

sampling saves time and money by collecting data from entire groups of individuals that are close together.

- Random sampling helps avoid bias in choosing a sample. Bias can still occur in the sampling process due to **undercoverage**, which happens when some members of the population cannot be chosen.
- The most serious errors in sample surveys, however, are ones that occur after the sample is chosen. The single biggest problem is **nonresponse**: when people can't be contacted or refuse to answer. Incorrect answers by respondents can lead to **response bias**. Finally, the wording of questions has a big influence on the answers.

## 4.1 TECHNOLOGY CORNER

TI-Nspire instructions in Appendix B; HP Prime instructions on the book's Web site.

10. Choosing an SRS

page 215

## Section 4.1 Exercises

1. **Students as customers** A high school's student newspaper plans to survey local businesses about the importance of students as customers. From an alphabetical list of all local businesses, the newspaper staff chooses 150 businesses at random. Of these, 73 return the questionnaire mailed by the staff. Identify the population and the sample.
2. **Student archaeologists** An archaeological dig turns up large numbers of pottery shards, broken stone tools, and other artifacts. Students working on the project classify each artifact and assign it a number. The counts in different categories are important for understanding the site, so the project director chooses 2% of the artifacts at random and checks the students' work. Identify the population and the sample.
3. **Sampling stuffed envelopes** A large retailer prepares its customers' monthly credit card bills using an automatic machine that folds the bills, stuffs them into envelopes, and seals the envelopes for mailing. Are the envelopes completely sealed? Inspectors choose 40 envelopes at random from the 1000 stuffed each hour for visual inspection. Identify the population and the sample.
4. **Customer satisfaction** A department store mails a customer satisfaction survey to people who make credit card purchases at the store. This month, 45,000 people made credit card purchases. Surveys are mailed to 1000 of these people, chosen at random, and 137 people return the survey form. Identify the population and the sample.
5. **Call the shots** An advertisement for an upcoming TV show asked: "Should handgun control be tougher? You call the shots in a special call-in poll tonight. If yes, call 1-900-720-6181. If no, call 1-900-720-6182. Charge is 50 cents for the first minute." Over 90% of people who called in said "Yes." Explain why this opinion poll is almost certainly biased.
6. **Explain it to the congresswoman** You are on the staff of a member of Congress who is considering a bill that would provide government-sponsored insurance for nursing-home care. You report that 1128 letters have been received on the issue, of which 871 oppose the legislation. "I'm surprised that most of my constituents oppose the bill. I thought it would be quite popular," says the congresswoman. Are you convinced that a majority of the voters oppose the bill? How would you explain the statistical issue to the congresswoman?
7. **Instant opinion** A recent online poll posed the question "Should female athletes be paid the same as men for the work they do?" In all, 13,147 (44%) said "Yes," 15,182 (50%) said "No," and the

remaining 1448 said “Don’t know.” In spite of the large sample size for this survey, we can’t trust the result. Why not?

8. **Sampling at the mall** You have probably seen the mall interviewer, approaching people passing by with clipboard in hand. Explain why even a large sample of mall shoppers would not provide a trustworthy estimate of the current unemployment rate.

9. **Sleepless nights** How much sleep do high school students get on a typical school night? An interested student designed a survey to find out. To make data collection easier, the student surveyed the first 100 students to arrive at school on a particular morning. These students reported an average of 7.2 hours of sleep on the previous night.

- (a) What type of sample did the student obtain?  
(b) Explain why this sampling method is biased. Is 7.2 hours probably higher or lower than the true average amount of sleep last night for all students at the school? Why?

10. **Online polls** In June 2008, *Parade* magazine posed the following question: “Should drivers be banned from using all cell phones?” Readers were encouraged to vote online at [parade.com](http://parade.com). The July 13, 2008, issue of *Parade* reported the results: 2407 (85%) said “Yes” and 410 (15%) said “No.”

- (a) What type of sample did the *Parade* survey obtain?  
(b) Explain why this sampling method is biased. Is 85% probably higher or lower than the true percent of all adults who believe that cell phone use while driving should be banned? Why?

11. **Do you trust the Internet?** You want to ask a sample of high school students the question “How much do you trust information about health that you find on the Internet—a great deal, somewhat, not much, or not at all?” You try out this and other questions on a pilot group of 5 students chosen from your class. The class members are listed below.

- (a) Explain how you would use a line of Table D to choose an SRS of 5 students from the following list. Explain your method clearly enough for a classmate to obtain your results.  
(b) Use line 107 to select the sample. Show how you use each of the digits.

