

Chapter 8 – Producing Data: Sampling

8.1 (a) The population is (all) college students who live in off-campus housing. **(b)** The sample is the 78 students who live in off-campus housing and returned the questionnaire.

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 (a) The population is all users of the software. Unless the company's market is primarily educational, the 1100 individuals (who are mostly faculty) will not represent the population. **(b)** The sample is the 186 people who completed the survey.

8.4 Because all the students surveyed are enrolled in a special senior honors class, these students may be more likely to be interested in joining the club (and more willing to pay \$35 to do so). The direction of bias is likely to overestimate the proportion of all psychology majors who are willing to pay to join this club. This is a convenience sample.

8.5 This is a biased sampling method for obtaining customer opinions on the taxi service. Those who take the time to write an online review are likely to do so because they are upset with the service they received. Customers who were satisfied with their service are not likely to write a review. The direction of bias is likely to overestimate the proportion of customers who have a negative opinion on the service.

8.6 Number from 01 to 27 alphabetically (down the columns). With the applet: Population = 1 to 27, select a sample of size 4, then click Reset and Sample. With Table B, enter at line 133 and choose 04 = Bluffs, 18 = Georgetown, 07 = Brownstone, and 13 = Del-Lynn.

8.7 Number from 01 to 24 alphabetically (down the columns). With the applet: Population = 1 to 24, select a sample of size 3, then click Reset and Sample. With Table B, enter at line 127 and choose 06 = Deis, 08 = Fernandez, and 11 = Gemayel.

8.8 (a) Assign five-digit labels to each record, from 00001 to 55914. **(b)** With Table B, enter at line 141 and choose 35964, 23822, 50842, 53372, 50232, and 44575.

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 all teens is larger than the sample size for Hispanic teens. Larger samples yield more information, which means more accuracy, which means a smaller margin of error.

8.11 Label the suburban townships from 01 to 30 alphabetically (down the columns). With Table B, enter at line 118 and choose 19 = Orland, 03 = Bloom, 25 = Riverside, and 04 = Bremen. Next, label the Chicago townships from 1 to 8 (down the columns). With Table B, enter at line 127 and choose 4 = Lake View, 3 = Lake.

8.12 Label the students in each class as shown in the table below. If Table B, starting at line 138, is used to choose the samples, the students selected are those listed in the table. (Answers may vary depending on which class rank was chosen first.)

Class	Labels	First four students in the sample
Freshman	0001 to 1127	0840, 0470, 1098, 0484
Sophomore	001 to 989	514, 472, 321, 819
Juniors	001 to 943	400, 036, 002, 428
Seniors	001 to 895	673, 596, 423, 822

8.13 (a) The population is all physicians practicing in the United States. The sample size is $n = 2379$. If the 2379 were randomly selected, we could draw conclusions, but there was too much nonresponse. **(b)** The nonresponse rate is $\frac{100,000-2379}{100,000} = 97.62\%$. We don't know the attitudes of the nonrespondents about health care reform, so the results may not be credible. **(c)** They only received 2379 responses.

8.14 Question A asks whether existing law should be overturned. Question B simply asks whether openly gay men and women should be allowed to serve. Anybody who answers "yes" to Question A would surely answer "yes" to Question B, but the converse is not true. Thus, Question A is slanted toward a more negative response on gays in the military.

8.15 (a) The sample was not randomly selected. **(b)** The true percentage is most likely lower. People who don't often spend time with NPR probably don't visit the NPR Facebook page.

8.16 (Answers will vary. One possible answer follows.) **(a)** "Should texting while driving be made illegal?" **(b)** The cell-only group would be more supportive of texting while driving, so the sample percent that favors making texting while driving illegal would decrease. This is, indeed, bias. We're likely to overestimate the percent of all adults who favor making texting while driving illegal. **(c)** If we did not take into account the number of landline phones into a residence, a residence with multiple landlines would have a higher chance of being selected for the sample.

8.17 (a) all customers who have purchased something in the last year.

8.18 (c) the 831 people who responded.

8.19 (c) 58.5%.

