

Sample Surveys Review SOLUTIONS/EXPLANATIONS – Multiple Choice questions

Correct answers are bolded. Explanations are in red.

1. Ann Landers, who wrote a daily advice column appearing in newspapers across the country, once asked her readers, "If you had to do it all over again, would you have children?" Of the more than 10,000 readers who responded, 70% said no. What does this show?
 - (A) **The survey is meaningless because of voluntary response bias.**
Explanation: Voluntary response bias is powerful – it means only people with strong opinions about the topic will respond. It will always give meaningless information for that reason, no matter what the "characteristics of her readers" include.
 - (B) No meaningful conclusion is possible without knowing something more about the characteristics of her readers. (See above)
 - (C) The survey would have been more meaningful if she had picked a random sample of the 10,000 readers who responded. This is nonsense. If she picks a random sample of the 10,000 people, these are the people who voluntarily responded. It then becomes a biased random sample and is still meaningless.
 - (D) The survey would have been more meaningful if she had used a control group. Also nonsense. How do you have a control group with a voluntary survey? "Hey, you, please DON'T take this survey."
 - (E) This was a legitimate sample, randomly drawn from her readers and of sufficient size to allow the conclusion that most of her readers who are parents would have second thoughts about having children. NOT random, hello.
2. Which of the following are true statements?
 - I. If bias is present in a sampling procedure, it can be overcome by dramatically increasing the sample size. BIG NO NO!!! If there is bias present, then taking a larger sample means you will just have an even larger biased sample.
 - II. There is no such thing as a "bad sample." There are TONS of ways to take a bad sample. Like asking students if candy and sodas should be sold in school vending machines. Their opinion is different from parents and doctors.
 - III. Sampling techniques that use probability techniques effectively eliminate bias. Wrong. There is no way to effectively eliminate bias. Researchers try very hard to REDUCE bias as much as possible, but there will always be at least some kind of bias present.
 - (A) I only
 - (B) II only
 - (C) III only
 - (D) **None of the statements are true.**
 - (E) None of the above gives the complete set of true responses.
3. Two possible wordings for a questionnaire on gun control are as follow:
 - I. The United States has the highest rate of murder by handguns among all countries. Most of these murders are known to be **crimes of passion or crimes provoked by anger** between acquaintances. Are you in favor of a 7-day cooling-off period between the filing of an application to purchase a handgun and the resulting sale?
 - II. The United States has one of the highest violent crime rates among all countries. Many people want to keep handguns in their homes for **self-protection**. Fortunately U.S. citizens are guaranteed **the right to bear arms** by the Constitution. Are you in favor of a 7-day waiting period between the filing of an application to purchase a needed handgun and the resulting sale?

One of these questions showed that 25% of the population favored a 7-day waiting period between application for purchase of a handgun and the resulting sale, while the other question showed that 70% of the population favored the waiting period. Which produced which result and why?

- (A) The first question probably showed 70% and the second question 25% because of the lack of randomization in the choice of pro-gun and anti-gun subjects as evidenced by the wording of the questions.
- (B) The first question probably showed 25% and the second question 70% because of a placebo effect due to the wording of the questions.
- (C) The first question probably showed 70% and the second question 25% because of the lack of a control group.

- (D) The first question probably showed 25% and the second question 70% because of response bias due to the wording of the questions.
- (E) **The first question probably showed 70% and the second question 25% because of response bias due to the wording of the questions.** As bolded in the question above, these ideas will provoke response bias because people will answer in the way they think the interviewer wants them to answer or it will simply make them think about that particular side of the issue.

4. Which of the following are true statements?

- I. Voluntary response samples often underrepresent people with strong opinions. **False - exactly the opposite!** Voluntary response bias refers to when **ONLY** the people with strong opinions respond to a survey. Because otherwise why would you spend your precious time responding?
- II. Convenience samples often lead to undercoverage bias. **True - undercoverage** refers to when parts of the population are left out. So if you survey your family and friends, you are leaving out **MANY** people. If you call people in the phone book you are leaving out people who don't have a phone number.
- III. Questionnaires with nonneutral wording are likely to have response bias. **True - nonneutral wording is meant to provoke someone to answer a question a certain way.**

(A) I and II

(B) I and III

(C) II and III

(D) I, II, and III

(E) None of the above give the complete set of true responses.

