FLS 6441 - Methods III: Explanation and Causation

Week 3 - Field Experiments

Jonathan Phillips

April 2019

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 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 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 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 1

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 depends on Potential Outcomes**

Independent Treatment Assignment

- Last week, we identified why it's hard to estimate causal effects:
- **The Treatment Assignment Mechanism depends on 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 depends on 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 depends on 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, $Y_1, Y_0 \perp D$:

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

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!

Independent Treatment Assignment

- What is the treatment assignment mechanism under **randomization**?
  - It has nothing to do with potential outcomes!
  - So we get a representative sample of Y_0 and Y_1

Independent Treatment Assignment

- What is the treatment assignment mechanism under **randomization**?
  - It has nothing to do with potential outcomes!
  - So we get a representative sample of Y_0 and Y_1
    - Every unit has **exactly the same** probability of treatment

Independent Treatment Assignment

- What is the treatment assignment mechanism under **randomization**?
  - It has nothing to do with potential outcomes!
  - So we get a representative sample of Y_0 and Y_1
    - Every unit has **exactly the same** probability of treatment
    - Potential outcomes are ’Completely Missing at Random’

Independent Treatment Assignment

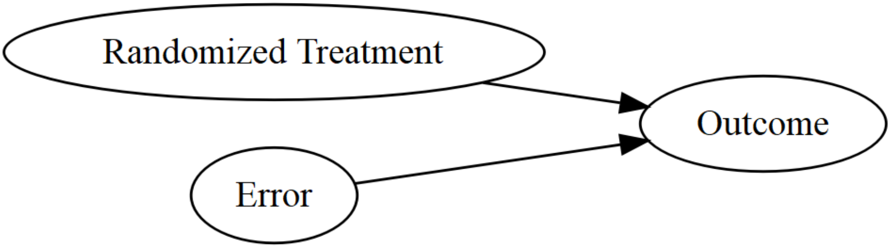
- What is the treatment assignment mechanism under **randomization**?
  - It has nothing to do with potential outcomes!
  - So we get a representative sample of Y_0 and Y_1
    - Every unit has **exactly the same** probability of treatment
    - Potential outcomes are 'Completely Missing at Random'
    - 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

- Why does randomization remove selection bias?
- Assume: $Y_{1i} = Y_{0i} + \alpha$, where α is the real constant treatment effect

$$A\hat{T}E = E(Y_1|D=1) - E(Y_0|D=0)$$

Independent Treatment Assignment

- Why does randomization remove selection bias?
- Assume: $Y_{1i} = Y_{0i} + \alpha$, where α is the real constant treatment effect

$$\hat{ATE} = E(Y_1|D=1) - E(Y_0|D=0)$$

$$\hat{ATE} = \underbrace{\alpha}_{\text{Real ATE}} + \underbrace{E(Y_0|D=1) - E(Y_0|D=0)}_{\text{Bias}}$$

- Now, use the Independence of Treatment Assignment:

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

$$\hat{ATE} = \underbrace{\alpha}_{\text{Real ATE}}$$

- This works for observable *and* unobservable variables

Independent Treatment Assignment

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

Independent Treatment Assignment

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

Independent Treatment Assignment

- But this 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 this 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 this 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

Balance in Repeateed Experiments

-

Section 2

Analysis

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

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 ($Y_i = \alpha + \beta D_i = \epsilon_i$):

	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

Repeated Experiments

- The results from one experiment are not perfect

Repeated Experiments

- The results from one experiment are not perfect
- Treatment effects are still *probabilistic* (random variables) so we may get the wrong answer by chance

Repeated Experiments

- The results from one experiment are not perfect
- Treatment effects are still *probabilistic* (random variables) so we may get the wrong answer by chance
- In repeated experiments, 95% of confidence intervals will cross the true treatment effect

```
## Error in eval(predvars, data, env):  object 'Y'
not found
```

Clustered Treatments

- Clustered sampling: To reduce data collection costs

Clustered Treatments

- Clustered sampling: To reduce data collection costs
- Clustered treatment: To reduce implementation costs, or to test specific theories

Clustered Treatments

- Clustered sampling: To reduce data collection costs
- Clustered treatment: To reduce implementation costs, or to test specific theories
- Eg. Holding Town Hall meetings does not make sense at the individual level

