

## Chapter 2

### Designing Observational Studies and Experiments

#### Homework 2.1

2. A sample is the part of a population from which data are collected.
4. A sampling method that consistently underemphasizes or overemphasizes some characteristic(s) of the population is said to be biased.
6.
  - a. Andrew Brannan, Roger Collins, Jerry Heidler, Warren Hill, and Darryl Scott.
  - b. County of conviction, race, age (in years), and time served (in years).
  - c. County of conviction: Laurens, Houston, Toombs, Lee, and Chatham. Race: Caucasian, African American, Caucasian, African American, and Caucasian. Age, all in years: 66, 55, 37, 54, and 31. Time served, all in years: 14, 37, 15, 23, and 7.
  - d. Andrew Brannan:  $66 - 14 = 52$  years. Roger Collins:  $55 - 37 = 18$  years. Jerry Heidler:  $37 - 15 = 22$  years. Warren Hill:  $54 - 23 = 31$  years. Darryl Scott:  $31 - 7 = 24$  years.
  - e. The youngest when convicted was Roger Collins, at age 18 years.
8.
  - a. Mars Polar Lander, Opportunity, Mars Reconnaissance Orbiter, Yinghuo-1, and Mars Orbiter Mission.
  - b. Mission, outcome, cost (in millions of dollars), and launch mass (in pounds).
  - c. Mission: lander, rover, orbiter, orbiter, and orbiter. Outcome: failure, success, success, failure, and success. Cost, all in millions of dollars: 110, 400, 720, 163, and 74. Mass, all in pounds: 640, 408, 4810, 29,100, and 2948.
  - d. Mars Polar Lander,  $110 / 640 = 0.172$ ; Opportunity,  $400 / 408 = 0.980$ ; Mars Reconnaissance Orbiter,  $720 / 4810 = 0.150$ ; Yinghuo-1,  $163 / 29,100 = 0.006$ ; Mars Orbiter Mission,  $74 / 2948 = 0.025$ ; all in millions of dollars per pound.
- e. Yinghuo-1, 0.006 million dollars per pound.
10.
  - a. The variable is whether people think the Affordable Care Act goes too far.
  - b. The sample is the 1000 likely voters who were polled.
  - c. The population is all likely voters.
12.
  - a. The variable is whether people think the United States does too much in solving the world's problems.
  - b. The sample is the 1501 adults who were surveyed.
  - c. The population is all American adults.
14.
  - a. The variable is whether people believe that police can protect them from violent crime.
  - b. The sample is the 776 Caucasians who were surveyed.
  - c. The population is all Caucasians.
16.
  - a. The variable is whether parents will limit their children's choices of college based on cost.
  - b. The sample is the 1000 parents who were surveyed.
  - c. The population is all parents with college-bound teenagers ages 16 to 18 years.
18.
  - a. The researchers were trying to answer whether simvastatin heals ulcers.
  - b. The sample is the 66 ulcer patients who were tested.
  - c. The population is all patients with ulcers.
  - d. The researchers concluded that the drug heals ulcers. The study is part of inferential statistics because it uses sample data to draw a conclusion about a population.

