Agenda

- Warm Up   10 min
- Check up time   10 min
  - ~~Checked 9/16/19, copies~~
- Types of bias   20 min
- Lecture, sampling methods   20 min
  - *Copies*
- Try it!   10 min
- *Chocolate time*   *10 min*
- FRQ practice   20 min
  - *Copies*
- Exit Pass   5 min

Warm Up

One of the earliest studies of the relationship between smoking and lung cancer was conducted by Hammond and Horn (1958). They tracked 187,766 men over a 44-month period, noting their smoking habits and whether or not they died of lung cancer.

1. Identify the explanatory and response variables in this study.

2. Is this an observation study, survey, or experiment? Explain.

3. Would you generalize the results of this study to the population of all adults? Why or why not?

Checkup time

Types of SAMPLING bias

- **Selection (or "undercoverage") bias**
  - Some of population is excluded or underrepresented in selection
  - Example: Population is teachers. Sample doesn't include math department.

- **Nonresponse bias**
  - No data is collected from individuals who have been selected.
  - Example: Mr. Colligan is given a survey, but he's too busy to fill it out.

- **Response bias**
  - Process distorts responses
  - Example: Study asks employees, "Considering how hard you work, don't you think it's ridiculous that you don't get free sandwiches every day?"

Example #1 (of 3)

- The cost of the 2010 U.S. Census was $4.5 billion, an average of $15.99 per American.
- The Census Bureau starts by mailing packets to everyone at their mailing address, as registered by the Internal Revenue Service.
- They call people who do not return the packets.
- They go to the addresses of people who do not return their packets or answer their phones.

"Representatives shall be apportioned among the several States according to their respective numbers, counting the whole number of persons in each State" –U.S. Constitution, Article 1, Section 2 (modified by 14th Amendment)

How the Census Bureau knows it missed a million kids

To check its survey count, the bureau added up the 21,120,000 birth records from the five years leading up to the 2010 census
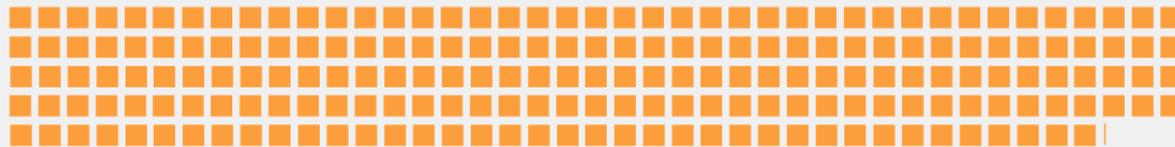
100k people
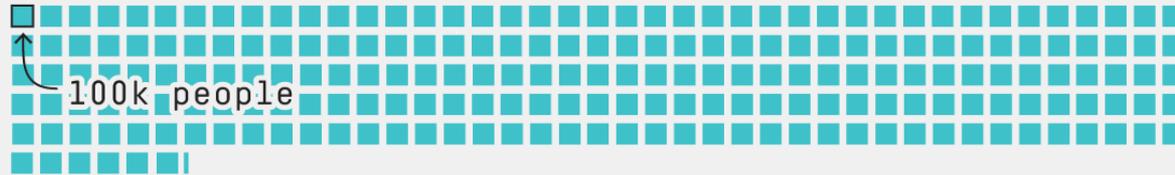
and adjusted that number by subtracting the 154,000 ■■| death records and adding 240,000 ■■■| to account for net international immigration.

This resulted in an estimated total of 21,206,000 children age 0-4.

The 2010 census survey counted 20,201,000 children age 0-4,

which is 1,005,000 ■■■■■■■■■■■■■| less than the estimated totals.

Example #2

- "Official 2016 Democratic Party Survey"

Which group of GOP backers do you think will have the biggest impact on the 2016 elections? (please choose one)

❏ Right-wing billionaires, such as the Koch brothers and Sheldon Adelson
❏ Ruthless political operatives, such as Karl Rove
❏ Extreme far-right organizations, such as the Tea Party Nation and the Club for Growth
❏ Conservative media outlets

How should the DNC react when Republicans and their special-interest backers target Democrats with lies and personal attacks? (please choose one)

❏ Hit back fast and hard with the truth, and hold the GOP's feet to the fire
❏ Wait until the dust settles, and then offer a low-key response
❏ Completely ignore lies and mudslinging
❏ Not sure

Example #3 (of 3)

The *2019 Global Health Survey*, from the *Partners in Health* organization.

Which one is this?

		What They Said	
		I've done drugs	I haven't done drugs
What They Did	Drugs	172	120
	No drugs	13	358

Sampling methods

SAMPLING METHODS → Example

- Anthony DeLuca lives in an apartment complex with 30 floors, and 20 apartments on each floor.

- The homeowners' association wants to sample the opinions of its members on the possibility of switching to a new management service.

- They hire Anthony to take a sample of 60 apartments.

Judgment sample ☹

- Individuals are chosen based on your judgment of who is representative of the population

- *Example:*
  - Anthony could decide to interview people that he believed were representative of all homeowners.

