Chapter 8 Solutions

8.1. (a) The population is (all) college students. (b) The sample is the 104 students at the researcher's college.

8.2. The population is all the artifacts discovered at the dig. The sample is those artifacts (2% of the population) that are chosen for inspection.

8.3. The population is all 45,000 people who made credit card purchases. The sample is the 137 people who returned the survey form.

8.4. It is a convenience sample; she is only getting opinions from students who are at the student center at a certain time of day. This might underrepresent some group: commuters, graduate students, or nontraditional students, for example.

8.5. (a) For example, a call-in poll, or a survey form published in the campus newspaper. (b) For example, a convenience sample method such as interviewing students as they enter the student center, or as they leave the parking lot.

8.6. Number from 01 to 33 alphabetically (down the columns). With the applet: Population = 1 to 33, select a sample of size 3, then click Reset and Sample. With Table B, enter at line 117 and choose
    16=Fairington, 32=Waterford Court, 18=Fowler.
See note on page 54 about using Table B.

8.7. Number from 01 to 28 alphabetically (down the columns). With the applet: Population = 1 to 28, select a sample of size 6, then click Reset and Sample. With Table B, enter at line 139 and choose
    04=Bonds, 10=Fleming, 17=Lumumba, 19=Nguyen, 12=Gomez, 13=Gupta.
See note on page 54 about using Table B.

8.8. (a) Assign five-digit labels to each record, from 00001 to 55914. (b) With Table B, enter at line 120 and choose 35476, 39421, 04266, 35435, and 43742. See note on page 54 about using Table B.

8.9. With the election close at hand, the polling organization wants to increase the accuracy of its results. Larger samples provide better information about the population.

8.10. The sample size for the general public is larger than the sample size for Pentecostals. (Larger samples yield more information, which means more accuracy, which means a smaller margin of error.)

    Note: *We estimate that there were between 500 and 800 people in the general-public sample, and 110 to 140 in the Pentecostal sample. These numbers are based on the usual formulas for margins of error; there may have been other details about this survey (not given in the exercise) that make those formulas inappropriate.*

8.11. Label the suburban townships from 01 to 30. With Table B, enter at line 101 and choose
19–Orland, 22–Proviso, 05–Calumet, 13–Maine, 25–Riverside, 28–Thornton
Then label the Chicago townships from 1 to 8; starting on line 110, we choose
3–Lake, 8–West Chicago, 4–Lake View, 7–South Chicago
See note on page 54 about using Table B.

8.12. Label the students in each class as shown in the table below. If Table B (starting at line 122) is used to choose the samples, the students selected are those listed in the table.

Class	Labels	First five students in sample
Freshmen	0001 to 1127	0529, 0908, 0815, 0727, 1025
Sophomores	001 to 989	602, 755, 892, 330, 634
Juniors	001 to 943	184, 281, 868, 710, 350
Seniors	001 to 895	143, 367, 494, 271, 758

8.13. The higher no-answer was probably the second period—more families are likely to be gone for vacations, or to be outside enjoying the warmer weather, and so on. Nonresponse of this type might underrepresent those who are more affluent (and are able to travel). In general, high nonresponse rates always make results less reliable, because we do not know what information we are missing.

8.14. Question A brought the higher numbers in favor of a tax cut; "new government programs" has considerably less appeal than the list of specific programs given in the second wording.

8.15. Features will vary depending on the Web site. The weakness of online polls is that they rely on voluntary response.

8.16. (b) The sample is the group about which we have information.

8.17. (a) The population is the group about which we wish to learn.

8.18. (b) A voluntary response sample is typically biased, so we can draw few conclusions about the population.

8.19. (a) Each member of the population needs a three-digit label, and we need 440 of them—not 441 labels, as in (b).

8.20. (c) Take two digits at a time, and ignore unused and duplicate labels.

8.21. (a) Take four digits at a time, and ignore unused labels.

8.22. (b) The available plots are stratified by terrain.

8.23. (b) Those without phones or with unlisted numbers—39% of the population—cannot be included in the sample.

8.24. (a) Accuracy is determined by sample size. The (smaller) sample of parents is a subset of the (larger) sample of all adults.

8.25. The population is adult U.S. residents, and the sample is the 1027 adults who were interviewed.

8.26. The population is the 1000 envelopes stuffed during a given hour. The sample is the 40 envelopes selected.

8.27. Numbering from 01 to 40 alphabetically (down the columns), we enter Table B at line 117 and choose

38–Washburn 16–Garcia 32–Rodriguez 18–Helling 37–Wallace
06–Cabrera 23–Morgan 19–Husain 03–Batista 25–Nguyen
See note on page 54 about using Table B.

8.28. With the applet: Population = 1 to 341, select a sample of size 25, then click Reset and Sample. Using line 129 of Table B, the first 5 codes are 288, 229, 131, 303, and 007.

8.29. (a) Assign labels 0001 through 1410. (b) Beginning at line 105, we choose plots 0769, 1315, 0094, 0720, and 0906.

8.30. (a) Label the names from 0001 to 7500. (b) Starting on line 142 of Table B, the first 10 pharmacists are

7282, 3297, 3408, 4178, 5426, 2817, 6176, 2169, 5388, 6041
See note on page 54 about using Table B.

