Chapter 1

Introduction to Statistics

1. What Is Statistics

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

2. Descriptive and Inferential Statistics

A **variable** is a characteristic or attribute that can assume different values.
Data are the values (measurements or observations) that the variables can assume. Variables whose values are determined by chance are called **random variables.** A collection of data values forms a **data set.** Each value in the data set is called a **data values** or a **datum.**

Data can be use in different ways. The body of knowledge called statistics is sometimes divided into two main areas, depending on how data are used. The two areas are **descriptive statistics** and **inferential statistics.**
In descriptive statistics the statistician tries to describe a situation.

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

The second area of statistics is called **inferential statistics**. Here, statistician tries to make inferences from **samples** to **populations**. Inferential statistics uses **probability,** the chance of an event occurring.

It is important to distinguish between a sample and a population

A **population** consists of all subjects (human or otherwise) that are being studied.

A **sample** is a group of subject selected from a population.

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

3. Variables and Types of Data

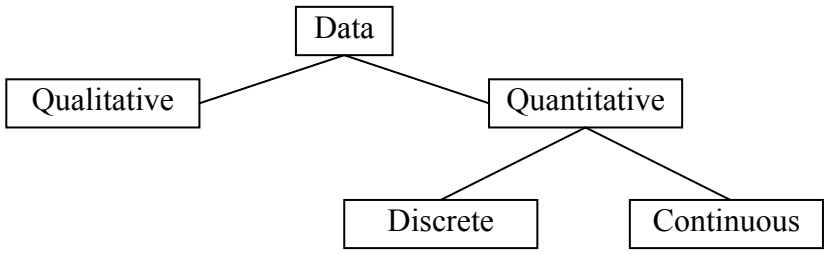
Variables can be classified as qualitative or quantitative. **Qualitative variables** are variables that can be placed into distinct categories, according to some characteristic or attribute. For example, if subject are classified according to gender (male or female), the variable gender is qualitative.

Quantitative variables are numerical and can be ordered or ranked. For example, the variable age, heights, weights, body temperature are numerical.

Quantitative variables can be further classified into two groups: discrete and continuous. *Discrete variables* can be assigned values such as 0, 1, 2, 3 and are said to be countable. Examples of discrete variables are the number of children in a family and the number of students in a classroom.

Discrete variables assume values that can be counted.
Continuous variables can assume all values between any two specific values. They are obtained by measuring.

The classification of variable can be summarized as follows:



In addition to being classified as qualitative or quantitative, variables can be classified by how they are categorize, counted, or measured. For example, can the data be organized into specific categories, such as area of residence (rural, suburban, or urban)? Can the data value be ranked, such as first place second place, etc.? Or are the values obtained from measurement, such as heights or temperature? This type of classification, i.e., how variables are categorized, counted, or measured, uses **measurement scales,** and four common types of scales are used: nominal, ordinal, interval, and ratio.

The **nominal level of measurement** classifies data into mutually exclusive (non-overlapping), exhausting categories in which no order or ranking can be imposed on the data.
Example: College instructors are classified according to subject they taught (e.g., English, Math, history, or psychology). Gender, eye color, political or religious affiliation, nationality and so on are nominal-level data.

The **ordinal level of measurement** classifies data into categories that can be ranked; however, precise differences between the ranks do not exist.
Example: Letter grades (A, B, C, D, F), rating scale (poor, good, excellent)

The **interval level of measurement** ranks data, and precise differences between units of measure do exist; however, there is no meaningful zero (no true zero).
Example: Body temperature (in degrees Celsius), IQ.

The **ratio level of measurement** possesses all the characteristics of interval measurement, and there exists a true zero. In addition, true ratios exist when the same variable is measured on two different members of the population.
Example: height, weight, time, salary and age etc..

4. 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.

Data can also be collected in other ways, such as *surveying records* or *direct observation* of situations.

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

Random Sampling
Random Samples are selected by using chance methods or random numbers. One such method is to number each subject in the population. Then place numbered cards in a bowl, mix them thoroughly, and select as many cards as needed. Statisticians also use another method of obtaining numbers. They generated random numbers with a computer or calculator. Before the invention of computers, random numbers were obtained from a table or random numbers.

Systematic Sampling
Researchers obtain *systematic samples* by numbering each subject of the population and then selecting every *k*th subject. For example, suppose there were 2000 subjects in the population and a sample of 50 subjects were needed. Since 2000÷50=40, then k=40, and every 40th subject would be selected; however, the first subject (numbered between 1 and 40) would be selected at random. Suppose subject 12 were the first subject selected; then the sample would consist of the subjects whose numbers were 12, 52, 92, etc., until 50 subjects were obtained.

Stratified Sampling
Researchers obtain *stratified samples* by dividing the population in groups (called strata) according to some characteristic that is important to the study, then sampling from each group.

Cluster Sampling
Researchers select *cluster samples* by using intact groups called clusters. Suppose researchers wish to survey apartment dwellers in a large city. If there are 10 apartment buildings in the city, the researcher can select at random 2 buildings from the 10 and interview all the residents of these buildings.