FLS 6441 - Methods III: Explanation and Causation

Week 3 - Field Experiments

Jonathan Phillips

April 2019

Section 1

Rest of the Course

Rest of the Course

- The rest of the course is mostly about:
  - **Design-Based Solutions** to the Fundamental Problem of Causal Inference: Which treatment assignment mechanisms **avoid these biases** and provide plausible counterfactuals

Rest of the Course

- The rest of the course is mostly about:
  - **Design-Based Solutions** to the Fundamental Problem of Causal Inference: Which treatment assignment mechanisms **avoid these biases** and provide plausible counterfactuals
  - How much can we learn with better research design?

Rest of the Course

- The rest of the course is mostly about:
  - **Design-Based Solutions** to the Fundamental Problem of Causal Inference: Which treatment assignment mechanisms **avoid these biases** and provide plausible counterfactuals
  - How much can we learn with better research design?
  - **Model-Based Solutions:** Not so much.

Rest of the Course

		Independence of Treatment Assignment	Researcher Controls Treatment Assignment?
Controlled Experiments	Field Experiments	✓	✓
	Survey and Lab Experiments	✓	✓
Natural Experiments	Randomized Natural Experiments	✓	
	Instrumental Variables	✓	
	Discontinuities	✓	
Observational Studies	Difference-in-Differences		
	Controlling for Confounding		
	Matching		
	Comparative Cases and Process Tracing		

Section 2

Independence

Independent Treatment Assignment

- Last week, we identified why it's hard to estimate causal effects:

Independent Treatment Assignment

- Last week, we identified why it's hard to estimate causal effects:
- **The Treatment Assignment Mechanism is almost never independent of Potential Outcomes**

Independent Treatment Assignment

- Last week, we identified why it's hard to estimate causal effects:
- **The Treatment Assignment Mechanism is almost never independent of Potential Outcomes**
- So estimates of the ATE are **biased**

Independent Treatment Assignment

- Last week, we identified why it's hard to estimate causal effects:
- **The Treatment Assignment Mechanism is almost never independent of Potential Outcomes**
- So estimates of the ATE are **biased**
- The solution?

Independent Treatment Assignment

- Last week, we identified why it's hard to estimate causal effects:
- **The Treatment Assignment Mechanism is almost never independent of Potential Outcomes**
- So estimates of the ATE are **biased**
- The solution?
- **Treatment Assignment Mechanisms that *ARE* independent of potential outcomes**

Independent Treatment Assignment

- Why does Independence of Treatment Assignment help us achieve causal inference?
  - We want to estimate:

$$E(Y_1) - E(Y_0) \tag{1}$$

Independent Treatment Assignment

- Why does Independence of Treatment Assignment help us achieve causal inference?
  - We want to estimate:

$$E(Y_1) - E(Y_0) \tag{1}$$

  - Our data provides:

$$E(Y_1|D=1) \ , E(Y_0|D=0) \tag{2}$$

Independent Treatment Assignment

- Why does Independence of Treatment Assignment help us achieve causal inference?
  - We want to estimate:

$$E(Y_1) - E(Y_0) \tag{1}$$

  - Our data provides:

$$E(Y_1|D=1)\ ,\ E(Y_0|D=0) \tag{2}$$

  - With independence, $Y_1, Y_0 \perp D$:

$$E(Y_1|D=1) \quad = \quad E(Y_1)\ ,$$

Independent Treatment Assignment

- Why does Independence of Treatment Assignment help us achieve causal inference?
  - We want to estimate:

$$E(Y_1) - E(Y_0) \tag{1}$$

  - Our data provides:

$$E(Y_1|D=1)\ ,\ E(Y_0|D=0) \tag{2}$$

  - With independence, $Y_1, Y_0 \perp D$:

$$E(Y_1|D=1) = E(Y_1)\ ,\ E(Y_0|D=0) = E(Y_0) \tag{3}$$

Independent Treatment Assignment

- Why does Independence of Treatment Assignment help us achieve causal inference?
  - We want to estimate:

$$E(Y_1) - E(Y_0) \tag{1}$$

  - Our data provides:

$$E(Y_1|D=1)\ ,\ E(Y_0|D=0) \tag{2}$$

  - With independence, $Y_1, Y_0 \perp D$:

$$E(Y_1|D=1) = E(Y_1)\ ,\ E(Y_0|D=0) = E(Y_0) \tag{3}$$

$$E(Y_1|D=1) - E(Y_0|D=0) = E(Y_1) - E(Y_0) \tag{4}$$

$$\tag{5}$$

- Potential outcomes in the treatment and control groups are now **unbiased** and representative of *all* the units

Independent Treatment Assignment

- What is the treatment assignment mechanism under **randomization**?

Independent Treatment Assignment

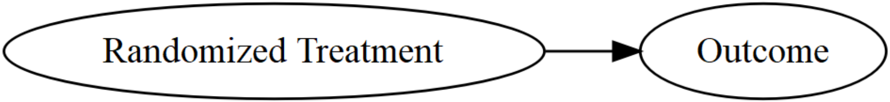
- What is the treatment assignment mechanism under **randomization**?
  - It has nothing to do with potential outcomes!
    - Every unit has **exactly the same** probability of treatment
    - No omitted variable bias is possible
    - No self-selection is possible
    - No reverse causation is possible

Independent Treatment Assignment

- This is the **entire** causal diagram:

Independent Treatment Assignment

- This is the **entire** causal diagram:



Independent Treatment Assignment

- But our logic works only based on **expectations** (averages)

Independent Treatment Assignment

- But our logic works only based on **expectations** (averages)
  - *On average*, potential outcomes will be balanced

