the way of results of the study.

5. Impossibility of sampling

Deriving a representative sample is difficult, when the universe is too small or too heterogeneous. In this case, census study is the only alternative. Moreover, in studies requiring a very high standard of accuracy, the sampling method may be unsuitable. There will be chances of errors even if samples are drawn most carefully.

? Do you think that a researcher can cope up with sampling limitations?

2.4 Characteristics of a Good Sample

✪ Section Overview

In the last section we have seen the advantage and limitations of sampling. In this section we will see features of good sampling. When sampling your sample, you should try to have samples with these good characters.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss good features of sample;
- ✓ Analyze whether a sample is good or bad; and
- ✓ Select sample that fits the good sampling features.

The followings are features of good sample:

(1) Goal-oriented: A sample design should be goal oriented. It is means and should be oriented to the research objectives and fitted to the survey conditions.

(2) Accurate representative of the universe: A sample should be an accurate representative of the universe from which it is taken. There are different methods for selecting a sample. It will be truly representative only when it represents all types of units or groups in the total population in fair proportions. In brief sample should be selected carefully as improper sampling is a source of error in the survey.

(3) Proportional: A sample should be proportional. It should be large enough to represent the universe properly. The sample size should be sufficiently large to provide statistical stability or reliability. The sample size should give accuracy required for the purpose of particular study.

(4) Random selection: A sample should be selected at random. This means that any item in the group has a full and equal chance of being selected and included in the sample. This makes the selected sample truly representative in character.

(5) Economical: A sample should be economical. The objectives of the survey should be achieved with minimum cost and effort.

(6) Practical: A sample design should be practical. The sample design should be simple i.e. it should be capable of being understood and followed in the fieldwork.

(7) Actual information provider: A sample should be designed so as to provide actual information required for the study and also provide an adequate basis for the measurement of its own reliability.

In brief, a good sample should be truly representative in character. It should be selected at random and should be adequately proportional. These, in fact, are the attributes of a good sample.

2.5 Sampling Techniques

✪ Section Overview

Dear learner, in the last section of this chapter you have learnt about sampling processes and advantages and disadvantages sampling. Again, in this section you will learn about sampling techniques.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss the sampling technique;
- ✓ Explain how probability sampling differs from probability sampling; and
- ✓ Determine which technique is good for one based on the type of research.

? What types of sampling techniques you know?

There are several different sampling techniques available, and they can be subdivided into two groups: probability sampling and non-probability sampling.

In probability (random) sampling, you start with a complete sampling frame of all eligible individuals from which you select your sample. In this way, all eligible individuals have a chance of being chosen for the sample, and you will be more able to generalise the results from your study. Probability sampling methods tend to be more time-consuming and expensive than non-probability sampling.

In non-probability (non-random) sampling, you do not start with a complete sampling frame, so some individuals have no chance of being selected. Consequently, you cannot estimate the effect of sampling error and there is a significant risk of ending up with a non-representative sample which produces non-generalisable results. However, non-probability sampling methods tend to be cheaper and more convenient, and they are useful for exploratory research and hypothesis generation.

A. Probability Sampling Methods

? How many types of probability sampling are there?

1. Simple random sampling

In this case each individual is chosen entirely by chance and each member of the population has an equal chance, or probability, of being selected. One way of obtaining a random sample is to give each individual in a population a number, and then use a table of random numbers to decide which individuals to include.1 For example, if you have a sampling frame of 1000 individuals, labelled 0 to 999, use groups of three digits from the random number table to pick your sample. So, if the first three numbers from the random number table were 094, select the individual labelled "94", and so on.

As with all probability sampling methods, simple random sampling allows the sampling error to be calculated and reduces selection bias. A specific advantage is that it is the most straightforward method of probability sampling. A disadvantage of simple random sampling is that you may not select enough individuals with your characteristic of interest, especially if that characteristic is uncommon. It may also be difficult to define a complete sampling frame and inconvenient to contact them, especially if different forms of contact are required (email, phone, post) and your sample units are scattered over a wide geographical area.

2. Systematic sampling

Individuals are selected at regular intervals from the sampling frame. The intervals are chosen to ensure an adequate sample size. If you need a sample size n from a population of size x, you should select every x/nth individual for the sample. For example, if you wanted a sample size of 100 from a population of 1000, select every 1000/100 = 10th member of the sampling frame.

Systematic sampling is often more convenient than simple random sampling, and it is easy to administer. However, it may also lead to bias, for example if there are underlying patterns in the order of the individuals in the sampling frame, such that the sampling technique coincides with the periodicity of the underlying pattern. As a hypothetical example, if a group of students were being sampled to gain their opinions on college facilities, but the Student Record Department's central list of all students was arranged such that the sex of students alternated between male and female, choosing an even interval (e.g. every 20th student) would result in a sample of all males or all females. Whilst in this example the bias is obvious and should be easily corrected, this may not always be the case.

3. Stratified sampling

In this method, the population is first divided into subgroups (or strata) who all share a similar characteristic. It is used when we might reasonably expect the measurement of interest to vary between the different subgroups, and we want to ensure representation from all the subgroups. For example, in a study of stroke outcomes, we may stratify the population by sex, to ensure

equal representation of men and women. The study sample is then obtained by taking equal sample sizes from each stratum. In stratified sampling, it may also be appropriate to choose non-equal sample sizes from each stratum. For example, in a study of the health outcomes of nursing staff in a country, if there are three hospitals each with different numbers of nursing staff (hospital A has 500 nurses, hospital B has 1000 and hospital C has 2000), then it would be appropriate to choose the sample numbers from each hospital proportionally (e.g. 10 from hospital A, 20 from hospital B and 40 from hospital C). This ensures a more realistic and accurate estimation of the health outcomes of nurses across the county, whereas simple random sampling would over-represent nurses from hospitals A and B. The fact that the sample was stratified should be considered at the analysis stage.

Stratified sampling improves the accuracy and representativeness of the results by reducing sampling bias. However, it requires knowledge of the appropriate characteristics of the sampling frame (the details of which are not always available), and it can be difficult to decide which characteristic(s) to stratify by.

4. Clustered sampling

In a clustered sample, subgroups of the population are used as the sampling unit, rather than individuals. The population is divided into subgroups, known as clusters, which are randomly selected to be included in the study. Clusters are usually already defined, for example individual GP practices or towns could be identified as clusters. In single-stage cluster sampling, all members of the chosen clusters are then included in the study. In two-stage cluster sampling, a selection of individuals from each cluster is then randomly selected for inclusion. Clustering should be considered in the analysis. The General Household survey, which is undertaken annually in England, is a good example of a (one-stage) cluster sample. All members of the selected households (clusters) are included in the survey.1

Cluster sampling can be more efficient that simple random sampling, especially where a study takes place over a wide geographical region. For instance, it is easier to contact lots of individuals in a few GP practices than a few individuals in many different GP practices. Disadvantages include an increased risk of bias, if the chosen clusters are not representative of the population, resulting in an increased sampling error.

B. Non-Probability Sampling Methods

? What types of non-probability sampling techniques are there?

1. Convenience sampling

Convenience sampling is perhaps the easiest method of sampling, because participants are selected based on availability and willingness to take part. Useful results can be obtained, but the results are prone to significant bias, because those who volunteer to take part may be different from those who choose not to (volunteer bias), and the sample may not be representative of other characteristics, such as age or sex. Note: volunteer bias is a risk of all non-probability sampling methods.

2. Quota sampling

This method of sampling is often used by market researchers. Interviewers are given a quota of subjects of a specified type to attempt to recruit. For example, an interviewer might be told to go out and select 20 adult men, 20 adult women, 10 teenage girls and 10 teenage boys so that they could interview them about their television viewing. Ideally the quotas chosen would proportionally represent the characteristics of the underlying population.

Whilst this has the advantage of being relatively straightforward and potentially representative, the chosen sample may not be representative of other characteristics that weren't considered (a consequence of the non-random nature of sampling).

3. Judgement (or Purposive) Sampling

Also known as selective, or subjective, sampling, this technique relies on the judgement of the researcher when choosing who to ask to participate. Researchers may implicitly thus choose a "representative" sample to suit their needs, or specifically approach individuals with certain characteristics. This approach is often used by the media when canvassing the public for opinions and in qualitative research.

Judgement sampling has the advantage of being time-and cost-effective to perform whilst resulting in a range of responses (particularly useful in qualitative research). However, in addition to volunteer bias, it is also prone to errors of judgement by the researcher and the findings, whilst being potentially broad, will not necessarily be representative.

4. Snowball sampling

This method is commonly used in social sciences when investigating hard-to-reach groups. Existing subjects are asked to nominate further subjects known to them, so the sample increases in size like a rolling snowball. For example, when carrying out a survey of risk behaviours amongst intravenous drug users, participants may be asked to nominate other users to be interviewed.

Snowball sampling can be effective when a sampling frame is difficult to identify. However, by selecting friends and acquaintances of subjects already investigated, there is a significant risk of selection bias (choosing a large number of people with similar characteristics or views to the initial individual identified).

2.6 Sample Size Determination

✪ Section Overview

Dear learner, it is meant that sample should be representative of the population in any means. Thus, to determine the sample size which is a very representative of the community we need to determine the sample size using different techniques like formula. This section will introduce you with how to determine sample size.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss about sample size;
- ✓ Determine sample size for a given population; and
- ✓ Analyze whether a given sample is representative of the community.

? What do you think of sample size?

The sample size is a term used in research for defining the number of subjects included in a sample size. By sample size, we understand a group of subjects that are selected from the general population and is considered a representative of the real population for that specific study.

Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample.

A critically important aspect of any study is determining the appropriate sample size to answer the research question. Studies should be designed to include a sufficient number of participants to adequately address the research question. Studies that have either an inadequate number of participants or an excessively large number of participants are both wasteful in terms of participant and investigator time, resources to conduct the assessments, analytic efforts and so on. These situations can also be viewed as unethical as participants may have been put at risk as part of a study that was unable to answer an important question. Studies that are much larger than they need to be to answer the research questions are also wasteful.

Sample size determination depend on:

i) **Degree of homogeneity:** The size of the population variance is an important parameter. The greater the dispersion in the population the larger the sample must be to provide a given estimation precession.

ii) **Degree of confidence required**: Since a sample can never reflect its population for certain, the researcher must determine how much precision/accuracy s/he needs. Precision is measured in terms of

- ➢ the margin of error.
- ➢ The degree of confidence (how sure you are)

iii) **Number of sub groups to be studied:** If the research is to make estimates on several subgroups of the population then the sample must be large enough for each of these subgroups to meet the desired quality level.

iv) **Cost:** cost considerations have a major implication. All studies have some budgetary constraint and hence cost dictates the size of the sample.

Overall:

For small populations (under 1000 a large sampling ratio (about 30%). Hence, a sample size of about 300 is required.

For moderately large population (10,000), a smaller sampling ratio (about 10%) is needed – a sample size around 1,000.
To sample from very large population (over 10 million), one can achieve accuracy using tiny sampling ratios (.025%) or samples of about 2,500.

The Common formula was Yamane Taro (1967).

$$\text{n} = \frac{\text{N}}{1 + N(e^2)} n$$

Where n-sample, N-population Size, e- level of precision.

$$\text{n} = \frac{pq(z)^2}{u^2}$$

Where n is the total sample size, P is the sample proportion, q is (1-P), u is the acceptable error term (let the error term be 0.05), (Z=1.96) is the standard normal variable in the accepted level of the error term, the level of confidence (α=0.05) will be used to check the level of significance.

