STA 235 - Causal Inference: Natural Experiments

Spring 2021

McCombs School of Business, UT Austin

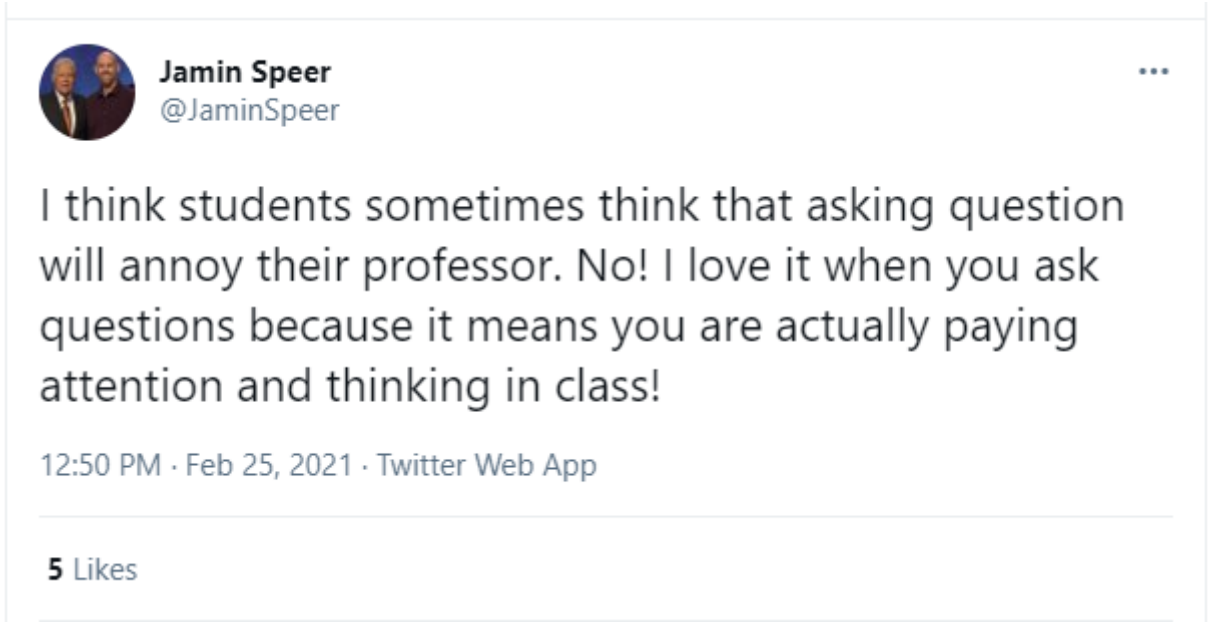
Reminders

Homework 2 is due next Monday (noon)

- Remember to ask your questions **before the weekend**
- I will make **additional OH** available if they get full.
- Remember to **contact me** if you need to meet another time.

Reminders (cont.)

Class participation increases learning



Last class

- **Selection on observables**
  - Assumption of non-random selection, but selection on observables.
- **Matching and Weighting**:
  - Use of other adjustment methods beyond regression.
  - Advantages and disadvantages of matching.
  - Use of weighting for approximating different populations.

Today

- **Natural Experiments**:
  - Identifying random assignment* in observational studies
  - Use exogeneous variation to identify causal effects.
- **Differences in Differences**:
  - Using two dimensions for identification.
  - Assumptions and shortcomings.

Is there randomness out there?

Finding "RCTs" in the wild

- Given that we can't run RCTs for everything, the next best thing is finding a source of random variation that, for all practical purposes, **would work as an RCT**

Natural Experiments

You, as a researcher, did not assign units to treatment levels

1. **Random**: Assignment to an intervention is random (e.g. lottery).
2. **As if random**: Assignment to an intervention is not random, but it's not correlated with potential confounders.

Context matters!

Potential outcomes in Observational Studies

- The same **potential outcomes framework** that we reviewed for RCTs also work with observational studies.

Steps to identify a Natural Experiment:

1) Identify treatment groups: What is the control status?

2) Identify your estimand of interest: Write it down in terms of PO!

3) Identify potential threats to causality: Is this as good as random?

Let's talk about the JITT example

- A retailer provides a 15% discount to first 1,000 customers, 10% to customers 1,001-2,000 (and no discount after).

Is this a natural experiment?

Mixed answers

Let's think about this more carefully

An example: Timely discounts

- **Two treatments:** 10% discount (1) and 15% discount (2) (control is no discount).
- **Outcome:** Total sales ($)
- **Estimand:** Average Treatment Effects,

$$ATE_1 = E[Y(1) - Y(0)]$$

and

$$ATE_2 = E[Y(2) - Y(0)]$$

How are people assigned to treatment?

An example: Timely discounts

- **Two treatments:** 10% discount (1) and 15% discount (2) (control is no discount).
- **Outcome:** Total sales ($)
- **Estimand:** Average Treatment Effects,

$$ATE_1 = E[Y(1) - Y(0)]$$

and

$$ATE_2 = E[Y(2) - Y(0)]$$

Could there be confounding? Why?

