

## Chapter 8 – Producing Data: Sampling

**8.1 (a)** The population is (all) college students. (Even though the political scientist only obtained a list of all 3456 undergraduates at her college, she wanted to know about all college students.) **(b)** The sample is the 104 students at the researcher's college who returned the questionnaire. Because she only has her own college to sample from, the population she can make conclusions about is students at her college.

**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.

**8.5** 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 willing to pay to join this club. This is a convenience sample.

**8.7** Number from 01 to 26 alphabetically (down the columns). With the applet: Population = 1 to 26, select a sample of size 5, then click Reset and Sample. With Table B, enter at line 141 and choose 23 = Rodriguez, 12 = Gemayel, 16 = Ippolito, 25 = Sgambellone, and 02 = Ahmadiani.

**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.

**8.16** 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 percentage that favors making texting while driving illegal would decrease. This is indeed bias. We're likely to overestimate the percentage 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** (b) the 152 voters returning the questionnaire

**8.19** (b) 5458, 0815, 0727, 1025, 6027

**8.20** (b) The poll uses voluntary response, so the results tell us little about the population of interest.

**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).

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

**8.32** **(a)** The population is (something like) adult residents of the United States. **(b)** The nonresponse rate is  $1169/2000 = 58.45\%$ . **(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.

**8.33** 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.

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

**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).

**8.43 (a)** Since  $200/5 = 40$ , we will choose one of the first 40 names at random. 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 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 1 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.