5. Each of the 29 NBA teams has 12 players. A sample of 58 players is to be chosen as follows. Each team will be asked to place 12 cards with its players' names into a hat and randomly draw out two names. The two names from each team will be combined to make up the sample. Which of the following sampling techniques is being used in this situation?

(A) Simple Random Sample - **No, a SRS refers ONLY to when the entire sampling frame (part of population you are sampling from) is numbered off or has their names put in a hat and you randomly choose from ALL the possible people. It's a very large and costly process usually. The technical "definition" of a SRS is that every possible sample of size n has the same probability of being chosen.** In this case, it would not be possible for, say, all of the Dallas Mavericks to be chosen.

(B) Stratified Sample - YES! All the NBA players have been divided in *strata* by teams. From each team they will randomly choose 2.

(C) Cluster Sample - **No.** This refers to when the population is divided into heterogeneous groups (clusters) and a certain number of **whole** clusters are chosen randomly. Then **everyone in the cluster** is included in the sample. This is different from Stratified because in each strata a certain number is picked out randomly.

(D) Multi-stage Sample - **No.** This refers to when more than one sampling procedure is used. For example, we could divide up GHS by grade level, then randomly pick a grade. Then, randomly pick a subject area, let's say English. Then from all the English classes we could randomly pick 1 (or however many). Then we could randomly pick a certain number of people from each class (stratifying) or the entire class (clustering).

(E) Systematic Sample - **No.** This refers to when the sampling frame is numbered off or listed in some kind of order, then a random starting place is chosen, and finally every n^{th} person is selected for the sample.

6. To survey the opinions of bleacher fans at Wrigley Field, a surveyor plans to select every one-hundredth fan entering the bleachers one afternoon. Will this result in a simple random sample of Cub fans who sit in the bleachers?

(A) Yes, because each bleacher fan has the same chance of being selected. - **No they don't. What about the people who arrive early (or late)?**

(B) Yes, but only if there is a single entrance to the bleachers. - **Nope. Doesn't matter. It's still not the definition of a SRS because not every sample has the same probability of being chosen.**

(C) Yes, because the 99 out of 100 bleacher fans who are not selected will form a control group. - **Um, no. Control groups are for experiments, not surveys.**

(D) Yes, because this is an example of systematic sampling, which is a special case of simple random sampling. **No, because systematic sampling is NOT SRS. SRS is its own thing and very specific.**

(E) No, because not every sample of the intended size has an equal chance of being selected. Correct because this is the definition of a simple random sample.

7. Which of the following are true statements about sampling error?

- I. Sampling error can be eliminated only if a survey is both extremely well designed and extremely well conducted.
– False. Sampling error refers to the natural variation between samples. For example, when we did the Random M&M's activity everyone ended up with a different average pile size because your calculator gave you different random numbers. That's sampling error and it can never be eliminated.
- II. Sampling error concerns natural variation between samples, is always present, and can be described using probability. True – Just what I said above.
- III. Sampling error is generally smaller when the sample size is larger. True – if in the Random M&M's activity we had all taken a random sample of 25 piles of M&M's instead of only 5 piles then our averages would have been close to the actual mean of 7.37. This is different from saying, "You have to have a very large sample to accurately estimate the population", which is not true. The analogy in the textbook is helpful: if you want to taste your soup to make sure it has enough salt, you don't have to eat the whole pot or even a whole cup. You just take a spoonful to get an idea. But this answer choice is a general statement that the larger the sample the smaller the sampling error. I'm sure that if you indeed ate the whole pot of soup you would know for sure if it needed more salt. But then it wouldn't even be an issue anymore.

- (A) I and II
- (B) I and III
- (C) II and III**
- (D) I, II, and III
- (E) None of the above gives the complete set of true responses

8. What fault do all these sampling designs have in common?

- I. The *Wall Street Journal* plans to make a prediction for a presidential election based on a survey of its readers.
This has voluntary response bias, so only the readers with strong opinions will care enough to respond.
- II. A radio talk show asks people to phone in their views on whether the United States should pay off its huge debt to the United Nations. Same as the first – voluntary response bias.
- III. A police detective, interested in determining the extent of drug use by teenagers, randomly picks a sample of high school students and interviews each one about any illegal drug use by the student during the past year.
This is a great example of response bias because no student (or at least no intelligent student) will admit to a cop face to face that they use illegal drugs. So the student will more than likely respond they don't do drugs because that's the response that will make them look better in the eyes of the cop.