20. a. The researchers were trying to answer whether autistic adults are less able to process social rewards than monetary rewards.
- b. The sample is the 20 adults who were in the study.
- c. The population is all adults.
- d. The conclusion is that adults with autism are less able to process social rewards than adults without autism. It is part of inferential statistics because it uses data from a sample to make a statement about the population.
22. a. The researchers were trying to answer whether women who are more sexually confident are also more likely to achieve sexual satisfaction.
- b. The sample is the 45 women who took the online survey.
- c. The population is all women.
- d. The conclusion is that women who are more sexually confident are also more likely to achieve sexual satisfaction. It is part of inferential statistics because it draws a conclusion about a population, based on data taken from a sample.
24. **Using a TI-84:** Mariah, Rani, May, and Brenton. **Using StatCrunch:** Rani, Brenton, Kali, and Shea.
26. a. **Using a TI-84:** Samuel, Paola, Joshua, Win, and Phoebe. **Using StatCrunch:** Win, Taja, Nathan, Samuel, and Jeffrey.
- b. **Using a TI-84:**  $3 \div 5 = \frac{3}{5}$ . **Using StatCrunch:**  $2 \div 5 = \frac{2}{5}$ .
- Using this result to describe the sample is part of descriptive statistics because it does not draw conclusions about a larger group.
- c. **Using a TI-84:** Samuel, Jeffrey, Phoebe, Win, and Arnold. **Using StatCrunch:** Karen, Win, Monique, Arnold, and Jeffrey.
- d. **Using a TI-84:**  $3 \div 5 = \frac{3}{5}$ . **Using StatCrunch:**  $2 \div 5 = \frac{2}{5}$ .
- e. **Using a TI-84:** Yes. **Using StatCrunch:** Yes. However, not all randomly selected samples of size 5 will give the same results. Because of the randomness of choosing the sample, different samples could be collected.
28. a. **Using a TI-84:** Dimitrios, Aksana, Jessica, Luis, Fan, Chris, and Gauri. **Using StatCrunch:** Gauri, Chris, Aksana, Fadi, Devin, Jose, and Julia.
- b. **Using a TI-84:**  $4 \div 7 = \frac{4}{7}$ . **Using StatCrunch:**  $5 \div 7 = \frac{5}{7}$ .
- c.  $9 \div 14 = \frac{9}{14}$
- d. **(Either technology)** No. The difference between the answers in b. and c. is due to sampling error.
- e. For many random samples of size 7, the proportion of students who think it is more important to improve student success would not be the same on each sample and would not all be the same as the proportion for all 14 students. This is due to sampling error.
30. a.  $10571 \div 20329 = 0.520$
- b.  $523 \div 1000 = 0.523$
- c. No, the result from part (b) does not equal the result from part (a). This is due to sampling error.
- d. It would be inferential statistics because it draws a conclusion about a population based on data from a sample.
32. Do you access Facebook every day or not access Facebook every day?
34. Do you have a regular exercise program or not have one?

36. The method favors students who take evening classes, so it has sampling bias.
38. The wording of the question is not clear (since it asks whether they post daily and also if they like Facebook), so it has response bias.
40. Because 9 out of 12 subjects did not respond, the method has nonresponse bias. It also has response bias because an adult who neglects their children is unlikely to say “yes.”
42. Because 93% of those who were contacted did not give a response, the method has nonresponse bias.
44. The method has response bias because the scale of numbers for the response is not consistent.
46. This method has sampling bias because it favors cars that pass by during the morning rush hour.
48. The method has response bias because the question addresses more than one issue. It also has sampling bias, because it excludes people who do not watch this TV show.
50. a. The survey is likely to have nonresponse bias because participation is voluntary. It probably also has sampling bias, since it favors diners who want to complain about their experience.
- b. The survey likely has less nonresponse bias because of the incentive. It likely has less sampling bias because of the incentive. It may have more response bias, since the future discount likely improves the customer’s satisfaction with the restaurant.
52. Using samples involves less time, less money, and less labor than taking a census.
54. Answers may vary.
56. Answers may vary.
58. Sampling error refers to the random nature of the sample; nonsampling error refers to the design of the sampling process.
4. False. Convenience sampling should never be used because such samples usually do not represent the population well.
6. Cluster sampling is the method because the 40 blocks are randomly selected, but every adult resident of each block is surveyed.
8. Systematic sampling is the method because every 100<sup>th</sup> car fuel tank after the first selected tank is tested.
10. Convenience sampling is the method because the employee only surveys the Americans whom she can contact easily.
12. Stratified sampling is the method because registered voters are randomly sampled within each of three strata: Republicans, Democrats, and Independents.
14. Simple random sampling is the method because sample members are selected at random from the whole population.
16. The method is systematic sampling because the pollster surveys every 10<sup>th</sup> person after the first to be selected.
18. The method is simple random sampling because members are randomly selected from all the paying guests in the past month.
20. a.  $420 \div 50 = 8.4$ ; round down to 8.
- b. **Using a TI-84:** 4. **Using StatCrunch:** 5.
- c. **Using a TI-84:**  
 $4, 4 + 8 = 12, 12 + 8 = 20, 20 + 8 = 28,$   
 $28 + 8 = 36$
- Using StatCrunch:**  
 $5, 5 + 8 = 13, 13 + 8 = 21, 21 + 8 = 29,$   
 $29 + 8 = 37.$
22. a.  $47,756 \div 150 \approx 318.4$ ; round down to 318.
- b. **Using a TI-84:** 130. **Using StatCrunch:** 168.
- c. **Using a TI-84:**  
 $130, 130 + 318 = 448, 448 + 318 = 766,$   
 $766 + 318 = 1084, 1084 + 318 = 1402$

