Chapter 10 – Sample Surveys

Section 10.1

1. Texas A&M.

The A&M administrators should take a survey. They should sample a part of the student body, selecting respondents with a randomization method. They should be sure to draw a sufficiently large sample.

2. Satisfied workers.

They gathered data from only a part of the large population of employees, and they selected that part at random. A sample size of several hundred is a reasonable size. It's not really clear, though, whether 437 employees were contacted, or if a larger number were contacted, and only 437 completed the survey. There may be nonresponse bias in the sample.

Section 10.2

3. A&M again.

The proportion in the sample is a statistic. The proportion of all students is the parameter of interest. The statistic estimates that parameter, but is not likely to be exactly the same.

4. Satisfied respondents.

The survey result is a statistic. It estimates the true proportion of satisfied workers, but does not give that value precisely.

Section 10.3

5. Sampling students.

This is not an SRS. Although each student may have an equal chance to be in the survey, groups of friends who choose to sit together will either all be in or out of the sample, so the selection is not independent.

6. Sampling satisfaction.

Yes. Each employee has an equal and independent chance of being sampled.

Section 10.4

7. Sampling A&M students.

a) This is a cluster sample, with each selected dormitory as a cluster.

b) This is a stratified sample, stratified by class year.

c) This systematic sample, with a randomized starting point.

8. Satisfactory satisfaction samples.

a) This is a stratified sample, stratified by duration of employment.

b) This is a systematic sample, without a random starting point.

c) This is a multi-stage sample.

Section 10.6

9. Survey students.

Several terms are poorly defined. The survey needs to specify the meaning of "family" for this purpose and the meaning of "higher education." The term "seek" may also be poorly defined (for example, would applying to college but not being admitted qualify for seeking more education?)

10. Happy employees.

The survey is likely to be biased because employees won't want to express unhappiness in front of their supervisors or their co-workers. This is a form of response bias.

Section 10.7

11. Student samples.

a) This would suffer from voluntary response bias. Only those students who saw the advertisement could even be part of the sample, and only those who choose to (and are able to) go to the website would actually be in the sample.

b) This would be a convenience sample, as well as suffer from voluntary response bias.

12. Surveying employees.

a) This is a convenience sample. It would also suffer from voluntary response bias. Furthermore, it also may contain response bias. Employees may worry that bosses or other workers are watching to see who takes a survey, which may cause them to answer differently than they would have otherwise.

b) There is likely to be substantial nonresponse bias.

Chapter Exercises.

13. Roper.

a) Roper is not using a simple random sample. The samples are designed to get 500 males and 500 females. This would be very unlikely to happen in a simple random sample.

b) They are using stratified sample, with two strata, males and females.

14. Student center survey.

a) The students are not using a simple random sample. The samples are designed to get 50 students from each grade level. This would be very unlikely to happen in a simple random sample.

b) They are using a stratified sample, with four strata, one for each class year.

15. Drug tests.

a) This is a cluster sample, with teams being the clusters.

b) Cluster sampling is a reasonable solution to the problem of randomly sampling players because an entire team can be sampled easily. It would be much more difficult to randomly sample players from many different teams on the same day.

16. Gallup.

a) The population of interest is all adults in the United States aged 18 and older.

b) The sampling frame is U.S. adults with telephones.

c) Some members of the population (e.g. many college students) don't have land-line telephones, so they could never be chosen in the sample. This may create a bias.

17. Medical treatments.

a) **Population** – Unclear, but possibly all U.S. adults.

b) **Parameter** – Proportion who have used and have benefitted from alternative medical treatments.

c) **Sampling Frame** – All Consumers Union subscribers.

d) **Sample** – Those subscribers who responded.

e) **Method** – Not specified, but probably a questionnaire mailed to all subscribers.

f) **Left Out** – Those who are not Consumers Union subscribers.

g) **Bias** – Voluntary response bias. Those who respond may have strong feelings one way or another.

