No Phones or Bathtubs

Many statistical analyses must consider changing characteristics of populations over time. Here are some observations of life in the United States from 100 years ago:

- 8% of homes had a telephone.
- 14% of homes had a bathtub.
- The mean life expectancy was 47 years.
- The mean hourly wage was 22 cents.
- There were approximately 230 murders in the entire United States.

Although these observations from 100 years ago are in stark contrast to the United States of today, statistical analyses should always consider changing population characteristics that might have more subtle effects.

2-1 Review and Preview

Chapter 1 presented some critically important concepts, including context of data, source of data, sampling method, conclusions, and practical implications. Like the data in Table 2-1, many samples of data are large, so understanding them requires that we organize, summarize, and represent the data in a way that allows us to gain insight. We can organize and summarize data numerically in tables or visually in graphs, as described in this chapter. Of course, our ultimate goal is not the mere generation of tables or graphs; instead, we want to use tables and graphs as keys that unlock the hidden and important characteristics of data. In this chapter we are mainly concerned with the *distribution* of a data set, which is one of the following five characteristics that are typically most important. This chapter focuses mainly on the distribution of data. Chapter 3 presents methods for investigating the other characteristics.

Characteristics of Data

1. **Center:** A representative value that indicates where the middle of the data set is located.
2. **Variation:** A measure of the amount that the data values vary.
3. **Distribution:** The nature or shape of the spread of the data over the range of values (such as bell-shaped).
4. **Outliers:** Sample values that lie very far away from the vast majority of the other sample values.
5. **Time:** Any change in the characteristics of the data over time.

Study Hint: Blind memorization is not effective in remembering information. To remember the above characteristics of data, it may be helpful to use a memory device (or mnemonic) for the first five letters **CVDOT.** Remembering the sentence "**C**omputer **V**iruses **D**estroy **O**r **T**erminate" is an easy way to help us remember the five key characteristics of data.

Critical Thinking and Interpretation: Going Beyond Formulas and Manual Calculations

In the modern statistics course, it is not so important to memorize formulas or manually perform complex arithmetic calculations. Instead, we get results by using technology (a calculator or computer software), and then we focus on making practical sense of results through critical thinking. This chapter includes detailed steps for important procedures, but it is not necessary to master those steps in all cases. However, we recommend that in each case you perform a few manual calculations before using technology. This will enhance your understanding and help you acquire a better appreciation of the results obtained from the technology.

2-2 Frequency Distributions

Key Concept When one is working with large data sets, a *frequency distribution* (or *frequency table*) is often helpful in organizing and summarizing data. A frequency distribution helps us to understand the nature of the *distribution* of a data set.

DEFINITION A **frequency distribution** (or **frequency table**) shows how data are partitioned among several categories (or *classes*) by listing the categories along with the number (frequency) of data values in each of them.