Clustered Treatments

- Clustered sampling: To reduce data collection costs
- Clustered treatment: To reduce implementation costs, or to test specific theories
- Eg. Holding Town Hall meetings does not make sense at the individual level
- If treatment (or sampling) are clustered, we will have dependencies in our errors - closer people are more similar

Clustered Treatments

- Clustered sampling: To reduce data collection costs
- Clustered treatment: To reduce implementation costs, or to test specific theories
- Eg. Holding Town Hall meetings does not make sense at the individual level
- If treatment (or sampling) are clustered, we will have dependencies in our errors - closer people are more similar
- So standard errors must be clustered at the level of treatment (eg. village)

Clustered Treatments

- Clustered sampling: To reduce data collection costs
- Clustered treatment: To reduce implementation costs, or to test specific theories
- Eg. Holding Town Hall meetings does not make sense at the individual level
- If treatment (or sampling) are clustered, we will have dependencies in our errors - closer people are more similar
- So standard errors must be clustered at the level of treatment (eg. village)
- In general, inference is more efficient with more higher-level units (more villages, less people per village)

Clustered Treatments

- Clustered sampling: To reduce data collection costs
- Clustered treatment: To reduce implementation costs, or to test specific theories
- Eg. Holding Town Hall meetings does not make sense at the individual level
- If treatment (or sampling) are clustered, we will have dependencies in our errors - closer people are more similar
- So standard errors must be clustered at the level of treatment (eg. village)
- In general, inference is more efficient with more higher-level units (more villages, less people per village)
    - But there is usually a cost trade-off

Controlling for Covariates

- Do we need to control for covariates in experiments?

Controlling for Covariates

- Do we need to control for covariates in experiments?
- If randomization worked and the sample size is large, usually not

Controlling for Covariates

- Do we need to control for covariates in experiments?
- If randomization worked and the sample size is large, usually not
- Three reasons to include controls:
  - Small sample, but note causal inference is now model-dependent

Controlling for Covariates

- Do we need to control for covariates in experiments?
- If randomization worked and the sample size is large, usually not
- Three reasons to include controls:
  - Small sample, but note causal inference is now model-dependent
  - Chance/residual imbalance on a specific variable which we want to adjust for

Controlling for Covariates

- Do we need to control for covariates in experiments?
- If randomization worked and the sample size is large, usually not
- Three reasons to include controls:
  - Small sample, but note causal inference is now model-dependent
  - Chance/residual imbalance on a specific variable which we want to adjust for
  - To improve precision, i.e. reduce the standard errors on β
    - The more variation in Y we can explain with covariates, the more certain we can be on the effectof D

Other Quantities of Interest

- Average Treatment Effects are just one summary statistic

Other Quantities of Interest

- Average Treatment Effects are just one summary statistic
- Can be influenced by outliers

Other Quantities of Interest

- Average Treatment Effects are just one summary statistic
- Can be influenced by outliers
- What if an average effect of +5% income leaves half the population hugely rich and half very poor?

Other Quantities of Interest

- Average Treatment Effects are just one summary statistic
- Can be influenced by outliers
- What if an average effect of +5% income leaves half the population hugely rich and half very poor?
- Average treatment effects are easiest (difference-in-means equals mean-difference)

Other Quantities of Interest

- Average Treatment Effects are just one summary statistic
- Can be influenced by outliers
- What if an average effect of +5% income leaves half the population hugely rich and half very poor?
- Average treatment effects are easiest (difference-in-means equals mean-difference)
- But we can also estimate Quantile treatment effects, eg. the effect of treatment on the bottom 10% of the distribution

Section 3

Assumptions

Assumptions

1. Compliance with Randomization procedure
2. Randomization produced balance on potential outcomes
3. SUTVA
4. Excludability

1. Compliance with Randomization procedure

- Randomization is unpopular

1. Compliance with Randomization procedure

- Randomization is unpopular
- Need to verify treatment allocation
    - Transparency, documentation

1. Compliance with Randomization procedure

- Randomization is unpopular
- Need to verify treatment allocation
  - Transparency, documentation
- And treatment compliance
  - Did anyone assigned to control manage to get treatment?

