It's Friday!

Turn in your Take Home LCQ

Dead trees

In Rocky Mountain National Park, many mature pine trees along Highway 34 are dying due to infestation by pine beetles. Scientists would like to use a sample of size 200 to estimate the proportion of the approximately 5000 pine trees along the highway that have been infested.

(a) Explain why it wouldn't be practical for scientists to obtain an SRS in this setting.

Every tree would have to be identified and numbered.

(not practical to number every tree along highway and then search for selected trees)

(b) A possible alternative would be to use the first 200 pine trees along the highway as you enter the park. Why isn't this a good idea?

This convenience sampling method is not a good idea because these trees are unlikely to be representative of the popul.

- trees more likely damaged near people?
- beetles afraid of people?

(c) Describe how to select a systematic random sample of 200 pine trees along Highway 34.

5000

$\frac{5000}{200} = 25$ ↑ sample size

- Select every 25th tree walking along highway.
- To choose a starting point, select a number 1 to 25. We would select that tree and every 25th tree thereafter until 200 have been selected.

Ch. 4 So far...

Bad Sampling → shows bias
- Convence
- Voluntary Response

which can underestimate or overestimate true values

Random Sampling → will help Avoid these problems
- SRS
- Stratified
- Cluster
- Systematic Random

Ch. 4 So far...

Bad Sampling → shows bias
- Convence
- Voluntary Response

Random Sampling → will help Avoid these problems
- SRS
- Stratified
- Cluster
- Systematic Random

but... there are other problems that are hard to avoid.

Sample Surveys: What Else Can Go Wrong? (pages 233–235)

AP Stats - Lesson 4.1: Day 3: What is wrong with these surveys?

Identify what is wrong in each of these surveys. Be sure to explain.

look at question #1

Undercoverage

1. The mayor of Springfield is interested in finding out the average age of people in the city. He obtains a list of all of the landline telephones in the city, and then contacts a simple random sample of 300 people. He uses the data from the sample to estimate the average age of all the people in the city.

   a. What is wrong with this survey?

   only people with land lines are included.

   b. Do you think the Mayor will over or underestimate the true mean age of people in Springfield? Why?

   overestimation of true mean age because older people are probably more likely to own a landline

1. The mayor of Springfield is interested in finding out the average age of people in the city. He obtains a list of all of the landline telephones in the city, and then contacts a simple random sample of 300 people. He uses the data from the sample to estimate the average age of all the people in the city.

   a. What is wrong with this survey?

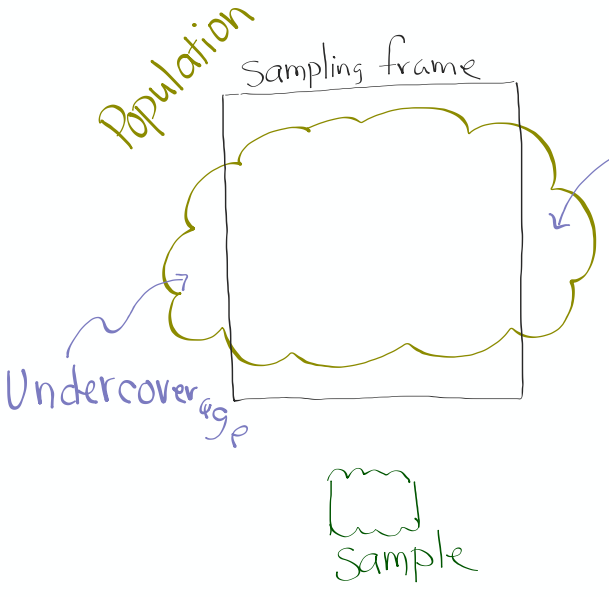
   He is only contacting people with a land line. People without land lines aren't surveyed.

   b. Do you think the Mayor will over or underestimate the true mean age of people in Springfield? Why?

   Overestimate - We would think people with only land lines are older.