**Homework 2.2**

2. We should always round down when calculating  $k$  for systematic sampling.

**Using StatCrunch:**

$$168,168 + 318 = 486, 486 + 318 = 804,$$

$$804 + 318 = 1122, 1122 + 318 = 1440.$$

24. From the police department, survey  $0.62(70) = 43$  employees. From the fire department, survey  $0.29(70) = 20$  employees. From the judicial department, survey  $0.09(70) = 6$  employees. Since  $43 + 20 + 6 = 69$ , one more person should be selected at random from one of the three departments, also selected at random, to meet the goal of a sample size of 70.

26. The total number of students in the four schools is  $1936 + 1466 + 899 + 83 = 4384$ . The proportions are: Franklin High School,  $1936 \div 4384 = 0.442$ ; Centennial High School,  $1466 \div 4384 = 0.334$ ; Fred J. Page High School,  $899 \div 4384 = 0.205$ ; Middle College High School,  $83 \div 4384 = 0.019$ .

The numbers of students in the sample from each high school, respectively, are:

$$0.442(50) \approx 22; 0.334(50) \approx 17;$$

$$0.205(50) \approx 10; 0.019(50) \approx 1.$$

28. The total number of applicants to the five graduate business majors is  $85 + 368 + 109 + 90 + 83 = 735$ . The proportions are: Accounting,  $85 \div 735 = 0.116$ ; Finance,  $368 \div 735 = 0.501$ ; Information Risk and Operations Management,  $109 \div 735 = 0.148$ ; Management,  $90 \div 735 = 0.122$ ; Marketing,  $83 \div 735 = 0.113$ .

The numbers of applicants in the sample from each major, respectively, are:

$$0.116(100) \approx 12; 0.501(100) \approx 50;$$

$$0.148(100) \approx 15; 0.122(100) \approx 12;$$

$$0.113(100) \approx 11.$$

30. The proportions of each of the strata: Female undergraduate,  $10,588 \div 29,135 = 0.363$ ; female graduate,  $4475 \div 29,135 = 0.154$ ; female professional,  $1421 \div 29,135 = 0.049$ ; male undergraduate,  $7762 \div 29,135 = 0.266$ ; male graduate,  $3736 \div 29,135 = 0.128$ ; male professional,  $1153 \div 29,135 = 0.040$ .

The numbers of students in the sample from each of the strata, respectively:

$$0.363(1200) \approx 436; 0.154(1200) \approx 185;$$

$$0.049(1200) \approx 59; 0.266(1200) \approx 319;$$

$0.128(1200) \approx 154; 0.040(1200) = 48$ . Because of rounding, the sample would actually have 1201 students.

32. **Using a TI-84:** Republicans Reagan, Farnsworth, Biggs, Yarbrough, Yee; Democrats Tovar, Bedford, Bradley, McGuire.

**Using StatCrunch:** Republicans Farnsworth, Crandell, Melvin, Yee, Worsley; Democrats Bradley, Hobbs, McGuire, Gallardo.

34. The number of clusters is  $75 \div 25 = 3$ . **Using a TI-84:** Red Sox, Royals, Athletics. **Using StatCrunch:** Royals, Tigers, Indians.

36. Stratified sampling is being used, where the strata are farmers and city or suburban residents because each of the two strata is sampled separately.
38. Cluster sampling would require the least money and effort because surveying each resident on a selected block involves less travel time than a simple random sample. The city would decide on a sample size, identify a frame of all the blocks in Los Angeles, then divide the desired sample size by the smallest number of residents per block. The required number of blocks would be randomly selected, and then every resident on the selected blocks would be surveyed.
40. Simple random sampling is the best method, since Barnes & Noble® has a frame and the surveying can be done using e-mail. The company would choose a desired sample size, then randomly select that many online customers from the frame.
42. a. If each city block is treated as a cluster, the city would decide on a sample size, identify a frame of all the city blocks in Kansas City, and then divide the desired sample size by the smallest number of residents per block. That many blocks would be randomly selected, and every resident on the selected blocks would be interviewed in person.

