Module 5

Sampling

1. What does it mean for a sample to be *representative* of a population?

2. What is the best way to generate a representative sample?

3. In terms of sampling, what is *bias*?

    a. What are some examples of biased sampling methods?

4. In the 1970's the popular advice columnist Ann Landers published a letter from a young couple about to be married. The letter said, "so many of our friends seem to resent their children. They envy us and our freedom to go and come as we please. Then there's the matter of money. They say their kids keep them broke. Will you please ask your readers the question: 'If you had it to do over again – would you have children?'" Nearly 10,000 of her readers responded. 70% said "no."

    The *Good Housekeeping* magazine published the results of the poll with a sidebar that said, "All of us at *Good Housekeeping* know that no mother will be able to read Ann Landers' report without passionately agreeing or disagreeing. We would like to know what your reaction is. Won't you therefore, take a minute or two to let us know how you would answer the question: 'If you had it to do over again, would you have children?' "

The results for the *Good Housekeeping* poll were 95% "yes."

Which explanation best explains the difference in the results for these two polls?

    a. Confidentiality: in a column about this poll Ann Landers wrote, "people tell me what they wouldn't dare tell anyone else."
    b. Voluntary response: those who responded have strong opinions that may not represent the population of parents in the U.S.
    c. Convenience sampling: both surveys underrepresent the opinions of people who do not read newspapers or magazines.
    d. Sampling variability due to chance: no two samples are ever the same, so we expect results from samples to vary.

5. In response to the Ann Landers survey, the newspaper *Newsday* hired a reputable polling firm to conduct a national survey using the same question, "If you had it to do over again, would you have children?" The firm used random selection techniques to identify a sample of 1373 U.S. parents. 91% of the sample said "yes".

    Which poll gives the most reliable representation of the opinion of U.S. parents?

    a. the *Good Housekeeping* poll because it has the most believable result
    b. all are equally reliable because each result came from a survey using the same question.
    c. Ann Landers poll because the sample is the largest
    d. the *Newsday* poll because the sample is randomly selected

6. Suppose that your college administration wants to charge a $100 fee per term for a reserved parking space on campus. The administration wants to know the percentage of students at the college who would support this fee. Which sampling plan will best represent the opinions of students at your college?

   a. Stand in one of the most crowded parking lots and interview students as they get out of their cars.
   b. Choose four 8 a.m. classes at random. Survey all of the students in each of these classes.
   c. When students register online, the question appears in a pop-up window that must be answered in order to proceed with registering. If students register in person, make them answer the question before their registration is processed.
   d. Hand out a survey to all students who have a parking ticket.

How Size Affects Random Samples

Recall the spinners we studied in the lab. We asked how many times a three-color spinner would land on Blue if it were spun 45 times. In theory it should happen 15 times (33.3%) but we often got results that were either much higher (over 40%) or much lower (below 26%).

7. What were your results?

Recall that we also set the number of spins for 3,000 and recorded the percent of spins that came up Blue. The lowest percentage was above 32% and the highest was below 35%.

8. What happened to the percentage of blue spins as we increased the total number of spins?

9. 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.