Exercise. Calculate sample size by using the above formula for population size 11354 at 0.05 margin of error.

Activity 2

1. Discuss about sample.
2. Discuss the advantages and disadvantages of sampling.
3. Discuss the features of good research.
4. Discuss the difference between probability and non -probability sampling.
5. Discuss the main four things that sample size determination is based on

Chapter Three: Collection of Quantitative Data

✪ Unit Introduction

Dear learner, this chapter will introduce you to collection of quantitative data, data collection principles and scales of measurements. In any type of research, we are expected to collect data in appropriate ways. The quality of the collected data can affect the overall result. The data collection needs to follow some mandatory principle. A researcher also needs to now types of data from the angle of scale of measurements. This chapter has four parts: collection of quantitative data, data collection principles, scale of measurements and methods, tools and instruments of quantitative data collection.

Unit objectives:

Dear learner, by the end of this section you should be able to:

- ✓ Understand the concept of data collection in quantitative research;
- ✓ Discuss different data collection types;
- ✓ Explain principle of data collection;
- ✓ Identify different scales of measurements; and
- ✓ Discuss about methods, tools and instruments of quantitative data collection.

✹ Unit pre-test questions

6. What is meant by data?
7. What is data collection?
8. How quantitative data collection differs from the qualitative one
9. What data collection types?
10. What are data collection principles?
11. What are types of data based on their scale of measurement?
12. Are quantitative and qualitative researches the same?
13. What methods, tools and instruments are used in quantitative data collection?

3.1 Collection of quantitative data

✪ Section Overview

Data is a very important thing in research. Without data nothing will proceed with result. Thus, we need to collect our data in different ways. Data collection has also its own procedures. This chapter will introduce you with these things and other related issue.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss about data;
- ✓ Discuss about quantitative data collection; and
- ✓ Differentiate between types of data collection

? What do you of data?

Data refers to any group of facts, measurements, or observations used to make conclusion about the problem of investigation. It can range from material created in a laboratory, to information obtained in social-science research, such as a filled-out questionnaire, video and audio recordings, or photographs, etc. It can be of two types based on the type of research under taken; quantitative or qualitative. It can also be classified in different type based on different things. For example, data can be classified in to primary and secondary based on the source it is collected from.

? What is meant by quantitative data?

Quantitative data is defined as the value of data in the form of counts or numbers where each data-set has a unique numerical value associated with it. It is information about quantities; that is, information that can be measured and written down with numbers. Some examples of quantitative data are your height, your shoe size, and the length of your fingernails. Speaking of which, it might be time to call Guinness. Usually, there are measurement units associated with the data, e.g. meters, in the case of the height of a person. It makes sense to set boundary limits to such data, and it is also meaningful to apply arithmetic operations to the data. The difference between quantitativeand qualitative data is that, quantitative data can be counted, measured, and expressed using numbers while qualitative data is descriptive and conceptual.

? What do you know of data collection?

Data collection is defined as the "process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer queries, stated research questions, test hypotheses, and evaluate outcomes." It is the processes of collection of facts.

? What ere the two main form in which data can be collected?

Quantitative data is usually collected for statistical analysis using surveys, polls or questionnaires sent across to a specific section of a population. Quantitative data collection methods rely on random sampling and structured data collection instruments that fit diverse experiences into predetermined response categories. The core forms in which data can be collected are primary and secondary data. While the former is collected by a researcher through first-hand sources, the latter is collected by an individual other than the user. Sources of data can Primary and secondary data sources. Primary data sources include information collected and processed directly by the researcher, such as observations, surveys, interviews, and focus groups. Secondary data sources include information retrieved through preexisting sources: research articles, Internet or library searches, etc. Data can also be collected from internal or external sources. **Internal Source**- When data are collected from reports and records of the organization itself, it is known as the internal source. **External Source**- When data are collected from outside the organization, it is known as the external source. There are

also two sources of data in Statistics**: Statistical sources** refer to data that are collected for some official purposes and include censuses and officially conducted surveys. **Non-statistical** sources refer to the data that are collected for other administrative purposes or for the private sector.

Types of Data Collection

Data collection falls under two broad categories; Primary data collection and secondary data collection.

- **Primary Data Collection**

Primary data collection is the gathering of raw data collected at the source. It is a process of collecting the original data collected by a researcher for a specific research purpose. It could be further analyzed into two segments; qualitative research and quantitative data collection methods. Primary data can be collected in a number of ways. However, the most common techniques are self-administered surveys, interviews, field observation, and experiments. Primary data collection is quite expensive and time consuming compared to secondary data collection.

- **Secondary Data Collection**

Secondary data collection, on the other hand, is referred to as the gathering of second-hand data collected by an individual who is not the original user. It is the process of collecting data that is already existing, be it already published books, journals and/or online portals. In terms of ease, it is much less expensive and easier to collect. Your choice between Primary data collection and secondary data collection depend on the nature, scope and area of your research as well as its aims and objectives.

3.2 Principles of Data collection

- **Section Overview**

Data collection is often the most time consuming and expensive part of research. Before collecting any data, it is useful to stop and take stock of the situation, to make sure money and time is not wasted. This section teaches you about these principles of data collection to be more effective.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss different data collection principles; and
- ✓ Guided by the principle for data collection

? What principle of data collection do you know?

The basic principles of data collection include keeping things as simple as possible; planning the entire process of data selection, collection, analysis and use from the start; and ensuring that any data collected is valid, reliable and credible. It is also important that ethical issues are considered.

- **Keeping things simple**

It is important not to undertake any data collection or analysis methodology that is more complex or expensive than is necessary. The key, therefore, is to keep things as simple as possible.

- **Planning the whole process**

It is always important to know why information is needed before collecting it. A common mistake is to collect information before working out how it will be analysed or used. Sometimes, this means that the information cannot be properly analysed and used because it has not been collected in the right way, at the right time or in the right place. Some basic questions to ask before collecting any information are as follows.

- What information do you intend to gather?
- Where will you get this information, and how will it be collected?
- Why is the information needed, and what questions is the information going to answer?
- Who will use the information once collected?
- How will the information be analysed?
- How will any analyses be used?

If the answers to any of these questions are unknown or uncertain then it is important to find out the answers before going any further. Huge amounts of time, money and energy are wasted every year because information is collected that is never analysed or used. “If you collect information just because you think it might be useful at some stage in the future then there is a very good chance it will never be used. The golden rule is ‘if data is not being used then stop collecting it’. The time otherwise spent collecting the data can then be used for something more productive.

Another key principle is to collect information on a ‘need to know’ basis rather than a ‘like to know’ basis. This means being very clear about what information, is needed rather than collecting information on all kinds of issues that might be interesting.

- **Ensuring reliability, credibility and validity**

As far as possible, all information collected and used in research should be reliable, valid and credible.

♣ Data is considered **reliable** when there is confidence that similar results would be obtained if the data collection exercise was repeated within the same period, using the same methods. If

data is reliable it means it is not too heavily dependent on the skills and honesty of the person collecting it.

♣ Data is **valid** when it measures or describes what it set out to measure or describe. Data is not valid if it is misused. For example, information collected on attendance at a training session would be valid if used to show that the training session was held and people turned up. But information on attendance would not be valid if used to claim that participants had increased their awareness or understanding of an issue. Another common mistake is to get information from just one or two stakeholders and then to use this information as if it represents the views of a much wider population.

♣ Data is considered **credible** when it is believable, and is consistent with a 'common sense' view of the world. But just because data is not credible does not mean it is inaccurate. It simply means that it needs further checking. For example, if a small pilot project claimed to have data that showed it had greatly increased the living standards of farmers in a region the data may not be considered credible at first. But further data collection and analysis might confirm the findings and explain why such large changes had occurred. In that case the new data would be considered credible.

One method that is often used to improve the reliability, validity and credibility of information collected is triangulation. Triangulation means crosschecking information through using different methods of collection, talking to different stakeholders or using different people to collect data.

The ethics of data collection

Anyone engaged in formal research is expected to be familiar with the ethics of data collection and use. Some of these are described below.

♣ **Avoidance of harm** is a key principle whenever data is collected. People should not be put in a position where they might suffer because of the information they provide. For example, villagers supplying information about government services; women supplying information about domestic violence; children supplying information about bullying; or even staff providing information on leadership culture within an organization could all be considered potentially at harm. Measures should always be taken to mitigate the possibility of harm. If this is not possible then the data should not be collected.

♣ The **benefits and costs** to different person need to be considered. For example, there may seem little harm in getting together a group of farmers to engage in a focus group discussion about farming methods. But in some cases, this might mean taking them away from their fields at harvest time. Where possible, it is important to balance the costs and benefits of data collection.

♣ Participation in research activities should always be **voluntary**, and people should not be pressured into taking part. In fact any attempt to pressurise people into engaging with research

almost always backfires, because people are usually unwilling to tell the truth in situations where they feel forced to participate.

♣ **Confidentiality** needs to be respected. Some people may be willing to express opinions provided they are not quoted, or the information is not used widely. If this is the case then this needs to be clearly recorded alongside any notes taken. The information should not then be disseminated or used without the consent of the person who supplied the information. However, it is normally acceptable to use the information to shape judgements or come to conclusions.

♣ Likewise, **anonymity** needs to be respected. Some people or organisations are willing for their opinions or stories to be used provided they are not personally named. In these cases, it is fine to record and disseminate the information, but the person or organisation supplying the information needs to remain anonymous. This might mean taking active steps to make sure that others cannot find out who the person or organisation supplying the information was.

♣ If researcher wants to use a story for publicity purposes, or wants to publish a photograph of informants then **informed consent** should always be sought. This means recording whether participants are happy to be quoted; whether they are happy for their real names to be used; and whether they are happy for their photographs to be used, and, if so, how.

♣ **Cultural sensitivities** should always be respected. This means considering differences in culture, local behaviour and norms, religious beliefs and practices, sexual orientation, gender roles, disability, age, ethnicity and other social differences when planning data collection activities or communicating findings.

3.3 Scales of measurement

✪ Section Overview

Dear learner, in the previous sections of this chapter you have learnt quantitative data collection and data collection principle. In this section you will learn about scale of measurement. Scales of measurement refer to ways in which variables/numbers are defined and categorized. Each scale of measurement has certain properties which in turn determines the appropriateness for use of certain statistical analyses.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss about scale of measurements; and
- ✓ Differentiate among nominal, ordinal, interval and scale of measurement.

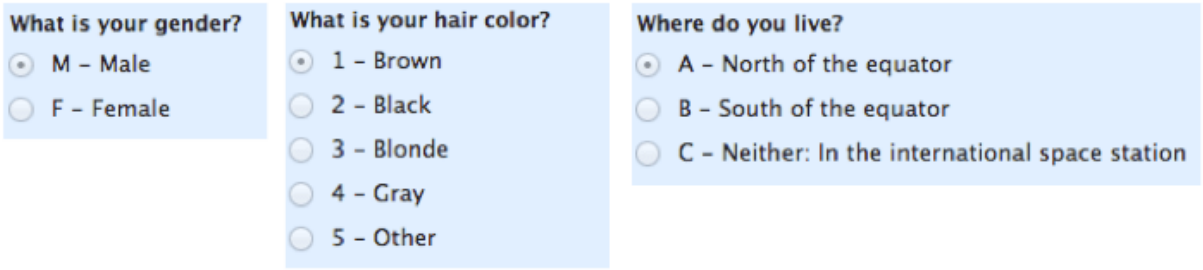
? What do you think of measurement and scale of measurement?

