# CHAPTER 1

## 1. INTRODUCTION

### Definition and classifications of statistics

### Definition:

We can define statistics in two ways.

1. Plural sense (lay man definition). Statistics are the raw data themselves, like statistics of births, statistics of deaths, statistics of students, statistics of imports and exports, etc.
   It is an aggregate or collection of numerical facts.
2. Singular sense (formal definition)
   Statistics is defined as the science of collecting, organizing, presenting, analyzing and interpreting numerical data for the purpose of assisting in making a more effective decision.

### Classifications:

Depending on how data can be used statistics is some times divided in to two main areas or branches.

1. **Descriptive Statistics**: is concerned with summary calculations, graphs, charts and tables.

2. **Inferential Statistics**: is a method used to generalize from a sample to a population. For example, the average income of all families (the population) in Ethiopia can be estimated from figures obtained from a few hundred (the sample) families.

- It is important because statistical data usually arises from sample.
- Statistical techniques based on probability theory are required.

### Stages in Statistical Investigation

There are five stages or steps in any statistical investigation.

1. **Collection of data**: the process of measuring, gathering, assembling the raw data up on which the statistical investigation is to be based.
   - Data can be collected in a variety of ways; one of the most common methods is through the use of survey. Survey can also be done in different methods, three of the most common methods are:
     - Telephone survey
     - Mailed questionnaire
     - Personal interview.

   **Exercise:** discuss the advantage and disadvantage of the above three methods with respect to each other.

2. **Organization of data**: Summarization of data in some meaningful way, e.g. table form
3. **Presentation of the data**: The process of re-organization, classification, compilation, and summarization of data to present it in a meaningful form.
4. **Analysis of data**: The process of extracting relevant information from the summarized data, mainly through the use of elementary mathematical operation.
5. **Inference of data**: The interpretation and further observation of the various statistical measures through the analysis of the data by implementing those methods by which conclusions are formed and inferences made.
   - Statistical techniques based on probability theory are required.

## Definitions of some terms

a. **Statistical Population**: It is the collection of all possible observations of a specified characteristic of interest (possessing certain common property) and being under study. An example is all of the students in AAU 3101 course in this term.
b. **Sample**: It is a subset of the population, selected using some sampling technique in such a way that they represent the population.
c. **Sampling:** The process or method of sample selection from the population.
d. **Sample size:** The number of elements or observation to be included in the sample.
e. **Census:** Complete enumeration or observation of the elements of the population. Or it is the collection of data from every element in a population
f. **Parameter:** Characteristic or measure obtained from a population.
g. **Statistic:** Characteristic or measure obtained from a sample.
h. **Variable:** It is an item of interest that can take on many different numerical values.

## Types of Variables or Data:

1. **Qualitative Variables** are nonnumeric variables and can't be measured. Examples include gender, religious affiliation, and state of birth.
2. **Quantitative Variables** are numerical variables and can be measured. Examples include balance in checking account, number of children in family. Note that quantitative variables are either discrete (which can assume only certain values, and there are usually "gaps" between the values, such as the number of bedrooms in your house) or continuous (which can assume any value within a specific range, such as the air pressure in a tire.)

## Applications, Uses and Limitations of statistics

### Applications of statistics:

- In almost all fields of human endeavor.
- Almost all human beings in their daily life are subjected to obtaining numerical facts e.g. abut price.
- Applicable in some process e.g. invention of certain drugs, extent of environmental pollution.
- In industries especially in quality control area.

### Uses of statistics:

The main function of statistics is to enlarge our knowledge of complex phenomena. The following are some uses of statistics:

1. It presents facts in a definite and precise form.
2. Data reduction.
3. Measuring the magnitude of variations in data.
4. Furnishes a technique of comparison
5. Estimating unknown population characteristics.
6. Testing and formulating of hypothesis.
7. Studying the relationship between two or more variable.
8. Forecasting future events.

### Limitations of statistics

As a science statistics has its own limitations. The following are some of the limitations:

- Deals with only quantitative information.
- Deals with only aggregate of facts and not with individual data items.
- Statistical data are only approximately and not mathematical correct.
- Statistics can be easily misused and therefore should be used be experts.

## INTRODUCTION TO METHODS OF DATA COLLECTION

There are two sources of data:

1. Primary Data
    - Data measured or collect by the investigator or the user directly from the source.
    - Two activities involved: planning and measuring.
        a) Planning:
            - Identify source and elements of the data.
            - Decide whether to consider sample or census.
            - If sampling is preferred, decide on sample size, selection method,… etc
            - Decide measurement procedure.
            - Set up the necessary organizational structure.
        b) Measuring: there are different options.
            - Focus Group
            - Telephone Interview
            - Mail Questionnaires
            - Door-to-Door Survey
            - Mall Intercept
            - New Product Registration
            - Personal Interview and
            - Experiments are some of the sources for collecting the primary data.
2. Secondary Data
    - Data gathered or compiled from published and unpublished sources or files.
    - When our source is secondary data check that:
        - The type and objective of the situations.
        - The purpose for which the data are collected and compatible with the present problem.
        - The nature and classification of data is appropriate to our problem.
        - There are no biases and misreporting in the published data.

Note: Data which are primary for one may be secondary for the other.