- b. To conduct stratified sampling, the data collectors would first choose a total sample size, then identify what proportions of registered voters are Democrats, Republicans, Independents, and so on, and compute the sample size for each of the strata by multiplying the total sample size by the respective proportions. The required number for each of the strata would then be randomly selected.
  - c. Cluster sampling would be easier than stratified because the data collectors would only need to visit the selected blocks in person.
  - d. Stratified sampling would probably give better results if the sample size is small because it is more likely to get a sample that represents the whole city.
44. a. The police used systematic sampling when the traffic was heavier because they stopped every fourth car.
- b. Sampling every third and fourth car is not systematic sampling because it violates the pattern of selecting every  $k$ th person, animal, or thing.
- c. In lighter traffic, the police could have pulled over every other car. They would still be stopping two cars out of every four, but they would be using systematic sampling.
46. Answers may vary.
48. Answers may vary.
50. Answers may vary.

**Homework 2.3**

- 2. In a double-blind study, neither the individuals nor the researcher in touch with the individuals know who is in the treatment group(s) and who is in the control group.
  - 4. A lurking variable is a variable that causes both the explanatory and response variables to change during the study.
  - 6. a. The treatment groups are the second and third groups because they receive training in addition to that which the first group receives.
  - b. The study is an experiment because each participant is assigned to one of the treatment and control groups.
  - c. Random assignment means that the researchers use random sampling to decide which participants are in which groups. For example, the researchers could create a frame of all 50 older adults, randomly choose 17 of them to be in the second group, randomly choose another 17 for the third group, and assign the other 16 to the first group.
  - d. The sample is the 50 older adults in the study. The population is all older adults.
8. a. The explanatory variable is the type of training that participants did. The response variable is walking speed when an older person is performing a mental task at the same time.
- b. The researchers concluded that the training methods for improving walking speed that include both physical and mental tasks are more effective than those used in the first group when an older adult is performing a mental task at the same time. Causality can be concluded because the participants were randomly assigned to the treatment and control groups.
- c. The first group's confidence may have increased because the group's training was easier than that of the second and third groups.
10. a. It makes sense that the study is observational because the researchers cannot randomly assign anyone to have a major bone fracture.
- b. It would be unethical to randomly assign an older adult to "treatment" when that treatment requires a major bone fracture.
- c. The sample is the people whose records were studied. The population is all adults over 60.
- d. The explanatory variable is whether or not the person had a major bone fracture. The response variable is the death rate.

- e. The conclusion is that the death rate for older adults who have had a major fracture is higher than the death rate for older adults who have never had a major fracture. Only an association can be concluded because there was no random assignment to treatment or control groups.
12. a. The study is observational because there is no random assignment.
- b. Since a placebo has no proven medical effect, it would be unethical for the doctors to administer it instead of prescribing an effective remedy for an acute cough.
- c. The sample is the 241 children in the study. The population is all children with an acute cough.
- d. The researchers concluded that children who took levodropropizine recovered better from coughs than children who took other cough syrups. Only an association can be concluded because there was no random assignment to treatment and control groups.
- e. Researchers could be influenced, consciously or unconsciously, by earning a salary from the company that manufactures levodropropizine. The two researchers' disclosing that they work for the company encourages other researchers who do not work for the company to repeat the experiment and see if they get similar results.
14. a. The treatment group is the one that received a gift card plus monetary rewards based on their class work. The control group is the one that received only a gift card.
- b. The study is an experiment because the researchers randomly assigned students to the treatment and control groups.
- c. Random assignment means that the researchers chose some students at random to receive the treatment and others to be the control group.
- d. The sample is the 1019 students in the study. The population is all low-income community college students who are parents.
16. a. It would be impossible to use a placebo for a monetary reward. A participant would quickly discover whether they have real or fake money.
- b. In order to be double-blind, the participants would have to not know whether they will receive monetary rewards, which would remove the incentive to earn more credits.
- c. The explanatory variable is whether or not students would receive an additional monetary reward based on the credits they earn. The response variable is the number of credits the students earned.
- d. The researchers concluded that monetary rewards increase the number of credits earned by low-income community college students who are parents. Causality can be concluded because students were randomly assigned to the treatment and control groups.
- e. Mistakenly giving monetary rewards to some of the control group introduces a possible lurking variable; believing that they will be rewarded no matter how many classes they pass could decrease students' motivation to do well.
18. a. This is an observational study because there is no differentiation into treatment and control groups and no random assignment.
- b. The sample is the 210 motorists whose behavior was observed. The population is all motorists in Chicago.
- c. The explanatory variable is whether the crosswalk was marked or unmarked. The response variable is whether or not the motorist stopped.
- d. The researchers concluded that motorists are more likely to follow the Must Stop law at marked than at unmarked crosswalks. Only association can be concluded because there was no random assignment into treatment and control groups.

