Name: Key AP STATISTICS

Sampling & Surveys Exercises (Ch. 4.1)

1. A large retailer prepares its customers' monthly credit card bills using an automatic machine that folds the bills, stuffs them into envelopes, and seals the envelopes for mailing. Are the envelopes completely sealed? Inspectors choose 40 envelopes from the 1000 stuffed each hour for visual inspection. Identify the population and the sample.

   Population: All envelopes in the population Sample: 40 envelopes

2. A high school's student newspaper plans to survey local businesses about the importance of students as customers. From an alphabetical list of all businesses, the newspaper staff chooses 150 businesses at random. Of these, 73 return the questionnaire mailed by the staff. Identify the population and the sample.

   Population: All businesses in the population Sample: 73 who return the questionnaire

3. How much sleep do high school students get on a typical school night? A counselor designed a survey to find out. To make data collection easier, the counselor surveyed the first 100 students to arrive at school on a particular morning. These students reported an average of 7.2 hours of sleep on the previous night.

   a) What type of sample did the counselor obtain? convenience sample :(

   b) Explain why this sampling method is biased. Is 7.2 hours probably higher or lower than the true average amount of sleep last night for all students at the school? Why?

This is a convenience sample, which always produces a biased result. The counselor should have done a random sample.

The average (mean) of 7.2 hours may be less than the TRUE average for the whole school because these students woke up earlier to arrive at school early.

4. Many websites include customer reviews of products, restaurants, hotels, and so on. The manager of a hotel was upset to see that 26% of reviewers on a travel website gave the hotel "1 star" – the lowest possible rating. Explain how bias in the sampling method could affect the estimate.

Because this is from a website where someone volunteers to give feedback, this is called a voluntary response sample. If a person takes the time to comment or leave a review, then they probably feel very strongly about the hotel (negatively). The true percentage of all the hotel's customers would probably rate higher.

5. You are on the staff of a member of Congress who is considering a bill that would provide government-sponsored insurance for nursing-home care. You report that 1128 letters have been received on the issue, of which 871 oppose the legislation. "I'm surprised that most of my constituents oppose the bill. I thought it would be quite popular," says the congresswoman. Are you convinced that a majority of the voters oppose the bill? How would you explain the statistical issue to the congresswoman?

871/1128
$\hat{p} = .77$

NO, not convinced that a majority of the voters oppose the bill. This was also a voluntary response sample. People who are strongly against this issue will take the time to write/call in. It is likely that the true proportion of all voters who oppose this bill is less than 77%.

6. Suppose 1000 iPhones are produced at a factory today. Management would like to ensure that the phones' display screens meet their quality standards before shipping them to retail stores. Since it takes about 10 minutes to inspect an individual phone's display screen, managers decide to inspect a sample of (20) phones from the day's production.

   a) An eager employee suggests that it would be easy to inspect the last 20 iPhones that were produced today. Why isn't this a good idea?

   This is a convenience sample. The last 20 phones produced isn't necessarily a representation of all the phones.

   b) Another employee recommends inspecting every fiftieth iPhone that is produced. Explain why this sampling method is (not) and SRS.

   Although at first this seems like a good idea, it's not a random sample because every phone does not have an equal chance of being selected. This is called a "systematic" sample.

7. Laying fiber-optic cable is expensive. Cable companies want to make sure that, if they extend their lines out to less dense suburban or rural areas, there will be sufficient demand and the work will be cost-effective. They decide to conduct a survey to determine the proportion of households in a rural subdivision that would buy the service. They select a simple random sample of 5 blocks in the subdivision and survey each family that lives on one of those blocks.

What is the name for this kind of sampling method? cluster sampling.
Broke into groups - blocks in the neighborhood - and then surveyed all the families on those chosen blocks.

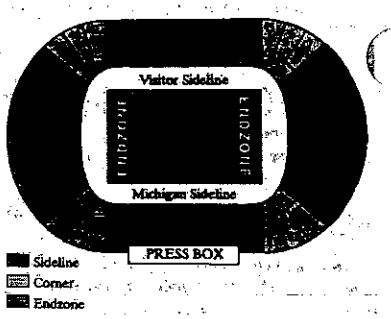
8. A hotel has 30 floors with 40 rooms per floor. The rooms on one side of the hotel face the water, while rooms on the other side face a golf course. There is an extra charge for the rooms with a water view. The hotel manager wants to select 120 rooms and survey the registered guest in each of the selected rooms about his or her overall satisfaction with the property.

30 floors
40 rooms
1200 rooms

a) Describe how to obtain a sample of 120 rooms using <u>stratified</u> random sampling. Explain your choice of strata and why this method might be preferred to simple random sampling (SRS).

b) Describe how to obtain a sample of 120 rooms using <u>cluster</u> sampling. Explain your choice of clusters and why this method might be preferred to simple random sampling (SRS).

9. Michigan Stadium, also known as "The Big House," seats over 100,000 fans for a football game. The University of Michigan athletic department plans to conduct a survey about concessions that are sold during games. Tickets are most expensive for seats near the field and on the sideline. The cheapest seats are high up in the end zones.



a) The athletic department is considering a stratified random sample. What would you recommend as the strata? Why?

