

Summer AP Statistics Assignment ANSWER KEY:

Examples from the chapter

Example, page 208: Sampling Hardwoods and Humans

- a) population: All pieces of wood
sample: 5 pieces of wood
- b) population: All adult U.S. residents
sample: 1500 adults that responded

Example, page 210: Illegal Immigrants and Driver's Licenses

Mr. Dobbs used a voluntary response sample. Those who voted were his viewers and more likely to support his views. The 97% is probably way more than the actual percent of people in the regular population who would support this candidate.

Check Your Understanding, page 211:

1. convenience – This could lead him to think that the oranges are of better quality than they really are.
2. voluntary response – Only those who feel strongly about it will respond.

Example, page 218: Sampling at a School Assembly

- a) Simple Random Sample (SRS) – Put all seat numbers in a hat and pull out 80 of them. The people in those 80 seats must complete the survey.
Disadvantage? Certain grades could be over or under represented.
- b) Stratified Random Sample (stratified) – The students are seated in grade levels already (groups). Within each group, randomly select 20 seats to survey by putting the numbers in a hat.
Disadvantage? More complicated to administer
- c) Cluster Sample (cluster) – Use each column in the auditorium as a group. There are students from each grade in the columns. There are 40 seats in each column. Randomly select 2 whole columns to survey.
Disadvantage? May select more of one type of student and less of another type of student.

Check Your Understanding, page 219:

1. You would have to identify 200 different seats, go to those seats in the arena and find the people who are sitting there. This would take time and may not have anyone occupying the seats throughout the game.
2. Use lettered rows as the strata because each lettered row is the same distance from the court and would contain only seats with the same ticket price. All of the people in any given stratum would have paid roughly the same amount for the tickets.
3. Use the numbered sections as the clusters because they include all different seat prices. Each section contains seats with many different ticket prices so the people in a section would mirror the characteristics of the population as a whole.

Check Your Understanding, page 224:

1. a) undercoverage – sampling error
b) non-response – non-sampling error
c) convenience – sampling error
2. This question made it sound as if diapers are not that big of a problem, so fewer people will want to ban them than if it was asked differently.



Name: Key

AP STATISTICS

Sampling & Surveys Exercises (Ch. 4.1)

1. A high school's student newspaper plans to survey local businesses about the importance of students as customers. From telephone book listings, 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 local businesses Sample: 73 businesses that returned the questionnaire

2. 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 (1000) Sample: 40 envelopes chosen

3. 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?

$\hat{p} = .77$
No, not convinced that a majority of the voters oppose the bill. This was a voluntary response sample. People who are strongly against this issue will take the time to write in. It is likely that the true proportion who oppose is less than 77%.

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.

This is also 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 who would be negative to the hotel is probably less than 26%.

5. How much sleep do high school students get on a typical school night? An interested student designed a survey to find out. To make data collection easier, the student 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 student 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 bias result. The average (or mean) of 7.2 hours is probably less than the true average for the whole school because these students woke up earlier to arrive at school early.

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 also a convenience sample. The last 20 phones would not be a representation of all the phones.

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

This is called a systematic sample. (not in the textbook though)
Every phone does not have an equal chance of being selected.

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.

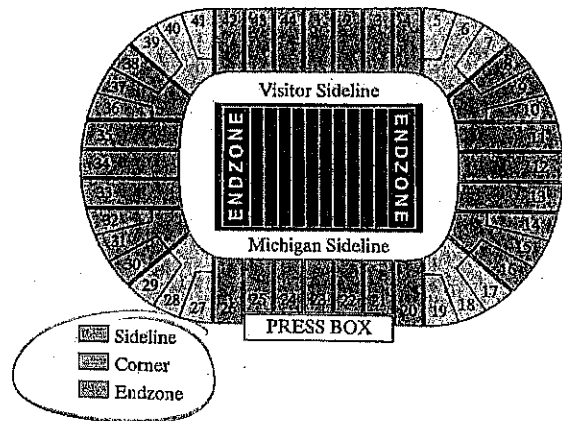
What is the name for this kind of sampling method? cluster sampling

Broke into groups (blocks in the neighborhood)
& then surveyed all the families on those chosen blocks.
This saves time & \$ for the company.

8. A corporation employs 2000 male and 500 female engineers. A stratified random sample of 200 male and 50 female engineers gives every individual in the population the same chance to be chosen for the sample. Is it an SRS? Explain.

This is not a simple random sample "SRS" because they did not start with one large group of people. The corporation broke the engineers into male & female groups first & then chose from those groups.

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 map of the stadium is shown.



- 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, & endzone seats.

- 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 tickets, corner tickets, and endzone tickets. A cluster sample is easier because the people in your sample are all sitting together.

* Also could use section #15 as "clusters"

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 under-coverage might lead to bias in this study. Explain the likely direction of the bias.

This is an example of under-coverage because students who do not live in the dorms cannot be part of the sample. Some students may live off campus and be less likely to eat in the cafeteria, so the estimate that the director gets is probably 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. Two female statistics students asked a random sample of 60 high school boys if they have ever cried during a movie. Thirty of the boys were asked directly and the other 30 were asked anonymously by means of a "secret ballot." When the responses were anonymous, 63% of the boys said "Yes," whereas only 23% of the other group said "Yes." Explain why the two percentages are so different.

When asked in person, the boys may claim that they don't cry because they are embarrassed to admit that they do. This affects the results of the survey. When given an anonymous survey, boys are more likely to get true responses.

13. The following are potential survey questions:

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

"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?" ← "leading question"

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

Comment on the errors made in these survey questions.

too much technical wording

For the first question, it is slanted in favor of warning labels because the first sentence states that some have developed brain cancer. The second question is also slanted in favor of nation health insurance by stating that it would reduce costs. The third question is not very clear (somewhat confusing to someone who knows nothing about environmental causes) and also a bit leading.

14. Survey #1:

Ideally, the sampling frame in a survey should list every individual in the population, but in practice, this is often difficult. Suppose that a sample of households in a community is selected at random from the telephone directory.

Survey #2:

Suppose you want to know the average amount of money spent by the fans attending opening day for the Cleveland Indians baseball season. You get permission from the team's management to conduct a survey at the stadium, but they will not allow you to bother the fans in the club seating or box seats (the most expensive seating). Using a computer, you randomly select 500 seats from the rest of the stadium. During the game, you ask the fans in those seats how much they spent that day.

Provide a reason why each of these surveys might yield a biased result.

Survey #1 - You are only selecting households from the telephone directory. This is under coverage because not every house hold (or person) is listed.

Survey #2 - This is also under coverage because you are leaving out a group of people. This result is biased because only the lower priced seat's opinions are included.

C 15. 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 16. 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: different 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 17. 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 18. 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?

- NO
- 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.

*Broke into groups: Rep, Dem, Ind.
Then picked some from each group.*

19. 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?

- 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. ?
- 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 is not an example of voluntary response.

20. 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 ...

- 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.

47