8.20 (a) response bias. (The sampled participants may report their diet to be healthier than it truly was.)

8.21 (b) a stratified random sample. (Plots are stratified by terrain.)

8.22 (a) 001, 002, 003, ..., 439, 440. (Each member of the population needs a three-digit label, and we need 440 of them [not 441, as in part (b)]).

8.23 (c) 04, 18, 07, 13, 02, 05. (Notice that, in part [b], “07” appears in the sample twice. Option [a] includes numbers not in the 01 to 30 range.)

8.24 (b) undercoverage.

8.25 (b) The result for the entire sample is more accurate because it comes from a larger sample. (People over 65 are a subset of the original sample.)

8.26 (a) The population is all adults, aged 18 and older, living in all 50 U.S. states and the District of Columbia. **(b)** The sample is the 1,015 adults who were interviewed.

8.27 (a) The population is all Hispanic residents of Denver. The sample is the 200 adults who answered the questions at the selected mailing addresses. **(b)** This may suffer from response bias, since the officer doing the questioning was also Hispanic.

8.28 (a) It is possible that people who are more likely to answer the phone share similar beliefs on the direction of the city. If this is the case, randomizing which adult is interviewed within the household will provide a more representative sample. **(b)** The population of the survey is all adults who live in Chicago. Both landline and cellular phones should be included, because certain subpopulations (such as older adults) may only have landline phones; whereas other subpopulations (such as young adults) may only have cellular phones. **(c)** Yes, many residents may not be reached. A resident who initially received a cell phone number while living in a different geographical region will have an exchange that is not a Chicago exchange. This could potentially bias the results. For instance, if people have a non-Chicago exchange and are currently residents, then they may have moved to Chicago because they like the city. Their opinion of the city is likely to be higher, since they voluntarily moved there. (Answers will vary.)

8.29 (a) The population of interest is adult residents of Greenville, South Carolina. The response rate is $726/2461 = 0.295$, or 29.5%. **(b)** $100(436/689) = 0.6328$, or 63.28% of those who responded to the age question are between 18 and 64. $100(401330/461299) = 0.87$, or 87% of the population is between 18 and 64. This is not surprising, since cell phone numbers were not included in the sample and younger adults are more likely to only have a cell phone and no landline. **(c)** This sampling method is likely to give biased information, since the sample is not representative of the population of interest. The bias may underestimate, since younger adults who are more likely to use the trail are underrepresented in the survey. (Answers may vary.)

8.30 (a) The accuracy is greater because the sample size was larger. **(b)** The margin of error would be larger, because the sample sizes within these groups would have been smaller. Random sampling across the entire population can yield a sample where the sample size within a minority group is very small. The purpose of oversampling is to ensure a large enough sample within each group in order to yield a smaller margin of error for each group.

8.31 Question A drew the 47%. Perhaps the other version, with the “too high” option, prompted respondents with an option that more clearly reflected their views.

8.32 (a) The sample size is $12212 + 2038 + 22721 = 36971$. **(b)** This poll may give unreliable information because it was a voluntary response poll.

8.33 The questions were worded very differently. The U.S. has the Second Amendment, which allows guns (and the U.S. has a very active National Rifle Association, which supports gun ownership). Canada has very different gun laws; gun ownership is generally forbidden. Given these facts, opinions would differ in any case (even if the questions were worded the same).

8.34 The response rate was $5029/45,956 = 0.1094$; the nonresponse rate was $0.8906 = 89.1\%$.

8.35 (a) Assign labels 0001 through 5024. Using Table B and starting at line 119, select: 1887, 2099, 4826, 3547, and 4216. **(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.

8.36 People likely claim to wear their seat belts because they know they should (and it is the law); they may be 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).

8.37 (a) $300/30000 = 0.01$; $100/10000 = 0.01$. **(b)** To be a simple random sample, *every* possible sample of size 400 must have the same chance of being selected. This is not the case, since the only possible samples are those with 300 undergraduates and 100 graduate students.

8.38 The higher “ring-no-answer” was probably the second period—more families are likely to be gone for vacation 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.

