CHAPTER 1

The Nature of Probability and Statistics

DEFINITION:

Statistics: *is the science of conducting studies to collect, organize, summarize, analyze, and draw conclusions from data.*

Population: *consists of all subjects (human or otherwise) that are being studied.*

Sample: *is a group of subjects selected from a population.*

Variable: *is a characteristic or attribute that can assume different values.*

Random Variable: *is a variables whose values are determined by chance.*

Data: *Data are the values (measurements or observations) that the variables can assume.*

The body of knowledge called statistics is sometimes divided into two main areas, depending on how data are used. The two areas are

1. Descriptive statistics

2. Inferential statistics

Descriptive statistics: *consists of the collection, organization, summarization, and presentation of data.*

Inferential statistics: *consists of generalizing from samples to populations, performing estimations and hypothesis tests, determining relationships among variables, and making predictions.*

Variables and Types of Data:

*Variables can be classified as **qualitative** or **quantitative**.*

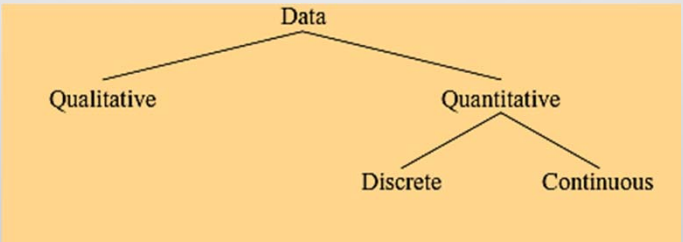
Qualitative variables: *are variables that can be placed into distinct categories, according to some characteristic or attribute. For example, if subjects are classified according to gender (male or female).*

Quantitative variables: *are numerical and can be ordered or ranked. For example, the variable age is numerical, and people can be ranked in order according to the value of their ages. Other examples of quantitative variables are heights, weights, and body temperatures.*

Quantitative variables can be further classified into two groups: ***discrete*** *and* ***continuous****.*

Discrete variables: *assume values that can be counted.*

Continuous variables: *can assume an infinite number of values between any two specific values. They are obtained by measuring. They often include fractions and decimals.*



The variables are classified as qualitative or quantitative, and can be classified by the types of scales are used: nominal, ordinal, interval, and ratio.

The nominal level *of measurement classifies data into mutually exclusive (nonoverlapping) categories in which no order or ranking can be imposed on the data.*

The ordinal level *of measurement classifies data into categories that can be ranked; however, precise differences between the ranks do not exist.*

Data Collection and Sampling Techniques:

** Data can be collected in a variety of ways. One of the most common methods is through the use of **surveys**. Surveys can be done by using a variety of methods. Three of the most common methods are the **telephone survey**, the **mailed questionnaire**, and the **personal interview**.*

** To obtain samples that are unbiased—i.e., that give each subject in the population an equally likely chance of being selected—statisticians use four basic methods of sampling: **random**, **systematic**, **stratified**, and **cluster sampling**.*

1) ***Random samples:*** *are selected by using chance methods or random numbers. One such method is to number each subject in the population.*

2) ***Systematic samples*** *are obtain by numbering each subject of the population and then selecting every kth subject.*

3) ***Stratified samples:*** *are obtain by dividing the population into groups (called strata) according to some characteristic that is important to the study, then sampling from each group.*

4) ***Cluster samples:*** *are obtain by dividing the population into groups called clusters by some means such as geographic area or schools in a large school district, etc. Then the researcher randomly selects some of these clusters and uses all members of the selected clusters as the subjects of the samples.*

Observational and Experimental Studies:

*There are several different ways to classify statistical studies. This section explains two types of studies: **observational studies** and **experimental studies**.*

In an **observational study: the researcher merely observes what is happening or what has happened in the past and tries to draw conclusions based on these observations.*

*** In an **experimental study**, the researcher manipulates one of the variables and tries to determine how the manipulation influences other variables.*

*** Statistical studies usually include one or more **independent variables** and one **dependent variable**.*

- *The **independent variable** in an experimental study is the one that is being manipulated by the researcher. The independent variable is also called the explanatory variable.*

- *The dependent variable is the resultant variable and called the outcome variable.*