Measurement is the process of assigning numbers to objects or observations, the level of measurement being a function of the rules under which the numbers are assigned.

Scales of measurement refer to ways in which variables/numbers are defined and categorized. Each scale of measurement has certain properties which in turn determines the appropriateness for use of certain statistical analyses. The four scales of measurement are nominal, ordinal, interval, and ratio.

A. Nominal level

The nominal type differentiates between items or subjects based only on their names or (meta)categories and other qualitative classifications they belong to. They are used for labeling variables, without any quantitative value. "Nominal" scales could simply be called "labels." Numbers may be used to represent the variables but the numbers do not have numerical value or relationship. Examples: gender, nationality, ethnicity, language, genre, style, biological species, and form.



Note: a sub-type of nominal scale with only two categories (e.g. male/female) is called "dichotomous." Other sub-types of nominal data are "nominal with order" (like "cold, warm, hot, very hot") and nominal without order (like "male/female").

Characteristics of Nominal Scale:

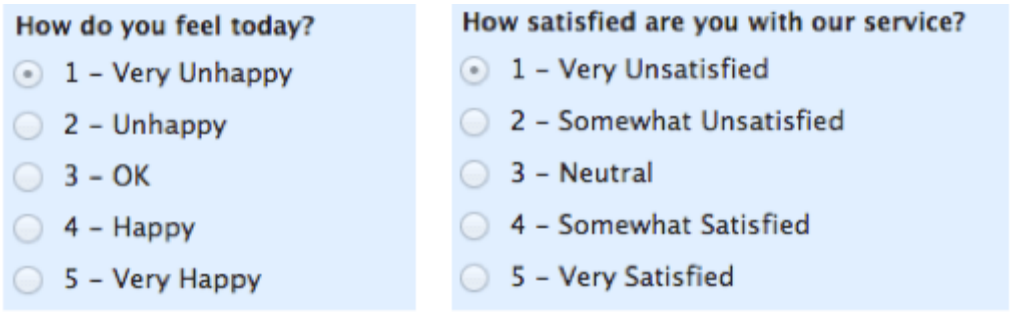
- A nominal scale variable is classified into two or more categories. In this measurement mechanism, the answer should fall into either of the classes.
- It is qualitative. The numbers are used here to identify the objects.
- The numbers don't define the object characteristics. The only permissible aspect of numbers in the nominal scale is "counting."
- From central tendencies, only mode can be calculated for nominal scale data.

C. Ordinal scale

The ordinal type allows for rank order (1st, 2nd, 3rd, etc.) by which data can be sorted, but still does not allow for relative degree of difference between them. Examples: data consisting of a spectrum of values, such as 'completely agree', 'mostly agree', 'mostly disagree', 'completely disagree' when measuring opinion.

The ordinal scale places events in order, but there is no attempt to make the intervals of the scale equal in terms of some rule. Rank orders represent ordinal scales and are frequently used

in research relating to qualitative phenomena. A student's rank in his graduation class involves the use of an ordinal scale. One has to be very careful in making statement about scores based on ordinal scales. For instance, if Bety's position in his class is 10 and Hawi's position is 40, it cannot be said that Bety's position is four times as good as that of Hawi. Ordinal scales only permit the ranking of items from highest to lowest. Ordinal measures have no absolute values, and the real differences between adjacent ranks may not be equal. All that can be said is that one person is higher or lower on the scale than another, but more precise comparisons cannot be made. Thus, the use of an ordinal scale implies a statement of 'greater than' or 'less than' (an equality statement is also acceptable) without our being able to state how much greater or less. The real difference between ranks 1 and 2 may be more or less than the difference between ranks 5 and 6. Since the numbers of this scale have only a rank meaning, the appropriate measure of central tendency is the median. With ordinal scales, the order of the values is what's important and significant, but the differences between each one is not really known. Take a look at the example below. In each case, we know that a #4.is better than a #3 or #2, but we don't know–and cannot quantify–how much better it is. Ordinal scales are typically measures of non-numeric concepts like satisfaction, happiness, discomfort, etc.



Characteristics of the Ordinal Scale:

- The ordinal scale shows the relative ranking of the variables
- It identifies and describes the magnitude of a variable
- Along with the information provided by the nominal scale, ordinal scales give the rankings of those variables
- The interval properties are not known
- The surveyors can quickly analyze the degree of agreement concerning the identified order of variables
- From the central tendencies only mode and median can be calculated for ordinal scale data.

Examples:

- Ranking of school students – 1st, 2nd, 3rd, etc

- Evaluating the frequency of occurrences
    - Very often
    - Often
    - Not often
    - Not at all
- Assessing the degree of agreement
    - Totally agree
    - Agree
    - Neutral
    - Disagree
    - Totally disagree

D. Interval scale

The interval type allows for the degree of difference between items, but not the ratio between them. Interval scales are numeric scales in which we know both the order and the exact differences between the values. The classic example of an interval scale is Celsius temperature because the difference between each value is the same. For example, the difference between 60 and 50 degrees is a measurable 10 degrees, as is the difference between 80 and 70 degrees. Interval scales are nice because the realm of statistical analysis on these data sets opens up. For example, central tendency can be measured by mode, median, or mean; standard deviation can also be calculated. Interval scales not only tell us about order, but also about the value between each item.

Here's the problem with interval scales: they don't have a "true zero." For example, there is no such thing as "no temperature," at least not with celsius. In the case of interval scales, zero doesn't mean the absence of value, but is actually another number used on the scale, like 0 degrees Celsius. Negative numbers also have meaning. Without a true zero, it is impossible to compute ratios. With interval data, we can add and subtract, but cannot multiply or divide.

consider this: 10 degrees C + 10 degrees C = 20 degrees C. No problem there. 20 degrees C is not twice as hot as 10 degrees C, however, because there is no such thing as "no temperature" when it comes to the Celsius scale. When converted to Fahrenheit, it's clear: 10C=50F and 20C=68F, which is clearly not twice as hot. I hope that makes sense. Bottom line, interval scales are great, but we cannot calculate ratios, which brings us to our last measurement scale.

Characteristics of Interval Scale:

- The interval scale is quantitative as it can quantify the difference between the values
- It allows calculating the mean and median of the variables
- To understand the difference between the variables, you can subtract the values between the variables

- The interval scale is the preferred scale in Statistics as it helps to assign any numerical values to arbitrary assessment such as feelings, calendar types, etc.
- From central tendencies mode, median and mean can be calculated for interval scale data.

Examples:

- Age group: 0-6, 7-14, 15-22, 23-30, 31-40
- Students mark: Above 94, 85-94, 75-84, 65-74, 55-64, 45-54, Below 45

E. Ratio scale

Ratio scales tell us about the order, they tell us the exact value between units, and they also have an absolute zero–which allows for a wide range of both descriptive and inferential statistics to be applied. Everything above about interval data applies to ratio scales, plus ratio scales have a clear definition of zero. Good examples of ratio variables include height, weight, and duration. Ratio scales provide a wealth of possibilities when it comes to statistical analysis. These variables can be meaningfully added, subtracted, multiplied, divided (ratios). Central tendency can be measured by mode, median, or mean; measures of dispersion, such as standard deviation and coefficient of variation can also be calculated from ratio scales.

In summary, nominal variables are used to "name," or label a series of values. Ordinal scales provide good information about the order of choices, such as in a customer satisfaction survey. Interval scales give us the order of values + the ability to quantify the difference between each one. Finally, Ratio scales give us the ultimate–order, interval values, plus the ability to calculate ratios since a "true zero" can be defined.

Characteristics of Ratio Scale:

- Ratio scale has a feature of absolute zero
- It doesn't have negative numbers, because of its zero-point feature
- It affords unique opportunities for statistical analysis. The variables can be orderly added, subtracted, multiplied, divided. Mean, median, and mode can be calculated using the ratio scale.
- Ratio scale has unique and useful properties. One such feature is that it allows unit conversions like kilogram – calories, gram – calories, etc.

Examples:

What is your weight in Kgs?

- Less than 55 kgs
- 55 – 75 kgs
- More than 95 kgs

In summary, nominal variables are used to "name," or label a series of values. Ordinal scales provide good information about the order of choices, such as in a customer satisfaction survey. Interval scales give us the order of values + the ability to quantify the difference between each one. Finally, Ratio scales give us the ultimate–order, interval values, plus the ability to calculate ratios since a "true zero" can be defined. More to simply express see the following table

Level of Measurement	**Features**
Nominal	Named variable
Ordinal	Named + ordered variable
Interval	Named+ ordered+ proportionate interval between variables
Ratio	Named+ ordered+ proportionate interval between variables + can accommodate absolute zore

Exercise:

1. Which measurement scale does the following describe?

Name of the scale	**????**	**????**	**????**	**????**
Nature of assignment	Simply assigned	Events in order	Precise	Some order
Quantitative value	No	Intervals not equal and no absolute value	Yes	Yes
Comparison	No	No	Yes	Yes, but not exactly

2. Which level of measurements are the following statements?

A. 1. sex of household head -1 female, 2 male –
B. Irrigation – 1- yes 0- no –
C. Climatic condition – drought -1-yes 0 no –
D. Use of improved seed- yes no
E. Distance to market –
F. Education – 1- primary 2-upp-

Sources of Error in Measurement

Measurement should be precise and unambiguous in an ideal research study. To do so, the researcher must be aware about the sources of error in measurement. The following are the possible sources of error in measurement.

(a) Respondent: -may be reluctant to express strong negative feelings, may have very little knowledge but may not admit his ignorance, fatigue, boredom, anxiety, etc.

(b) Situation: Situational factors may also come in the way of correct measurement.
Respondent may distort responses if someone is present
If the respondent feels that secrecy is not assured, he/she may be reluctant to express certain feelings.
(c) Measurer: The interviewer can distort responses by rewording or reordering questions. Her/his behaviour, style and looks may encourage or discourage certain replies from respondents.

3.4 Methods, tools and instruments of collecting quantitative data

✪ Section Overview

Dear learner, in the previous sections of this chapter you have learnt quantitative data collection, data collection principle and scale of measurements which are seen to be very important in collection of data in a research. In this section you will learn about methods, tools and instruments of quantitative data collection. Different data collection methods are used in quantitative data collection. They have their own advantage and disadvantage. Let we see them below.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss distinguish among methods, tools and instruments of data collection;
- ✓ Identify quantitative data collection methods;
- ✓ Discuss advantages and disadvantages of different quantitative data collection methods; and
- ✓ Determine to use which data collection method in a given circumstance by calculating the advantage and disadvantages of different methods.

Quantitative data collection methods rely on random sampling and structured data collection instruments that fit diverse experiences into predetermined response categories. They produce results that are easy to summarize, compare, and generalize.

Quantitative research is concerned with testing hypotheses derived from theory and/or being able to estimate the size of a phenomenon of interest. Depending on the research question, participants may be randomly assigned to different treatments. If this is not feasible, the researcher may collect data on participant and situational characteristics in order to statistically control for their influence on the dependent, or outcome, variable. If the intent is to generalize from the research participants to a larger population, the researcher will employ probability sampling to select participants.