Undercoverage

1. The mayor of Springfield is interested in finding out the average age of people in the city. He obtains a list of all of the landline telephones in the city, and then contacts a simple random sample of 300 people. He uses the data from the sample to estimate the average age of all the people in the city.
   a. What is wrong with this survey?
   He is only contacting people with a land line. People without land lines aren't surveyed.
   b. Do you think the Mayor will over or underestimate the true mean age of people in Springfield? Why?
   Overestimate - We would think people with only land lines are older.



the sample doesn't look like the population

2. The administration at a school wants to know the proportion of students that did all of their homework last night. They select a simple random sample of 100 students and send an email to each of them asking if they did all of their homework last night. Of the 40 responses, 36 of the students said that they did all of their homework last night (90%).

   a. What is wrong with this survey?

   - less than 100 responded
   - Students got to choose
     - emailers more likely to do HW

   Students could lie

   b. Do you think the administration will over or underestimate the true proportion of students who did all of their homework last night? Why?

2. The administration at a school wants to know the proportion of students that did all of their homework last night. They select a simple random sample of 100 students and send an email to each of them asking if they did all of their homework last night. Of the 40 responses, 36 of the students said that they did all of their homework last night (90%).

   a. What is wrong with this survey?

   - Only 40 of 100 responded
   - Students might lie and say they did when they didn't.

   b. Do you think the administration will over or underestimate the true proportion of students who did all of their homework last night? Why?

   Overestimate. Students might lie because admin is doing the asking. Or they might not respond if they didn't do HW.

their homework last night. They select a simple random sample of 100 students and send an email to each of them asking if they did all of their homework last night. Of the 40 responses, 36 of the students said that they did all of their homework last night (90%).

a. What is wrong with this survey?

Nonresponse
- Only 40 of 100 responded.
- Students might lie and say they did when they didn't.

Response bias

b. Do you think the administration will over or underestimate the true proportion of students who did all of their homework last night? Why?

Overestimate. Students might lie because admin is doing the asking. Or they might not respond if they didn't do HW.

3. Boy Scout Peter M. wants to know the proportion of people in his neighborhood who support the Boy Scouts. He takes a random sample of 30 homes and visits them dressed in his uniform.

a. What is wrong with this survey?

Response bias
Possibly say support because of uniform.
- may not get people at home

b. Do you think Peter will over or underestimate the true proportion of his neighbors who support the Boy Scouts? Why?

Overestimation

Underestimate

3. Boy Scout Peter M. wants to know the proportion of people in his neighborhood who support the Boy Scouts. He takes a random sample of 30 homes and visits them dressed in his uniform.

a. What is wrong with this survey?

He is influencing responses. People don't want to tell him they don't support Boy Scouts because they know he is one.

b. Do you think Peter will over or underestimate the true proportion of his neighbors who support the Boy Scouts? Why?

Overestimate. People may say they support Scouts even if they don't.

3. Boy Scout Peter M. wants to know the proportion of people in his neighborhood who support the Boy Scouts. He takes a random sample of 30 homes and visits them dressed in his uniform.

a. What is wrong with this survey?

He is influencing responses. People don't want to tell him they don't support Boy Scouts because they know he is one.

b. Do you think Peter will over or underestimate the true proportion of his neighbors who support the Boy Scouts? Why?

Overestimate. People may say they support Scouts even if they don't.

Response Bias

Big Ideas:

Big Ideas:

Undercoverage

when the same members of the population cannot or are less likely to be chosen

Ex. land lines

Big Ideas:

Undercoverage: when the same members of the population cannot or are less likely to be chosen

Ex. land lines

Nonresponse: When an individual is selected to be in a sample but chooses not to participate or isn't reached

* different from voluntary response

Non response can only occur after a sample has been chosen.

Nate Silver, Statistician
Maintains website www.fivethirtyeight.com
Author of book: "The Signal and the Noise"

"Response rates to political polls are dismal. Even polls that make every effort to contact a representative sample of voters now get no more than 10 percent to complete their surveys - down from about 35% in the 1990's".

How do you write survey questions that accurately measure public opinion?

