

- 4.8. (a) response error (nonsampling error)  
 (b) data entry error (nonsampling error)  
 (c) This sounds like a voluntary response sample, which is a biased sampling design (sampling error).

- 4.12. (a) The response rate is

$$\frac{1727}{41000} \approx 0.042 \quad (\text{or about } 4.2\%).$$

- (b) If this sample was a SRS, the margin of error would be

$$\text{margin of error} = \frac{1}{\sqrt{1727}} \approx \frac{1}{41.56} \approx 0.024 \quad (\text{that is, about } 2.4\%).$$

- (c) The margin of error only accounts for random sampling error. There are nonsampling errors present in this study:

- inaccurate sampling frame: parents were selected from two data bases. Parents listed in these data bases are probably more likely to an interest in their child's education. This leads to undercoverage.
- Nonresponse; see part (a).

The margin of error does not account for these two sources of nonsampling error. Finally, I have serious doubts this sample of  $n = 1727$  parents is an SRS. In this case, the formula we used to approximate the margin of error in part (b) doesn't apply.

- 4.19. (a) No one likes having things they own taken away from them, and the second statement refers to one's "right" as a US citizen. This will lead individuals to pick the second option.

- (b) The phrases "skyrocketing gasoline prices," "very small amount," and "reducing our dependence on foreign oil" are all swaying respondents to agree. A better way to ask this question is, more simply,

*Should the Alaskan wilderness be used for oil exploration?*

- (c) The word "excessive" is not needed. This will sway respondents to answer "Yes." Otherwise, I think the question is clear as is.

- (d) The word "drugs" could mean different things to different people. Technically, I "use drugs" because I drink coffee (caffeine is a drug). I don't think this is what the questioners had in mind. A better way to ask this question would be to identify precisely which drug (or drugs) are of interest to the questioner. For example,

*Do you use marijuana?*

Of course, questions like this may lead to response error. Some respondents may not feel comfortable answering honestly. Randomized response is an alternative survey method to use here.

**4.28.** (a) A stratified random sample would consist of two SRS's: one of in-state students (200) and one of out-of-state students (200). Assign each in-state student a code from 1, 2, 3, ..., 4900. We can use R's `sample` function to select the sample:

```
> in.state = seq(1,4900,1)
> sample(in.state,200,replace=F)
 [1] 1185 2300 1007 1687 3608 2548 3844 4747 4755 1868 1878 2115 2860 3914 2879 4560
[17] 3033 1516 2135 2931 4368 4390 340 3918 119 7 3686 3657 3232 4625 3239 2110
[33] 3709 1747 3114 1097 3129 1848 2546 743 2375 2405 672 2948 3245 4362 4461 3989
[49] 456 1192 3346 1056 1398 2787 2469 504 3732 3021 483 4396 373 966 4014 3808
[65] 871 2667 4192 396 2801 372 2828 4496 3407 4338 3387 1493 4823 2266 1040 1845
[81] 908 3318 2437 3053 2810 2441 1069 400 4786 237 939 3664 2679 3995 4262 4577
[97] 4436 2361 1397 2929 693 4732 1685 2921 1623 3136 4082 886 1055 963 3162 3901
[113] 3880 410 3758 4399 4850 560 3946 2766 1926 1737 3048 1385 3015 3723 3584 111
[129] 4692 2142 3476 3714 1701 3539 4652 1822 99 2835 1170 4617 894 3348 3507 2728
[145] 4729 3671 4417 4744 2552 3210 905 58 653 2093 4869 3216 4789 2250 55 836
[161] 3977 866 1064 2260 3715 4285 1814 1547 3060 1197 4730 3891 3803 2899 1469 3866
[177] 563 991 2829 1139 3509 4420 4037 4898 1460 132 609 595 4875 1704 2958 4110
[193] 3526 1582 3212 3647 1009 18 4465 2914
```

These codes would then identify which in-state students are selected for the SRS. The first five in-state students would be those whose codes are 1185, 2300, 1007, 1687, and 3608. We would then repeat this for out-of-state students. The two SRSs combined form the stratified random sample.

(b) The chance any one of the in-state students would be selected is simply

$$\frac{1}{4900} \approx 0.000204.$$

Recall that in a SRS, each individual has the same chance of being selected into the sample (equally likely). This ensures representativeness on average. Similarly, the chance any one of the out-of-state students would be selected is

$$\frac{1}{2100} \approx 0.000476.$$

**4.36.** (a) Assign each student a code: 1, 2, 3, ..., 20000. Use R's `sample` function to choose the sample of 200 students like we did in Problem 4.28 for the in-state students. This would give a **simple random sample** (SRS) of students enrolled at the university.

(b) A **systematic sample** chooses individuals differently than a SRS. With our students and codes from part (a), we could select

- the 50th student on the list,
- the 150th student on the list,
- the 250th student on the list,

and so on until we have selected 200 students in total. The last student selected would be the 19950th student on the list. The term “systematic” means we select individuals

in a pre-determined way. Once the first student is selected, taking every 100th student after that will eventually complete the sample. To make this as fair as possible, the code of the first student selected should be selected randomly among the first 100 students. Then, take every 100th student after that to complete the systematic sample.

(c) We would do this in the same way we did in Problem 4.28(a) for the in-state students and out-of-state students.

- Label each on-campus student 1, 2, 3, ..., 8000. Use R's `sample` function to select a SRS of 50 students.
- Label each off-campus student 1, 2, 3, ..., 12000. Use R's `sample` function to select a SRS of 150 students.

The two SRSs combined form the **stratified random sample**.

**4.37.** This is more of a “free-response” exercise. The exercise is valuable because it makes you think carefully about the logistics of actually carrying out a sample survey.

(a) Specifying the exact population would be done in advance. One way is to restrict attention to full-time undergraduate students only (i.e., those students who are taking 12 credit hours in the current semester). You could also include part-time students, but they are less likely to be taking lab courses (they are taking fewer hours).

(b) A simple random sample (SRS) could be used if you wanted to look at full-time undergraduate students as a whole. A stratified random sample would be an option if you wanted to first stratify the undergraduate population into different groups (or strata) beforehand, for example,

- male students, female students
- students in different colleges (e.g., CAS, Engineering and Computing, Business, Journalism, Music, etc.)

This would allow you ensure fair representation across all strata and do comparisons among them.

(c) Carrying out surveys is difficult because you have to work with people! Here are some difficulties you might encounter:

- How would you get a sampling frame? You need this to select your sample. The Registrar's Office might have a list of all full-time undergraduates who attend USC. Do you think they will give you that list? If not, where would you get it from?
- Even if you overcame the huge obstacle above (getting the frame), how would you contact the students? Email? Very few respond to email surveys. You will likely have a high nonresponse rate.
- How will you keep participants' responses confidential? This is usually mandated anytime a study involves human subjects.