Typical quantitative data gathering strategies include:

- Administering surveys with closed-ended questions (e.g., face-to face and telephone interviews, mail questionnaires, etc.)
- Experiments/clinical trials.
- Observing and recording well-defined events (e.g., counting the number of patients waiting in emergency at specified times of the day).
- Obtaining relevant data from management information systems.

Interviews

An interview is a face-to-face conversation between two individuals with the sole purpose of collecting relevant information to satisfy a research purpose. In Quantitative research (survey research), interviews are more structured than in Qualitative research. In a structured interview, the researcher asks a standard set of questions and nothing more.

➢ Telephone interviews

Advantages:

 Less time consuming
 Less expensive
 Researcher has ready access to anyone who has a landline telephone.
 Higher response rate than the mail questionnaire.
 Can be fully automated using CATI (Computer Assisted Telephone Interviewing) saving data processing time.

Disadvantages:

 The response rate is not as high as the face-to-face interview.
 The sample may be biased as only those people who have landline phones are contacted (excludes people who do not have a phone, or only have cell phones).

➢ Face-to-face interviews

Advantages:

- Enables the researcher to establish rapport with potential participants and therefore gain their cooperation.
- Yields the highest response rates in survey research.
- Allows the researcher to clarify ambiguous answers and when appropriate, seek follow- up information.

Disadvantages:

- Impractical when large samples are involved
- Can be time consuming and expensive.

- **Computer Assisted Personal Interviewing (CAPI):** Is a form of personal interviewing, but instead of completing a questionnaire, the interviewer brings along a laptop or hand-held computer to enter the information directly into the database?

Advantages:

- Saves time involved in processing the data.
- Saves the interviewer from carrying around hundreds of questionnaires.

Disadvantages:

- Can be expensive to set up.
- Requires that interviewers have computer and typing skills.

Questionnaire: This is the process of collecting data through an instrument consisting of a series of questions and prompts to receive a response from individuals it is administered to. Questionnaires are designed to collect data from a group. For clarity, it is important to note that a questionnaire isn't a survey, rather it forms a part of it. A survey is a process of data gathering involving a variety of data collection methods, including a questionnaire. On a questionnaire, there are three kinds of questions used. They are; fixed-alternative, scale, and open-ended. With each of the questions tailored to the nature and scope of the research.

Questionnaires often make use of checklist and rating scales. These devices help simplify and quantify people's behaviors and attitudes. A checklist is a list of behaviors, characteristics, or other entities the researcher is looking for. Either the researcher or survey participant simply checks whether each item on the list is observed, present or true or vice versa. A rating scale is more useful when a behavior needs to be evaluated on a continuum. They are also known as Likert scales.

Mail questionnaires:

Advantages:

- Can be sent to a large number of people.
- Saves the researcher time and money compared to interviewing.
- People are more truthful while responding to the questionnaires regarding controversial issues in particular due to the fact that their responses are anonymous.
- Allow the respondent to answer at their leisure.

Disadvantages:

- In most cases, the majority of people who receive questionnaires don't return them. Therefore: Over-sampling may be necessary if doing a one-time mail out in order to get enough completed questionnaires to be generalizable to the population. o Follow-up reminders to participants encouraging them to complete the questionnaire may be necessary, thereby increasing the time and cost to conduct the study. o May need to offer incentives to increase response rate.

• Time – mail surveys take longer than other types of surveys. Web-based questionnaires: A new and inevitably growing methodology is the use of Internet based research. This would mean receiving an e-mail on which you would click on an address that would take you to a secure web-site to fill in a questionnaire.

Advantages:

• This type of research is often quicker and less detailed.

• Very cost effective.

Disadvantages:

• Excludes people who do not have a computer or are unable to access a computer.

• Need to have access to email addresses.

• Many worksites have screening mechanisms in place blocking access to employee emails.

• The validity of such surveys may be in question as people might be in a hurry to complete it and so might not give accurate responses.

Merits	**Demerits**
Can be administered in large numbers and is cost-effective	Answers may be dishonest or the respondents lose interest midway
It can be used to compare and contrast previous research to measure change.	Questionnaires can'tproduce qualitative data
Easy to visualize and analyze.	Questions might be left unanswered.
Questionnaires offer actionable data.	Respondents may have a hidden agenda.
Respondent identity is protected.	Not all questions can be analyzed easily.
Questionnaires can cover all areas of a topic.	Relatively inexpensive.

Activity 2

1. Discuss quantitative data collection.
2. Discuss about data collection principles.
3. Discuss about the types of scale of measurement.
4. Discuss methods of quantitative data collection.
5. Discuss the advantage of using interview vs questionnaires in quantitative data collection.

Chapter Four: Quantitative Data Analysis

Unit Introduction

Dear learner, in the last three chapters of this course, hopefully you have got good things about quantitative research, sampling and quantitative data collection in detail. In this chapter you will learn about, how to analyse the quantitative data collected and how to present it. Data collection is not an end by itself in a research. After collecting quality data, a research needs to analysis in appropriate way and present it in appropriate way too. This chapter has three parts: presentation of quantitative data, descriptive statistical methods in quantitative research, and inferential statistical method in quantitative research.

Unit objectives:

Dear learner, by the end of this section you should be able to:

- ✓ Understand the concept of quantitative data analysis;
- ✓ Present quantitative data in different way;
- ✓ Explain why descriptive statistical method is important;
- ✓ Explain why inferential statics is important;
- ✓ Diffentaite between descriptive and inferential statistics; and
- ✓ Determine when to use descriptive and inferential statistical methods in research.

Unit pre-test questions

1. What is data analysis?
2. How can quantitative data be analysed?
3. Why is descriptive statistical method important in quantitative data analysis?
4. What is inferential statistics and its importance?
5. When should a researcher use descriptive or inferential statistics in analysing quantitative data?

? What do you know about data analysis?

Data analysis is defined as a process of cleaning, transforming, and modeling data to discover useful information for business decision-making. The purpose of Data Analysis is to extract useful information from data and taking the decision based upon the data analysis. It is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. An essential component of ensuring data integrity is the accurate and appropriate analysis of research findings.

Quantitative data analysis may include the calculation of frequencies of variables and differences between variables. A quantitative approach is usually associated with finding evidence to either support or reject hypotheses you have formulated at the earlier stages of your research process.

4.1 Presentation of quantitative data

✪ Section Overview

Dear learner, in the previous sections of this chapter you have learnt about quantitative data analysis. The data analysed need to be presented in an appropriate way for the reader. The way our data is presented can encourage or discourage the reader. Thus, we need to know what type of data presentation we need to use in different condition and for different types of data. This section will help you to learn about quantitative data presentation.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Determine ways of data presentation; and
- ✓ Choose appropriate data presentation method for a given type of data

? How can a researcher present a quantitative data?

It is likely that there will be occasions when you have numerical information that you want to include in your work, for example figures and other statistics from secondary sources (such as books, journal articles or newspaper reports); the results of experiments; or data that you have collected and analysed as part of a project or dissertation. Such information can be used to illustrate an argument or convey complex or detailed information in a concise manner.

There are three main methods of presenting such information:

1. It can be incorporated into the **main body of text** - Numbers are most effective in the main body of the text of an essay, report or dissertation when there are only two values to compare.
   For example:86% of male students said they regularly ate breakfast compared to 62% of female students.
2. It can be presented separately as a **table** -Tables are used to present numerical data in a wide variety of publications from newspapers, journals and textbooks to the sides of grocery packets. They are the format in which most numerical data are initially stored and analysed and are likely to be the means you use to organise data collected during experiments and dissertation research.
   When to use tables
   Tables are an effective way of presenting data:
   - when you wish to show how a single category of information varies when measured at different points (in time or space). For example, a table would be an appropriate way of showing how the category unemployment rate varies between different countries in the EU (different points in space);

- when the dataset contains relatively few numbers. This is because it is very hard for a reader to assimilate and interpret many numbers in a table . In particular, avoid the use of complex tables in talks and presentations when the audience will have a relatively short time to take in the information and little or no opportunity to review it at a later stage;
- when the precise value is crucial to your argument and a graph would not convey the same level of precision. For example, when it is important that the reader knows that the result was 2.48 and not 2.45;
- when you don't wish the presence of one or two very high or low numbers to detract from the message contained in the rest of the dataset. For example if you are presenting information about the annual profits of an organisation and don't want the underlying variability from one year to the next to be swamped by a large loss in a particular year.

3. It can be used to construct a **graph or chart -** Graphs are a good means of describing, exploring or summarising numerical data because the use of a visual image can simplify complex information and help to highlight patterns and trends in the data. They are a particularly effective way of presenting a large amount of data but can also be used instead of a table to present smaller datasets.

Determining which of these methods is the most appropriate depends upon the amount of data you are dealing with and their complexity. The choice about whether to use text, tables or graphs requires careful consideration if you are to ensure that your reader or audience understands your argument and is not left struggling to interpret data that are poorly presented or in an inappropriate format. It is crucial to remember that when using a table or graph the associated text should describe what the data reveal about the topic; you should not need to describe the information again in words.

4.1.1 Frequency distribution

The frequency (f) of a particular observation is the number of times the observation occurs in the data. The distribution of a variable is the pattern of frequencies of the observation. Frequency distributions are portrayed as frequency tables, histograms, frequency distributions can show either the actual number of observations falling in each raw or the percentage of observations. If the frequency distribution shows the percentage of observation, the distribution is called a relative frequency distribution. Frequency distribution tables can be used for both categorical and numeric variables. A frequency distribution is an overview of all distinct values in some variable and the number of times they occur. That is, a frequency distribution tells how frequencies are distributed over values. Frequency distributions are mostly used for summarizing categorical variables. The definition of frequency is how often something happens. An example of frequency is a person blinking their eyes 47 times in one minute.

Example 1 – Constructing a frequency distribution table

A survey was taken on for 2o homes. In each of 20 homes, people were asked how many cars were registered to their households. The results were recorded as follows:

1, 2, 1, 0, 3, 4, 0, 1, 1, 1, 2, 2, 3, 2, 3, 2, 1, 4, 0, 0

Use the following steps to present this data in a frequency distribution table.

Divide the results (x) into intervals, and then count the number of results in each interval. In this case, the intervals would be the number of households with no car (0), one car (1), two cars (2) and so forth.
Make a table with separate columns for the interval numbers (the number of cars per household), the tallied results, and the frequency of results in each interval. Label these columns Number of cars, Tally and Frequency.