BY COURTNEY KENNEDY

FACTTANK
NEWS IN THE NUMBERS

AUGUST 4, 2017

Personal finance questions elicit slightly different answers in phone surveys than online

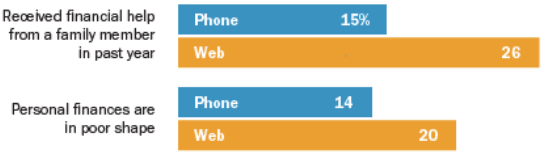
BY GEORGE ELLIOTT MORRIS AND COURTNEY KENNEDY

People polled by telephone are slightly less likely than those interviewed online to say their personal finances are in "poor shape" (14% versus 20%, respectively), a Pew Research Center survey experiment has found.

The experiment, conducted in February and March, is part of a line of research at the Center looking into

More Americans report financial stress when interviewed online than by phone

% who say ...



Note: Web sample size: 2,066; phone sample size: 1,778.
Source: Survey conducted Feb. 28-March 12, 2017.

PEW RESEARCH CENTER

Big Ideas:

Undercoverage: when the same members of the population cannot or are less likely to be chosen

Ex. land lines

Nonresponse: When an individual is selected to be in a sample but chooses not to participate or isn't reached

* different from voluntary response

Non response can only occur after a sample has been chosen.

Response Bias

Pattern of inaccurate responses.

- Could be due to wording, interviewer, lying, etc.

Some people make up answers

Jimmy Kimmell

"Which do you agree with: Obamacare or the Affordable Care Act?"

Check your understanding

Check Your Understanding:

1. Each of the following is a possible source of bias in a sample survey. Name the type of bias that could result.

(a) The sample is chosen at random from a telephone directory.

(b) Some people cannot be contacted in five calls.

(c) Interviewers choose people walking by on the sidewalk to interview.

(a) The sample is chosen at random from a telephone directory.

Undercoverage, only people with numbers in the phone directory can be chosen.

(b) Some people cannot be contacted in five calls.

Nonresponse, some of the sample can't be reached. so they don't respond.

(c) Interviewers choose people walking by on the sidewalk to interview.

Convenience sample, all the people on the sidewalk could have some commonality.

2. A survey paid for by makers of disposable diapers found that 84% of the sample opposed banning disposable diapers.

> Here is the actual question: "It is estimated that disposable diapers account for less than 2% of the trash in today's landfills. In contrast, beverage containers, third-class mail, and yard wastes are estimated to account for about 21% of the trash in landfills. Given this, in your opinion, would it be fair to ban disposable diapers?"

Do you think the estimate of 84% is less than, greater than, or about equal to the percent of all people in the population who would oppose banning disposable diapers? Explain your reasoning.

2. A survey paid for by makers of disposable diapers found that 84% of the sample opposed banning disposable diapers.

   Here is the actual question: "It is estimated that disposable diapers account for less than 2% of the trash in today's landfills. In contrast, beverage containers, third-class mail, and yard wastes are estimated to account for about 21% of the trash in landfills. Given this, in your opinion, would it be fair to ban disposable diapers?"

   Do you think the estimate of 84% is less than, greater than, or about equal to the percent of all people in the population who would oppose banning disposable diapers? Explain your reasoning.

84% is likely greater. The wording of the question makes it sound like diapers are not a problem in landfills. The question is leading.

2. A survey paid for by makers of disposable diapers found that 84% of the sample **opposed banning** disposable diapers.

Here is the actual question: "It is estimated that disposable diapers account for less than 2% of the trash in today's landfills. In contrast, beverage containers, third-class mail, and yard wastes are estimated to account for about 21% of the trash in landfills. Given this, in your opinion, would it be fair to ban disposable diapers?"

Do you think the estimate of 84% is less than, greater than, or about equal to the percent of all people in the population who would oppose banning disposable diapers? Explain your reasoning.

84% is likely greater. The wording of the question makes it sound like diapers are not a problem in landfills. The question is leading.

Be careful with this. It means they think diapers should be allowed to be disposable.

AP Exam Tip

If you are asked to describe how issues with the collection of survey data lead to bias, you're expected to address two ideas:

1. describe how the members of the sample might respond differently from the rest of the population.

2. explain how this difference would lead to an underestimate or overestimate.

4.1....25, 27, 29, 31, 33, 35-40

AND

Complete the "Read Ahead Notes for section 4.2 Day 1"