

Chapter 1 – Stats Starts Here

Section 1.1

- 1. Grocery shopping.** Discount cards at grocery stores allow the stores to collect information about the products that the customer purchases, what other products are purchased at the same time, whether or not the customer uses coupons, and the date and time that the products are purchased. This information can be linked to demographic information about the customer that was volunteered when applying for the card, such as the customer's name, address, sex, age, income level, and other variables. The grocery store chain will use that information to better market their products. This includes everything from printing out coupons at the checkout that are targeted to specific customers to deciding what television, print, or Internet advertisements to use.
- 2. Online shopping.** Amazon hopes to gain all sorts of information about customer behavior, such as how long they spend looking at a page, whether or not they read reviews by other customers, what items they ultimately buy, and what items are bought together. They can then use this information to determine which other products to suggest to customers who buy similar items, to determine which advertisements to run in the margins, and to determine which items are the most popular so these items come up first in a search.

Section 1.2

- 3. Super Bowl.** When collecting data about the Super Bowl, the games themselves are the *who*.
- 4. Nobel laureates.** Each year is a case, holding all of the information about that specific year. Therefore, the year is the *who*.

Section 1.3

- 5. Grade level.**
 - a)** If we are, for example, comparing the percentage of first-graders who can tie their own shoes to the percentage of second-graders who can tie their own shoes, grade-level is treated as categorical. It is just a way to group the students. We would use the same methods if we were comparing boys to girls or brown-eyed kids to blue-eyed kids.
 - b)** If we were studying the relationship between grade-level and height, we would be treating grade level as quantitative.

2 *Part I Exploring and Understanding Data*

6. ZIP codes.

- a) ZIP codes are categorical in the sense that they correspond to a location. The ZIP code 14850 is a standardized way of referring to Ithaca, NY.
- b) ZIP codes generally increase as the location gets further from the east coast of the United States. For example, one of the ZIP codes for the city of Boston, MA is 02101. Kansas City, MO has a ZIP code of 64101, and Seattle, WA has a ZIP code of 98101.

7. **Voters.** The response is a categorical variable.

8. **Job hunting.** The answer is a categorical variable.

9. **Medicine.** The company is studying a quantitative variable.

10. **Stress.** The researcher is studying a quantitative variable.

Chapter Exercises

11. **The News.** Answers will vary.

12. **The Internet.** Answers will vary.

13. **Gaydar.** *Who* – 40 undergraduate women. *What* – Ability to differentiate short men from tall men. *Population of interest* – All women.

14. **Hula-hoops.** *Who* – An unknown number of participants. *What* – Heart rate, oxygen consumption, and rating of perceived exertion. *Population of interest* – All people.

15. **Bicycle Safety.** *Who* – 3,500 cars. *What* – Distance from car to bicycle and whether or not the bicycle rider is wearing a helmet. *Population of interest* – All cars passing bicyclists.

16. **Investments.** *Who* – 30 similar companies. *What* – 401(k) employee participation rates (in percent). *Population of interest* – All similar companies.

17. **Honesty.** *Who* – Workers who buy coffee in an office. *What* – amount of money contributed to the collection tray. *Population of interest* – All people in honor system payment situations.

18. **Osteopenia.** *Who* – 24 patients. *What* – The effects the treatments have on osteopenia. *Population of interest* – All people with these forms of osteopenia.

19. **Not-so-diet soda.** *Who* – 515 participants in the study. *What* – candy bar consumption and waist size change. *Population of interest* – All people.

20. **Molten iron.** *Who* – 10 crankshafts at Cleveland Casting. *What* – The pouring temperature (in degrees Fahrenheit) of molten iron. *Population of interest* – All crankshafts at Cleveland Casting.

- 21. Weighing hummingbirds.** *Who* – 34 hummingbirds. *Cases* – Each hummingbird is a case. *What* – Weight, wingspan, body length, and sex. *When* – Not specified. *Where* – Not specified. *Why* – To estimate weight from easier-to-measure variables. *How* – Researchers collected data on 34 hummingbirds they were able to catch. *Variables* – Weight. *Type* – Quantitative. *Units* – Not specified. *Variable* – Wingspan. *Type* – Quantitative. *Units* – Not specified. *Variable* – Body length. *Type* – Quantitative. *Units* – Not specified. *Variable* – Sex. *Type* – Categorical.
- 22. Schools.** *Who* – Students. *What* – Age (probably in years, though perhaps in years and months), race or ethnicity, number of absences, grade level, reading score, math score, and disabilities/special needs. *When* – This information must be kept current. *Where* – Not specified. *Why* – Keeping this information is a state requirement. *How* – The information is collected and stored as part of school records. *Variables* – There are seven variables. Race or ethnicity, grade level, and disabilities/special needs are categorical variables. Number of absences, age, reading test score, and math test score are quantitative variables. *Concerns* – What tests are used to measure reading and math ability, and what are the units of measure for the tests?
- 23. Arby's menu.** *Who* – Arby's sandwiches. *What* – type of meat, number of calories (in calories), and serving size (in ounces). *When* – Not specified. *Where* – Arby's restaurants. *Why* – These data might be used to assess the nutritional value of the different sandwiches. *How* – Information was gathered from each of the sandwiches on the menu at Arby's, resulting in a census. *Variables* – There are three variables. Number of calories and serving size are quantitative variables, and type of meat is a categorical variable.
- 24. Age and party.** *Who* – 1180 Americans. *What* – Region, age (in years), political affiliation, and whether or not the person voted in the 2006 midterm Congressional election. *When* – First quarter of 2007. *Where* – United States. *Why* – The information was gathered for presentation in a Gallup public opinion poll. *How* – Phone Survey. *Variables* – There are four variables. Region, political affiliation, and whether or not the person voted in 1998 are categorical variables, and age is a quantitative variable.
- 25. Babies.** *Who* – 668 training sessions. *Cases* – Each of the 668 sessions is a case. *What* – Runner's age, type of training, runner's weight, training level, training duration, runner's improvement, runner's sex. *When* – 1998–2000. *Where* – Not specified. *Why* – Researchers were investigating the impact of training on the runner's sprint performance. *How* – Not specified exactly. *Variable* – Runner's age. *Type* – Quantitative. *Units* – Not specified, probably years. *Variable* – Duration of training. *Type* – Quantitative. *Units* – Weeks. *Variable* – Weight of the runners. *Type* – Quantitative. *Units* – Not specified, probably pounds. *Variable* – Level of training. *Type* – Categorical. *Variable* – type of training. *Type* – Categorical. *Variable* – Sex. *Type* – Categorical. *Variable* – runner's improvement. *Type* – Categorical.