Read the list of data from left to right and place a tally mark in the appropriate row. For example, the first result is a 1, so place a tally mark in the row beside where 1 appears in the interval column (Number of cars). The next result is a 2, so place a tally mark in the row beside the 2, and so on. When you reach your fifth tally mark, draw a tally line through the preceding four marks to make your final frequency calculations easier to read. Add up the number of tally marks in each row and record them in the final column entitled Frequency. Your frequency distribution table for this exercise should look like this:

Table1. Frequency table for the number of cars registered in each household

Number of cars (x)	Tally	Frequency(f)	% (relative frequency)
0	𝍸	4	20%
1	𝍸 I	6	30%
2	𝍸	5	25%
3	III	3	15%
4	II	2	10%

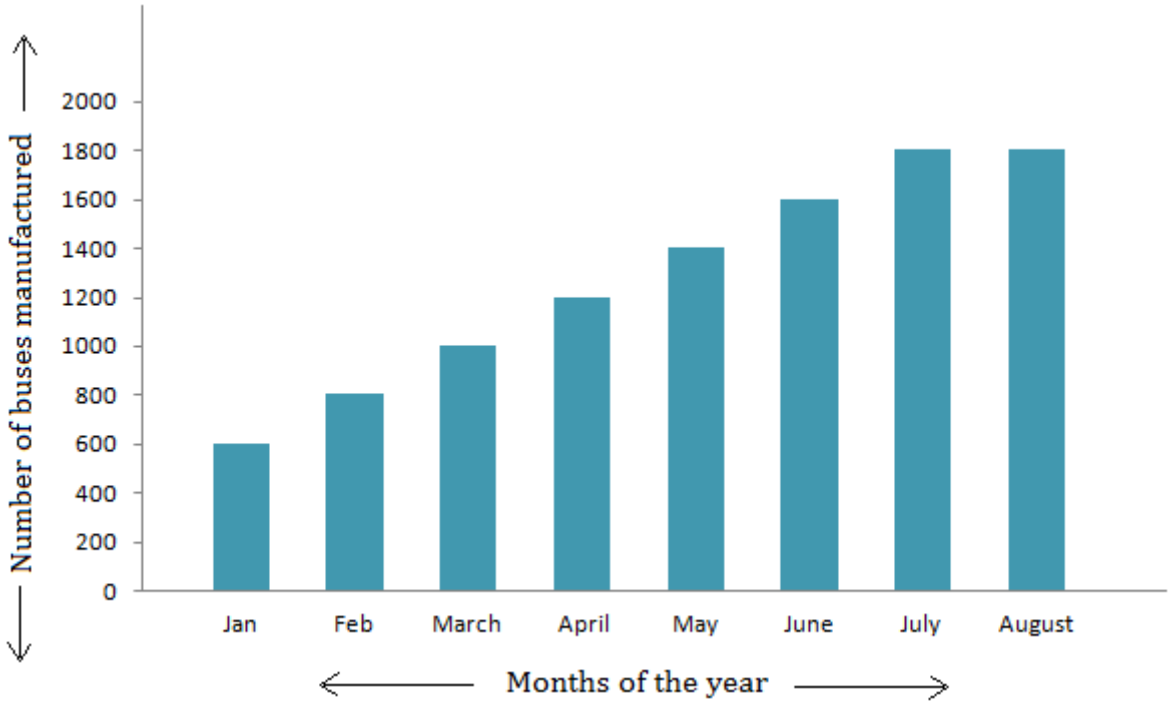
By looking at this frequency distribution table quickly, we can see that out of 20 households surveyed, 4 households had no cars, 6 households had 1 car, etc

Exercise: create a frequency table for the following distribution and calculate the relative frequency.
2, 11, 23, 5, 17, 17, 23, 13, 3, 13, 11, 15, 23. 5, 23

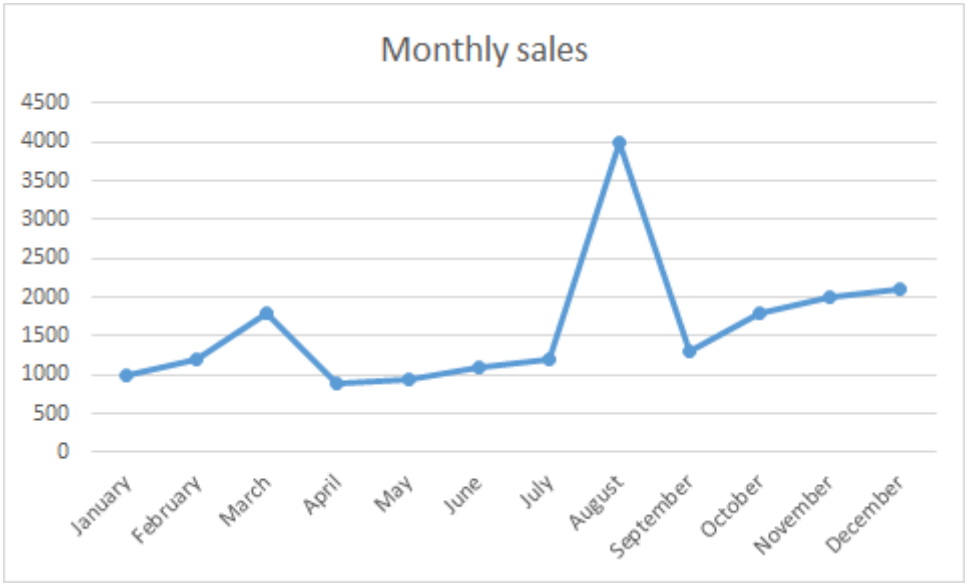
4.1.2. Diagrams and graphs

Data can be represented in many ways. The 4 main types of graphs are a bar graph or bar chart, line graph, pie chart, and diagram.

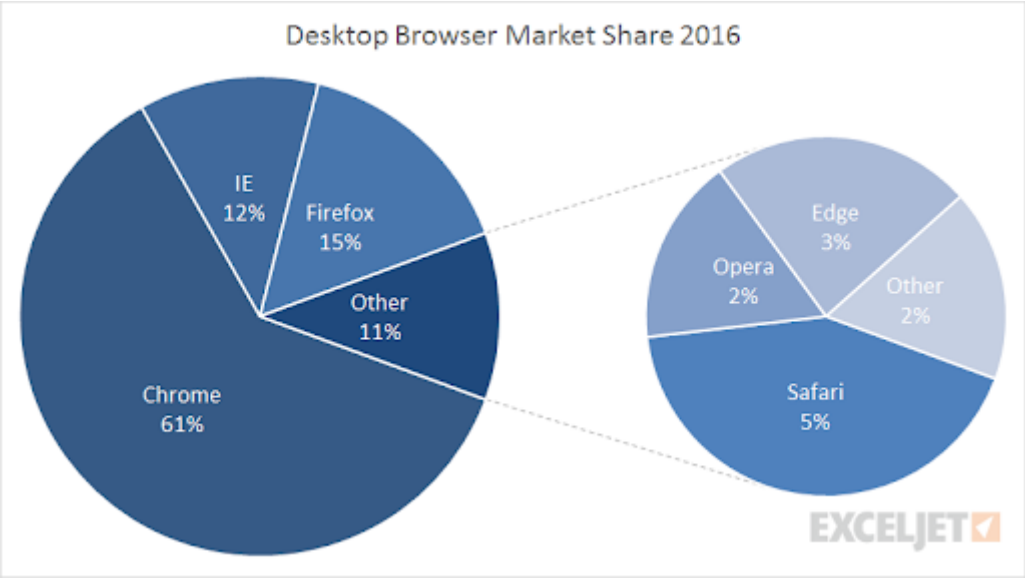
Bar graphs are used to show relationships between different data series that are independent of each other. In this case, the height or length of the bar indicates the measured value or frequency. Below, you can see the example of a bar graph which is the most widespread visual for presenting statistical data.



Line graphs represent how data has changed over time. This type of charts is especially useful when you want to demonstrate trends or numbers that are connected. For example, how sales vary within one year. In this case, financial vocabulary will come in handy. Besides, line graphs can show dependencies between two objects during a particular period.



Pie charts are designed to visualize how a whole is divided into various parts. Each segment of the pie is a particular category within the total data set. In this way, it represents a percentage distribution.



4.2 Descriptive statistical methods in quantitative research

✪ Section Overview

Dear learner, in the previous sections of this chapter you have learnt about quantitative data analysis and quantitative data presentation. Hopefully it has added something to your knowledge about data analysis and presentation in quantitative data. In this section we will look at how to analysis quantitative data in descriptive statistics, types of descriptive statistics and their benefits.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss about descriptive statistics and its importance;
- ✓ Differentiate among different types of descriptive statics; and
- ✓ Discuss the advantages of types of descriptive statistics.

? What do you think of descriptive statistics?

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. It is the ways of summarizing the scores of samples is called descriptive statistics. Descriptive statistics refer to the description of the sample. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data. The type of statistical methods used for this purpose is called descriptive statistics. They include both numerical (e.g. mean, mode, variance) and graphical tools (e.g. histogram, boxplot) which allow to summarize a set of data and extract important information such as central tendencies and dispersion.

Descriptive statistics are used to describe or summarize data in ways that are meaningful and useful. For example, it would not be useful to know that all of the participants in our example wore blue shoes. Central tendency describes the central point in a data set. Variability describes the spread of the data.

Descriptive statistics involves summarizing and organizing the data so they can be easily understood. Descriptive statistics, unlike inferential statistics, seeks to describe the data, but do not attempt to make inferences from the sample to the whole population. Here, we typically describe the data in a sample.

Descriptive statistics are brief descriptive coefficients that summarize a given data set, which can be either a representation of the entire or a sample of a population. Descriptive statistics are broken down into measures of central tendency and measures of variability (spread).

4.2.1 Measures of central tendency

? What is measure of central tendency?

A measure of central tendency is a summary statistic that represents the center point or typical value of a dataset. The goal of measures of central tendency is to come up with one single number that best describes a distribution of scores. In statistics, the three most common measures of central tendency are the mean, median, and mode. Each of these measures calculates the location of the central point using a different method. Each of these measures describes a different indication of the typical or central value in the distribution.

- **Mode**-The mode is the most commonly occurring value in a distribution, the most frequent number—that is, the number that occurs the highest number of times.
  Example: The mode of {4, 2, 4, 3, 2, 2} is 2 because it occurs three times, which is more than any other number.
  A set of numbers with two modes is bimodal, a set of numbers with three modes is trimodal, and a set of numbers with four or more nodes is multimodal.
  Finding the Mode
  To find the mode, or modal value, it is best to put the numbers in order. Then count how many of each number. A number that appears most often is the mode.
  Example:
  3, 7, 5, 13, 20, 23, 39, 23, 40, 23, 14, 12, 56, 23, 29
  In order these numbers are:
  3, 5, 7, 12, 13, 14, 20, 23, 23, 23, 23, 29, 39, 40, 56
  This makes it easy to see which numbers appear most often.
  In this case the mode is 23.

  Another Example: {19, 8, 29, 35, 19, 28, 15}
  Arrange them in order: {8, 15, 19, 19, 28, 29, 35}

19 appears twice, all the rest appear only once, so 19 is the mode.

More Than One Mode
We can have more than one mode.
Example: {1, 3, 3, 3, 4, 4, 6, 6, 6, 9}
3 appears three times, as does 6.
So there are two modes: at 3 and 6
Having two modes is called "bimodal".
Having more than two modes is called "multimodal".

N.B there may be a case in which mode may not appear. Example {19, 8, 29, 35, 43, 28, 15} – No mode for there is no number occurred beyond one time.

Grouping

In some cases (such as when all values appear the same number of times) the mode is not useful. But we can group the values to see if one group has more than the others.
Example: {4, 7, 11, 16, 20, 22, 25, 26, 33}
Each value occurs once, so let us try to group them.
We can try groups of 10:
0-9: 2 values (4 and 7)
10-19: 2 values (11 and 16)
20-29: 4 values (20, 22, 25 and 26)
30-39: 1 value (33)
In groups of 10, the "20s" appear most often, so we could choose 25 (the middle of the 20s group) as the mode.
You could use different groupings and get a different answer.

