

# Chapter 1

## Basic Vocabulary

Everything's got its own lingo, and statistics is no exception. The trick is to get a handle on the lingo right from the get-go, so when it comes time to work problems, you'll pick up on cues from the wording and get going in the right direction. You can also use the terms to search quickly in the table of contents or the index of this book to find the problems you need to dive into in a flash. It's like with anything else: As soon as you understand what the language means, you immediately start feeling more comfortable.

### The Problems You'll Work On

In this chapter, you get a bird's-eye view of some of the most common terms used in statistics and, perhaps more importantly, the context in which they're used. Here's an overview:

- » The big four: population, sample, parameter, and statistic
- » The statistics terms you'll calculate, such as the mean, median, standard deviation, z-score, and percentile
- » Types of data, graphs, and distributions
- » Data analysis terms, such as confidence intervals, margin of error, and hypothesis tests

### What to Watch Out For

Pay particular attention to the following:

- » Pick out the big four in every situation; they'll follow you wherever you go.
- » Really get the idea of a distribution; it's one of the most confusing ideas in statistics, yet it's used over and over — so nail it now to avoid getting hammered later.
- » Focus not only on the terms for the statistics and analyses you'll calculate but also on their interpretation, especially in the context of a problem.

## Picking Out the Population, Sample, Parameter, and Statistic

**1–4** You're interested in knowing what percent of all households in a large city have a single woman as the head of the household. To estimate this percentage, you conduct a survey with 200 households and determine how many of these 200 are headed by a single woman.

1. In this example, what is the population?
2. In this example, what is the sample?
3. In this example, what is the parameter?
4. In this example, what is the statistic?

## Distinguishing Quantitative and Categorical Variables

**5–6** Answer the problems about quantitative and categorical variables.

5. Which of the following is an example of a quantitative variable (also known as a numerical variable)?
  - (A) the color of an automobile
  - (B) a person's state of residence
  - (C) a person's zip code
  - (D) a person's height, recorded in inches
  - (E) Choices (C) and (D)
6. Which of the following is an example of a categorical variable (also known as a qualitative variable)?
  - (A) years of schooling completed
  - (B) college major
  - (C) high-school graduate or not
  - (D) annual income (in dollars)
  - (E) Choices (B) and (C)

# Getting a Handle on Bias, Variables, and the Mean

**7–11** You're interested in the percentage of female versus male shoppers at a department store. So one Saturday morning, you place data collectors at each of the store's four entrances for three hours, and you have them record how many men and women enter the store during that time.

- 7.** Why can collecting data at the store on one Saturday morning for three hours cause bias in the data?
- (A) It assumes that Saturday shoppers represent the whole population of people who shop at the store during the week.
  - (B) It assumes that the same percentage of female shoppers shop on Saturday mornings as any other time or day of the week.
  - (C) Perhaps couples are more likely to shop together on Saturday mornings than during the rest of the week, bringing the percentage of males and females closer than during other times of the week.
  - (D) The subjects in the study weren't selected at random.
  - (E) All of these choices are true.
- 8.** Because a variable is a characteristic of each individual on which data is collected, which of the following are variables in this study?
- (A) the day you chose to collect data
  - (B) the store you chose to observe
  - (C) the gender of each shopper who comes in during the time period
  - (D) the number of men entering the store during the time period
  - (E) Choices (C) and (D)
- 9.** In this study, \_\_\_\_\_ is a categorical variable, and \_\_\_\_\_ is a quantitative variable.
- 10.** Which chart or graph would be appropriate to display the proportion of males versus females among the shoppers?
- (A) a bar graph
  - (B) a time plot
  - (C) a pie chart
  - (D) Choices (A) and (C)
  - (E) Choices (A), (B), and (C)
- 11.** How would you calculate the mean number of shoppers per hour?

## Understanding Different Statistics and Data Analysis Terms

**12–17** Answer the problems about different statistics and data analysis terms.

- 12.** Which of the following data sets has a median of 3?
- (A) 3, 3, 3, 3, 3
  - (B) 2, 5, 3, 1, 1
  - (C) 1, 2, 3, 4, 5
  - (D) 1, 2, 4, 4, 4
  - (E) Choices (A) and (C)
- 13.** Susan scores at the 90th percentile on a math exam. What does this mean?

14. You took a survey of 100 people and found that 60% of them like chocolate and 40% don't. Which of the following gives the distribution of the "chocolate versus no chocolate" variable?
- (A) a table of the results
  - (B) a pie chart of the results
  - (C) a bar graph of the results
  - (D) a sentence describing the results
  - (E) all of the above
15. Suppose that the results of an exam tell you your  $z$ -score is 0.70. What does this tell you about how well you did on the exam?
16. A national poll reports that 65% of Americans sampled approve of the president, with a margin of error of 6 percentage points. What does this mean?
17. If you want to estimate the percentage of all Americans who plan to vacation for two weeks or more this summer, what statistical technique should you use to find a range of plausible values for the true percentage?

## Using Statistical Techniques

18–19 You read a report that 60% of high-school graduates participated in sports during their high-school years.

18. You believe that the percentage of high-school graduates who played sports is higher than what was reported. What type of statistical technique do you use to see whether you're right?
19. You believe that the percentage of high-school graduates who played sports in high school is higher than what's in the report. If you do a hypothesis test to challenge the report, which of these  $p$ -values would you be happiest to get?
- (A)  $p = 0.95$
  - (B)  $p = 0.50$
  - (C)  $p = 1$
  - (D)  $p = 0.05$
  - (E)  $p = 0.001$

## Working with the Standard Deviation

20 Solve the problem about standard deviation.

20. Which data set has the highest standard deviation (without doing calculations)?
- (A) 1, 2, 3, 4
  - (B) 1, 1, 1, 4
  - (C) 1, 1, 4, 4
  - (D) 4, 4, 4, 4
  - (E) 1, 2, 2, 4