Anderson	Deng	Glaus	Nguyen	Samuels
Arroyo	De Ramos	Helling	Palmiero	Shen
Batista	Drasin	Husain	Percival	Tse
Bell	Eckstein	Johnson	Prince	Velasco
Burke	Fernandez	Kim	Puri	Wallace
Cabrera	Fullmer	Molina	Richards	Washburn
Calloway	Gandhi	Morgan	Rider	Zabidi
Delluci	Garcia	Murphy	Rodriguez	Zhao

12. **Apartment living** You are planning a report on apartment living in a college town. You decide to select three apartment complexes at random for in-depth interviews with residents.

- (a) Explain how you would use a line of Table D to choose an SRS of 3 complexes from the list below. Explain your method clearly enough for a classmate to obtain your results.  
(b) Use line 117 to select the sample. Show how you use each of the digits.

Ashley Oaks	Chauncey Village	Franklin Park	Richfield
Bay Pointe	Country Squire	Georgetown	Sagamore Ridge
Beau Jardin	Country View	Greenacres	Salem Courthouse
Bluffs	Country Villa	Lahr House	Village Manor
Brandon Place	Crestview	Mayfair Village	Waterford Court
Briarwood	Del-Lynn	Nobb Hill	Williamsburg
Brownstone	Fairington	Pemberly Courts	
Burberry	Fairway Knolls	Peppermill	
Cambridge	Fowler	Pheasant Run	

13. **Sampling the forest** To gather data on a 1200-acre pine forest in Louisiana, the U.S. Forest Service laid a grid of 1410 equally spaced circular plots over a map of the forest. A ground survey visited a sample of 10% of these plots.<sup>13</sup>

- (a) Explain how you would use your calculator or Table D to choose an SRS of 141 plots. Your description should be clear enough for a classmate to carry out your plan.  
(b) Use your method from (a) to choose the first 3 plots.

14. **Sampling gravestones** The local genealogical society in Coles County, Illinois, has compiled records on all 55,914 gravestones in cemeteries in the county for the years 1825 to 1985. Historians plan to use these records to learn about African Americans in Coles County’s history. They first choose an SRS of 395 records to check their accuracy by visiting the actual gravestones.<sup>14</sup>

- (a) Explain how you would use your calculator or Table D to choose the SRS. Your description should be clear enough for a classmate to carry out your plan.  
(b) Use your method from (a) to choose the first 3 gravestones.

15. **Random digits** Which of the following statements are true of a table of random digits, and which are false? Briefly explain your answers.

- (a) There are exactly four 0s in each row of 40 digits.  
(b) Each pair of digits has chance 1/100 of being 00.  
(c) The digits 0000 can never appear as a group, because this pattern is not random.

16. **Random digits** In using Table D repeatedly to choose random samples, you should not always begin at the same place, such as line 101. Why not?

17. **iPhones** Suppose 1000 iPhones are produced at a factory today. Management would like to ensure that the phones' display screens meet their quality standards before shipping them to retail stores. Since it takes about 10 minutes to inspect an individual phone's display screen, managers decide to inspect a sample of 20 phones from the day's production.

- Explain why it would be difficult for managers to inspect an SRS of 20 iPhones that are produced today.
- An eager employee suggests that it would be easy to inspect the last 20 iPhones that were produced today. Why isn't this a good idea?
- Another employee recommends a different sampling method: Randomly choose one of the first 50 iPhones produced. Inspect that phone and every fiftieth iPhone produced afterward. (This method is known as **systematic random sampling**.) Explain carefully why this sampling method is *not* an SRS.

18. **Dead trees** On the west side of Rocky Mountain National Park, many mature pine trees are dying due to infestation by pine beetles. Scientists would like to use sampling to estimate the proportion of all pine trees in the area that have been infected.

- Explain why it wouldn't be practical for scientists to obtain an SRS in this setting.
- A possible alternative would be to use every pine tree along the park's main road as a sample. Why is this sampling method biased?
- Suppose that a more complicated random sampling plan is carried out, and that 35% of the pine trees in the sample are infested by the pine beetle. Can scientists conclude that exactly 35% of *all* the pine trees on the west side of the park are infested? Why or why not?

