

Agenda

Homework (AP)

Pg.226 #1-7

- Warm Up 10 min
- Project walkthrough 20 min
- Population vs. sample 25 min
- Types of bias 20 min
 - *Copies of biased survey*
- Hand back tests? 10 min
 - ~~*Update*~~
- Exit Pass 5 min
- *Get chocolate for tomorrow's lesson*
- *Bring up links:*
 - http://inventorspot.com/articles/running_numbers_13954
(scroll past breasts)
 - <https://www.youtube.com/watch?v=mlq3s4ub1-Y>

Warm Up (AP)

1. Briefly explain the difference between these terms:

- a. Observational study
- b. Experiment

tinyurl.com/602seating

2. (Test #5) A study found a correlation of $r = -0.61$ between the gender of a worker and his or her income. Why should these researchers lose their jobs?
3. (Test #7) A scatterplot of a variable Y versus a variable X produces these results. What must r be?



4. Guess the number of plastic bottles used in the United States every 5 minutes.

Gallery Walk

- Windows present first
 - Presenters spread around the room, evenly spaced
 - Have Project with you.
- Doors:
 - Say hi. Introduce yourself.
 - Think of a unique question to ask Presenters. Make your question with sincere (rather than comical) intentions.
 - Out of everyone currently presenting, who has the **strongest** relationship? What are their variables?
 - Out of everyone currently presenting, who has the **weakest** relationship? What are their variables?
- *Switch after ~10 minutes*

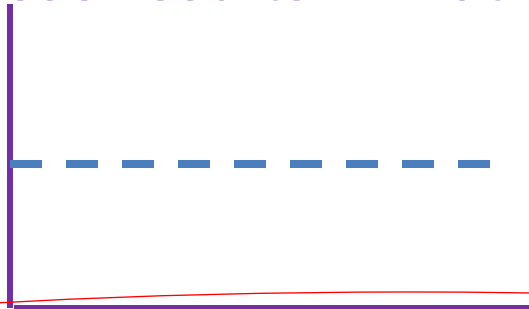
Unit 4

Test next
Friday 2/21

- Observational studies vs. experiments, conclusions drawn from each
- Sampling
 - SRS (simple random sample) → Everyone gets equal chance of selection
 - Stratified random sampling, and the advantages of stratifying
 - Types of bias
- Designing experiments
 - Randomization
 - Blocking
 - Confounding
 - Blinding

Warm Up

1. Briefly explain the distinction between the following terms:
 - a. Observational study
 - b. Experiment
2. (Test #5) A study found a correlation of $r = -0.61$ between the gender of a worker and his or her income. Why should these researchers lose their jobs?
3. (Test #7) A scatterplot of a variable Y versus a variable X produces these results. What must r be?



4. Guess the number of plastic bottles used in the United States every five minutes.

Video

Sum: Forty Tales from the Afterlives,
by David Eagleman

56% of Americans believe Elvis is alive

Elvis is reported to have died in his Graceland mansion on August 16th, 1977. On the 12th anniversary of this event, a Dallas record company wanted to learn the opinions of all adult Americans on the issue of whether Elvis was really dead.

They sponsored a national call-in survey. Listeners of more than 100 radio stations were asked to call a number, at a charge of \$2.50, to voice their opinion concerning whether Elvis was really dead. 56% of callers thought Elvis was alive

Bad bad sampling

Jacque Bourque is a medical researcher interested in the population of HIV positive individuals. They work at a clinic in Santa Barbara, and their current research project involves evaluating the health of HIV positive individuals before the onset of AIDS. Their sample consists of those HIV positive patients at the clinic who have voluntarily agreed to be part of the research project.

Did they use a representative sample? Why or why not?

“Representative”

- Population → the *entire* group of people or objects (observational units) of interest
 - μ = mean of population
 - σ = standard deviation of population
- Sample → a small *part* of the population
 - \bar{x} = mean of sample
 - s = standard deviation of sample
- “Representative” → has similar characteristics to the population
 - *Example.* This class might be a representative sample of:
 - Teenagers’ shoe sizes.
 - ~~Teenagers’ GPA’s.~~

Window/Door

For each variable, indicate whether this class would be a representative sample of all teenagers. Justify briefly.

GPA. *No, not representative. This is a 4th-level math class. You're all relatively high-achieving, and have higher GPA's.*

1. Hours slept last night.
2. Number of siblings.
3. Whether you prefer to call or text your friends.
4. Political viewpoint.
5. Gender.
6. Total income that you'll receive over your lifetime.

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 ever 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

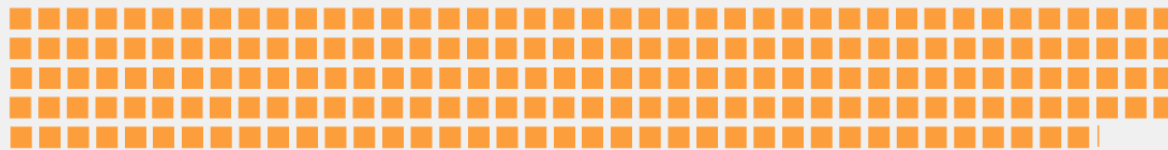


and adjusted that number by subtracting the **154,000** death records and adding **240,000** to account for net international immigration.