8.39 (a) How the sample was obtained can contribute to bias in the results, if the sampling is not done randomly and fairly. We need to try to avoid undercoverage, for example. **(b)** (Answers will vary.) For example, exactly how the 655 Internet users were selected is not given.

8.40 For community pharmacy, the labels will range from 00001 to 10266; for hospital or other health care facility: 0001 to 2363; for academic or government: 001 to 307; for industry: 001 to 474; and for corporate office, professional practice, or clinic: 001 to 130. For the community pharmacy sample, starting with line 125 in Table B, select pharmacists 06565, 00795, 08727, 09517, 06489, 05007, 04197, 08796, 07051, and 09547. For the

industry sample, starting at line 133, select pharmacists 457, 404, 180, 333, 020, 193, 181, 320, 161, and 337.

8.41 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 112 in Table B, we choose:

Forest type	Labels	Parcels selected
Climax 1	01 to 36	04, 11, 19, 35
Climax 2	01 to 72	27, 30, 57, 62, 56, 02, 06
Climax 3	01 to 31	08, 02, 25
Secondary	01 to 42	11, 17, 14, 29

8.42 (a) The sample size for the public is much larger, so the survey is more accurate for this group. **(b)** It's likely that people working in health-related fields have opinions about health care issues that differ from those of the public. The researchers probably want to examine this.

8.43 (a) Because $200/5 = 40$, we will choose one of the first 40 names at random. Beginning on line 127 in Table B, the addresses selected are 25, 65, 105, 145, and 185. (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 addresses 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 chosen, 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 five 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.

8.44 (a) Randomly select a number between 1 and 10 (using Table B or a random number generator). Then sample every tenth plot thereafter according to the numbering scheme. It is more effective to consider the plots numbered 1 through 10 as being numbered 0 through 9, if using the table. **(b)** Using Table B, line 131, the first digit is 0. If the plots are labeled 0 through 9, this selects plot 0, using the selection scheme from part (a). We select plots numbered 1, 11, 21, 31, 41, (Answers will vary if plots are labeled 1 through 10.)

8.45 (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 landline phones, and those who do not wish to have their phone numbers published. **(b)** Those with unlisted landline 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.)

8.46 The polls give very different results because of bias in the survey methodology. The call-in poll suffers from voluntary response bias since only people who feel very strongly about the issue will pay \$1 to participate in the survey. The results of the scientific polling organization are likely more reliable.

8.47 (a) The wording is clear, but it will almost certainly be slanted toward a high positive response. (Would anyone hear the phrase “skyrocketing gas prices” and *not* be inclined to agree that drilling for oil in Alaska is a good idea?) **(b)** The wording is clear, and it makes the case for a national health care system, so it will slant responses toward “yes.” (Note: Some students will say that making a case for a national health care system will slant the responses toward “no.”) **(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.)

8.48 (Answers will vary considerably. See the textbook for several examples.) **(a)** One example: “Should colleges do away with the ‘tenure’ system, which effectively allows lazy and incompetent faculty members to stay in highly paid, easy jobs with summers off?” **(b)** One example: “Do you favor the tenure system for college professors or not?” **(c)** One example: “Tenure protects college professors from ‘unjust’ firing for academic freedom purposes, for example. Have you personally experienced any professors who, in your opinion, overstepped the academic freedom boundaries?”

8.49 In Canada, as in many places, elected officials aren’t necessarily qualified to make decisions about statistics. In this case, the minister is terribly misguided. Critics of the proposal are worried that the sample will not be representative of the population—presumably because people who fill out the optional long-form questions will be systematically different from those who don’t. Larger samples do not address such problems of bias.

8.50 (a) They gave greater weight to polls with larger sample sizes because larger sample sizes yield a smaller margin of error. **(b)** More recent polls will represent the most current information, which is important since people’s beliefs may change over time. The population on July 7, 2016, was all likely voters of the United State presidential election. The population may change, because people can change their mind as to whether or not they will vote. **(c)** Polls with partisan sponsors likely suffer from bias that tends to favor the candidate most closely aligned with their beliefs.

8.51 and **8.52** are Web-based exercises.