- (A) All the designs make improper use of stratification.
- (B) All the designs have errors that can lead to strong bias.**
- (C) All the designs confuse *association* with *cause and effect*.
- (D) None of the designs satisfactorily controls for sampling error.
- (E) None of the designs makes use of chance in selecting a sample.

9. A state auditor is given an assignment to choose and audit 26 companies. She lists all the companies whose name begins with A, assigns each a number, and uses a random number table to pick one of these numbers and thus one company. She proceeds to use the same procedure for each letter of the alphabet and then combines the 26 results into a group for auditing. Which of the following are true statements based on this scenario?

- I. Her procedure makes use of chance. True – it uses chance. Simple statement.
- II. Her procedure results in a simple random sample. FALSE! – She is stratifying by first letter in the name of the company. If I were a company who was evading taxes I wouldn't want to have a name that begins with X or Z. My chances of being selected are most likely way higher than those companies who have a name that starts with A.
- III. Each company has an equal probability of being audited. – Nope, like what I mentioned above.

- (A) I and II
- (B) I and III
- (C) II and III
- (D) I, II, and III
- (E) None of the above gives the complete set of true responses.**

10. A researcher planning a survey of heads of households in a particular state has census lists for each of the 23 counties in that state. The procedure will be to obtain a random sample of heads of households from each of the counties rather than grouping all the census lists together and obtaining a sample from the entire group. Which of the following is a true statement about the resulting stratified sample?
- It is not a simple random sample. – **True. This is a stratified sample because she's dividing the state up by counties.**
 - It is easier and less costly to obtain than a simple random sample. – **True. To take all the lists and put them together, then to put everyone's name in a hat or number everyone off, and then to choose a random sample from these would be quite an undertaking!**
 - It gives comparative information that a simple random sample wouldn't give. – **True. Perhaps some counties are poorer than others. Now each geographic region has a voice in the sample.**
- (A) I only
(B) I and II
(C) I and III
(D) I, II, and III
(E) None of the above gives the complete set of true responses.
11. To find out the average occupancy size of student-rented apartments, a researcher picks a simple random sample of 100 such apartments. Even after one follow-up visit, the interviewer is unable to make contact with anyone in 27 of these apartments. Concerned about nonresponse bias, the researcher chooses another simple random sample and instructs the interviewer to continue this procedure until contact is made with someone in a total of 100 apartments. The average occupancy size in the final 100-apartment sample is 2.78. Is this estimate probably too low or too high?
- Too low, because of undercoverage bias.
 - Too low, because convenience samples overestimate average results.
 - Too high, because of undercoverage bias. – Think of why the interviewer was unable to make contact with anyone in those 27 apartments... no one was home, which was the point of the survey (average occupancy). So if he continues around until he finds random apartments with people in them, chances are those apartments have more people if he's catching someone at home. This will overestimate the average occupancy and is excluding the lower occupancy apartments (undercoverage bias).**
 - Too high, because convenience samples overestimate average results.
 - Too high, because voluntary response samples overestimate average results.
12. To conduct a survey of long-distance calling patterns, a researcher opens a telephone book to a random page, closes his eyes, puts his finger down on the page, and then reads off the next 50 names. Which of the following are true statements regarding this scenario?
- The survey design incorporates chance. **True, there is chance involved in choosing a random page.**
 - The procedure results in a simple random sample. – **NOPE! He has control over where his finger is pointing.**
 - The procedure could easily result in selection bias. **Absolutely! This is Undercoverage bias (a type of selection bias) because people without a phone or those who have unlisted numbers are excluded.**
- (A) I and II
(B) I and III
(C) II and III
(D) I, II, and III
(E) None of the above give the complete set of true responses.
13. Which of the following are true statements about sampling?
- Careful analysis of a given sample will indicate whether or not it is random. – **False. You have to look at the sampling procedure to know if it was randomized or not. If you just hand me some surveys you collected I don't know if the respondents were selected randomly just based on the data.**
 - Sampling error implies an error, possibly very small but still an error, on the part of the surveyor. – **False. As in #7, sampling error refers to the natural variation that occurs between different samples.**
 - Data obtained while conducting a census are **always** more accurate than data obtained from a sample, no matter how careful the design of the sample. – **False. Tricky question too. Though taking a census will give you very accurate data compared to a sample, if the sample survey was biased then the whole thing is bunk. Be careful with these types of statements that are absolutes (using the word "always" is a red flag).**
- (A) I only
(B) II only
(C) III only
(D) None of the statements are true.
(E) None of the above gives the complete set of true responses.