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 (of 3)

- “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.

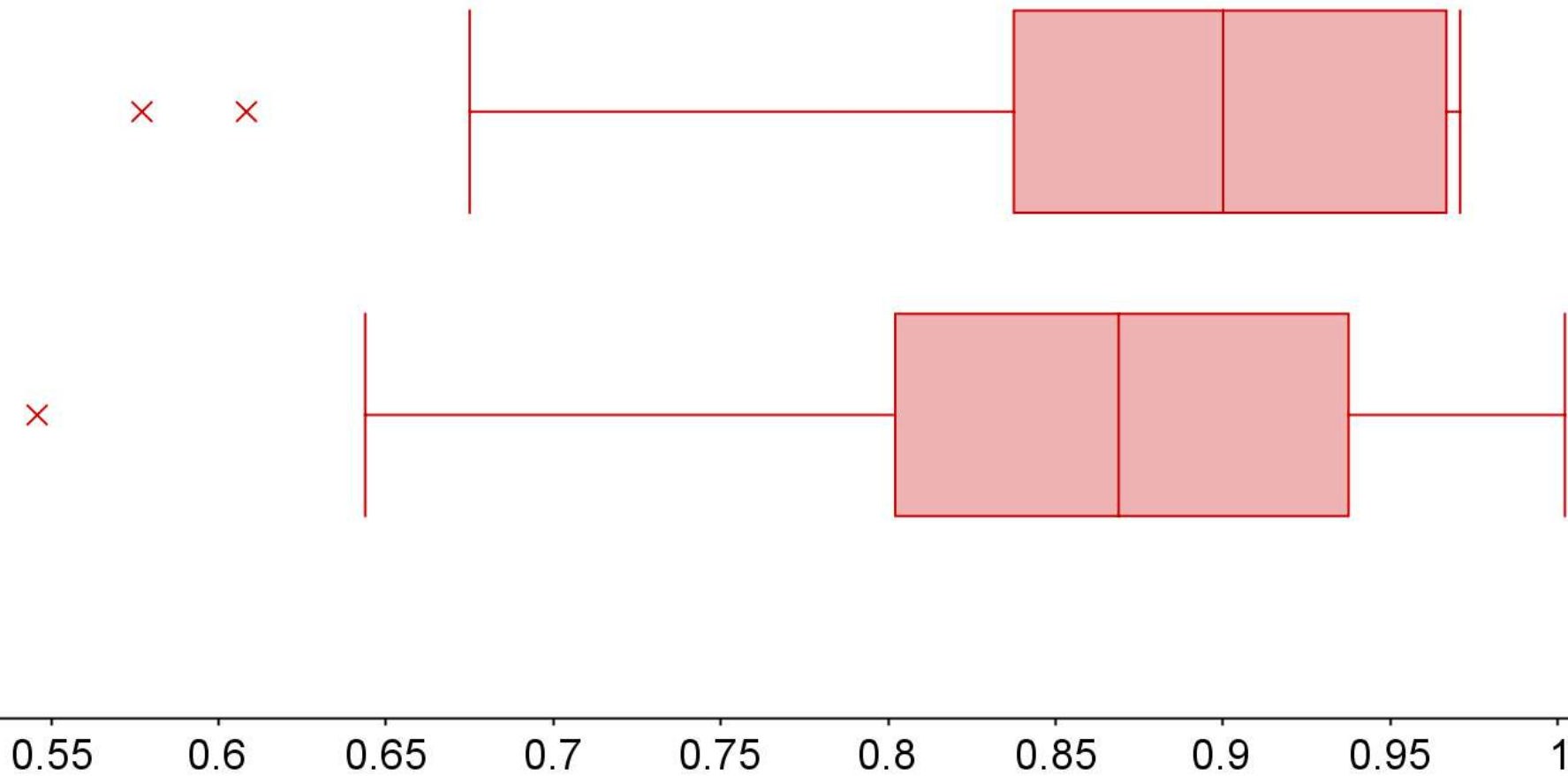
I made this up.