Grouping also helps to find what the typical values are when the real world messes things up.

Advantages of Mode

- The mode is the only average that can be used if the data set is not in numbers, for instance the colours of cars in a car park.
- It is easy to understand and simple to calculate.
- It is not affected by extremely large or small values.
- It can be located just by inspection in ungrouped data and discrete frequency distribution.
- It can be useful for qualitative data.
- It can be computed in an open-end frequency table.
- It can be located graphically.

Disadvantage of Mode

- The mode may easily be unrepresentative of the bulk of the data, and so produce a misleading picture. Imagine if we had the following set of scores: 3, 4, 4, 5, 6, 7, 8, 8, 96, 96, 96. Here the mode is 96 - but most of the scores are low numbers, and so 96 is unrepresentative of them.
- There may be more than one mode in a set of scores. For example, in the set of scores 3,3,3,4,4,4,6,6,6, there are three modes!

- There can be more than one mode, and there can also be no mode which means the mode is not always representative of the data.

❖ **Median**

The Median is the "middle" of a sorted list of numbers, middle number in a list of numbers.
To find the Median, place the numbers in value order and find the middle.
Example: find the Median of 12, 3 and 5
Put them in order:
3, 5, 12
The middle is 5, so the median is 5.
Example:
3, 13, 7, 5, 21, 23, 39, 23, 40, 23, 14, 12, 56, 23, 29
When we put those numbers in order we have:
3, 5, 7, 12, 13, 14, 21, 23, 23, 23, 23, 29, 39, 40, 56
There are fifteen numbers. Our middle is the eighth number:
3, 5, 7, 12, 13, 14, 21, 23, 23, 23, 23, 29, 39, 40, 56
The median value of this set of numbers is 23.
(It doesn't matter that some numbers are the same in the list.)
Two Numbers in the Middle
But, with an even amount of numbers things are slightly different.
In that case we find the middle pair of numbers, and then find the value that is half way between them. This is easily done by adding them together and dividing by two.
Example:
3, 13, 7, 5, 21, 23, 23, 40, 23, 14, 12, 56, 23, 29
When we put those numbers in order we have:
3, 5, 7, 12, 13, 14, 21, 23, 23, 23, 23, 29, 40, 56
There are now fourteen numbers and so we don't have just one middle number, we have a pair of middle numbers:
3, 5, 7, 12, 13, 14, 21, 23, 23, 23, 23, 29, 40, 56
In this example the middle numbers are 21 and 23.
To find the value halfway between them, add them together and divide by 2:
21 + 23 = 44
then 44 ÷ 2 = 22
So the Median in this example is 22.
(Note that 22 was not in the list of numbers ... but that is OK because half the numbers in the list are less, and half the numbers are greater.)

A quick means to find middle is: count how many numbers, add 1 then divide by 2.
Example1: there are 45 number, to find middle, 45+ 1/2= 23. Thus, the median of this is the 23 number.
Example2:66 plus 1 is 67, then divide by 2 and we get 33.5

33 and a half? That means that the 33rd and 34th numbers in the sorted list are the two middle numbers. So to find the median: add the 33rd and 34th numbers together and divide by 2.

- In a simple way, if the data set has an
  - Odd number of entries: median is the middle data entry. Example: a median for 1,2,3,4,5, the median is 3
  - Even number of entries: median is the mean of the two middle data entries. Example: a median for 1,2,3,4,5,6, is 3.5 for we add the two middle numbers 3 and 4 and dived it by 2.

The median is usually preferred in these situations because the value of the mean can be distorted by the outliers. However, it will depend on how influential the outliers are. If they do not significantly distort the mean, using the mean as the measure of central tendency will usually be preferred.

Advantages of Median

Advantage of the median:

- The median is less affected by outliers and skewed data than the mean
- It is resistant to the distorting effects of extreme high or low scores
- Can be used with ordinal interval and ratio scale data

Limitation of the median:

- The median cannot be identified for nominal data, as it cannot be logically order.
- It is more vulnerable to sampling fluctuations than the mean
- It is less mathematically useful than the mean.
- If there is an even number of numbers, the median is found by averaging the two middle numbers. This means the median value may not actually be a number in the original data set.

❖ **Mean**

The Arithmetic Mean is the average of the numbers: a calculated "central" value of a set of numbers. Add up all the numbers, then divide by how many numbers there are. Applicable for interval and ratio data, not applicable for nominal or ordinal data.

Population Mean: $\mu = \frac{\sum x}{N}$

Sample Mean: $\overline{x} = \frac{\Sigma x}{n}$

Example: calculate the mean for the following distribution

a. 24, 13, 19, 26, 1
b. 33, 3, 29, 5, 11, 7
c.

Advantages of Mean

- The mean takes account of all values to calculate the average.
- The mean can be used for both continuous and discrete numeric data.

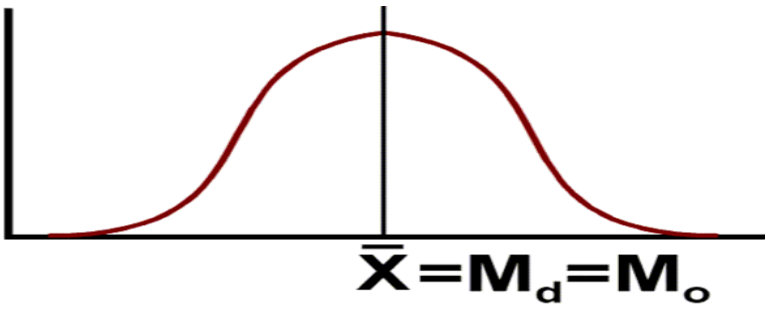
- It is the only measure of central tendency that uses the information from every single score.
- It has certain mathematical advantages; it is very common in statisticalformula, in one form or another.
- It is the measure which is most resistant to sampling fluctuation.

Limitations of the mean:

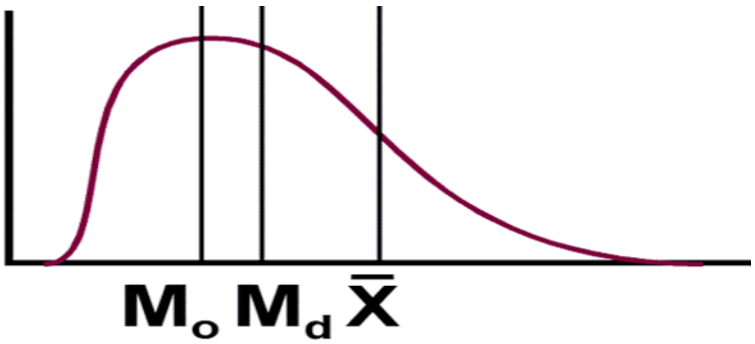
- The mean cannot be calculated for categorical data, as the values cannot be summed.
- As the mean includes every value in the distribution the mean is influenced by outliers and skewed distributions (e.g., "mean income" in the U.K. is a highly misleading statistic, because the few millionaires that contribute to this mean have a disproportionate effect, biasing the mean upwards from what it would otherwise be).Just one or two high or low scores can seriously distort the mean.
- The mean can only be used with interval or ratio data; it cannot be used with ordinal or nominal data.

Tips on measures of central tendencies

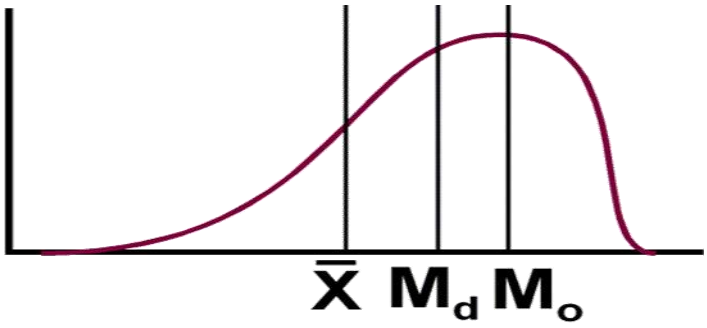
- With perfectly bel shaped distributions, the mean, median, and mode are identical.



- With positively skewed data, the mode is lowest, followed by the median and mean.



- With negatively skewed data, the mean is lowest, followed by the median and mode.



4.2.2. Measures of variability/despersion

? What do you think of measure of dispersion?

The dispersion of a distribution reveals how the observations are spread out or scattered on each side of the center. To measure the dispersion, scatter, or variation of a distribution is as important as to locate the central tendency. If the dispersion is small, it indicates high uniformity of the observations in the distribution. Absence of dispersion in the data indicates perfect uniformity. This situation arises when all observations in the distribution are identical. If this were the case, description of any single observation would be enough**.**

The study of dispersion bears its importance from the fact that various distributions may have exactly the same averages, but substantial differences in their variability. See the following example.

Mean of X = $\frac{80+90+100+110+120}{5}=100$

Mean of Y = $\frac{0+50+100+150+200}{5}=100$

Median of X and median of Y = 100 . Same Mean and same Median. But Y is more dispersed than X. that is why despersion is important more.

There are five types of measures of varaince/dispersion. These are : range, quartile deviation, mean deviation, varaince and standard daviation

I. **Range-** The Range is the difference between the lowest and highest values.

Range = **(Max. data entry) – (Min. data entry**).

Example: In {4, 6, 9, 3, 7} the lowest value is 3, and the highest is 9. So the range is 9 − 3 = 6.
To find the range, first order the data from least to greatest. Then subtract the smallest value from the largest value in the set.The range is the size of the smallest interval (statistics) which contains all the data and provides an indication of statistical dispersion. Since it only depends on two of the

observations, it is most useful in representing the dispersion of small data sets.The range also represents the variability of the data. Datasets with a large range are said to have large variability, while datasets with smaller ranges are said to have small variability.To calculate data, the data should be quantitative.

Example:

The wait time to see a bank teller is studied at 2 banks.

Bank A:	5.2	6.2	7.5	8.4	9.2
Bank B:	6.6	6.8	7.5	7.7	7.9

Find the mean, median, and range for each bank.

Exercise: Find the range of the following two sets**,** do they vary the same?

Set A: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Set B: 1, 10, 10, 10, 10, 10, 10, 10, 10, 10

Advantages of Range

- It is simple to undestant and easy to calculate
- It is less time consuming

Disadvantages og Range

- It is not based on each and evvery item of the distribution
- It is very much affec ted by extreme value
- The value of range is affected more by sampling fluctuationsrange can't be computed in case of open ended distribution

II. Quartile Deviation

Quartiles in statistics are values that divide your data into quarters.

The Quartile Deviation(QD) is the product of half of the difference between the upper and lower quartiles. Mathematically we can define as:

Quartile Deviation(QD) = (Q3 – Q1) / 2.

Quartile deviation is based on the difference between the first quartile and the third quartile in the frequency distribution and the difference is also known as the interquartile range, the difference divided by two is known as quartile deviation or semi interquartile range.

Quartile Deviation defines the absolute measure of dispersion.The measure of dispersion depending upon the lower and upper quartiles is known as the quartile deviation. The difference between the upper and lower quartile is known as the Interquartile range. Half the interquartile range is known as Semi-interquartile range or quartile deviation.

To find the quartiles, we use the logic that the first quartile lies halfway between the lowest value and the median; and the third quartile lies halfway between the median and the largest value.

