

Lab 9b: Sampling

Stat 131A, Fall 2018

Learning Objectives:

- This lab assumes that you've watched Phil Stark's video from SticiGui, chapter 24.

https://www.youtube.com/watch?v=_nE-kvM2VYk&t=138s

General Instructions

- Write your solutions in an `Rmd` (R markdown) file.
 - Name this file as `lab09b-first-last.Rmd`, where `first` and `last` are your first and last names (e.g. `lab09b-gaston-sanchez.Rmd`).
 - Knit your `Rmd` file as an html document (default option).
 - Submit your `Rmd` and `html` files to bCourses, in the corresponding lab assignment.
-

Problem 1

Which research question is poorly stated because it does not specify a population?

- a. What is your commute time when you drive to school?
- b. What is the average commute time for faculty working at a community college in California?
- c. What proportion of college students take online classes?

Problem 2

Which of the following questions is stated as a cause-and-effect question?

- a. Do instructors who learn their students' names receive higher ratings on student evaluations?
- b. Is lack of health insurance linked to shorter life expectancy?
- c. Does tutoring correlate with improved performance on exams?
- d. Is maternal smoking associated with low birth weight?

Problem 3

Here is an excerpt from a study published by the Pew Research Center in 2011 titled “Teens and Digital Citizenship Survey.”

“In 2011 Teens and Digital Citizenship Survey . . . obtained telephone interviews with a nationally representative sample of 799 teens ages 12 to 17 years old and their parents living in the continental United States. A combination of landline and cellular random digit dial samples was used to represent all teens and their parents in the continental United States who have access to either a landline or cellular phone.”

The survey describes a “nationally representative sample.” Which feature of the sampling plan is necessary in order for the sample to be representative of the population? Check all that apply.

- a. A random sample, such as the random digit dials used to choose the phone numbers.
- b. A large sample, such as the 799 teens and their parents.
- c. A reputable organization conducting the study (such as the Pew Research Center).

Problem 4

In California in 2000, the ballot included an initiative to add “none of the above” to the list of options in all candidate races. Prior to the election, a Field Poll indicated that support for the measure was 10% below opposition for the measure.

The poll was a telephone survey of 1000 registered voters in California. According to the Field Poll, “The survey was completed by telephone in English and Spanish using random digit dialing methods.”

A spokesperson for the initiative was critical of the poll because only 1000 people were surveyed. He pointed out that there are 23.5 million registered voters in California.

Is his criticism valid?

- a. Yes, because it is impossible for 1,000 people to be representative of 23,500,000 people.
- b. Yes, because a telephone survey will not represent the opinions of those without telephones.
- c. No, because the sample was randomly selected.
- d. No, because he was a spokesperson for the initiative, so he is biased.

Problem 5

Refer to the previous question. Would his criticism be valid if this was a national initiative and only 1000 people were randomly selected from all registered voters in the nation? Why or why not?

Problem 6

Suppose that you want to estimate the proportion of students at your college that read the college newspaper. Which sampling method is the best for producing this estimate?

- a) Select 50 students at random from the college.
- b) Select 100 students at random from the college.
- c) Select 200 students at random from the list of students enrolled in Journalism courses.
- d) Select 300 students who pick up the school newspaper from the library.
- e) Either (a) or (b). Both are equally representative of the student population at the college.

Problem 7

If an attempt is made to include every individual from a population in a sample, then the investigation is called a census. Which of the following is an example of a census?

- a. Weigh every child at an elementary school for a study of physical fitness at the school.
- b. The U.S. Population Census.
- c. Poll students at a college by including a pop-up window in the online registration form. Students have to answer the survey questions before they can register. Try to catch the remaining students at the admissions window when they register in person.
- d. Survey parents whose children attend a local elementary school by randomly selecting 50 parents from the school's records.

Problem 8

Which of the following is an example of a census?

- a. Survey students about the college class schedule by randomly selecting 500 from the college's enrollment records.
- b. To determine the median income of all working U.S. adults, a marketing firm conducts a national survey of 50,000 randomly selected individuals across all 50 states.

- c. A large retailer wants to determine the proportion of its customers who browse for merchandise online but make the final purchase in person at a store. At the time of purchase, each cashier is required to ask the customer whether the customer shopped online for the merchandise first.
- d. To determine the proportion of the graduating class that is college ready, a local high school tests the reading comprehension of each 12th grader.

Problem 9

When conducting a survey, which of the following is a reason to use a random sample? Check all that apply.

- a. It helps make cause-and-effect connections.
- b. It helps to avoid bias.
- c. It is representative of the population.
- d. It systematically favors the correct choice over every other choice on each question.

Problem 10

Suppose that you want to estimate the proportion of students at your college that attend at least one of the college's sports events.

Which sampling plan will produce the most reliable results?

- a. Select 50 students at random from the college.
- b. Select 100 students at random from students attending one of the college's football games.
- c. Select 200 students at random from the list of student athletes.
- d. Select 300 students who purchase tickets to at least one of the college's sports events.

Problem 11

One way MSNBC solicits the opinions of its news consumers is through the organization's online Speak Out surveys. Using social media and its television and radio news shows, MSNBC encourages its news consumers to go online and respond to various opinion polls about trending topics. Typically the polls remain open. In one ongoing Speak Out poll, MSNBC asks whether marijuana should be legalized, decriminalized, or illegal. As of May 10th, 2015, more than 26,000 people had responded to the poll, and 77% indicated that marijuana should be legalized.

What can we conclude from the poll?

- a. Because the sample is large (more than 26,000), it is reasonable to conclude that approximately 77% of the public believes marijuana should be legalized in the United States.
- b. The result is surprising but reasonable because people view smoking marijuana as equivalent to drinking alcohol.
- c. This survey is biased by voluntary response and probably overestimates the percentage of the public that supports legalization of marijuana.
- d. The results are believable, so the survey is not biased.

Problem 12

You see a classmate in front of the art building occasionally stopping students and asking them questions. He happens to stop you, and you ask what he's doing. He says that he is working on an assignment for his statistics class. He is collecting student opinions about the college. However, he only has time before his art class to conduct the survey. You tell him that he should use a random sample instead of occasionally stopping students as they enter the art building.

He replies, "Why should I use a random sample? This is so much easier." Which of the following statements should you include in your explanation? Check all that apply.

- a. If you use a random sample, you are more likely to avoid bias in the survey results.
- b. In a random sample some members of the student population are less likely to be selected than others.
- c. A random sample will be representative of the student population.
- d. A random sample allows you to draw conclusions about the opinions of all students at your college.

Problem 13

You are working on an assignment for your statistics class. You need to estimate the proportion of students at your college who delay taking their first math class for at least one year. Which sampling plan will produce the most reliable results?

- a. Obtain a list of all students enrolled at the college. Randomly select 100 students from the list.
- b. Work with the math instructors to create a list of students currently taking a math class. Randomly select 200 students from the list.
- c. Visit the cafeteria on multiple days at different times. Randomly select students who are using the cafeteria until 300 are selected.

- d. Randomly select ten math classes. Survey every student in each of the ten classes.
(Sample size will be about 400.)