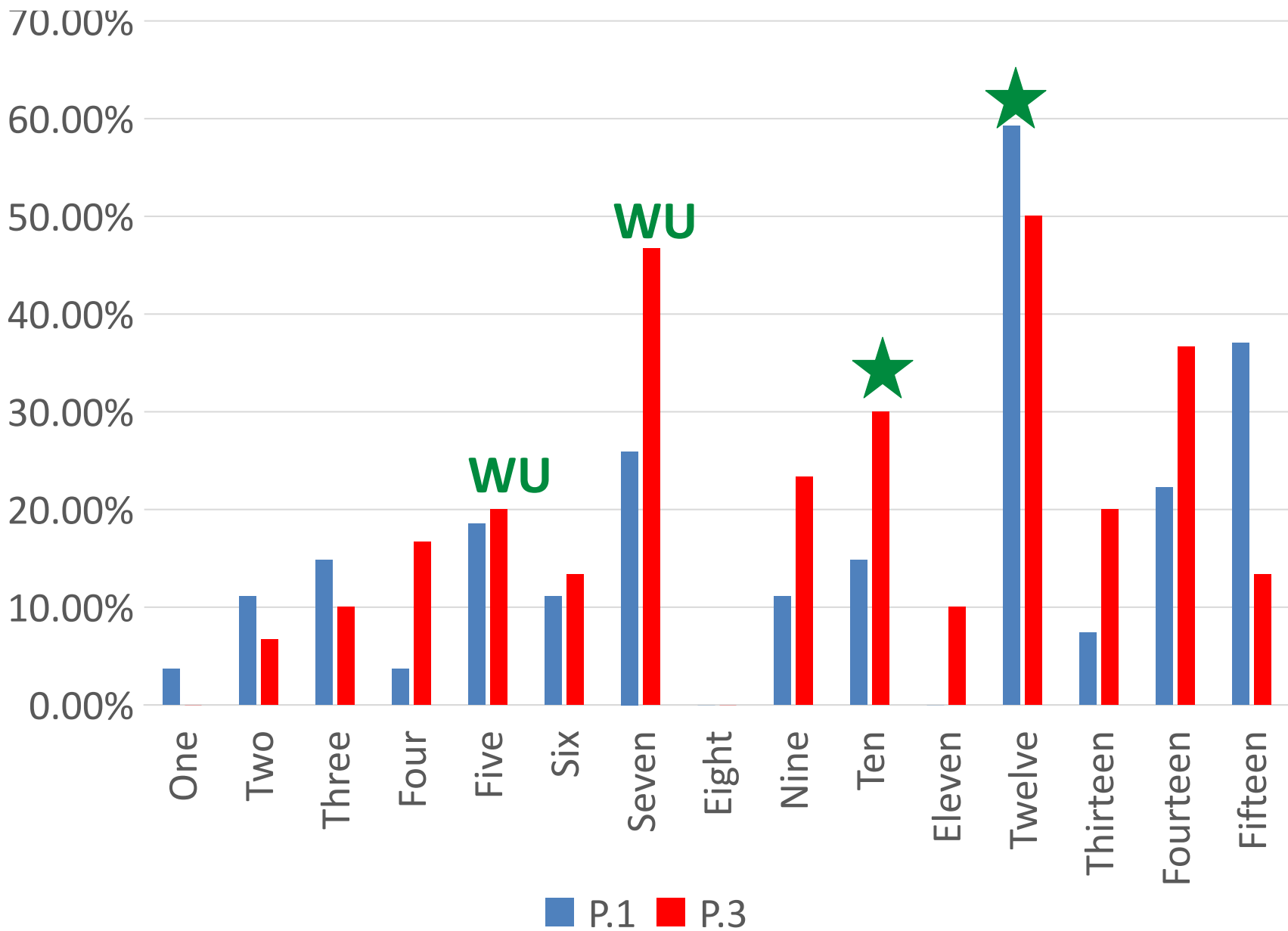
Which one is this?

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

Unit 3 Test

Regression





Common Mistakes, 1 of 2

10. The fraction of the variation in the values of y that is explained by the least-squares regression of y on x is

- A. *the correlation coefficient.*
- B. the slope of the least-squares regression line.
- C. the coefficient of determination.
- D. the intercept of the least-squares regression line.
- E. the residual.

12. The least-squares regression line is the line that...

- F. makes the square of the correlation in the data as large as possible.
- G. makes the sum of the squares of the vertical distances of the data points from the line as small as possible.
- H. *passes through the greatest number of data points.*
- I. best splits the data in half, with half of the points above the line and half below the line.
- J. all of the above.

Common Mistakes, 2 of 2

17. At summer camp, one of Carla's counselors told her that you can determine air temperature from the number of cricket chirps. $r = 0.461$ and $y = 56.23 + 0.1355x$, where x is the number of cricket chirps and y is air temperature.
- d.) Suppose that Carla counted 249 chirps on a day when the temperature was 55°F . If this point were the 13th data point, what effect, if any, would this 13th point have on the slope and the y intercept of Carla's regression line? Explain.

Exit Pass

The August 23, 1999, issue of the *Sacramento Bee* reported a study involving data volunteered by 17,251 users of the *abcnews.com* website. Users were asked whether or not they used the Internet to escape problems. About 6% of those responding confessed to some sort of addiction to the Internet.

1. Identify the population and sample used in this study.
2. Do you believe that 6% is a reasonable estimate of the number of Americans who have an Internet addiction? If not, indicate whether you believe this estimate is too high or too low. Explain.