

Faculty of Mathematical Science  
First Year  
An Introduction to Statistics and Probability  
S 102

### **Course outline**

1. Introduction on Statistical Science
2. Descriptive Statistics
  - 2.1 Tables and Graphical Presentations
  - 2.2 Numerical Summaries
    - i. Measures of Central Tendency
    - ii. Measures of Variation
    - iii. Measures of Position
3. Probability
  - 3.1 Random Variables
  - 3.2 Probability Rules
    - i. The Addition Rule
    - ii. The Multiplication Rule
  - 3.3 Conditional Probability
  - 3.4 Bayes Theorem
  - 3.5 Probability Functions
    - i. Discrete Distribution
    - ii. Continuous Distribution

### **References**

- Allan G. Bluman "Elementary Statistics "third edition

## **General Definitions**

### **Lecture 1**

#### **1 Lecture objectives**

- Demonstrate knowledge of statistical terms
- Differentiate between the two branches of statistics
- Identify types of data
- Differentiate between the population term and sample
- Identify the four basic sampling techniques

## 2 Why we study statistics?

- To be able read and understand the various statistical studies.
- To conduct research in your field.
- To become better consumers and citizens.

## 3 Statistical fields

Statistics is used in almost all fields of human endeavor, for example:

- In public health, an administrator might be concerned with the number of residents who contract a new strain of flu virus during a certain year.
- In sports, for example, a statistician may keep records of the number of yards a running back gains during a football game
- In education, a researcher might want to know if new methods of teaching are better than old ones.

## 4 Definitions

### 4.1 Statistics

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

*Methodological steps for statistical analysis in scientific research*

*DataCollection  $\longrightarrow$  DataDescription  $\longrightarrow$  DataAnalysis  $\longrightarrow$  DrawConclusion*

Statistics is divided into two main branches

#### 4.1.1 Descriptive statistics

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

#### 4.1.2 Inferential statistics

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

An area of inferential statistics called **hypothesis testing** is a decision-making process for evaluating claims about a population, based on information obtained from samples.

## 4.2 Data

Data are the values (measurements or observations) that the variable can assume.

Data can be used in different ways, depending on how data are used. *Data set* is collection of data values forms.

## 4.3 Variable

A variable is a characteristic or attribute that can assume different values.

If the values are determined by chance it called *random variable*.

## 4.4 Elements

An individual or something gathering information around it.

## 4.5 Observation or Measurement

Variable value of an element.

## 4.6 Population

Consists of all subjects possessing a common characteristic (human or otherwise) that being studied.

## 4.7 Sample

Sample is a subgroup or subset selected from the population subjects.

## 4.8 Parameter

A parameter is a characteristic or measure obtained by using all the data values from a specific population.

## 4.9 Statistic

A statistic is a characteristic or measure obtained by using the data values from a sample. (*not to be confused with Statistics*)

### Population vs Sample

- The population includes all objects of interest whereas the sample is only a portion of the population.
- Parameters are associated with populations and statistics with samples.
- Parameters are usually denoted using Greek letters ( $\mu, \sigma$ ) while statistics are usually denoted using Roman letters (x, s).

There are several reasons why we don't work with populations. They are usually large so it cost money and time, and it is often impossible to get data for every object we are studying.

We compute statistics, and use them to estimate parameters. The computation is the first part of the statistics (Descriptive Statistics) and the estimation is the second part (Inferential Statistics).

#### 4.10 Data collection method

Data can be collected in a variety of ways with some considerations: large staff attempting to collect this information, budget and time . One of the most common methods is through the use of surveys, Three of the most common methods are:

- The **telephone survey**, its has an advantage over personal interview surveys in that they are less costly. Also, people may be more candid in their opinions since there is no face-to-face contact. A major drawback to the telephone survey is that some people in the population will not have phones.
- The **mailed questionnaire**, it less expensive to conduct. Also, respondents can remain anonymous if they desire. Disadvantages of mailed questionnaire surveys include a low number of responses and inappropriate answers to questions. Another drawback is that some people may have difficulty reading or understanding the questions.
- The **personal interview**, the advantage of obtaining in-depth responses to questions from the person being interviewed. Disadvantage is that interviewers must be trained in asking questions, which makes it more costly than the other two survey methods and the interviewer may be biased in his or her selection of respondents.

Other ways, such as surveying records or direct observation of situations.

#### 4.11 Types of Sampling

We focus on four basic methods of sampling: random, systematic, stratified, and cluster sampling, to obtain samples that are unbiased

##### 4.11.1 Random samples

Every individual or item from the target frame (list of population) has an equal chance of being selected.

Random samples are selected by using chance methods or random numbers. One such method if population has small number of unit 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 (the number of sample).

Or generate random numbers with a computer or calculator to specify which member be select

#### 4.11.2 Systematic samples

are obtained by numbering each subject in 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  $2000 \div 50 = 40$ , 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

#### 4.11.3 Stratified samples

are obtained by dividing the population into groups (called *strata*) according to some characteristic that important to the study, then sampling from each group. Samples within the strata should be randomly selected. For example, suppose the president of a two-year college wants to learn how students feel about a certain issue. Furthermore, the president wishes to see if the opinions of the first-year students differ from those of the second-year students. The president will randomly select students from each group to use in the sample.

#### 4.11.4 Cluster samples

Here the population is divided 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