4 **Part I Exploring and Understanding Data**

- 26. Flowers.** *Who* – 385 species of flowers. *What* – Date of first flowering (in days). *When* – Not specified. *Where* – Southern England. *Why* – The researchers believe that this indicates a warming of the overall climate. *How* – Not specified. *Variables* – Date of first flowering is a quantitative variable. *Concerns* – Hopefully, date of first flowering was measured in days from January 1, or some other convention, to avoid problems with leap years.
- 27. Herbal medicine.** *Who* – experiment volunteers. *What* – herbal cold remedy or sugar solution, and cold severity. *When* – Not specified. *Where* – Major pharmaceutical firm. *Why* – Scientists were testing the efficacy of an herbal compound on the severity of the common cold. *How* – The scientists set up a controlled experiment. *Variables* – There are two variables. Type of treatment (herbal or sugar solution) is categorical, and severity rating is quantitative. *Concerns* – The severity of a cold seems subjective and difficult to quantify. Also, the scientists may feel pressure to report negative findings about the herbal product.
- 28. Orchards.** *Who* – American cherry orchards. *What* – Size of orchards (in square kilometers), number of months in existence, state, varieties of cherries grown, average price per ton (probably in dollars), gross sales (probably in dollars), and percent profit. *When* – Not specified. *Where* – United States. *Why* – Business analysts hoped to provide information that would be helpful to American cherry growers. *How* – Not specified. *Variables* – There are five quantitative variables and two categorical variables. Size of orchard, number of months in existence, average price per ton, gross sales, and percent profit are quantitative variables. State and variety of cherries grown are categorical variables.
- 29. Streams.** *Who* – Streams. *What* – Name of stream, substrate of the stream (limestone, shale, or mixed), acidity of the water (measured in pH), temperature (in degrees Celsius), and BCI (unknown units). *When* – Not specified. *Where* – Upstate New York. *Why* – Research was conducted for an Ecology class. *How* – Not specified. *Variables* – There are five variables. Name and substrate of the stream are categorical variables, and acidity, temperature, and BCI are quantitative variables.
- 30. Fuel economy.** *Who* – Every model of automobile in the United States. *What* – Vehicle manufacturer, vehicle type, weight (probably in pounds), horsepower (in horsepower), and gas mileage (in miles per gallon) for city and highway driving. *When* – This information is collected currently. *Where* – United States. *Why* – The Environmental Protection Agency uses the information to track fuel economy of vehicles. *How* – The data is collected from the manufacturer of each model. *Variables* – There are six variables. City mileage, highway mileage, weight, and horsepower are quantitative variables. Manufacturer and type of car are categorical variables.

- 31. Refrigerators.** *Who* – 353 refrigerators. *What* – Brand, cost (probably in dollars), size (in cu. ft.), type, estimated annual energy cost (probably in dollars), overall rating, and repair history (in percent requiring repair over the past five years). *When* – 2013. *Where* – United States.
Why – The information was compiled to provide information to the readers of *Consumer Reports*. *How* – Not specified. *Variables* – There are 7 variables. Brand, type, and overall rating are categorical variables. Cost, size, estimated energy cost, and repair history are quantitative variables.
- 32. Walking in circles.** *Who* – 32 volunteers. *What* – Sex, height, handedness, the number of yards walked before going out of bounds, and the side of the field on which the person walked out of bounds. *When* – Not specified. *Where* – Not specified. *Why* – The researcher was interested in whether people walk in circles when lost. *How* – Data were collected by observing the people on the field, as well as by measuring and asking the participants. *Variables* – There are 5 variables. Sex, handedness, and side of the field are categorical variables. Height and number of yards walked are quantitative variables.
- 33. Kentucky Derby 2014.** *Who* – Kentucky Derby races. *What* – Year, winner, jockey, trainer, owner, and time (in minutes, seconds, and hundredths of a second). *When* – 1875 – 2013. *Where* – Churchill Downs, Louisville, Kentucky. *Why* – It is interesting to examine the trends in the Kentucky Derby. *How* – Official statistics are kept for the race each year. *Variables* – There are 6 variables. Winner, jockey, trainer and owner are categorical variables. Date and duration are quantitative variables.
- 34. Indianapolis 500 .** *Who* – Indy 500 races. *What* – Year, driver, time (in minutes, seconds, and hundredths of a second), and speed (in miles per hour). *When* – 1911 – 2013.
Where – Indianapolis, Indiana. *Why* – It is interesting to examine the trends in Indy 500 races.
How – Official statistics are kept for the race every year. *Variables* – There are 4 variables. Driver is a categorical variable. Year, time, and speed are quantitative variables.