

# CHAPTER 1

## 1. INTRODUCTION

### 1.1 Definition and classifications of statistics

#### Definition:

We can define statistics in two ways.

1. Plural sense (lay man definition).

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 Dilla University 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. about 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

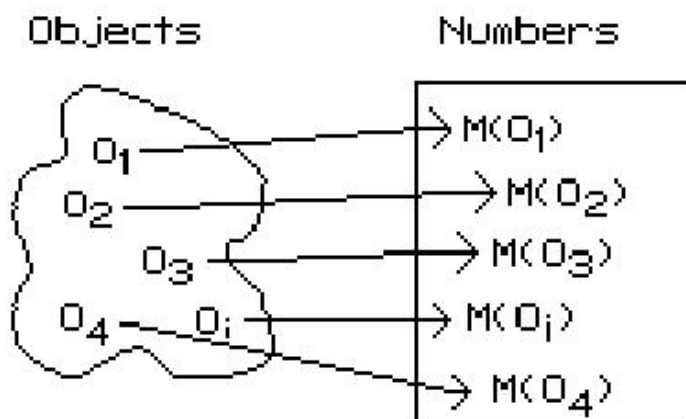
As a science statistics has its own limitations. The following are some of the limitations:

- 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 by experts.

### Scales of measurement

Proper knowledge about the nature and type of data to be dealt with is essential in order to specify and apply the proper statistical method for their analysis and inferences. Measurement scale refers to the property of value assigned to the data based on the properties of order, distance and fixed zero.

In mathematical terms measurement is a functional mapping from the set of objects  $\{O_i\}$  to the set of real numbers  $\{M(O_i)\}$ .



The goal of measurement systems is to structure the rule for assigning numbers to objects in such a way that the relationship between the objects is preserved in the numbers assigned to the objects. The different kinds of relationships preserved are called properties of the measurement system.

### **Order**

The property of order exists when an object that has more of the attribute than another object, is given a bigger number by the rule system. This relationship must hold for all objects in the "real world".

The property of ORDER exists

When for all  $i, j$  if  $O_i > O_j$ , then  $M(O_i) > M(O_j)$ .

### **Distance**

The property of distance is concerned with the relationship of differences between objects. If a measurement system possesses the property of distance it means that the unit of measurement means the same thing throughout the scale of numbers. That is, an inch is an inch, no matters where it falls - immediately ahead or a mile down the road.

More precisely, an equal difference between two numbers reflects an equal difference in the "real world" between the objects that were assigned the numbers. In order to define the property of distance in the mathematical notation, four objects are required:  $O_i, O_j, O_k$ , and  $O_l$ . The difference between objects is represented by the "-" sign;  $O_i - O_j$  refers to the actual "real world" difference between object  $i$  and object  $j$ , while  $M(O_i) - M(O_j)$  refers to differences between numbers.

The property of DISTANCE exists, for all  $i, j, k, l$

If  $O_i - O_j \geq O_k - O_l$  then  $M(O_i) - M(O_j) \geq M(O_k) - M(O_l)$ .

### **Fixed Zero**

A measurement system possesses a rational zero (fixed zero) if an object that has none of the attribute in question is assigned the number zero by the system of rules. The object does not need to really exist in the "real world", as it is somewhat difficult to visualize a "man with no height". The requirement for a rational zero is this: if objects with none of the attribute did exist would they be given the value zero. Defining  $O_0$  as the object with none of the attribute in question, the definition of a rational zero becomes:

The property of FIXED ZERO exists if  $M(O_0) = 0$ .

The property of fixed zero is necessary for ratios between numbers to be meaningful.

## **SCALE TYPES**

Measurement is the assignment of numbers to objects or events in a systematic fashion. Four levels of measurement scales are commonly distinguished: nominal, ordinal, interval, and ratio and each possessed different properties of measurement systems.

### **Nominal Scales**

Nominal scales are measurement systems that possess none of the three properties stated above.

- Level of measurement which classifies data into mutually exclusive, all inclusive categories in which no order or ranking can be imposed on the data.
- No arithmetic and relational operation can be applied.

Examples:

- Political party preference (Republican, Democrat, or Other,)
- Sex (Male or Female.)
- Marital status(married, single, widow, divorce)
- Country code
- Regional differentiation of Ethiopia.

### **Ordinal Scales**

Ordinal Scales are measurement systems that possess the property of order, but not the property of distance. The property of fixed zero is not important if the property of distance is not satisfied.

- Level of measurement which classifies data into categories that can be ranked. Differences between the ranks do not exist.
- Arithmetic operations are not applicable but relational operations are applicable.
- Ordering is the sole property of ordinal scale.

Examples:

- Letter grades (A, B, C, D, F).
- Rating scales (Excellent, Very good, Good, Fair, poor).
- Military status.

## Interval Scales

Interval scales are measurement systems that possess the properties of Order and distance, but not the property of fixed zero.

- Level of measurement which classifies data that can be ranked and differences are meaningful. However, there is no meaningful zero, so ratios are meaningless.
- All arithmetic operations except division are applicable.
- Relational operations are also possible.

Examples:

- IQ
- Temperature in  $^{\circ}\text{F}$ .

## Ratio Scales

Ratio scales are measurement systems that possess all three properties: order, distance, and fixed zero. The added power of a fixed zero allows ratios of numbers to be meaningfully interpreted; i.e. the ratio of Bekele's height to Martha's height is 1.32, whereas this is not possible with interval scales.

- Level of measurement which classifies data that can be ranked, differences are meaningful, and there is a true zero. True ratios exist between the different units of measure.
- All arithmetic and relational operations are applicable.

Examples:

- Weight
- Height
- Number of students
- Age

The following present a list of different attributes and rules for assigning numbers to objects. Try to classify the different measurement systems into one of the four types of scales. (Exercise)

1. Your checking account number as a name for your account.
2. Your checking account balance as a measure of the amount of money you have in that account.

3. The order in which you were eliminated in a spelling bee as a measure of your spelling ability.
4. Your score on the first statistics test as a measure of your knowledge of statistics.
5. Your score on an individual intelligence test as a measure of your intelligence.
6. The distance around your forehead measured with a tape measure as a measure of your intelligence.
7. A response to the statement "Abortion is a woman's right" where "Strongly Disagree" = 1, "Disagree" = 2, "No Opinion" = 3, "Agree" = 4, and "Strongly Agree" = 5, as a measure of attitude toward abortion.
8. Times for swimmers to complete a 50-meter race
9. Months of the year Meskerm, Tikimit...
10. Socioeconomic status of a family when classified as low, middle and upper classes.
11. Blood type of individuals, A, B, AB and O.
12. Pollen counts provided as numbers between 1 and 10 where 1 implies there is almost no pollen and 10 that it is rampant, but for which the values do not represent an actual counts of grains of pollen.
13. Regions numbers of Ethiopia (1, 2, 3 etc.)
14. The number of students in a college;
15. the net wages of a group of workers;
16. the height of the men in the same town;

<b>INTRODUCTION TO METHODS OF DATA COLLECTION</b>
---

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.