An example: Timely discounts

- **Two treatments:** 10% discount (1) and 15% discount (2) (control is no discount).
- **Outcome:** Total sales ($)
- **Estimand:** Average Treatment Effects,

$$ATE_1 = E[Y(1) - Y(0)]$$

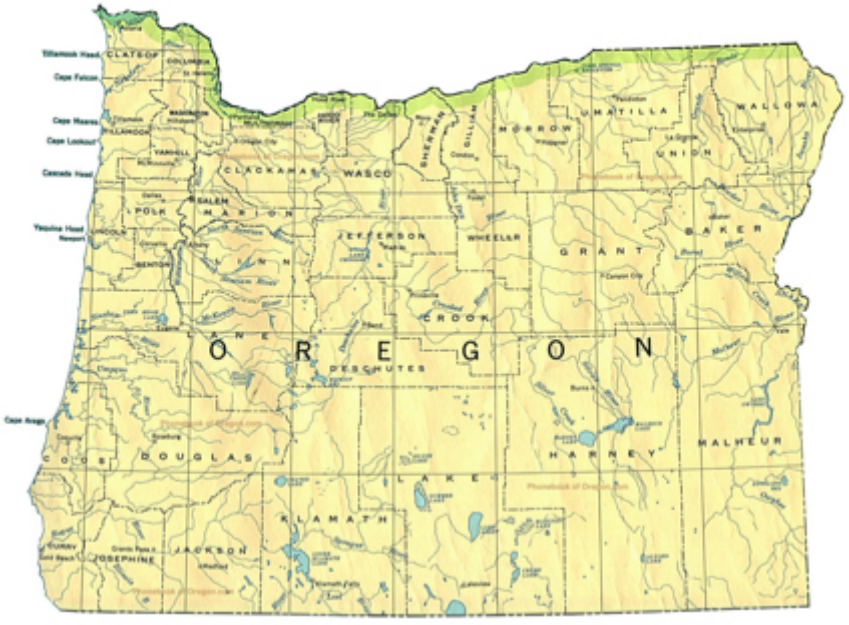
and

$$ATE_2 = E[Y(2) - Y(0)]$$

What if customers didn't know about the discount until they get there?

A true natural experiment: The Oregon Health Plan

- In 2008, Oregon implemented a **limited expansion of Medicaid**.
- **Target population**: Low-income adults.
- People selected through a **lottery**.



The Oregon Health Plan

What is the treatment in this case? What is *randomized?**

What is our estimand of interest?

What about external validity? For what population is this effect generalizable?

What is the first thing you would do with the data?

Let's go to R

Other natural experiments

- Natural experiments can arise from **anything**!
  - E.g. glitches in systems, allocation based on "random" variables.

Treatment assignment has to be [conditionally] independent of potential outcomes!

Additional Examples: Fish and Phones

- **Context**: Kerala, 1997.
- **Problem**: Fishermen out at sea cannot observe prices in the different markets, and due to time constraints can only visit one.
- **Identification Strategy**: Natural source of variation when cellphones were introduced and rolled out geographically.