  - But Anthony doesn't know all the 600 families, so he doesn't know who are "representative."

  - Most people tend to sample those whose opinions are thought to be "typical," which eliminates people whose opinions are extreme (but who nevertheless are members of the population and thus should be represented).

Voluntary sample ☹

- Individuals volunteer to be part of your sample.

- *Example:*
  - Anthony could leave a questionnaire in the lobby of the building and ask people to return it to his apartment.
  - Although easy for Anthony, many people who respond do so because they possess a strong opinion on the issue, giving a distorted view of the opinions of all the residents.

Convenience sample ☹

- Choose your sample based on what is convenient.

- *Example:*
  - Antonio might stand at the entrance to the building and ask people their opinions as they enter.
  - Although convenient, the people who are sampled will vary according to the time of day.
    - Midday, Antonio would encounter a preponderance of home-makers
    - Early evening, he would find more people returning from work
    - Later in the evening, perhaps younger, single people would dominate.

☺ Simple random sampling (SRS)

- Most common, simple, effective
- Random: "A sample chosen in such a way that every individual has an equal chance of being selected."

Hat vs. Calculator (**AP only**)

MATH, PRB, randInt(min.,max.,#)

- *Example:*
  - Anthony could take a list of the 600 apartments (which are numbered 101-120, 201-220,K) and randomly select 60 of these to sample.

☺ Systematic sampling

- Select an individual from the list at random
- After that, every k^th individual is selected, where *k* (the "**skip**") is a predetermined interval
- Efficient and useful, as long as the population is homogenous (evenly distributed)

- *Example*:
  - Anthony could take his list of apartments, randomly select one apartment from the first ten, (say the seventh or Apartment 107) and select every tenth apartment after that (Apartments 117, 127….or 207, 307….).
  - Problem: Every 20th apartment is probably in the same position on its floor.
    - Some aspect of positioning, such as corner apartments, might make some apartments cheaper or more expensive than the others.

☺ Stratified random sampling

- Divide the population into groups of individuals (called **strata**)
  - Strata are similar in some way that is important for the response variable.
- Choose a SRS from each strata and combine.
- Effective if population is <u>not</u> homogenous

- *Example:*
  - Anthony could randomly select two apartments on each floor (say 101 and 107 from the 1st floor, 206 and 219 on the 2nd floor, and so on).
  - Good plan if there are differences of opinion among floors.

☺ Cluster sampling

- Divide the population into groups (or **clusters**).
- Randomly select some of the clusters.
- <u>All</u> individuals in the chosen clusters are sampled.
- Very convenient, only with homogenous populations

- *Example*:
    - Anthony could randomly select three of the 30 floors and sample all 20 apartments on each floor.
    - Convenient and representative of the entire building
    - BUT only if opinions are consistent from floor to floor.

Window/Door → Sampling Methods

*You've heard that RCHS students tend to oppose the part of the dress code prohibiting prominent combinations of **red** and **black***. *You want to select a representative sample of the RCHS student body to survey about this issue.*

1. Describe how you could get a sample using each method, taking turns.
2. Summarize any weaknesses or strengths for each.

Chocolate time

1. CLOSE YOUR EYES and keep them closed. I don't care if Ms. Billingslea walks in or someone in the hallway shouts Fire. Don't open your eyes until I say so.
2. Hold out one hand, palm up.
3. I will give you a piece of chocolate. Eat it. Savor it.
4. Keep one hand out, palm up. Raise your other hand for a 2nd piece. Eat it. Savor it.
5. Remember which piece you liked more. Do <u>not</u> share your preference with anyone.

- I will ask for your preference.
  - Hold up 1 finger if you liked your 1st piece more.
  - Hold up 2 fingers if you liked your 2nd piece more.

Free-response questions

- Groups of 3.
- Complete on back of paper, or on separate paper titled "FRQ practice".
- Each student chooses one question. Compare answers. Critique.

FRQ 2004 #2

a. You're only asking the first 100 people who enter the dining hall, so many of them probably know each other, and they're probably early eaters. Not representative. Use systematic sampling instead, ask every 10th person.

b. Response bias. By stating a negative consequence after asking the question, the surveyors are decreasing the chances that participants will state that they want the food improved. Eliminate the 2nd half of the sentence, after the comma.

FRQ 2008 #2

a. You only get people with strong opinions.

b. This doesn't solve the problem at all....you'll still have non-response bias, just twice as many strongly opinionated people.

c. Face-to-face interviews. Include the forms in mandatory paperwork at the beginning of the school year. Provide incentives for participating.

FRQ 2005 #5

a. Non-response bias, people hang up their phones so you only get strong opinions.

OR undercoverage bias, you're only sampling people who have telephones, so it's not a representative sample.

c. Stratified sampling by state. Randomly select an equal number of households from each state,

Exit Pass

Principal Billingslea wants to know students' opinions on the current dress code of River City High School. She has asked you to survey a representative sample of the student body.

1. How would you get a representative sample? Describe in detail, with relevant vocabulary.