

## 1.1

## VARIABLES AND DATA

In Chapters 1 and 2, we will present some basic techniques in *descriptive statistics*—the branch of statistics concerned with describing sets of measurements, both *samples* and *populations*. Once you have collected a set of measurements, how can you display this set in a clear, understandable, and readable form? First, you must be able to define what is meant by measurements or “data” and to categorize the types of data that you are likely to encounter in real life. We begin by introducing some definitions—new terms in the statistical language that you need to know.

**Definition** A **variable** is a characteristic that changes or varies over time and/or for different individuals or objects under consideration.

For example, body temperature is a variable that changes over time within a single individual; it also varies from person to person. Religious affiliation, ethnic origin, income, height, age, and number of offspring are all variables—characteristics that vary depending on the individual chosen.

In the Introduction, we defined an *experimental unit* or an *element of the sample* as the object on which a measurement is taken. Equivalently, we could define an experimental unit as the object on which a variable is measured. When a variable is actually measured on a set of experimental units, a set of measurements or **data** result.

**Definition** An **experimental unit** is the individual or object on which a variable is measured. A single **measurement** or data value results when a variable is actually measured on an experimental unit.

If a measurement is generated for every experimental unit in the entire collection, the resulting data set constitutes the *population* of interest. Any smaller subset of measurements is a *sample*.

**Definition** A **population** is the set of all measurements of interest to the investigator.

**Definition** A **sample** is a subset of measurements selected from the population of interest.

## EXAMPLE

## 1.1

A set of five students is selected from all undergraduates at a large university, and measurements are entered into a spreadsheet as shown in Figure 1.1. Identify the various elements involved in generating this set of measurements.

**Solution** There are several *variables* in this example. The *experimental unit* on which the variables are measured is a particular undergraduate student on the campus, identified in column C1. Five variables are measured for each student: grade point average (GPA), gender, year in college, major, and current number of units enrolled. Each of these characteristics varies from student to student. If we consider the GPAs of all students at this university to be the population of interest, the five GPAs in column C2 represent a *sample* from this population. If the GPA of each undergraduate student at the university had been measured, we would have generated the entire *population* of measurements for this variable.