Q1 is the median (the middle) of the lower half of the data, and Q3 is the median (the middle) of the upper half of the data. (3, 5, 7, 8, 9), | (11, 15, 16, 20, 21). Q1 = 7 and Q3 = 16.

Then Subtract Q1 from Q3.

It is meant that, Q1 is the middle value in the first half of the data set. In the case of even number of data points in the first half of the data set, the middle value is the average of the two middle values. Q3 is the middle value in the second half of the data set.

Exercise: Calculate median, mean, range and quartile deviation of the following data set
23, 11, 13, 17, 9, 10,8, 3,4

A. Merits of Quartile Deviation:

- It can be easily calculated and simply understood.
- It does not involve much mathematical difficulties.
- As it takes middle 50% terms hence it is a measure better than Range and Percentile Range.
- It is not affected by extreme terms as 25% of upper and 25% of lower terms are left out.
- In case we are to deal with the center half of a series this is the best measure to use

Demerits or Limitation Quartile Deviation:

- As Q1 and Q3 are both positional measures hence are not capable of further algebraic treatment.
- Calculation are much more, but the result obtained is not of much importance.
- It is too much affected by fluctuations of samples.
- We can't call it a measure of dispersion as it does not show the scatterness around any average.
- In the case of distributions with high degree of variation, quartile deviation has less reliability.

Exercise: Find the range and interquartile range of the following distributions

1. 1,2,3,3,3,4,6,9,11,20
2. 9,7,6,6,5,4,4,1

III. Mean Deviation

In a statistical distribution or a set of data, the average of the absolute values of the differences between individual numbers and their mean.Mean deviation is a statistical measure of the average deviation of values from the mean in a sample. Calculate the average of the differences by adding them and dividing by the number of observations.
Mean Deviation = $\Sigma|x - \mu|/N$. Σ is Sigma, which means to sum up.
Or
To calculate mean deviation,

2. Find the mean of all values.
3. Substract t the value of each observation from the mean
4. Calculate the average of the differences by adding them and dividing by the number of the observation

Example: calculate mean daviation for the following data set
2,5,7, 10,12,14

Solution,
Mean = 2+5+7+10+12+14/6 = 8.3
Calculate the the diffenece between each value and aveage as follows 2-8.3=6.3, 5-8.3 =3.3, 7-8.3=1.3, 10-8.3=1.7, 12-8.3=3.7 and 14-8.3 = 5.7 is the diffence between each value,
N.B in this case the negative values are considered as positive.
Then, the average of the differences by addingand dividing by the number of the observation = 6.3+3.3+1.3+1.7+3.7+5.7/6= 3.66

The mean deviation is the first measure of dispersion that we will use that actually uses each data value in its computation. It is the mean of the distances between each value and the mean. It gives us an idea of how spread out from the center the set of values is.

Merits. Mean deviation is based on all the observations and is thus definitely a better measure of dispersion than the range and quartile deviation. Mean deviation is rigidly defined and is easy to understand and calculate. Mean deviation is rigidly understood.

Exercise: calculate mean deviation for the following data set
7,3,9,11, 5,8, 13

IV. Variance and Standard Deviation

The **variance**is a way to measure how far a set of numbers is spread out. Variance describes how much a random variable differs from its expected value. The variance is defined as the average of the squares of the differences between the individual (observed) and the expected value.

The variance is a squared value because it's convenient. Variance is not squared; it is the square of standard deviation. Variance is a measure of scatter; it is the average value of the squared distances measured from the mean. As such, the unit of variance is the square of the unit of the measured quantity.
The **standard deviation** is a number used to tell you how measurements for a group are spread out from the mean, or expected value. Low satndard deviation means that, most of the numbers are close to the average and high standard devaiation means that the numbers are more spread out. It is the square root of the variance. The standard deviation is expressed in the same units as the mean is, whereas the variance is expressed in squared units.
To calculate the variance,

1. First calculate mean
2. Then subtract the mean from each number.
3. Then square the results to find the squared differences.
4. Then add up all the squared differences.
5. Finally divide the sum by n minus 1. The result is the variance.

Where n is the total number of data points in your sample.

Variance: $S^2 = \frac{\sum(x_i - \bar{x})^2}{\cap - 1}$

s^2 = sample variance
xi = the value of the one observation
$\bar{x}$ = the mean value of all observations
n = the number of observations

Standard deviation for population :$\sigma = \frac{\sqrt{\sum(xi-\mu)^2}}{\mathbb{N}}$

σ = population standard deviation
N= the size of the population
Xi= each value from the population
μ =the population mean

Standard deviation for sample : $s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N-1}}$

Examples :Example: calculate varaince and standard daviation daviation for the following data set

Family #	1	2	3	4	5	6	7	8	9	10
Size(Xi)	3	3	4	4	5	5	6	6	7	7

$$\bar{x} = \frac{\sum x_i}{n} = \frac{50}{10} = 5$$

Family No.	1	2	3	4	5	6	7	8	9	10
Xi	3	3	4	4	5	5	6	6	7	7
xi-$\bar{x}$	-2	-2	-1	-1	0	0	1	1	2	2
(xi-$\bar{x}$) square	4	4	1	1	0	0	1	1	4	4
Xi square	9	9	16	16	25	25	36	36	49	49

$$s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} = \frac{20}{9} = 2.2,$$

$$s = \sqrt{2.2} = 1.48$$

Advantages of Variance

- The advantage of variance is that it treats all deviations from the mean the same regardless of their direction.
- The squared deviations cannot sum to zero and give the appearance of no variability at all in the data.

The drawback of variance

- It is not easily interpreted.
- It gives added weight to outliers, the numbers that are far from the mean and squaring these numbers can skew the data

N.B Variance can be negative and zero value means that all of the values within a data set are identical.

Merits of standard deviation

- It is the best measure of variation.
- It is rigidly defined and free from any ambiguity.
- Its calculation is based on all the observations of a series and it cannot be correctly calculated ignoring any item of a series.
- It strictly follows the algebraic principles, and it never ignores the + and – signs like the mean deviation.
- It is capable of further algebraic treatment as it has a lot of algebraic properties.
- It is not much affected by the fluctuations in sampling for which is widely used in testing the hypotheses and for conducting the different tests of significance viz. : test, t^2 test etc.
- It exhibits the scatter of dispersion of the various items of a series form its arithmetic mean and thereby justifies its name as a measure of dispersion.
- It enables us to make a comparative study of the two, or moiré series, and to tell upon their consistency, or stability through calculation of the important factors viz. co-efficient of variation, variance etc.
- It enables us to determine the reliability of the Mean of the two or more series when they show the identical means.
- It can be calculated through a good number of methods yielding the same results.

Demerits of standard deviation

- It is not understood by a common man.
- Its calculation is difficult as it involves many mathematical models and processes.
- It is affected very much by the extreme values of a series in as much as the squares of deviations of big items proportionately bigger than the squares of the smaller items.
- It cannot be used for comparing the dispersion of two, or more series given in different units.

Conclusion: The range and interquartile range are usually ineffective to measure the dispersion of a set of data. An useful measure that describes the dispersion of all the values is the variance or standard deviation.

Exercise: calculate varaince and standard deviation for the following data set. Here is the data set of 10 students score from 10.
4, 7, 8, 9. 8, 3, 0, 1, 2, 10

4.2.3. Measures of relationship

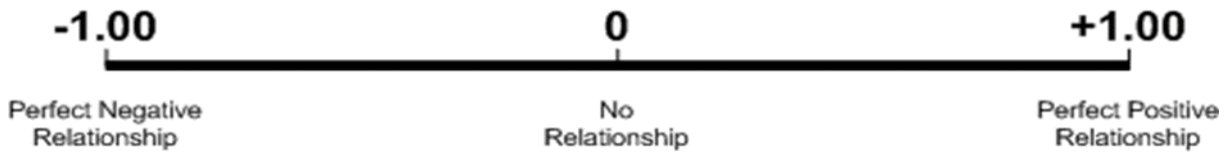
? What do you think of measure of relationship?

Measures of relation are statistical measures which show a relationship between two or more variables or two or more sets of data. For example, generally there is a high relationship or

correlation between parent's education and academic achievement. The major statistical measure of relationship is the correlation coefficient. Correlation coefficients are used to measure the strength of the relationship between two variables. Correlation can be positive, negative, and none (no correlation). Positive Correlation: Sign +ve relation is direct.

+ve implies one directional - an increase in one variable is associated with an increase in the other variable and a decrease in one variable is associated with a decrease in the other variable, as one variable increases so does the other. Negative Correlation: Sign –ve means the relation is indirect or inverse. -ve implies an increase in one variable is associated with a decrease in the other, as one variable increases, the other decreases. No Correlation: there is no apparent relationship between the variables.

The value of r ranges between (-1) and (+1). The value of r denotes the strength of the association



Correlation Coefficeint is represented bu ‘r’

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left(\sum x^2 - \frac{(\sum x)^2}{n}\right)\left(\sum y^2 - \frac{(\sum y)^2}{n}\right)}}$$

A sample of 6 children was selected, data about their age in years and weight in kilograms was recorded. Find the correlation between age and weight.

serial No	Weight (Kg)Y	Age X(years)
1	12	7
2	8	6
3	12	8
4	10	5
5	11	6
6	13	9

$$r = \frac{461 - \frac{41 \times 66}{6}}{\sqrt{\left[291 - \frac{(41)^2}{6}\right]\left[742 - \frac{(66)^2}{6}\right]}}$$

There are different types of Correlation Coefficient. These are: Pearson's Product-Moment; Spearman's Rank-Order; Point-biserial; and Phi-Coefficient

V. Pearson's Product-Moment

The Pearson product-moment correlation coefficient (Pearson's correlation, for short) is a measure of the strength and direction of association that exists between two variables measured on at least an interval scale.

For example, you could use a Pearson's correlation to understand whether there is an association between exam performance and time spent revising. You could also use a Pearson's correlation to understand whether there is an association between depression and length of unemployment.

The Pearson product-moment correlation was devised by Karl Pearson in 1895, and it is still the most widely used correlation coefficient. The Pearson product-moment correlation is an index of the degree of linear relationship between two variables that are both measured on at least an ordinal scale of measurement. The index is structured so the a correlation of 0.00 means that there is no linear relationship, a correlation of +1.00 means that there is a perfect positive relationship, and a correlation of -1.00 means that there is a perfect negative relationship. As you move from zero to either end of this scale, the strength of the relationship increases.

There are four assumption here

Assumption #1: Your two variables should be measured at the interval or ratio level (i.e., they are continuous). Examples of variables that meet this criterion include revision time (measured in hours), intelligence (measured using IQ score), exam performance (measured from 0 to 100), weight (measured in kg), and so forth.

Assumption #2: There is a linear relationship between your two variables.

Assumption #3: There should be no significant outliers. Outliers are simply single data points within your data that do not follow the usual pattern (e.g., in a study of 100 students' IQ scores, where the mean score was 108 with only a small variation between students, one student had a score of 156, which is very unusual, and may even put her in the top 1% of IQ scores globally).

Assumption #4: Your variables should be approximately normally distributed.

VI. Spearman's Rank-Order

It is a non-parametric measure of correlation. This procedure makes use of the two sets of ranks that may be assigned to the sample values of x and Y. Spearman Rank correlation coefficient could be computed in the following cases:

- Both variables are quantitative.
- Both variables are qualitative ordinal.
- One variable is quantitative and the other is qualitative ordinal.