- e. We cannot assume the conclusion is also true in Prairie City because the habits of motorists could be very different in a small farm town than in a large city. Motorists are also more likely to be acquainted with the pedestrians in a small town, which could influence their behavior.

20. **Using a TI-84:** Lenovo ThinkPad X240, Dell Latitude 7440, Dell XPS 13, Acer Aspire S7.  
**Using StatCrunch:** Lenovo ThinkPad X240, Acer Aspire S7, Samsung ATIV Book 9, HP Spectre 13 Ultrabook.

This is an example of random assignment because the ultrabooks were randomly assigned to the treatment and control groups.

22. **Using a TI-84:** Palm Beach State College, Virginia Military Institute, Kean University, SUNY College at Oneonta, Angelo State University, Langston University. **Using StatCrunch:** Eastern Illinois University, Lander University, Boise State College, Langston University, Kean University, Oakland University.

This is an example of random assignment because the colleges were randomly assigned to the treatment and control groups.

24. a. Because this is an observational study, the student cannot conclude causality. There is likely to be response bias; people may overstate the amount of exercise they get. Also, motivation to stay healthy could be a lurking variable affecting whether or not people smoke (or quit smoking) and whether or not they exercise.
- b. The student could find people who currently smoke and randomly assign the smokers to one of two groups, treatment or control. The treatment group would exercise on a regular basis, while the control group would not. It would be impossible for the study to be double-blind since the participants know what treatment they have; however, it could be a blind study if the researcher(s) in contact with the participants do not know which is in each group.
26. a. Although there is an explanatory variable, there is no control group (or random assignment). This is an observational study, and causality cannot be concluded.

There is also a confounding variable since the reward is not just monetary. Students could also be motivated to write a better project for the public honor of appearing in the newspaper.

- b. The students could be randomly assigned to one of two groups, treatment or control. The control group would have the same assignment but no prize for the best project. The treatment group will be told that the best project will win a \$25 prize. It could be a blind study if the projects are graded by another teacher who does not know which group the students were in.
28. a. Although there is an explanatory variable, there is no control group (or random assignment). This is an observational study and causality cannot be concluded. There could be sampling bias and response bias because the survey is online; it could favor people with greater incomes who could afford newer cars, and voluntary response could favor people whose mileage improved after using the additive. The financial state of the respondents could also be a lurking variable, since it could influence both the ability of car owners to respond online and their ability to keep the car in good running condition.
- b. The magazine could select a random sample of drivers and randomly assign them to either a treatment group or a control group. The treatment group would be given the additive and asked to use it, while the control group would not. The gas mileage of each car would be recorded at the beginning of the study, and again a month afterwards.
30. Researchers could recruit a sample of volunteers who suffer from insomnia and randomly assign them to either a treatment or a control group. Volunteers in the treatment group would receive the drug, and volunteers in the control group would receive a sugar pill. The study could be double-blind, with one researcher labeling pill vials but not giving the code to the researchers in contact with the patients. The extent of insomnia would be measured at the beginning of the study and again a month later.

## 32 ISM: A Pathway to Introductory Statistics

32. A sample of students who have the same class with the same professor could be randomly assigned to a treatment or control group. Each student could be given a recording of the professor's lectures, but only the treatment group would be instructed to listen to the lectures while they sleep. The professor (or someone else who grades the tests) could be blind to which students are in which group.
34. With random assignment, the frame is a sample, and sampling divides the individuals into treatment group(s) and a control group. Stratified sampling does not involve treatment and control groups; it defines groups with similar characteristics (strata) that already exist in the population and creates frames for each of the strata.
36. The key difference in the designs of an experiment and an observational study is the presence of both treatment group(s) and a control group to which individuals are randomly assigned. Random assignment to one of the groups makes it possible to isolate the effects of the treatment from other factors.
38. Answers may vary.
40. The student is correct. The objects do not know whether they are in a treatment or control group.