18. Social life.

a) **Population** – Unclear, but possibly all U.S. young adults.

b) **Parameter** – Proportion who don't or don't really have a social life outside of the Internet.

c) **Sampling Frame** – Visitors to the website. The surveyors did not have a defined sampling frame in mind when designing the survey.

d) **Sample** – Website visitors that chose to respond to the survey.

e) **Method** – Nonrandom voluntary Internet questionnaire.

f) **Left Out** – Those who did not visit gamefaqs.com and those who did visit, but did not respond to the survey.

g) **Bias** – Voluntary response bias. Participants chose whether or not to participate. Additionally, undercoverage limits the scope of any conclusions to the population of young adults, since only those that visited gamefaqs.com could have even chosen to be in the sample.

19. Mayoral race.

a) **Population** – City voters.

b) **Parameter** – Not clear. They might be interested in the percentage of voters favoring various issues.

c) **Sampling Frame** – All city residents

d) **Sample** – As many residents as they can find in one block from each district. No randomization is specified, but hopefully a block is selected at random within each district.

e) **Method** – Multistage sampling, stratified by district and clustered by block.

f) **Left Out** – People not home during the time of the survey.

g) **Bias** – Convenience sampling. Once the block is randomly chosen as the cluster, every resident living in that block should be surveyed, not just those that were conveniently available. A random sample of each block could be also be taken, but we wouldn't refer to that as "cluster" sampling, but rather multi-stage, with stratification by district, a simple random sample of one block within each district, and another simple random sample of residents within the block.

20. Soil samples.

a) **Population** – Soil around a former waste dump.

b) **Parameter** – Proportion with elevated levels of harmful substances, or perhaps a measurement of the actual levels of harmful substances.

c) **Sampling Frame** – Accessible soil around the dump.

d) **Sample** – 16 soil samples.

e) **Method** – Not clear. There is no indication of random sampling.

f) **Left Out** – Soil in inaccessible areas around the dump.

g) **Bias** – possible convenience sample. Since there is no indication of randomization, the samples may have been taken from easily accessible areas. Soil in these areas may be more or less polluted than the soil in general.

21. Roadblock.

a) **Population** – All cars.

b) **Parameter** – Proportion of cars with up-to-date (or out-of-date) registrations, insurance, or safety inspections.

c) **Sampling Frame** – Cars on that road.

d) **Sample** – Cars stopped by the roadblock.

e) **Method** – Cluster sample of an area, stopping all cars within the cluster.

f) **Left Out** – Drivers that did not take that road, or traveled that road at a different time.

g) **Bias** – Undercoverage. The cars stopped might not be representative of all cars because of time of day and location. The locations are probably not chosen randomly, so might represent areas in which it is easy to set up a roadblock, resulting in a convenience sample.

22. Snack foods.

a) **Population** – Snack food bags.

b) **Parameter** – Proportion passing inspection, or perhaps weight of bags.

c) **Sampling Frame** – All bags produced each day.

d) **Sample** – 10 bags, one from each of 10 randomly selected cases.

e) **Method** – Multistage sampling. Presumably, they take a simple random sample of 10 cases, followed by a simple random sample of one bag from each case.

f) **Left Out** – Nothing.

g) **Bias** – No indication of bias.

23. Milk samples.

a) **Population** – Dairy farms.

b) **Parameter** – Whether or not the milk contains dirt, antibiotics, or other foreign matter.

c) **Sampling Frame** – All dairy farms

d) **Sample** – Not specified, but probably a random sample of farms.

e) **Method** – not specified

f) **Left Out** – Nothing.

g) **Bias** – Unbiased, as long as the day of inspection is randomly chosen. This might not be the case, however, since the farms might be spread out over a wide geographic area. Inspectors might tend to visit farms that are near one another on the same day, resulting in a convenience sample.

24. Mistaken poll.

The station's faulty prediction is more likely to be the result of bias. Only people watching the news were able to respond, and their opinions were likely to be different from those of other voters. The sampling method may have systematically produced samples that did not represent the population of interest.