8.31. If one always begins at the same place, then the results would not really be random.

8.32. (a) False. Such regularity holds only in the long run. If it were true, you could look at the first 39 digits and know whether or not the 40th was a 0. (b) True. All pairs of digits (there are 100, from 00 to 99) are equally likely. (c) False. Four random digits have chance 1/10000 to be 0000, so this sequence will occasionally occur. 0000 is no more or less random than 1234 or 2718, or any other four-digit sequence.

8.33. (a) The population is (something like) adult residents of the United States. (b) The nonresponse rate is $\frac{1169}{2000} = 58.45\%$. (c) This question will likely have response bias; specifically, many people will give an inaccurate count of how many movies they have seen in the past year.

8.34. Online polls, call-in polls, and voluntary response polls in general tend to attract responses from those who have strong opinions on the subject, and therefore are often not representative of the population as a whole. On the other hand, there is no reason to believe that randomly chosen adults would overrepresent any particular group, so the responses from such a group give a more reliable picture of public opinion.

8.35. The response rate was $\frac{5029}{45,956} \doteq 0.1094$, so the nonresponse rate was $0.8906 \doteq 89.1\%$.

8.36. (a) Assign labels 0001 through 5024, enter the table at line 104, and select: 1388, 0746, 0227, 4001, and 1858.

See note on page 54 about using Table B. (b) More than 171 respondents have run red lights. We would not expect very many people to claim they *have* run red lights when they have not, but some people will deny running red lights when they have.

8.37. People likely claim to wear their seat belts because they know they should; they are embarrassed or ashamed to say that they do not always wear seat belts. Such bias is likely in most surveys about seat belt use (and similar topics).

8.38. Each student has a 10% chance: 3 of 30 over-21 students, and 2 of 20 under-21 students. This is not an SRS because not every group of 5 students can be chosen; the only possible samples are those with 3 older and 2 younger students.

8.39. Assign labels 001 through 290 to the men, and 001 through 110 to the women. Entering the table at line 130, first choose the men (174, 095, and 178), and then continue on to choose the women (019, 007, and 041). See note on page 54 about using Table B.

8.40. Sample separately in each stratum; that is, assign separate labels, then choose the first sample, then continue on in the table to choose the next sample, etc. Beginning with line 102 in Table B, we choose:

Forest type	Labels	Parcels selected
Climax 1	01 to 36	19, 27, 26, 17
Climax 2	01 to 72	09, 55, 32, 22, 69, 56, 52
Climax 3	01 to 31	13, 07, 02
Secondary	01 to 42	27, 40, 01, 18

8.41. (a) We will choose one of the first 40 at random and then the addresses 40, 80, 120, and 160 places down the list from it. Beginning on line 120, the addresses selected are 35, 75, 115, 155, and 195. (Only the first number is chosen from the table.) (b) All addresses are equally likely; each has chance 1/40 of being selected. To see this, note that each of the first 40 has chance 1/40 because one is chosen at random. But each address in the second 40 is chosen exactly when the corresponding address in the first 40 is, so each of the second 40 also has chance 1/40. And so on.

This is not an SRS because the only possible samples have exactly one address from the first 40, one address from the second 40, and so on. An SRS could contain any 5 of the 200 addresses in the population. Note that this view of systematic sampling assumes that the number in the population is a multiple of the sample size.

8.42. **(a)** This design would omit households without telephones, those with only cell phones, and those with unlisted numbers. Such households would likely be made up of poor individuals (who cannot afford a phone), those who choose not to have phones, and those who do not wish to have their phone numbers published. **(b)** Those with unlisted numbers would be included in the sampling frame when a random-digit dialer (RDD) is used. (Additionally, RDDs exclude cell phones, although students may not be aware of this fact. For a discussion of this issue, see http://www.mysterypollster.com/main/2004/10/arianna_huffing.html.)

8.43. **(a)** Random-digit dialing randomly generates a phone number and dials it; this means that it can conceivably contact any person who has a landline telephone. (Of course, it could also call businesses, fax machines, modems, etc.) **(b)** If those performing surveys always asked their questions of the person who answered the phone, then those who are more likely to answer the phone would be overrepresented in the sample. Such people may be more likely to have some particular personality trait, for example; this could introduce bias into the sample.

8.44. **(a)** The wording is clear, but will almost certainly be slanted toward a high positive response. (Would anyone hear the phrase "brain cancer" and *not* be inclined to agree that a warning label is a good idea?) **(b)** The question makes the case for a national health care system, and so will slant responses toward "yes." **(c)** This survey question is most likely to produce a response similar to: "Uhh...yes? I mean, no? I'm sorry, could you repeat the question?" (And, if the person is able to understand the question, it is slanted in favor of day-care subsidies.)

8.46. **(a)** The population is Ontario residents; the sample is the 61,239 people interviewed. **(b)** The sample size is very large, so if there were large numbers of both sexes in the sample—this is a safe assumption because we are told this is a "random sample"—these two numbers should be fairly accurate reflections of the values for the whole population.

8.47. **(a)** The sample size for Hispanics was smaller. Smaller sample sizes give less information about the population, and therefore lead to larger margins of error (with the same confidence level). **(b)** The sample size was so small, and the margin of error so large, that the results could not be viewed as an accurate reflection of the population of Cubans.