### Chapter 2 Review Exercises

1. a. The individuals are the countries: Bahrain, Iraq, Israel, Kuwait, and Saudi Arabia.
- b. The variables are government, population (in millions), 2012 military expenditure (in billions of dollars), and oil production (in billions of barrels per day).
- c. For the variable government: monarchy, republic, republic, monarchy, and monarchy. For the variable population, all in millions: 1.3, 31.9, 7.7, 2.7, and 26.9. For 2012 military expenditure, all in billions of dollars: 0.92, 5.69, 15.54, 5.95, and 54.22. For oil production, all in billions of barrels per day: 0.05, 2.99, 0.20, 2.69, and 9.90.
- d. Bahrain,  $0.92 \div 0.05 = 18.4$ ; Iraq,  $5.69 \div 2.99 = 1.903$ ; Israel,  $15.54 \div 0.20 = 77.7$ ; Kuwait,  $5.95 \div 2.69 = 2.212$ ; Saudi Arabia,

$54.22 \div 9.90 = 5.477$ ; all in dollars per barrel per day.

- e. Israel has the greatest ratio of 2012 military expenditure to oil production, 77.7 dollars per barrel per day.
2. a. The variable is whether American adults experience a lot of happiness and enjoyment.
  - b. The sample is the 500 American adults who were telephoned.
  - c. The population is all American adults.
3. a. **Using a TI-84:** Antoine, Jacob, Ruben, Sandra, Dante, Jose. **Using StatCrunch:** Mario, Sandra, Antoine, Jacob, Alyssa, John.
  - b. **Using a TI-84:**  $3 \div 6 = \frac{1}{2}$ . **Using StatCrunch:**  $3 \div 6 = \frac{1}{2}$ . Using this result to describe the sample is part of descriptive statistics because it does not generalize the results of the sample to describe the population.
  - c. The proportion who prefer comedies is  $\frac{5}{12} = 0.417$ .
  - d. **Using a TI-84 or Using StatCrunch:** No, the sample proportion who prefer comedies does not equal the population proportion who prefer comedies. The difference is due to sampling error.
  - e. No, the proportion in each sample would not equal the proportion of all 12 students who prefer comedies. The difference is due to sampling error.
  - f. If two researchers perform the same study with different simple random samples of the same size, their inferences will not necessarily be the same because the sample data are not the same.
4. Choose the number of questions you usually ask during one hour of your prestatistics class: 0, 1, 2, 3, or more than 3.



5. The method has sampling bias; the sampling favors people who visit the militia group site.
6. The method has sampling bias, response bias, and nonresponse bias. The sampling favors people who are often in the financial district; some people may exaggerate their salary; 55 of those who were approached declined to answer.
7. Answers may vary.
8. The method is cluster sampling because the researcher selects 50 blocks at random and then surveys each adult resident of those blocks.
9. The method is simple random sampling because Human Resources creates a frame of all U.S. employees and selects at random from that frame.
10. The method is convenience sampling because the pollster only surveys people who are easy to find, without attention to any random selection.
11. The method is stratified sampling because the researchers identify two strata (people with landlines and people with cell phones), and randomly select numbers in each of the strata.
12. The method is systematic sampling because the manager surveys every eighth person leaving the store after the first person is randomly selected.
13. a.  $105,000 \div 800 = 131.25$ ; round down to 131.  
b. **Using a TI-84:** 57. **Using StatCrunch:** 47.  
c. **Using a TI-84:** 57,  $57 + 131 = 188$ ;  
 $188 + 131 = 319$ ;  $319 + 131 = 450$ ;  
 $450 + 131 = 581$ . **Using StatCrunch:** 47,  
 $47 + 131 = 178$ ;  $178 + 131 = 309$ ;  
 $309 + 131 = 440$ ;  $440 + 131 = 571$ .
14. The total number of employees (in thousands) is  $82 + 57 + 30 + 19 + 8 + 3 = 199$ . The proportions are: commercial airplanes,  $82 \div 199 \approx 0.412$ ; defense, space, and security,  $57 \div 199 \approx 0.286$ ; corporate,  $30 \div 199 \approx 0.151$ ; engineering, operations, and technology,  $19 \div 199 \approx 0.095$ ; shared services group,  $8 \div 199 \approx 0.040$ ; other,  $3 \div 199 \approx 0.015$ .  
The numbers of employees in the sample from each group, respectively, are:  $0.412(80) \approx 33$ ;  
 $0.286(80) \approx 23$ ;  $0.151(80) \approx 12$ ;  
 $0.095(80) \approx 8$ ;  $0.040(80) \approx 3$ ;  $0.015(80) \approx 1$ .
15. **Using a TI-84:** Democrats Gerratana, Ayala, and Duff; Republicans Frantz and Linares.  
**Using StatCrunch:** Democrats Looney, Stillman, and Bartolomeo; Republicans Guglielmo and Welch.
16. If the clusters are city blocks, cluster sampling would require the least time and effort. The city would create a frame of all the blocks in the city, select some at random, and then survey each resident of the selected blocks.
17. a. The treatment groups are the three groups who receive the drug; the control group is the group who receives a placebo.  
b. The study is an experiment because the researchers randomly assigned participants to one of the three treatment groups or to the control group.  
c. Random assignment means that the researchers randomly assigned patients to one of the four groups. To accomplish the random assignment, create a frame of the 560 MDD adults. For each treatment group, randomly select 140 different MDD adults to be in the group. The remaining 140 MDD adults are the control group.  
d. The sample is the 560 MDD adults in the study. The population is all adults with MDD.
18. a. The placebo could be a sugar pill.  
b. Neither the study participants nor the researchers in contact with them know which treatment (or placebo) the participants receive. One researcher could have labeled the pill vials with numbers to identify the pills, but not told the code to another researcher who was in contact with the participants.  
c. The explanatory variable is the dosage of the drug the person receives. The response variable is the person's HRSD score.