25. Another mistaken poll.

The newspaper's faulty prediction was more likely to be due to sampling error. The description of the sampling method suggests that samples should be representative of the voting population. Random chance in selecting the individuals who were polled means that sample statistics will vary from the population parameter, perhaps by quite a bit.

26. Parent opinion, part 1.

a) This is a voluntary response sample. Only those who see the ad, feel strongly about the issue, and have web access will respond.

b) This is cluster sampling, but probably not a good idea. The opinions of parents in one school may not be typical of the opinions of all parents.

c) This is an attempt at a census, and will probably suffer from nonresponse bias.

d) This is stratified sampling. If the follow-up is carried out carefully, the sample should be unbiased.

27. Parent opinion, part 2.

a) This sampling method suffers from voluntary response bias. Only those who see the show and feel strongly will call.

b) Although this method may result in a more representative sample than the method in part **a)**, this is still a voluntary response sample. Only strongly motivated parents attend PTA meetings.

c) This is multistage sampling, stratified by elementary school and then clustered by grade. This is a good design, as long as the parents in the class respond. There should be follow-up to get the opinions of parents who do not respond.

d) This is systematic sampling. As long as a starting point is randomized, this method should produce reliable data.

28. Churches.

a) This is a multistage design, with a cluster sample at the first stage and a simple random sample for each cluster.

b) If any of the three churches you pick at random are not representative of all churches, then your sample will reflect the makeup of that church, not all churches. Also, choosing 100 members at random from each church could introduce bias. The views of the members of smaller churches chosen in the sample will be weighted heavier in your sample than the views of members of larger churches, especially if the views of the members of that small church differ from the views of churchgoers at large. The hope is that random sampling will equalize these sources of variability in the long run.

29. Playground.

The managers will only get responses from people who come to the park to use the playground. Parents who are dissatisfied with the playground may not come.

30. Roller coasters.

a) This is a systematic sample.

b) The sampling frame is patrons willing to wait in line for the roller coaster. The sample should be representative of the people in line, but not of all the people at the park.

c) This sample is likely to be representative of those waiting in line for the roller coaster, especially if those people at the front of the line (after their long wait) respond differently from those at the end of the line.

d) Many people may see the long line, and choose not to go on the ride. These members of the population are excluded.

31. Playground, act two.

The first sentence points our problems that the respondent may not have noticed, and might lead them to feel they should agree. The last phrase mentions higher fees, which could make people reject improvements to the playground.

32. Wording the survey.

a) Responses to these questions will differ. Question 1 will probably get "no" answers, and Question 2 will probably get "yes" answers. This is response bias, based on the wording of the questions.

b) A question with neutral wording might be: "Do you think standardized tests are appropriate for deciding whether a student should be promoted to the next grade?"

33. Banning ephedra.

a) This is a voluntary response survey. The large sample will still be affected by any biases in the group of people that choose to respond.

b) The wording seems fair enough. It states the facts, and gives voice to both sides of the issue.

c) The sampling frame is, at best, those who visit this particular site, and even then depends of their volunteering to respond to the question.

d) This statement is true.

34. Survey questions.

a) The question is biased toward "yes" answers because of the word "pollute". A better question might be: "Should companies be responsible for any costs of environmental clean up?"

b) The question is biased toward "no" because of the preamble "18-year-olds are old enough to serve in the military. A better question might be: "Do you think the drinking age should be lowered from 21?"

35. More survey questions.

a) The question seems unbiased.

b) The question is biased toward "yes" because of the phrase "great tradition". A better question: "Do you favor continued funding for the space program?"

36. Phone surveys.

a) A simple random sample is difficult in this case because there is a problem with undercoverage. People with unlisted phone numbers and those without phones are not in the sampling frame. People who are at work, or otherwise away from home, are included in the sampling frame. These people could never be in the sample itself.

b) One possibility is to generate random phone numbers and call at random times, although obviously not in the middle of the night! This would take care of the undercoverage of people at work during the day, as well as people with unlisted phone numbers, although there is still a problem avoiding undercoverage of people without phones.

c) Under the original plan, those families in which one person stays home are more likely to be included. Under the second plan, many more are included. People without phones are still excluded.

d) Follow-up of this type greatly improves the chance that a selected household is included, increasing the reliability of the survey.

e) Random dialers allow people with unlisted phone numbers to be selected, although they may not be willing participants. There is a reason the number is unlisted. Time of day will still be an issue, as will people without phones.