The Spearman rank-order correlation coefficient (Spearman's correlation, for short) is a nonparametric measure of the strength and direction of association that exists between two variables measured on at least an ordinal scale. It is denoted by the symbol rs (or the Greek letter ρ, pronounced rho).

The Spearman correlation has the same range as the Pearson correlation, and the numbers mean the same thing. A zero correlation means that there is no relationship, whereas correlations of +1.00 and -1.00 mean that there are perfect positive and negative relationships, respectively.The Spearman correlation is less sensitive than the Pearson correlation to strong outliers that are in the tails of both samples. That is because Spearman's limits the outlier to the value of its rank.

VII. Phi-Coefficient

In statistics, the phi coefficient (or mean square contingency coefficient and denoted by φ or rφ) is a measure of association for two binary variables. Introduced by Karl Pearson, this measure is similar to the Pearson correlation coefficient in its interpretation.The Phi Coefficient is a measure of association between two binary variables (i.e. living/dead, black/white, success/failure). The Phi coefficient is an index of the degree of relationship between two variables that are measured on a nominal scale. Because variables measured on a nominal scale are simply classified by type.

For example, suppose you want to study the relationship between religious background and occupations. You have a classification systems for religion that includes Catholic, Protestant, Muslim, orthodox and other. You have also developed a classification for occupations that include Unskilled Laborer, Skilled Laborer, Clerical, Middle Manager, Small Business Owner, and Professional/Upper Management. You want to see if the distribution of religious preferences differ by occupation, which is just another way of saying that there is a relationship between these two variables. And Phi coefficient will applied. The Phi Coefficient is not used nearly as often as the Pearson and Spearman correlations.

4.3. Inferential statistical methods in quantitative research

✪ Section Overview

Dear learner, in the previous section of this chapter you have learnt about descriptive statistical methods in quantitative research. It is meant that descriptive statistics are only used for description not for inferential. If a researcher need to infer the data gained for the sample to the population, he/she needs to use inferential statistics. This section will introduce you with inferential statistics in quantitative research.

Section Objectives

Dear learner, by the end of this section you should be able to:

- ✓ Discuss about inferential statistics and its importance;
- ✓ Differentiate among different types of inferential statics;
- ✓ Discuss about estimation and hypothesis test; and
- ✓ Construct hypothesis and test it.

? What is infrential statistics and how it differs from the descriptive one?

Inferential statistics are the statistical procedures that are used to reach conclusions about associations between variables. They differ from descriptive statistics in that they are explicitly designed to test hypotheses.Inferential statistics are used when data is viewed as a subclass of a specific population.They are used to draw conclusions about the population from the sample data is called inferential statistics. Inferential statistics draw inferences about the population.

There are many different types of statistical analysis. Choosing the correct analytical approach for your situation can be a daunting process. You should plan your statistical approach at the start of your project, before you collect any data. Different statistical tests have different requirements and planning in advance has various benefits:

- Knowing the statistical approach will allow you to plan the way you collect your data.
- You will save time because you'll only collect relevant data.
- You will save effort.

Parameter:
Any value calculated for the population will be a constant and is called a parameter. Parameter is an unknown quantity in a sample study
Statistic:
Any sample value is defined as statistic and it is a variable form one sample to another. Statistic is a known quantity.

Parametric and Non-parametric tests
Parametric tests are those that make assumptions about the parameters of the population distribution from which the sample is drawn. This is often the assumption that the population data are normally distributed. **Non-parametric tests** are "distribution-free" and, as such, can be used for **non**-Normal variables.
Parametric test is based on the fact that the variables are measured on an interval scale, whereas in the non-parametric test, the same is assumed to be measured on an ordinal scale. It's safe to say that most people who use statistics are more familiar with parametric analyses than nonparametric analyses. Nonparametric tests are also called distribution-free tests because they don't assume that your data follow a specific distribution. You may have heard that you should use nonparametric tests when your data don't meet the assumptions of the parametric test

Reasons to use parametric tests
1: Parametric tests can perform well with skewed and non-normal distributions
Parametric tests can perform well with continuous data that are non-normal if you satisfy the sample size guidelines
2: Parametric tests can perform well when the spread of each group is different
For nonparametric tests that compare groups, a common assumption is that the data for all groups must have the same spread (dispersion). If your groups have a different spread, the nonparametric tests might not provide valid results.
3: Statistical power
Parametric tests usually have more statistical power than nonparametric tests. Thus, you are more likely to detect a significant effect when one truly exists.

Reasons to use non- parametric tests

1: Your area of study is better represented by the median

This is the best reason to use a nonparametric test and the one that isn't mentioned often enough. The fact that you can perform a parametric test with non-normal data

For example, the center of a skewed distribution, like income, can be better measured by the median where 50% are above the median and 50% are below.

2: You have a very small sample size

When you have a really small sample, you might not even be able to ascertain the distribution of your data because the distribution tests will lack sufficient power to provide meaningful results.

3: You have ordinal data, ranked data, or outliers that you can't remove

Typical parametric tests can only assess continuous data and the results can be significantly affected by outliers. Conversely, some nonparametric tests can handle ordinal data, ranked data, and not be seriously affected by outliers.0.

4.6.1 Strategies of inferential statistics

Inferential statistics arise out of the fact that sampling naturally incurs sampling error and thus a sample is not expected to perfectly represent the population. The methods of **inferential statistics** are (1) the estimation of parameter(s) and (2) testing of **statistical** hypotheses.

4.6.1.1 Estimation: any analytical method to find out unknown parameter through known staistics.

Estimator: the function of the sample value to estimate parameter.

Estimate: the numerical value obtained by substituting the data values in the estimator.

There are two types of estimation:

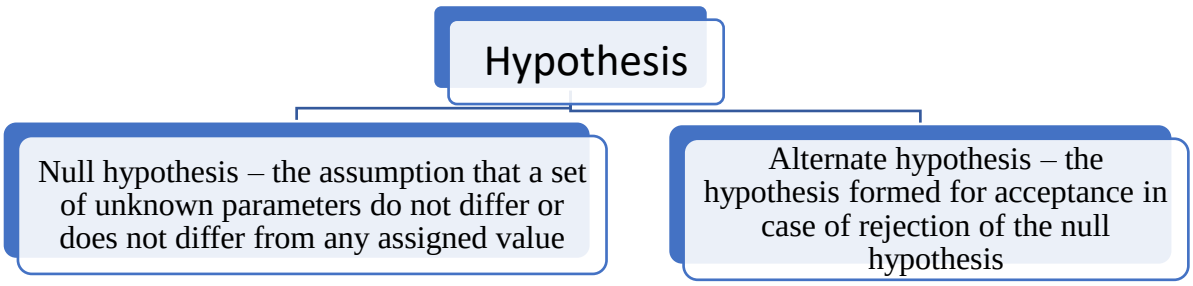
VIII. **Point Estimation:** The estimation of a parameter by an estimator is called point estimation. When the standard error of the estimator is small then estimator is said to be more precise.

IX. **Interval Estimation:** is defined as estimation of lower and upper limits for a parameter.

4.6.1.2 Hypothesis testing

?What Is Hypothesis ?

It is the tentative answers to research questions is called hypothesis. The assumption of an unknown parameter is called hypothesis In simple terms, a hypothesis refers to a supposition which is to be accepted or rejected. In hypothesis testing, an analyst tests a statistical sample, with the goal of accepting or rejecting a null hypothesis. The test tells the analyst whether or not his primary hypothesis is true. All analysts use two different hypotheses: the null hypothesis and the alternative hypothesis. The null hypothesis is the hypothesis the analyst believes to be true. Analysts believe the alternative hypothesis to be untrue, making it effectively the opposite of a null hypothesis. Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true.



The statistics used to accept or reject null hypothesis is called test statistics. Various sampling distributions for constructing test statistics are z, t, F or χ^2 depending upon

- type of data - qualitative or quantitative
- sample size - large or small

Rejection or acceptance of the null hypothesis based on the estimates of the test statistic is called test of hypothesis

- Reject the null hypothesis when it is true – Type I Error
- Accept the null hypothesis when it is false – Type II Error

The followings are types of test statistics we use to test our hypothesis.

❖ T and Z tests

In the parametric test, there can be two types of test, t-test and z-test.

T-test is used to examine how the means taken from two independent samples differ. T-test follows t-distribution, which is appropriate when the sample size is small, and the population standard deviation is not known.

Assumptions of T-test:

- All data points are independent.
- The sample size is small. Generally, a sample size exceeding 30 sample units is regarded as large, otherwise small but that should not be less than 5, to apply t-test.
- Sample values are to be taken and recorded accurately.

Z-test refers to a statistical analysis used to test the hypothesis that proportions from two independent samples differ greatly. The researcher adopts z-test, when the population variance is known, in essence, when there is a large sample sizeIn this way, it is assumed to be known, despite the fact that only sample data is available and so normal test can be applied.

Assumptions of Z-test:

- All sample observations are independent
- Sample size should be more than 30.

Key Differences Between T-test and Z-test

- The t-test can be understood as a statistical test which is used to compare and analyses whether the means of the two population is different from one another or not when the standard deviation is not known.
- As against, Z-test is a parametric test, which is applied when the standard deviation is known, to determine, if the means of the two datasets differ from each other.

X. Chi-square and Alaysis od Variance (ANOVA)

The chi square and Analysis of Variance (ANOVA) are both inferential statistical tests.

Chi square is used when we have two categorical variables (e.g., gender and alive/dead) and want to determine if one variable is related to another.

Chi-Sqaure test is based on the proportions of the two or more groups. Simply it deals with categorical variables (Nominal Scale). Eg. Association between Smoking (Yes/No) vs. Drinking Cofee (Yes/No)

In **ANOVA**, we have two or more group means (averages) that we want to compare.

Anova is based on the means of more than two groups (use t-test when there are two groups). It deals with variable (Interval Scale) with a categorical grouping variable to compare.

Eg. Whether the average income of Managerial, Technical and Administrative staffs differs or the same.

Activity 2

1. What is data analysis?
2. How can we anlysize quantitative data?
3. How can we present quantitative data?
4. What is descriptive statistics?
5. What are the main types of descriptive statistics and their advantages?
6. What is inferential statistics?
7. Why and when inferential statistics is important?
8. What is hypothesis test?
9. When should null hypothesis be accepted or rejected?

References

Aron, Arthur and Elaine N. Aron (2002) Statistics for the Behavioral and Social Sciences. Second edition. Pearson Education, INC.: New Jersey.

Cohen (1988) ‘’Practical Statistics’’ Ch-2 in his practical Statistics, London. Edward Adnold.

Gupta C. B. & V. Gupta (2004) An Introduction to Statistical Methods. Vicas Publishing House. New Delhi

Howitt, Denis and Ducan Cramer (2005) Introduction to Statistics in Psychology. Third edition. Pearson Education Limited: Harlow, England.

Kazmier Pohl (1984) Basic Statistics for Business and Statistics, McGraw-Hill Book Company.

Manish K. Bhatia, Ritugeet K., Shikha C., and Pawanpreet K. (2007) Research Methodology and Statistical Methods. Kalian Publishers, New Delhi.

Sheldon M. Ross (2010) Introductory Statistics

Spiegel (1999) Schaum’s outline of theory and problems of statistics. SI (Metric) Edition McGraw-Hill, New York.