CHAPTER 1: DECISION-MAKING USING STATISTICS

What is statistics?

There are two common definitions of **statistics**[1]:

(1) **Turning data into information.** This is a good definition of *descriptive statistics* – the topic of the first part of this text.

(2) **Making inferences about populations from samples.** This describes *inferential statistics* – the topic of the latter part of this text.

These two definitions are quite different, but between them, they capture much of what you will learn in an introductory statistics course.

To reach an understanding of the second definition, an understanding of the first definition is needed; that is why we will study descriptive statistics before inferential statistics. We will begin our discussion of how to turn data into information by introducing some terms and concepts that you'll hear throughout this course.

Major Themes of STAT 3090: Introductory Business Statistics

- **Data Analysis (Chapters 1-4):** A discussion on gathering, displaying, and summarizing data (the first definition of statistics).
- **Probability (Chapters 5-8):** A study of the laws of chance.
- **Statistical Inference (Chapters 9, 10, 13):** The science of drawing statistical conclusions from specific data using a knowledge of probability (the second definition of statistics).

[1] *Introductory Business Statistics* by Thomas K. Tiemann. This book is licensed under a Creative Commons Attribution 3.0 License.

Populations versus Samples

In statistics, our ultimate goal is to uncover truths about a population. A _____ is the set of all subjects or elements about which we are interested in making decisions (inferences). A **frame** or **data frame** is a list of the members of the population.

A _____ is a subset of the population which is used to gain insight about the population. Samples are used to represent the larger population of interest.

A sample is chosen because the population of interest is almost always too large to survey, and typically it would be too expensive to collect data from the entire population.

One exception to this is a **census**, in which we collect data from every subject in a population.

Only **random samples** are likely to be useful in statistics, as they give us the best chance of obtaining a sample that is representative of the population. This is why we will study probability.

Example 1.1: Automated Workers

According to a Gallup poll conducted in April 2017, 26% of U.S. workers say that it is at least somewhat likely that their job will be eliminated by new technology, automation, artificial intelligence or robots within the next 20 years. About 13% of workers say this will happen within the next five years. Results for this Gallup poll were based on telephone interviews of a random sample of 1,127 workers, aged 18 and older, living in all 50 U.S. states and the District of Columbia.

Identify the **population** and the **sample** in the above study.

Population:

Sample:

Parameters versus Statistics

A _____ is a numerical descriptive measure or numerical summary of a population. It is usually unknown since the population is too large to have a full collection of data, but it is often inferred from the sample data. Since parameters are descriptions of the population, a population can have many parameters.

A _____ is a numerical descriptive measure of a sample. It is calculated from the data in the sample, and thus is a known value.

Example 1.2: Automated Workers (continued)

According to a Gallup poll conducted in April 2017, 26% of U.S. workers say that it is at least somewhat likely that their job will be eliminated by new technology, automation, artificial intelligence or robots within the next 20 years. About 13% of workers say this will happen within the next five years. Results for this Gallup poll were based on telephone interviews of a random sample of 1,127 workers, aged 18 and older, living in all 50 U.S. states and the District of Columbia.

Identify a **statistic** and define the **parameter** of interest associated with it.

Statistic:

Parameter:

Descriptive versus Inferential

_____ **statistics** is analysis of data that simply describes the sample or population of interest. It involves the collection, organization, analysis, and presentation of data.

_____ **statistics** is an estimate or prediction or some other generalization about a population based on information contained in a sample.

The big idea is that there is some unknown population parameter about which we wish to know something. We estimate it approximately or answer a question about it by using information gathered from a sample and summarized by a statistic.

Example 1.3: Automated Workers (revisited)

According to a Gallup poll conducted in April 2017, 26% of U.S. workers say that it is at least somewhat likely that their job will be eliminated by new technology, automation, artificial intelligence or robots within the next 20 years. About 13% of workers say this will happen within the next five years. Results for this Gallup poll were based on telephone interviews of a random sample of 1,127 workers, aged 18 and older, living in all 50 U.S. states and the District of Columbia.

Is the purpose of the Gallup poll described to perform **descriptive** or **inferential** statistics?

Practice Exercises for Chapter 1

1. Americans' perceptions of the fairness of their taxes have varied over the years, tending to be lower in years when a Democratic president was in office and higher in years when there was a Republican president. In April 2017, the Gallup News Service conducted a survey of 1019 American adults aged 18 years or older. The results of the survey indicated that 61% of American adults believe that their federal income tax is fair. *Source: gallup.com*

   (a) What was the objective of this survey?

   (b) Describe the population of interest.

   (c) Describe the sample.

   (d) Is 61% a parameter or a statistic?

   (e) Was the Gallup organization conducting inferential or descriptive statistics?

2. A manufacturer of computer chips claims that less than 9% of its products are defective. When 2,000 chips are drawn from a large production, 7.75% were found to be defective.

   (a) Describe the population of interest.

   (b) Identify any parameters or statistics.

   (c) Is this an example of inferential or descriptive statistics?

3. A personnel director is interested in determining how effective a new mathematics course will be in improving the math computation skills of her company's employees. The director randomly selected twenty employees and determined the average math computation score both before and after instruction in the course.

   (a) Identify the population.

   (b) What characteristic of the population is being measured?

   (c) Is this an example of inferential or descriptive statistics?