37. Cell phone survey.

Cell phones are more likely to be used by younger individuals. This will result in an undercoverage bias. As cell phone use grows, this will be less of a problem. Also, many cell phone plans require the users to pay airtime for incoming calls. That seems like a sure way to irritate the respondent, and result in response bias toward negative responses.

38. Arm length.

a) Answers will vary. My arm length is 3 hand widths and 2 finger widths.

b) The parameter estimated by 10 measurements is the true length of your arm. The population is all possible measurements of your arm length.

c) The population is now the arm lengths of your friends. The average now estimates the mean of the arm lengths of your friends.

d) These 10 arm lengths are unlikely to be representative of the community, or the country. Your friends are likely to be of the same age, and not very diverse.

39. Fuel economy.

a) The statistic calculated is the mean mileage for the last six fill-ups.

b) The parameter of interest is the mean mileage for the vehicle.

c) The driving conditions for the last six fill-ups might not be typical of the overall driving conditions. For instance, the last six fill-ups might all be in winter, when mileage might be lower than expected.

d) The EPA is trying to estimate the mean gas mileage for all cars of this make, model, and year.

40. Accounting.

a) Assign numbers 001 to 120 to each order. Generate 10 random numbers between 001 and 120, and select those orders to recheck.

b) The supervisor should perform a stratified sample, randomly checking a certain percentage of each type of sales, retail and wholesale.

41. Happy workers?

a) A small sample will probably consist mostly laborers, with few supervisors, and maybe no project managers. Also, there is a potential for response bias based on the interviewer if a member of management asks directly about discontent. Workers who want to keep their jobs will likely tell the management that everything is fine!

b) Assign a number from 001 to 439 to each employee. Use a random number table or software to select the sample.

c) The simple random sample might not give a good cross section of the different types of employees. There are relatively few supervisors and project managers, and we want to make sure their opinions are noted, as well as the opinions of the laborers.

d) A better strategy would be to stratify the sample by job type. Sample a certain percentage of each job type.

e) Answers will vary. Assign each person a number from 01 to 14, and generate 2 usable random numbers from a random number table or software.

42. Quality control.

a) Select three cases at random, then select one jar randomly from each case.

b) Generate three random numbers between 61 and 80, with no repeats, to select three cases. Then assign each of the jars in the case a number between 01 and 12, and generate one random number for each case to select the three jars, one from each case.

c) This is not a simple random sample, since there are groups of three jars that cannot be the sample. For example, it is impossible for three jars in the same case to be the sample. This would be possible if the sample were a simple random sample.

43. A fish story.

What conclusions they may be able to make will depend on whether fish with discolored scales are equally likely to be caught as those without. It also depends on the level of compliance by fisherman. If fish are not equally likely to be caught, or fishermen more disposed to bring discolored fish, the results will be biased.

44. Another fish story.

No, sampling the dead fish will not give them a good estimate of the mean amount of radiation. They will only see fish that have died. Many fish may have received lower doses of radiation and survived. They won't be represented in the sample, so the mean will be biased high.

45. Sampling methods.

a) This method would probably result in undercoverage of those doctors that are not listed in the Yellow Pages. Using the "line listings" seems fair, as long as all doctors are listed, but using the advertisements would not be a typical list of doctors.

b) This method is not appropriate. This cluster sample will probably contain listings for only one or two types of businesses, not a representative cross-section of businesses.

46. More sampling methods.

a) A petition may pressure people into support. Additionally, some people may not be home on a Saturday, especially those who have taken their kids out to play in a distant park! We are undercovering a group made up of people who probably have a specific opinion.

b) If the food at the largest cafeteria is representative, this should be OK. However, those who really don't like the food won't be eating there. That group is undercovered.