"Strata" are just the groups that you break the whole population into. The most logical groups would be the sideline, corner, and endzone seats. (There are other options as well.) maybe section #'s?

b) Explain why a cluster sample might be easier to obtain. What would you recommend for the clusters? Why?

The clusters (or groups) could be the same groups that you are using for stratified... sideline, corner, endzone. A cluster sample is easier because the people in your sample are all sitting together.

10. The director of student life at a small college wants to know what percent of students eat regularly in the cafeteria. To find out, the director selects an SRS of 300 students who live in the dorms. Describe how undercoverage might lead to bias in this study. Explain the likely direction of the bias.

This is an example of undercoverage because students who do not live in the dorms cannot be part of the sample. Some students who live off campus could be less likely to eat in the cafeteria, so the estimate that the director gets is probably a higher % than if the question was asked to ALL students.

11. A total of 300 people participated in a free 12-week weight-loss course at a community health clinic. After one year, administrators emailed each of the 300 participants to see how much weight they had lost since the end of the course. Only 56 participants responded to the survey. The mean weight loss for this sample was 13.6 pounds. Describe how nonresponse might lead to bias in this study. Explain the likely direction of the bias.

The participants who did not lose much weight (or gained weight) may be reluctant to respond to the survey. This produces an average (mean) that may be larger than if everyone (all 300 participants) had given their weight loss.

12. Comment on each of the following as a potential sample survey question. Is the question clear? Is it slanted toward a desired response?

a) "Some cell phone users have developed brain cancer. Should all cell phones come with a warning label explaining the danger of using cell phones?"

This is slanted in favor of warning labels since the first statement says that some users have developed brain cancer. This is a leading question.

b) "Do you agree that a national system of health insurance should be favored because it would provide health insurance for everyone and would reduce administrative costs?"

This is also a leading question. This is slanted in favor of providing health insurance because it states that it would reduce admin. costs.

c) "In view of escalating environmental degradation and incipient resource depletion, would you favor economic incentives for recycling of resource-intensive consumer goods?"

This question is somewhat confusing to someone who does not know anything about environmental causes. The wording is a bit leading and also with too many confusing words to the general population.

C 13. A popular website places opinion poll questions next to many of its news stories. Simply click your response to join the sample. One of the questions was "Do you plan to diet this year?" More than 30,000 people responded, with 68% saying "Yes." Which of the following is true?

a) About 68% of Americans plan to diet.
b) The poll used a convenience sample, so the results tell us little about the population of all adults.
c) The poll uses voluntary response, so the results tell us little about the population of all adults.
d) The sample is too small to draw any conclusion.
e) None of these.

B 14. To gather information about the validity of a new standardized test for 10th grade students in a particular state, a random sample of 15 high schools was selected from the state. The new test was administered to every 10th grade student in the selected high schools. What kind of sample is this?

Groups? high schools in the state

a) simple random sample
b) cluster sample
c) voluntary response sample
d) stratified random sample
e) systematic random sample

D 15. When we take a **census**, we attempt to collect data from ...

a) A stratified random sample.
b) A convenience sample.
c) A voluntary response sample.
d) Every individual in the population.
e) Every individual selected in an SRS.

B 16. Suppose that 35% of the voters in a state are registered as Republicans, 40% as Democrats, and 25% as Independents. A newspaper wants to select a sample of 1000 registered voters to predict the outcome of the next election. If it randomly selects 350 Republicans, randomly selects 400 Democrats, and randomly selects 250 Independents. Did this sample procedure result in a simple random sample of registered voters from this state?

a) Yes, it was a simple random sample.
b) No, it was a stratified random sample.
c) No, it was a cluster sample.
d) No, it was a systematic random sample.
e) No, it was a convenience sample.

Groups? Dem, Rep, Ind — Then picked some from each group

17. A local news agency conducted a survey about unemployment by randomly dialing phone numbers during the work day until it gathered responses from 1000 adults in its state. In the survey, 19% of those who responded said they were not currently employed. In reality, only 6% of the adults in the state were not currently employed at the time of the survey. Which of the following best explains the difference in the two percentages?

D

a) The difference is due to sampling variability. We shouldn't expect the results of a random sample to match the truth about the population every time.
b) The difference is due to response bias. Adults who are employed are likely to lie and say that they are unemployed. what?
c) The difference is due to under-coverage bias. The survey included only adults and did not include teenagers who are eligible to work.
d) The difference is due to nonresponse bias. Adults who are employed are less likely to be available for the sample than adults who are unemployed.
e) The difference is due to voluntary response. Adults are able to volunteer as a member of the sample. This problem is NOT voluntary response.

18. A simple random sample of 1200 adult Americans is selected, and each person is asked the following question: "In light of the huge national deficit, should the government at this time spend additional money to send humans to Mars?" Only 39% of those responding answered "Yes." This survey ...

C

a) Is reasonably accurate because it used a large simple random sample.
b) Needs to be larger because only about 24 people were drawn from each state.
c) Probably understates the percent of people who favor sending humans to Mars.
d) Is very inaccurate, but neither understates nor overstates the percent of people who favor sending humans to Mars. Because simple random sampling was used, it is unbiased.
e) Probably overstates the percent of people who favor sending humans to Mars.

The wording of the question influences people to say "no".