

SECTION 4.1

Exercises

- pg 208**
- Students as customers** A high school's student newspaper plans to survey local businesses about the importance of students as customers. From telephone book listings, the newspaper staff chooses 150 businesses at random. Of these, 73 return the questionnaire mailed by the staff. Identify the population and the sample.
 - 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.
 - 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 from the 1000 stuffed each hour for visual inspection. Identify the population and the sample.
 - 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.
 - Call the shots** A newspaper advertisement for an upcoming TV show said: "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." Explain why this opinion poll is almost certainly biased.
 - 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?
- pg 210**
- pg 213**

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?

- pg 214**
- 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. The July 13, 2008, issue of *Parade* reported the results: 2407 (85%) said "Yes" and 410 (15%) said "No."
 - What type of sample did the *Parade* survey obtain?
 - 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?
 - 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.
 - What type of sample did the student obtain?
 - 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?
 - 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.
 - 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 at top right.
 - 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.¹²
- (a) Explain how you would use technology or Table D to choose an SRS of 141 plots. Your description should be clear enough for a classmate to obtain your results.
- (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.¹³
- (a) Explain how you would use technology or Table D to choose the SRS. Your description should be clear enough for a classmate to obtain your results.
- (b) Use your method from (a) to choose the first 3 gravestones.

15. Random digits In using Table D repeatedly to choose random samples, you should not always begin at the same place, such as line 100. Why not?

16. 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.

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.

- (a) Explain why it would be difficult for managers to inspect an SRS of 20 iPhones that are produced today.
- (b) 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?
- (c) Another employee recommends inspecting every fiftieth iPhone that is produced. 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.

- (a) Explain why it wouldn't be practical for scientists to obtain an SRS in this setting.
- (b) 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?
- (c) 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 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. How will you label the two strata? Use Table D, beginning at line 123, to choose a stratified random sample of 4 students and 2 faculty members.

- 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. How would you label the two strata from which you will sample? Use Table D, starting at 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 near the field and on the sideline. The cheapest seats are high up in the end zones (where one of the authors sat as a student). A map of the stadium is shown.



- (a) The athletic department is considering a stratified random sample. What would you recommend as the strata? Why?
- (b) 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: what is the chance? 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) Suppose there are 65 blocks in the subdivision. Use technology or Table D to select 5 blocks to be sampled. Explain your method clearly.

26. Systematic random sample Sample surveys often use a systematic random sample to choose a sample of apartments in a large building or housing units in a block at the last stage of a multistage sample. Here is a description of how to choose a systematic random sample.

Suppose that we must choose 4 addresses out of 100. Because $100/4 = 25$, we can think of the list as four lists of 25 addresses. Choose 1 of the first 25 addresses at random using Table D. The sample contains this address and the addresses 25, 50, and 75 places down the list from it. If the table gives 13, for example, then the systematic random sample consists of the addresses numbered 13, 38, 63, and 88.

- (a) Use Table D to choose a systematic random sample of 5 addresses from a list of 200. Enter the table at line 120.
- (b) Like an SRS, a systematic random sample gives all individuals the same chance to be chosen. Explain why this is true. Then explain carefully why a systematic sample is not an SRS.
27. **Sampling frame** Ideally, the sampling frame in a sample survey should list every individual in the population, but in practice, this is often difficult. Suppose that a sample of households in a community is selected at random from the telephone directory. Explain how this sampling method results in undercoverage that could lead to bias.
28. **Sampling frame** Refer to the previous exercise. It is more common in telephone surveys to use random digit dialing equipment that selects the last four digits of a telephone number at random after being given the exchange (the first three digits). Explain how this sampling method results in undercoverage that could lead to bias.
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.
- (a) Provide a reason why this survey might yield a biased result.
- (b) Explain whether the reason you provided in (a) is a sampling error or a nonsampling error.
30. **What kind of error?** Which of the following are sources of *sampling error* and which are sources of *nonsampling error*? Explain your answers.
- (a) The subject lies about past drug use.
- (b) A typing error is made in recording the data.
- (c) Data are gathered by asking people to mail in a coupon printed in a newspaper.
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.¹⁴
- (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.¹⁵ 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: do you think more or fewer than 171 of the 880 respondents really ran a red light? Why?
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.¹⁶ 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. Archaeologists plan to examine a sample of 2-meter-square plots near an ancient Greek city for artifacts visible in the ground. They choose separate random samples of plots from floodplain, coast, foothills, and high hills. What kind of sample is this?

- (a) A cluster sample
- (b) A convenience sample
- (c) A simple random sample
- (d) A stratified 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 exercises. 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. (d) 14, 03, 10, 22, 06.
- (b) 14, 31, 03, 10, 22. (e) 14, 03, 10, 22, 11.
- (c) 14, 03, 10, 22, 22.

40. When we take a census, we attempt to collect data from

- (a) a stratified random sample.
- (b) every individual selected in an SRS.
- (c) every individual in the population.
- (d) a voluntary response sample.
- (e) a convenience sample.

An example of a nonresponse error that can reduce the accuracy of a sample survey is

- (a) using voluntary response to choose the sample.
- (b) using the telephone directory as the sampling frame.
- (c) interviewing people at shopping malls to obtain a sample.
- (d) variation due to chance in choosing a sample at random.
- (e) inability to contact many members 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. Since 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 average teenager needs 9.3 hours of sleep per night but gets only 6.3 hours.¹⁷ 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 $\text{Sleep debt} = 2.23 + 3.17(\text{days})$. Do the students have reason to be skeptical of the research study’s reported results? 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.