- d. The conclusion of the study is that Lu AA21004 successfully lowers MDD adults' HRSD scores. The researchers can conclude causality because adults were randomly assigned to the treatment and control groups.
- e. The researchers concluded that the drug tends to lower MDD adults' HRSD scores, but that might not mean that the drug tends to reduce depression in MDD adults.
19. a. The study is observational because mothers were not randomly assigned to the group with eating disorders or the group without eating disorders.
- b. It would be impossible to use random assignment in this study because mothers could not start having an eating disorder, or stop having an eating disorder, due to a researcher telling them to.
- c. The sample is the mothers who were observed. The population is all mothers with first-born infants.
- d. The explanatory variable is whether or not the mother has an eating disorder. The response variable is the level of negative emotions expressed toward the infants during mealtimes.
- e. The conclusion of the study is that mothers with eating disorders express more negative emotions toward their first-born infants during mealtimes than mothers without eating disorders. It only describes an association; an observational study cannot conclude causality.
20. **Using a TI-84:** Elon University, Campbell University, Wellesley College, Columbia College, University of Mount Union. **Using StatCrunch:** Nichols College, Columbia College, Mills College, Rider University, Villanova University. Yes, it is an example of random assignment because the colleges were randomly assigned to the two groups.
21. a. The coordinator did not use random assignment, so she cannot conclude causality. Also, the attendance should include a time requirement because a student who attended the math center for only five minutes once in the entire semester should not be considered a student who used the center. Motivation could be a lurking variable: students who attend the math center might be more motivated, and study harder, than other students.
- b. The coordinator could randomly assign some students to a treatment group and others to a control group. The students in the treatment group would, for example, attend the math center for one hour per weekday during the entire semester, while the students in the control group would not attend the math center. After the semester is over, the coordinator could compare the proportion of the treatment group who passed their math classes that semester with the proportion of the control group who passed their math classes that semester.
22. The company could randomly assign some bald people to a treatment group and some to a control group. The treatment group would take the drug and the control group would take a sugar pill. The study could be double-blind. The company would then measure the extent of the individuals' hair growth after 8 months.

## Chapter 2 Test

1. a. The individuals are Delaware, Hawaii, Mississippi, Texas, and Wisconsin.
- b. The variables are region, number of workers (in thousands), and number of workers in unions (in thousands).
- c. Region: East, West, South, South, and Midwest. Number of workers: 370, 549, 1040, 10,877, and 2569, all in thousands. Number of workers in unions: 38, 121, 38, 518, and 317, all in thousands.
- d. Delaware,  $38 \div 370 \approx 0.1027 = 10.3\%$ ;  
Hawaii,  $121 \div 549 \approx 0.2204 = 22.0\%$ ;  
Mississippi,  $38 \div 1040 \approx 0.0365 = 3.7\%$ ;  
Texas,  $518 \div 10,877 \approx 0.04762 = 4.8\%$ ;  
Wisconsin,  $317 \div 2569 \approx 0.1234 = 12.3\%$ .
- e. Hawaii has the largest percentage of workers in unions, 22.0%.

