Data Science for Everyone

Week 6

Angela Lai

New York University

Outline

- Logistics
- Coding/demo
- HW 1: Difficult Questions
- Open Questions

Logistics

- Homework 2 due at 8 p.m. this Monday, March 9
- When is lab 4 due?

Logistics

- Homework 2 due at 8 p.m. this Monday, March 9
- When is lab 4 due? Same day, same time! 8 p.m. on March 9

Homework

- Careful with formatting! Save as HTML and THEN print to PDF
- Homework 1 and lab 3 grades are out

Midterm

- Review in lecture on Monday
- TAs are holding office hours on Tuesday, March 10, from 2-5 p.m. in 660 and 665, here in this building

Homework 1 Concept Review: 2c, d

Recall:

A theory is a conjecture about the causes of some phenomena of interest.

Observable implications: The theory predicts something about the world which we can check, with data.

Falsifiable: we can imagine some data evidence that would be incompatible with our theory being true.

Homework 1 Concept Review: 2c, d

2(c) What was one observable implication of the alternative prevailing theory – miasma theory – that was not supported by the evidence?

2(c) Snow's theory was falsifiable. If his theory was incorrect, what is an example of a piece of evidence that could have proved him wrong?

Homework 1 Concept Review

Recall:

Independent variable X: a treatment/control that may explain Y

Dependent variable Y: an outcome we want to explain

Confounder Z: causes/affects **both X and Y**

Homework 1 Concept Review: 3

Question 3 asked about an article titled ”New Study examines the effect of ‘open’ adoption on families.”

What’s the independent variable?

Homework 1 Concept Review

Example to get us thinking about the intuition:

In aggregate, we observe that the incidence of prostate cancer among nurses is considerably lower than the incidence among soldiers. Should we conclude that being in the army has a causal effect on prostate cancer?

What are X and Y? Then what's a plausible confounder?

Homework 1 Concept Review: 5a, b

Read the article, “Op-Ed: We have studied every mass shooting since 1966. Here’s what we’ve learned about the shooters,” in which the authors describe four commonalities they observe among the perpetrators of mass shootings.

5 (a) This study is an example of selecting on the dependent variable. What is the dependent variable in this case, and how do we know they are selecting on it?

5 (b) What would the scientists need to do in order to alleviate this problem of selecting on the dependent variable?

Open Lab

Questions?