19. **Who goes to the convention?** A club has 30 student members and 10 faculty members. The students are

Abel	Fisher	Huber	Miranda	Reinmann
Carson	Ghosh	Jimenez	Moskowitz	Santos
Chen	Griswold	Jones	Neyman	Shaw
David	Hein	Kim	O'Brien	Thompson
Deming	Hernandez	Klotz	Pearl	Utts
Elashoff	Holland	Liu	Potter	Varga

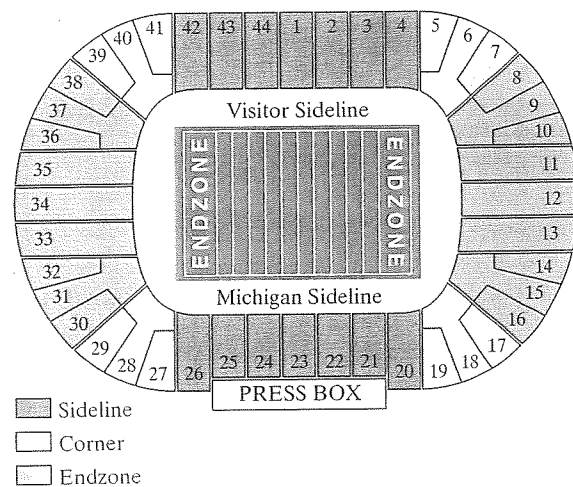
The faculty members are

Andrews	Fernandez	Kim	Moore	West
Besicovitch	Gupta	Lightman	Phillips	Yang

The club can send 4 students and 2 faculty members to a convention. It decides to choose those who will go by random selection. Describe a method for using Table D to select a stratified random sample of 4 students and 2 faculty. Then use line 123 to select the sample.

20. **Sampling by accountants** Accountants often use stratified samples during audits to verify a company's records of such things as accounts receivable. The stratification is based on the dollar amount of the item and often includes 100% sampling of the largest items. One company reports 5000 accounts receivable. Of these, 100 are in amounts over \$50,000; 500 are in amounts between \$1000 and \$50,000; and the remaining 4400 are in amounts under \$1000. Using these groups as strata, you decide to verify all the largest accounts and to sample 5% of the midsize accounts and 1% of the small accounts. Describe a method for using Table D to select a stratified random sample of the midsize and small accounts. Then use line 115 to select only the first 3 accounts from each of these strata.

21. **Go Blue!** Michigan Stadium, also known as "The Big House," seats over 100,000 fans for a football game. The University of Michigan athletic department plans to conduct a survey about concessions that are sold during games. Tickets are most expensive for seats on the sidelines. The cheapest seats are in the end zones (where one of the authors sat as a student). A map of the stadium is shown.



- The athletic department is considering a stratified random sample. What would you recommend as the strata? Why?
  - Explain why a cluster sample might be easier to obtain. What would you recommend for the clusters? Why?
22. **How was your stay?** A hotel has 30 floors with 40 rooms per floor. The rooms on one side of the hotel face the water, while rooms on the other side face a golf course. There is an extra charge for the rooms with a water view. The hotel manager wants to survey 120 guests who stayed at the hotel during a convention about their overall satisfaction with the property.