2.
  - a. The variable is whether an adult intends to buy wearable technology in the next 12 months.
  - b. The sample is the 2011 American adults who were surveyed.
  - c. The population is all American adults.
3. The study has response bias and nonresponse bias. The complex wording of the question may lead customers to give an answer that is not consistent with their opinion (response bias), and the nonresponse rate of 92% indicates nonresponse bias.
4. **Using a TI-84:** Jamie, Jared, Isabel, Lisa.  
**Using StatCrunch:** Jamie, Brianna, Dan, Michael.
5. The method is cluster sampling. The farmer randomly selects 8 subsections, then measures the total yield from each of those subsections.
6.
  - a.  $k = 500 \div 80 = 6.25$ ; round down to 6.
  - b. **Using a TI-84:** 1. **Using StatCrunch:** 2.
  - c. **Using a TI-84:**  $1 + 6 = 7$ ;  $7 + 6 = 13$ ;  $13 + 6 = 19$ ;  $19 + 6 = 25$ . **Using StatCrunch:**  $2, 2 + 6 = 8, 8 + 6 = 14, 14 + 6 = 20, 20 + 6 = 26$ .
7. The total number of students is given: 21,471. The proportions are: female undergraduates,  $4833 \div 21,471 \approx 0.225$ ; female graduate students,  $1792 \div 21,471 \approx 0.083$ ; male undergraduates,  $9725 \div 21,471 \approx 0.453$ ; male graduate students,  $5121 \div 21,471 \approx 0.239$ .  
  
The numbers of students in the sample from each of the strata, respectively, are:  
 $0.225(500) \approx 113$ ;  $0.083(500) \approx 42$ ;  
 $0.453(500) \approx 226$ ;  $0.239(500) \approx 119$ .
8.
  - a. The treatment groups are the 4 groups taking different drug dosages; the control group is the group receiving a placebo.
  - b. The study is an experiment because patients were randomly assigned to the treatment and control groups.
- c. Random assignment means that the researchers randomly assigned the patients to the groups. To accomplish this, create a frame of the 361 patients. Then for each of the 4 treatment groups, randomly select 72 patients to be in the group. The remaining 73 adults should be in the control group.
  - d. The sample is the 361 patients in the study. The population is all Japanese adults with type 2 diabetes.
9.
  - a. The placebo could be a sugar pill.
  - b. Neither the patients nor the researcher(s) in contact with the patients knew which patients were in each group; one researcher could have labeled the pill vials with numbers to identify the pills, but not tell the code to another researcher who was in contact with the individuals.
  - c. The explanatory variable is the dosage of the drug. The response variable is the glycated hemoglobin level.
  - d. The conclusion of the study is that the drug successfully lowers glycated hemoglobin levels in Japanese patients with type 2 diabetes. Because treatments and control were randomly assigned, the researchers can claim causality.
  - e. Researchers could be influenced, consciously or unconsciously, by earning a salary from the company that manufactures ipragliflozin. Reporting that they work for the company encourages other researchers who do not work for the company to repeat the experiment and see if they get similar results.
10.
  - a. The researcher did not use random assignment. The players who run every day may also practice basketball longer and harder than players who do not run every day; motivation may be a lurking variable.

36 ISM: A Pathway to Introductory Statistics

- b.** The researcher could randomly assign players to a treatment group and a control group. Players in the treatment group would run for one hour every day, and players in the control group would not run. The researcher could have an assistant oversee the training, so the researcher would be blind to which players are in which group. After the players have run daily for one month, the researcher would compare the scoring of the two groups in the next month, while the players in the treatment group continued to run daily.
- 11.** The owner could randomly assign the sales force to a treatment group and a control group. The treatment group would attend a workshop about emotions for a weekend. The control group would not attend the workshop. The owner could be blind to which employees attended the workshop. One month later, the owner would compare the monthly sales by the treatment and control groups.