

Section 1.1

Definition: Statistics is the study of how to collect, organize, analyze, and interpret numerical information from data.

It is important to know the context of the data we collect.

Here is some data:

x	0	2	3	6	7	8
y	98	90	88	72	61	50

The data is meaningless without *context*. Here is the *context* for the data:

	x	0	2	3	6	7	8
	y	98	90	88	72	61	50

Definition: Individuals are the people or objects included in the study.

Definition: A **variable** is a characteristic of the individual to be measured or observed.

Variables can be categorized as quantitative and qualitative.

Definition: A **quantitative variable** has a value or numerical measurement for which operations such as addition or averaging make sense.

Definition: A **qualitative variable** describes an individual by placing the individual into a category or group, such as male or female.

Exercise 1. Data from a medical study contain values of many variables for each of the people who were the subjects of the study.

Is gender (female or male) a qualitative or a quantitative variable?

Is age a qualitative or a quantitative variable?

Class Exercise 1. Other variables from the study are given below. Which are qualitative and which are quantitative?

- Race (Asian, black, white, or other)
- Smoker (yes or no)
- Systolic blood pressure (millimeters of mercury)
- Level of calcium in the blood (micrograms per milliliter)

Sometimes, *qualitative variables* are referred to as *categorical variables*. Some variables can answer questions only about categories. If the values of a variable are words rather than numbers, it's a good bet that it is categorical. But some variables can answer both kinds of questions. Amazon could ask for your *age* in years. That seems quantitative, and would be if they want to know the average age of those customers, who visit their site after 3 a.m. But suppose they want to decide which CD to offer you in a special deal - one by Raffi, Blink182, Carly Simon, or Montovani - and need to be sure they have adequate supplies on hand to meet the demand. Then thinking of your age in one of the categories child, teen, adult, or senior might be more useful. If it isn't clear whether a variable is categorical or quantitative, think about *Why* you are looking at it and what you want it to tell you.

Exercise 2. Let's suppose that I wanted to know the proportion of students in our class who are taking night classes this semester. I will use the students in the first row to make an estimate.

The number of students in the first row who take night classes is .

The number of students in the first row is .

So, the proportion of students in the first row who take night classes is .

Is this an accurate estimate for the entire class?

The number of students in the class who take night classes is .

The number of students in the class is .

The proportion of students in the class who take night classes is .

Definition: In population data, the data are from *every* individual of interest.

Definition: In sample data, the data are from *only some* of the individuals of interest.

Class Exercise 2. For each of the following questions, identify the population and the sample:
(a) A business magazine mailed a questionnaire to the human resource directors of all of the Fortune 500 companies, and received responses from 23% of them. Those responding reported that they did not find that such surveys intruded significantly on their workday.

(b) On August 9, 2014, Michael Brown was shot and killed by a police officer in Ferguson, Missouri. The Gallup Poll interviewed 1000 randomly selected American citizens on August 11, and reported that when "asked whether they trust in justice, 45% of Americans don't trust in justice, 37% do trust the justice system, while 18% are undecided."

Exercise 3. Let's suppose that I wanted to know the proportion of Evergreen Valley College students who are taking night classes this semester and I wanted to use this class as an estimate. Is our class representative of the entire college?

Remark: It is important that the sample data is representative of the population from which the data are drawn.

Class Exercise 3. Let's suppose I wanted to estimate the proportion of California community college students who wanted to transfer to San Jose State; the sample I use consists of Evergreen Valley College students. Would Evergreen Valley College students be representative of all California community college students? Why or why not?

Another way to categorize data is according to one of the four *levels of measurement*.

Definition: The **nominal level of measurement** applies to data that consist of names, labels, or categories. There are no implied criteria by which the data can be ordered from smallest to largest.

Example: Car color.

Definition: The **ordinal level of measurement** applies to data that can be arranged in order. However, differences between data values either cannot be determined or are meaningless.

Example: Shirt size.

Definition: The **interval level of measurement** applies to data that can be arranged in order. In addition, differences between data values are meaningful.

Example: Temperature in degrees Celsius.

Definition: The **ratio level of measurement** applies to data that can be arranged in order. In addition, both differences between data values and ratios of data values are meaningful. Data at the ratio level have a true zero.

Example: Number of pens a student owns.

Exercise 4. Classify the following as nominal data, ordinal data, interval data, or ratio data.

(a) The political party affiliation (Democrat, Republican, or Independent) in a sample of 50 voters

(b) The size of a car (subcompact, compact, mid-size, or full-size) rented by each of a sample of 30 business travelers.

(c) The temperature (in degrees Celsius) at which each in a sample of 20 pieces of heat-resistant plastic begins to melt.

(d) The current unemployment rate (measured as a percentage) for each of the 50 states.

Class Exercise 4. Classify the following as nominal data, ordinal data, interval data, or ratio data.

(a) A test taker's ranking (best, worst, etc.) of four brands of barbecue sauce for a panel of 10 testers

(b) The number of convicted murderers who receive the death penalty each year over a 10-year period.

(c) The defective status (defective or not) of each of 100 computer chips manufactured by Intel.

(d) Hair color.

(e) Level of service satisfaction with car repair (Very Satisfied, Satisfied, Not Satisfied).

(f) Height of players on the Evergreen Valley College soccer team.

Definition: **Descriptive statistics** involves methods of organizing, picturing, and summarizing information from samples or populations.

Definition: Inferential statistics involves methods of using information from a sample to draw conclusions regarding the population.

Student Feedback

My teaching methods are (I hope) continually subject to improvement. If you have any comments, suggestions, or ideas, please email them to me at Sithparran.Vanniasagaram@evc.edu .

Homework

All of the questions were taken from the textbook. You may or may not complete all of the problems, but you are expected to spend **at least three hours** trying to complete the homework. You are definitely encouraged to work with your classmates on the homework, but you should write answers by yourself.

If your goal is to get a C in the class, you should complete all of the C problems. If your goal is to get a B in the class, you should complete all of the B and C problems. If your goal is to get an A in the class, you should complete all of the A, B, and C problems. Even if your goal is to just get a C, I hope that you will attempt the A and B problems.

Almost all of the C problems and some of the B problems can be completed by looking at the class notes; most of the B problems and almost all of the A problems will require the use of the textbook or other resources. I will never grade an A problem for homework, but I may grade a B problem. I only give letters of recommendations (tutoring center, job, transferring to four-year universities, etc.) to students who **attempt** the A problems.

C problems

Section 1.1: 7, 9, 11, 13

B problems

Section 1.1: 1,3.

A problems

Section 1.1: 5, 15