Independent Treatment Assignment

- But our logic works only based on **expectations** (averages)
  - *On average*, potential outcomes will be balanced
  - That's more likely in larger samples

Independent Treatment Assignment

- But our logic works only based on **expectations** (averages)
  - *On average*, potential outcomes will be balanced
  - That's more likely in larger samples
  - Less likely in small samples; by chance, potential outcomes may be biased

Independent Treatment Assignment

- But our logic works only based on **expectations** (averages)
  - *On average*, potential outcomes will be balanced
  - That's more likely in larger samples
  - Less likely in small samples; by chance, potential outcomes may be biased
  - We have no way of *verifying* if potential outcomes are biased

Section 3

Analyzing Field Experiments

Analyzing Field Experiments

- If treatment is random we know that:

$$E(Y_1|D=1)-E(Y_0|D=0)=E(Y_1)-E(Y_0)$$

Analyzing Field Experiments

- If treatment is random we know that:

$$E(Y_1|D=1)-E(Y_0|D=0)=E(Y_1)-E(Y_0)$$

- What is $E(Y_1|D=1)$?

Analyzing Field Experiments

- If treatment is random we know that:

$$E(Y_1|D=1)-E(Y_0|D=0)=E(Y_1)-E(Y_0)$$

- What is $E(Y_1|D=1)$?
- What is $E(Y_0|D=0)$?

Analyzing Field Experiments

- If treatment is random we know that:

$$E(Y_1|D=1)-E(Y_0|D=0)=E(Y_1)-E(Y_0)$$

- What is $E(Y_1|D=1)$?
- What is $E(Y_0|D=0)$?
- This is easy!

Analyzing Field Experiments

- If treatment is random we know that:

$$E(Y_1|D=1) - E(Y_0|D=0) = E(Y_1) - E(Y_0)$$

- What is $E(Y_1|D=1)$?
- What is $E(Y_0|D=0)$?
- This is easy!
- Just the difference in outcome means between treatment and control units

Analyzing Field Experiments

- If treatment is random we know that:

$$E(Y_1|D=1)-E(Y_0|D=0)=E(Y_1)-E(Y_0)$$

- What is $E(Y_1|D=1)$?
- What is $E(Y_0|D=0)$?
- This is easy!
- Just the difference in outcome means between treatment and control units
  - And a simple T-test for statistical significance

Analyzing Field Experiments

- If treatment is random we know that:

$$E(Y_1|D=1)-E(Y_0|D=0)=E(Y_1)-E(Y_0)$$

- What is $E(Y_1|D=1)$?
- What is $E(Y_0|D=0)$?
- This is easy!
- Just the difference in outcome means between treatment and control units
  - And a simple T-test for statistical significance
  - NO modelling assumptions (“non-parametric”)

Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test

Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test

$$Y_i \sim \alpha + \beta D_i = \epsilon_i$$

Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test

$$Y_i \sim \alpha + \beta D_i = \epsilon_i$$

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i + \epsilon_i$$

Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test

$$Y_i \sim \alpha + \beta D_i = \epsilon_i$$

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) D_i + \epsilon_i$$

Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test

Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test
- T-test Results:

	estimate	statistic	p.value
1	0.27065	2.69475	0.00706

Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test
- T-test Results:

	estimate	statistic	p.value
1	0.27065	2.69475	0.00706

- Regression Results:

	term	estimate	std.error	statistic	p.value
1	(Intercept)	0.03459	0.07110	0.48647	0.62664
2	treatment	0.27065	0.10044	2.69472	0.00706

Analyzing Field Experiments

- How do we randomize?
  - Hard! We can't just 'pick' treated units off the top of our heads

Analyzing Field Experiments

- How do we randomize?
  - Hard! We can't just 'pick' treated units off the top of our heads
  - Computers are deterministic

Analyzing Field Experiments

- How do we randomize?
  - Hard! We can't just 'pick' treated units off the top of our heads
  - Computers are deterministic
  - The best we can do is to use atmospheric noise or radioactive decay

Analyzing Field Experiments

- How do we randomize?
  - Hard! We can't just 'pick' treated units off the top of our heads
  - Computers are deterministic
  - The best we can do is to use atmospheric noise or radioactive decay
- In the real world, randomization is hard
  - Pressure to help the most needy

Analyzing Field Experiments

- How do we randomize?
  - Hard! We can't just 'pick' treated units off the top of our heads
  - Computers are deterministic
  - The best we can do is to use atmospheric noise or radioactive decay
- In the real world, randomization is hard
  - Pressure to help the most needy
  - Political pressure

Analyzing Field Experiments

- How do we randomize?
  - Hard! We can't just 'pick' treated units off the top of our heads
  - Computers are deterministic
  - The best we can do is to use atmospheric noise or radioactive decay
- In the real world, randomization is hard
  - Pressure to help the most needy
  - Political pressure
  - We don't want to be guinea pigs!

Analyzing Field Experiments

- How do we randomize?

Analyzing Field Experiments

- How do we randomize?
- So how do we confirm that randomization has succeeded?

Analyzing Field Experiments

- How do we randomize?
- So how do we confirm that randomization has succeeded?
  - We can't directly test potential outcomes

1. **Qualitative research:** to reconstruct the treatment process

Analyzing Field Experiments

- How do we randomize?
- So how do we confirm that randomization has succeeded?
  - We can't directly test potential outcomes

1. **Qualitative research:** to reconstruct the treatment process
2. **Balance tests:** We can directly test other variables between treatment and control
   - Randomization balances *all* variables, not just potential outcomes

Section 4

Implementing Field Experiments

Section 5

Designing Field Experiments