14. Consider the following three events:

- I. Although 18% of the student body are minorities, in a random sample of 20 students, 5 are minorities. – **Sampling error – it just so happens that when they took a random sample, 25% were minorities. But hey, that's randomization.**
- II. In a survey about sexual habits, an embarrassed student deliberately gives the wrong answers. – **Response bias**
- III. A surveyor mistakenly records answers to one question in the wrong space. – **Just a human error/mistake (no technical stat word for this)**

Which of the following correctly characterizes the above?

- (A) I, sampling error; II response bias; III, human mistake**
- (B) I, sampling error; II, nonresponse bias; III, hidden error
- (C) I, hidden bias; II, voluntary sample bias; III, sampling error
- (D) I, undercoverage error; II, voluntary error; III, unintentional error
- (E) I, small sample error; II, deliberate error; III, mistaken error

15. A researcher plans a study to examine the depth of belief in God among **the adult population**. He obtains a simple random sample of 100 adults as they leave church one Sunday morning. All but one of them agree to participate in the survey. Which of the following are true statements?

- I. Proper use of chance as evidenced by the simple random sample makes this a well-designed survey. – **Nope. First off, it's highly doubtful he took a real SRS as people left church. "Hey, can I put your name in this hat so I can see if you get chosen to take this survey? It'll just take another 30 minutes or so because I need 50 more people." Not to mention the fact that he only took a sample from one church to estimate depth of belief in God among "the adult population." Estimating all adults from one church?... bad design.**
- II. The high response rate makes this a well-designed survey. – **Nope. I'm sure if it's a survey about belief in God they would be more inclined to participate because they just got finished at a church service. A high response rate doesn't say anything about the quality of the survey, just that many people decided to participate. Maybe they would've looked bad had they turned it down, so we could argue there is response bias present.**
- III. Selection bias makes this a poorly designed survey. – **True. Sums it up quite well.**

- (A) I only
- (B) II only
- (C) III only**
- (D) I and II
- (E) None of these statements is true.

Free Response Questions

1. A questionnaire is being designed to determine whether most people are or are not in favor of legislation protecting the habitat of the spotted owl. Give two examples of poorly worded questions, one biased toward each response.

Are you in favor of protecting the habitat of the spotted owl, which is almost extinct and in need of government help?

Are you in favor of protecting the habitat of the spotted owl, no matter how much unemployment and poverty this may cause among hard-working loggers?

2. To obtain a sample of 25 students from among the 500 students present in school one day, a surveyor decides to pick every twentieth student waiting in line to attend a required assembly in the gym.

- a. Explain why this procedure will not result in a simple random sample of the students present that day.
This is a systematic random sample and not a SRS because not every group of 25 students has the same chance of being selected. For example, some classes will arrive at the same time and therefore only one or no students from a certain class will be selected. A whole class could not possibly be selected this way.
- b. Describe a procedure that will result in a simple random sample of the students present that day. (Be specific)
Number off all the students from 001-500, and select three digits at a time from a random number table (RNT), skipping all repeats and numbers greater than 500, until 25 students have been selected.
OR
Put all the students' names in a hat, mix them well, then select 25 names for the sample.
OR
Number off all the students from 1-500 and use a random number generator (RNG) to select 25 students.

