Chapter 1
Statistics, Data and Stat Thinking

1.1 The Science of Statistics

Definition 1.1 Statistics: Statistics is a science, which deals with the collection of data, analysis of data, and making inferences about the population using the information contained in the sample.

1.2 Types of Statistical Applications in Business

Definition 1.2 Descriptive Statistics: Descriptive statistics consists of procedures used to summarize the information in a set of measurements and to describe the characteristics of the set. See study 1 on page 4 that describes the descriptive statistics.

Definition 1.3 Inferential Statistics: Inferential statistics consists of procedures used to make inferences about the population characteristics from information contained in the sample. See study 2 on page 4 that describes the inferential statistics.

1.3 Fundamental Elements of Statistics

Statistical methods are particularly useful for studying, analyzing, and learning about populations of experimental units.

Definition 1.4 An Experimental Unit: An experimental unit is an object (person, thing, etc) upon which we collect data.

Definition 1.5 Population: A finite or infinite collection of measurements or individuals that comprises the totality of all possible measurements within the context of a particular statistical study.

Definition 1.6 Variable: A variable is a characteristic or property of an individual population unit. For examples, income, age, weight of a group of people.
Definition 1.7 Sample: A sample is a subset of measurements selected from the population of interest.

Definition 1.8 Statistical Inference: A statistical inference is an estimate, prediction, or some other generalization about a population based on information contained in the sample. The objective of statistical inference is to make inferences (predictions or decisions) about certain characteristics (for example, mean, standard deviation) of one or more populations based on information contained in the samples from these populations.

Some examples of Population and Sample

Example 1: Different types of writing can sometimes be distinguished by the lengths of the words used in the novel. A student is interested in this fact and wants to study the lengths of words used by Mr. William in his novels. Suppose, the student opens a William’s novel at random and records the lengths of each of the first 100 words on the page.

(a) What is the population in this study?
(b) What is the sample?
(c) What is the variable measured?

Answer:
(a) All the words in the novel from which the sample was drawn.
(b) First 100 words on a particular page (randomly selected).
(c) Length of the words.

Example 2: Suppose FIU wants to estimate the average time a student takes to find a proper parking spot. During Fall 2004, an administrator randomly asked 200 students and recorded their parking times and found that it takes on average 10 minutes to find a parking spot.

(a) What is the response that will be measured in this survey? OR What is the variable of interest?
(b) Define the population of interest to the experimenter.
(c) Describe the sample.
(d) How is the inference expressed?

Answer:
(a) Parking time.
(b) All 2004 students those will use parking lots.
(c) The sample is 200 students.
(d) It is found that on average it takes 10 minutes to find a parking spot.

Example 3: A Gallup Youth poll was conducted to determine the topics that teenagers most want to discuss with their parents during weekend. The findings show that 50% would like more discussion about their education, 30% would like to talk about their financial problem and 20% would like to talk about social problem. The survey was based on a national sampling of 1000 teenagers, selected at random from all U.S. teenagers.

(a) Describe the population of interest.
(b) Describe the sample.
(c) Is the sample representative of the population?
(d) What is the variable of interest?
(e) How is the inference expressed?

Answer:
(a) All US teenagers.
(b) The sample is 1000 teenagers.
(c) Yes, because it is a random sample. A random sample is a good sample.
(d) Family discussion
(e) 50% would like more discussion about their education, 30% would like to talk about their financial problem and 20% would talk about social problem.

Definition 1.9 Measure of Reliability: A measure of reliability is a statement about the degree of uncertainty associated with a statistical inference. Read example 1.1 and example 1.2 on page 7.

Four Elements of Descriptive Statistical Problems
1. The population or sample of interest
2. One or more variables (characteristics of the population) that are to be investigated
3. Tables, graphs, or numerical summary tools
4. Identification of patterns in the data

Five Elements of Inferential Statistical Problems
1. The population of interest
2. One or more variables (characteristics of the population) that are to be investigated
3. The sample of population units
4. The inference about the population based on the information contained in the sample
5. A measure of reliability for the inference

1.4 Not in the syllabus

1.5 Types of Data

Statistics is the science of data. Understanding data is very important in statistics. Generally there are two types of data: quantitative and qualitative.

Definition 1.10 Quantitative Data: Quantitative data are numerical measurements that represent an amount or a quantity. For examples, height, weight, price etc. See more examples on Page 13.

Definition 1.11 Qualitative Data: Qualitative data are non numerical measurements that represent an attribute or a quality. For examples, occupation, sex, marital status etc. See more examples on Page 13.
1.6 Collecting Data

1. Data from a designed of experiment (primary data)
2. Data from a survey (primary data)
3. Data from an observational study (primary data)
4. Data from a published source (secondary data)

**Definition 1.12 Representative Sample:** A representative sample exhibits characteristics typical of those possessed by the target population. The most common way to satisfy the representative sample requirement is to select a random sample. A random sample ensures that every subset of fixed size in the population has the same chance of being included in the sample.

**Definition 1.13 Random Sample:** A random sample of $n$ experimental units is a sample selected from the population in such a way that every different sample of size $n$ has an equal chance of selection.

Example 1.6, page 15

1.7 The Role of Statistics in Managerial Decision Making

**Definition 1.14 Statistical thinking:** Statistical thinking involves applying rational thought and the science of statistics to critically assess data and the inferences. Fundamental to the thought process is that variation exists in population and process data. To gain some insight into the role statistics plays in critical thinking, please see examples 1.8 and 1.9 on page 18.

**Definition 1.15 Selection Bias:** Selection bias results when a subset of the experimental units in the population is excluded so that these units have no chance of being selected for the sample.

**Definition 1.16 Nonresponse Bias:** Nonresponse bias results when the researchers conducting a survey or study are unable to obtain data on all experimental units selected or sample.

**Definition 1.17 Measurement Error:** Measurement error refers to inaccuracies in the values of the data recorded.

Exercise 1.20, page 23.