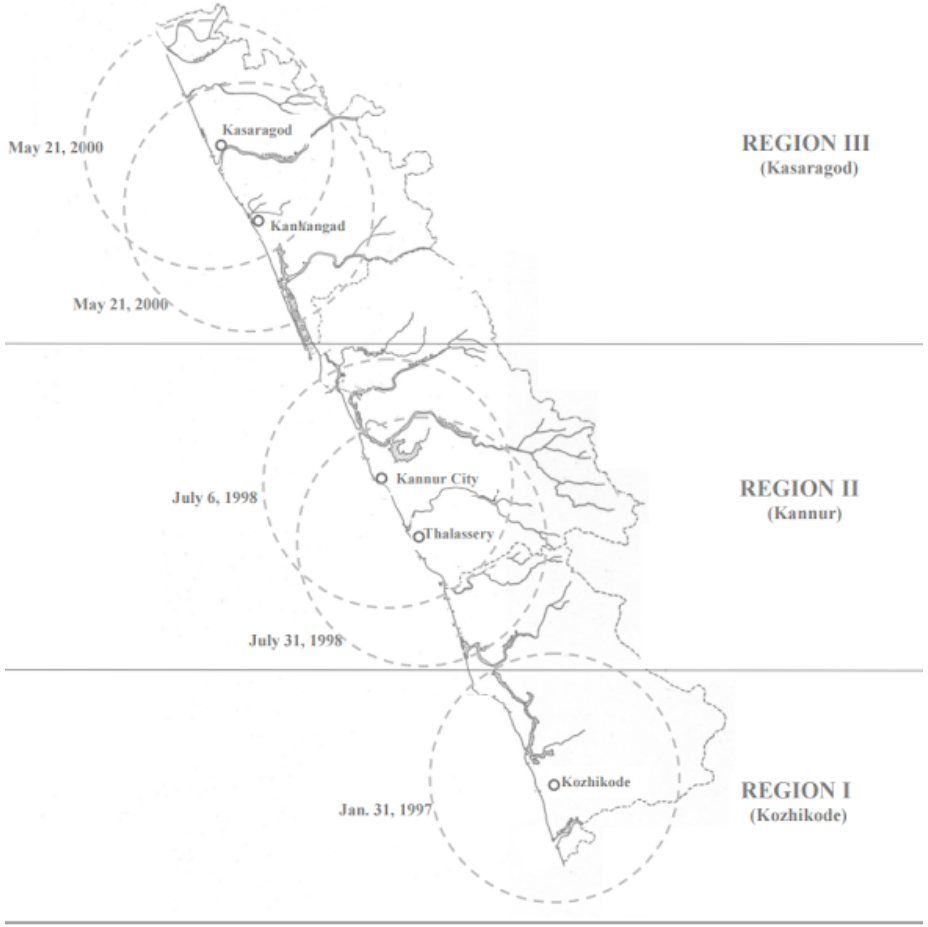


FIGURE II
SPREAD OF MOBILE PHONE COVERAGE IN KASARAGOD, KANNUR AND KOZHIKODE DISTRICTS

Additional Examples: Fish and Phones

- **Study**: Jensen (2007) uses the variation to study prices in time depending on location.
- **Results**: Significant decrease in price variation!

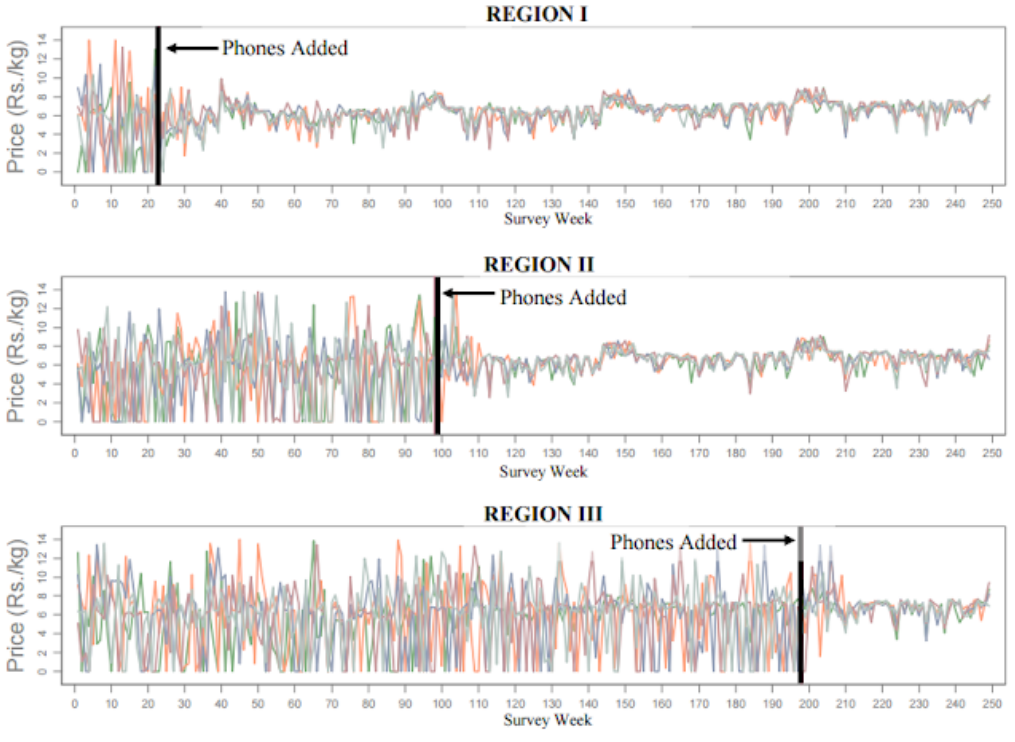


FIGURE IV
PRICES AND MOBILE PHONE SERVICE IN KERALA

Additional Examples: Natural disasters

- **Natural disasters** such as hurricanes or earthquakes usually make for great natural experiments
    - As long as you can **credibly** convince people they are not correlated with potential outcomes!
- Imagine that you have two cities, A and B, which both have equal probability of being affected by a hurricane:

City A: Large city

City B: Smaller town

- You want to study the effect of this natural disaster on unemployment. **Could you use this setting?**

Takeaway points

- We don't always need to **"randomize"**.
- We can exploit **natural variation**.
    - In **natural experiments** we are approximating an ideal RCT very straightforwardly.

References

- Angrist, J. and S. Pischke. (2015). "Mastering Metrics". *Chapter 2*.
- NBER (2021). "Oregon Health Insurance Experiment Background". *Summary of the policy*.
- Jensen, R. (2007). "The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector". *Quarterly Journal of Economics*.