3. A hot topic in government these days is welfare reform. Suppose a congresswoman wishes to survey her constituents concerning their opinions on whether the federal government should turn welfare over to the states. Discuss possible sources of bias with regard to the following four options:

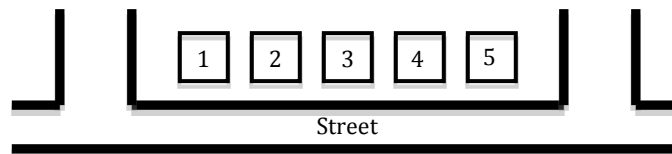
- (1) Conducting a survey via random telephone dialing into her district,
This would involve undercoverage bias because not everyone has a phone.
- (2) Sending out a mailing using a registered voter list,
This would also involve undercoverage bias because not everyone is a registered voter. This would also possibly result in nonresponse bias because most mailed surveys have low response rates.
- (3) Having a pollster interview everyone who walks past her downtown office, and
This is a convenience sample because there is no effort to randomize who is selected. It is simply whoever is closest (and most convenient). This is not a representative sample. This method would also perhaps result in response bias because if people are passing in front of the congresswoman's office they may be more likely to just say what they think the pollster's or the congresswoman's preferred position is OR they might say something rude out of spite if they dislike the congresswoman.
- (4) Broadcasting a radio appeal urging interested citizens in her district to call in their opinions to her office.
This is class voluntary response bias. She will surely get the strongest opinions and therefore not a representative sample. Radio and internet polling can rarely if ever be trusted because of voluntary response bias.

4. You and nine friends go to a restaurant and check your coats. You all forget to pick up the ticket stubs, and so when you are ready to leave, Hilda, the hatcheck girl, randomly gives each of you one of the ten coats. You are surprised that one person actually receives the correct coat. You would like to explore this further and decide to use a random number table to simulate the situation. Describe how the random number table can be used to simulate one trial of the coat episode. Explain what each of the digits 1 through 9 will represent.

Note: I won't ask you to do this on the test. Just know how to pull digits from a random number table.

Let the digits 0-9 represent the 10 people. I will select one digit at a time from a RNT, excluding repeats, until I have 10 different digits. Looking at the order of the digits will determine is a person received his or her jacket as follows: If the digit "1" appears in the first place, this denotes the first person receiving their jacket. If the digit "2" appears in the second place, this denotes the second person receiving their jacket, and so on up to zero in the 10th place, which would denote the 10th person receiving their jacket.

5. You are supposed to interview the residents of two of the five houses below. (I forgot the picture, but I drew it on the board for you guys. Here it is...)



- a. How would you choose which houses to interview?
I would use a RNG to select 2 houses from the digits 1-5.
OR
I would pull one digit at a time from a RNT, skipping all repeats and numbers over 5, until I have two different digits representing two different houses.
OR
I would put the numbers in a hat, mix them up, then select two.
- b. You plan to visit the homes at 9 a.m. If someone isn't home, explain the reason for and against substituting another house.
Selecting another house would cut costs and save time so I don't have to return later. However, this would no longer be a SRS and would include selection bias due to undercoverage because certain types of people (those that work outside the home during regular work hours) will not be home at 9:00am.
- c. Are there any differences you might expect to find among the residents based on the above sketch?
The people in the corner lot houses (1 and 5) might possibly be different in terms of income because corner lots tend to cost more.

6. A cable company plans to survey potential customers in a small city currently served by satellite dishes. Two sampling methods are being considered. Method A is to randomly select a sample of 25 city blocks and survey every family living on those blocks. Method B is to randomly select a sample of families from each of the five natural neighborhoods making up the city.
- What is the statistical name for the sampling technique used in Method A, and what is a possible reason for using it rather than an SRS?
Method A is a random cluster sample because the population is broken up into heterogeneous groups and an entire group (in this case a cluster of 25 city blocks) is selected for the sample. It is easier than a SRS because one does not have to deal with selecting individuals out of the entire population.
 - What is the statistical name for the sampling technique used in Method B, and what is a possible reason for using it rather than an SRS?
Method B is a stratified random sample because the population is broken up into homogenous groups (neighborhoods in this case) and a certain number of families are chosen out of each stratum. Stratified samples give comparative information that a SRS cannot offer due to the fact that each stratum (neighborhood) is represented in the sample.

Things to remember about sampling:

VOCABULARY!!! Go back and review your flashcards you made from Chapter 12.

Know each type of sampling technique – Census, SRS, Stratified, Cluster, Multi-stage, Systematic.

Know each type of bias... not just the definition, but the effects of each one on a sample survey.

Voluntary Response bias, Undercoverage (selection bias), Response bias, Nonresponse bias, convenience sampling (a biased type of sample; not representative)

Know how to select items using a Random Number Table.

What is sampling error? Remember: *Sampling error is to be expected, while bias is to be avoided.*