- (a) Explain why choosing a stratified random sample might be preferable to an SRS in this case. What would you use as strata?
  - (b) Why might a cluster sample be a simpler option? What would you use as clusters?
23. **Is it an SRS?** A corporation employs 2000 male and 500 female engineers. A stratified random sample of 200 male and 50 female engineers gives each engineer 1 chance in 10 to be chosen. This sample design gives every individual in the population the same chance to be chosen for the sample. Is it an SRS? Explain your answer.
24. **Attitudes toward alcohol** At a party there are 30 students over age 21 and 20 students under age 21. You choose at random 3 of those over 21 and separately choose at random 2 of those under 21 to interview about attitudes toward alcohol. You have given every student at the party the same chance to be interviewed. Why is your sample not an SRS?
25. **High-speed Internet** Laying fiber-optic cable is expensive. Cable companies want to make sure that if they extend their lines out to less dense suburban or rural areas, there will be sufficient demand and the work will be cost-effective. They decide to conduct a survey to determine the proportion of households in a rural subdivision that would buy the service. They select a simple random sample of 5 blocks in the subdivision and survey each family that lives on one of those blocks.
- (a) What is the name for this kind of sampling method?
  - (b) Give a possible reason why the cable company chose this method.
26. **Timber!** A lumber company wants to estimate the proportion of trees in a large forest that are ready to be cut down. They use an aerial map to divide the forest into 200 equal-sized rectangles. Then they choose a random sample of 20 rectangles and examine every tree that's in one of those rectangles.
- (a) What is the name for this kind of sampling method?
  - (b) Give a possible reason why the lumber company chose this method.
27. **Tweet, tweet!** What proportion of students at your school use Twitter? To find out, you survey a simple random sample of students from the school roster.
- (a) Will your sample result be exactly the same as the true population proportion? Explain.
  - (b) Which would be more likely to get your sample result closer to the true population value: an SRS of 50 students or an SRS of 100 students? Explain.
28. **Far from home?** A researcher wants to estimate the average distance that students at a large community college live from campus. To find out, she surveys a simple random sample of students from the registrar's database.
- (a) Will the researcher's sample result be exactly the same as the true population mean? Explain.
  - (b) Which would be more likely to get the researcher's sample result closer to the true population value: an SRS of 100 students or an SRS of 50 students? Explain.
29. **Baseball tickets** Suppose you want to know the average amount of money spent by the fans attending opening day for the Cleveland Indians baseball season. You get permission from the team's management to conduct a survey at the stadium, but they will not allow you to bother the fans in the club seating or box seats (the most expensive seating). Using a computer, you randomly select 500 seats from the rest of the stadium. During the game, you ask the fans in those seats how much they spent that day.
- Give a reason why this survey might yield a biased result. Explain the likely direction of the bias.
30. **Rise and shine** How long before school starts do students get out of bed, on average? Administrators survey a random sample of students on each school bus one morning.
- Give a reason why this survey might yield a biased result. Explain the likely direction of the bias.
31. **Nonresponse** A survey of drivers began by randomly sampling all listed residential telephone numbers in the United States. Of 45,956 calls to these numbers, 5029 were completed. The goal of the survey was to estimate how far people drive, on average, per day.<sup>15</sup>
- (a) What was the rate of nonresponse for this sample?
  - (b) Explain how nonresponse can lead to bias in this survey. Be sure to give the direction of the bias.
32. **Ring-no-answer** A common form of nonresponse in telephone surveys is "ring-no-answer." That is, a call is made to an active number but no one answers. The Italian National Statistical Institute looked at nonresponse to a government survey of households in Italy during the periods January 1 to Easter and July 1 to August 31. All calls were made between 7 and 10 P.M., but 21.4% gave "ring-no-answer" in one period versus 41.5% "ring-no-answer" in the other period.<sup>16</sup> Which period do you think had the higher rate of no answers? Why? Explain why a high rate of nonresponse makes sample results less reliable.
33. **Running red lights** The sample described in Exercise 31 produced a list of 5024 licensed drivers. The investigators then chose an SRS of 880 of these drivers to answer questions about their driving habits. One question asked was: "Recalling the last ten traffic lights you drove through, how many of them were red when you entered the intersections?" Of the 880 respondents, 171 admitted that at least one light had been red. A practical problem with this survey is that



people may not give truthful answers. What is the likely direction of the bias? Explain.

34. **Seat belt use** A study in El Paso, Texas, looked at seat belt use by drivers. Drivers were observed at randomly chosen convenience stores. After they left their cars, they were invited to answer questions that included questions about seat belt use. In all, 75% said they always used seat belts, yet only 61.5% were wearing seat belts when they pulled into the store parking lots.<sup>17</sup> Explain the reason for the bias observed in responses to the survey. Do you expect bias in the same direction in most surveys about seat belt use?

35. **Wording bias** Comment on each of the following as a potential sample survey question. Is the question clear? Is it slanted toward a desired response?

- (a) "Some cell phone users have developed brain cancer. Should all cell phones come with a warning label explaining the danger of using cell phones?"
- (b) "Do you agree that a national system of health insurance should be favored because it would provide health insurance for everyone and would reduce administrative costs?"
- (c) "In view of escalating environmental degradation and incipient resource depletion, would you favor economic incentives for recycling of resource-intensive consumer goods?"

36. **Checking for bias** Comment on each of the following as a potential sample survey question. Is the question clear? Is it slanted toward a desired response?

- (a) Which of the following best represents your opinion on gun control?
  - 1. The government should confiscate our guns.
  - 2. We have the right to keep and bear arms.
- (b) A freeze in nuclear weapons should be favored because it would begin a much-needed process to stop everyone in the world from building nuclear weapons now and reduce the possibility of nuclear war in the future. Do you agree or disagree?

**Multiple choice:** Select the best answer for Exercises 37 to 42.

37. The Web portal AOL places opinion poll questions next to many of its news stories. Simply click your response to join the sample. One of the questions in January 2008 was "Do you plan to diet this year?" More than 30,000 people responded, with 68% saying "Yes." You can conclude that

- (a) about 68% of Americans planned to diet in 2008.
- (b) the poll used a convenience sample, so the results tell us little about the population of all adults.
- (c) the poll uses voluntary response, so the results tell us little about the population of all adults.

- (d) the sample is too small to draw any conclusion.
- (e) None of these.

38. To gather information about the validity of a new standardized test for tenth-grade students in a particular state, a random sample of 15 high schools was selected from the state. The new test was administered to every 10th-grade student in the selected high schools. What kind of sample is this?

