STAT 2361
STATISTICS FOR BUSINESS AND ECCONOMICS
STAT 2311
STATISTICS 1
LECTURE NOTES

PREPEARED BY MOHAMMAD MADIAH

CHAPTER 1
Introduction to Statistics
Basic Concepts and Definitions

WHAT IS STATISTICS?

Statistics is defined as the methods used to collect, organize, present, analyze and interpret data for the purpose **of assisting in making more effective decisions.**

Data are the facts and figures collected, analyzed, and summarized for presentation and interpretation. All the data collected in a particular study are referred to as the **data set** for the study.

Elements are the entities on which data are collected.
A **variable** is a characteristic of interest for the elements.

A **population** is the collection of all elements in a particular study.

Measurements collected on each variable for every element in a study provide the data. The set of measurements obtained for a particular element is called an **observation**.

It is difficult to have information about all of the elements of the population of interest. Typically, we have taken measurements on a **sample** (subset) of the population.
A study (survey) that includes every element of the population is called a **census**, while collecting information from a sample is called a **sample survey.**

The purpose of sampling is to make decisions about the corresponding population. So, a selected sample should be a **representative sample,** that is; a sample that represents the characteristic of the population as closely as possible. A sample may be random or non-random. A sample is **random**: if it is selected in such a way that each element of the population has some chance of being selected.

A **population parameter** is a numerical measure of a summary characteristic of a population (population mean, population standard deviation…).

A **sample statistic** is a numerical measure of a summary characteristic of a sample (Sample mean…).

Types of Statistics:
1. **Descriptive Statistics** consists of methods for organizing, displaying, and describing data by using, **tables, graphs**, and **numerical measures.**

2. **Inferential Statistics** is methods that use sample results to help make a **decisions** or a **prediction** about population. For example, the average income of all families (the population) in the West Bank can be estimated from measures obtained from randomly selected families (the sample).

Types of Variables or Data:
1. **Qualitative Variables** are variables that can't assume a numerical value but can be classified into two or more categories (classes or labels). The data collected about such variable are called qualitative data. Examples include gender of person, hair color, and marital status
2. **Quantitative Variables** are variables that can be measured numerically. The data collected about such variable are called quantitative data. Examples include incomes, number of children in your family, and prices of homes.

   A quantitative variables are either **discrete** (which can assume only certain values, and there are usually "gaps" between values, such as the number of children in your family) or **continuous** (which can assume any numerical value within a specific interval or intervals, such as the time taken to serve a customer by a bank teller.)
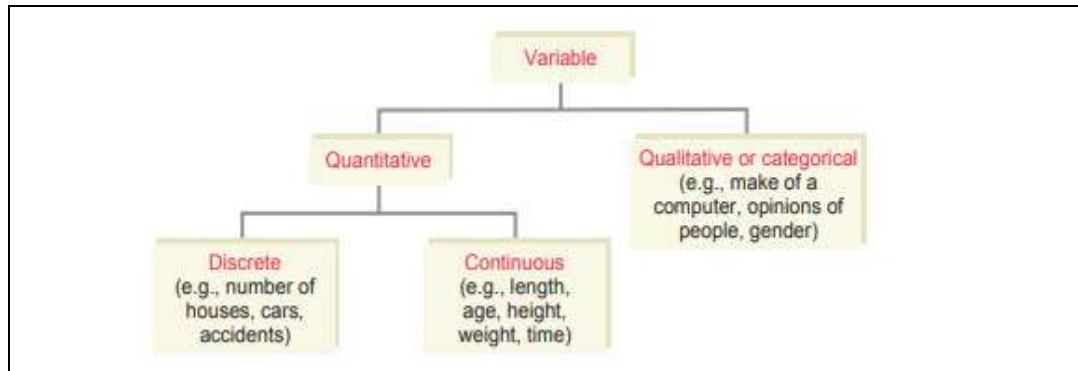
Sources of Data

1. **Secondary Data**: Data which are already available within a firm or organization. An example: statistical data about the unemployment rate can be obtained from the ministry of labor.

2. **Primary Data**: Data which must be collected.

Cross-sectional and Time series data

Cross-sectional data: Data collected at the same or approximately the same point in time
Time series data: Data collected at several successive periods of time.

Scales of Measurements

Data collection requires one of the following four scales (levels) of measurement: nominal, ordinal, interval, or ratio. The scale of measurement determines the ==amount of information contained== in the data and indicates the most appropriate data summarization and statistical analyses;

1. **Nominal Data:** The weakest data level and applies to data that are divided into different categories. Examples include gender, college, and marital status.
2. **Ordinal (Rank) Data***:* If we have data with the characteristic of the nominal scale and the order or rank of the data is meaningful, we have an ordinal scale. An example is a production evaluation: excellent, good, fair and poor.

3. **Interval Data**: If we have data with the characteristic of the ordinal scale plus the characteristic that the distance between two data items is expressed in terms of a fixed unit of measure, we have an interval measurement. Interval data are always numerical. Secondary school certificate (Tawjihi) scores are an example of interval-scaled data. For example, three students with tawjihi scores of 95, 87, and 93 can be ranked or ordered from the highest to the lowest. In addition, the differences between the scores are meaningful. For instance, student 1 scored 95 - 87 = 8 points more than student 2, while student 2 scored 87 - 93 = - 6 points less than student 3.

4. **Ratio Data**: The highest level of measurement, it has all prosperities of interval data and the ratio of two values is meaningful. Variables such as distance, height, weight, and time use the ratio scale of measurement. This scale requires that a zero value be included to indicate that nothing exists for the variable at the zero point.

- Nominal Scale - groups or classes – Gender
- Ordinal Scale - order matters - Ranks (top ten videos)
- Interval Scale - difference or distance matters – has arbitrary zero value -Temperatures (0F, 0C)
- Ratio Scale - Ratio matters – has a natural zero value – Salaries

Example

Table 1.7 shows data for eight cordless telephones (*Consumer Reports*, November 2012). The Overall Score, a measure of the overall quality for the cordless telephone, ranges from 0 to 100. Voice Quality has possible ratings of poor, fair, good, very good, and excellent. Talk Time is the manufacturer's claim of how long the handset can be used when it is fully charged.

a. How many elements are in this data set?
b. For the variables Price, Overall Score, Voice Quality, Handset on Base, and Talk Time, which variables are categorical and which variables are quantitative?
c. What scale of measurement is used for each variable?

TABLE 1.7   DATA FOR EIGHT CORDLESS TELEPHONES

Brand	Model	Price ($)	Overall Score	Voice Quality	Handset on Base	Talk Time (Hours)
AT&T	CL84100	60	73	Excellent	Yes	7
AT&T	TL92271	80	70	Very Good	No	7
Panasonic	4773B	100	78	Very Good	Yes	13
Panasonic	6592T	70	72	Very Good	No	13
Uniden	D2997	45	70	Very Good	No	10
Uniden	D1788	80	73	Very Good	Yes	7
Vtech	DS6521	60	72	Excellent	No	7
Vtech	CS6649	50	72	Very Good	Yes	7