- (a) A simple random sample
- (b) A stratified random sample
- (c) A cluster sample
- (d) A systematic random sample
- (e) A voluntary response sample

39. Your statistics class has 30 students. You want to call an SRS of 5 students from your class to ask where they use a computer for the online quizzes. You label the students 01, 02, ..., 30. You enter the table of random digits at this line:

14459 26056 31424 80371 65103 62253 22490 61181

Your SRS contains the students labeled

- (a) 14, 45, 92, 60, 56.
- (b) 14, 31, 03, 10, 22.
- (c) 14, 03, 10, 22, 22.
- (d) 14, 03, 10, 22, 06.
- (e) 14, 03, 10, 22, 11.



40. Suppose that 35% of the registered voters in a state are registered as Republicans, 40% as Democrats, and 25% as Independents. A newspaper wants to select a sample of 1000 registered voters to predict the outcome of the next election. If they randomly select 350 Republicans, randomly select 400 Democrats, and randomly select 250 Independents, did this sampling procedure result in a simple random sample of registered voters from this district?

- (a) Yes, because each registered voter had the same chance of being chosen.
- (b) Yes, because random chance was involved.
- (c) No, because not all registered voters had the same chance of being chosen.
- (d) No, because there were a different number of registered voters selected from each party.
- (e) No, because not all possible groups of 1000 registered voters had the same chance of being chosen.

41. A local news agency conducted a survey about unemployment by randomly dialing phone numbers until they had gathered responses from 1000 adults in their state. In the survey, 19% of those who responded said they were not currently employed. In reality, only 6% of the adults in the state were not currently employed



at the time of the survey. Which of the following best explains the difference in the two percentages?

- (a) The difference is due to sampling variability. We shouldn't expect the results of a random sample to match the truth about the population every time.
  - (b) The difference is due to response bias. Adults who are employed are likely to lie and say that they are unemployed.
  - (c) The difference is due to undercoverage bias. The survey included only adults and did not include teenagers who are eligible to work.
  - (d) The difference is due to nonresponse bias. Adults who are employed are less likely to be available for the sample than adults who are unemployed.
  - (e) The difference is due to voluntary response. Adults are able to volunteer as a member of the sample.
42. A simple random sample of 1200 adult Americans is selected, and each person is asked the following question: "In light of the huge national deficit, should the government at this time spend additional money to establish a national system of health insurance?" Only 39% of those responding answered "Yes." This survey
- (a) is reasonably accurate since it used a large simple random sample.
  - (b) needs to be larger since only about 24 people were drawn from each state.
  - (c) probably understates the percent of people who favor a system of national health insurance.
  - (d) is very inaccurate but neither understates nor overstates the percent of people who favor a system of national health insurance. Because simple random sampling was used, it is unbiased.
  - (e) probably overstates the percent of people who favor a system of national health insurance.
43.  **Sleep debt (3.2)** A researcher reported that the typical teenager needs 9.3 hours of sleep per night but gets only 6.3 hours.<sup>18</sup> By the end of a 5-day school week, a teenager would accumulate about 15 hours of "sleep debt." Students in a high school statistics class were skeptical, so they gathered data on the amount of sleep debt (in hours) accumulated over time (in days) by a random sample of 25 high school students. The resulting least-squares regression equation for their data is  $\text{Sleep debt} = 2.23 + 3.17(\text{days})$
- (a) Interpret the slope of the regression line in context.
  - (b) Are the students' results consistent with the researcher's report? Explain.
44.  **Internet charges (2.1)** Some Internet service providers (ISPs) charge companies based on how much bandwidth they use in a month. One method that ISPs use for calculating bandwidth is to find the 95th percentile of a company's usage based on samples of hundreds of 5-minute intervals during a month.
- (a) Explain what "95th percentile" means in this setting.
  - (b) Which would cost a company more: the 95th percentile method or a similar approach using the 98th percentile? Justify your answer.

## 4.2 Experiments

### WHAT YOU WILL LEARN

By the end of the section, you should be able to:

- Distinguish between an observational study and an experiment.
- Explain the concept of confounding and how it limits the ability to make cause-and-effect conclusions.
- Identify the experimental units, explanatory and response variables, and treatments in an experiment.
- Explain the purpose of comparison, random assignment, control, and replication in an experiment.
- Describe a completely randomized design for an experiment, including how to randomly assign treatments using slips of paper, technology, or a table of random digits.
- Describe the placebo effect and the purpose of blinding in an experiment.
- Interpret the meaning of statistically significant in the context of an experiment.
- Explain the purpose of blocking in an experiment. Describe a randomized block design or a matched pairs design for an experiment.