1. Compliance with Randomization procedure

- Randomization is unpopular
- Need to verify treatment allocation
  - Transparency, documentation
- And treatment compliance
  - Did anyone assigned to control manage to get treatment?
- **Design:** Double-blind assignment
- **Checks:** Qualitative fieldwork

2. Randomization Produced Balanced Potential Outcomes

- Impossible to Test!

2. Randomization Produced Balanced Potential Outcomes

- Impossible to Test!
- But we can test observable pre-treatment covariates

2. Randomization Produced Balanced Potential Outcomes

- Impossible to Test!
- But we can test observable pre-treatment covariates
- If covariates are the same in the treatment and control groups, this variable *cannot* explain any differences in outcomes

2. Randomization Produced Balanced Potential Outcomes

- Impossible to Test!
- But we can test observable pre-treatment covariates
- If covariates are the same in the treatment and control groups, this variable *cannot* explain any differences in outcomes
- **Check:** Normally a difference in means T-test

2. Randomization Produced Balanced Potential Outcomes

- Impossible to Test!
- But we can test observable pre-treatment covariates
- If covariates are the same in the treatment and control groups, this variable *cannot* explain any differences in outcomes
- **Check:** Normally a difference in means T-test
- **Check:** Or a Kolmogorov-Smirnov Test of identical distributions

2. Randomization Produced Balanced Potential Outcomes

- What if a balance test comes back with a p-value < 0.05?

2. Randomization Produced Balanced Potential Outcomes

- What if a balance test comes back with a p-value < 0.05?
- It probably will!
    1. We are testing many variables, so some differences arise by chance

2. Randomization Produced Balanced Potential Outcomes

- What if a balance test comes back with a p-value < 0.05?
- It probably will!
    1. We are testing many variables, so some differences arise by chance
    2. We have a large N, so we can detect very small differences

2. Randomization Produced Balanced Potential Outcomes

- What if a balance test comes back with a p-value < 0.05?
- It probably will!
  1. We are testing many variables, so some differences arise by chance
  2. We have a large N, so we can detect very small differences
- **Check:** For balance, **What matters are *substantive* differences, not p-values**

2. Randomization Produced Balanced Potential Outcomes

- What if a balance test comes back with a p-value < 0.05?
- It probably will!
  1. We are testing many variables, so some differences arise by chance
  2. We have a large N, so we can detect very small differences
- **Check:** For balance, **What matters are *substantive* differences, not p-values**
- Two safety nets:
  1. **Analysis:** We can still include covariates in our analysis, controlling for 'residual' imbalance

2. Randomization Produced Balanced Potential Outcomes

- What if a balance test comes back with a p-value < 0.05?
- It probably will!
  1. We are testing many variables, so some differences arise by chance
  2. We have a large N, so we can detect very small differences
- **Check:** For balance, **What matters are *substantive* differences, not p-values**
- Two safety nets:
  1. **Analysis:** We can still include covariates in our analysis, controlling for 'residual' imbalance
  2. **Analysis:** We are using p-values in our *analysis*, which take into account 'chance' imbalance

2. Randomization Produced Balanced Potential Outcomes

- What if a balance test comes back with a p-value < 0.05?
- It probably will!
  1. We are testing many variables, so some differences arise by chance
  2. We have a large N, so we can detect very small differences
- **Check:** For balance, **What matters are *substantive* differences, not p-values**
- Two safety nets:
  1. **Analysis:** We can still include covariates in our analysis, controlling for ’residual’ imbalance
  2. **Analysis:** We are using p-values in our *analysis*, which take into account ’chance’ imbalance

3. SUTVA

- Stable Unit Treatment Value Assumption = **No Spillovers**

3. SUTVA

- Stable Unit Treatment Value Assumption = **No Spillovers**
- Technically, treatment of unit j does not affect the potential outcomes for unit i

3. SUTVA

- Stable Unit Treatment Value Assumption = **No Spillovers**
- Technically, treatment of unit j does not affect the potential outcomes for unit i

$$(Y_{1i}, Y_{0i}) \perp D_j$$

3. SUTVA

- Stable Unit Treatment Value Assumption = **No Spillovers**
- Technically, treatment of unit j does not affect the potential outcomes for unit i

$$(Y_{1i}, Y_{0i}) \perp D_j$$

$$Y_i(D_i, D_j, D_k, D_l, D_m, D_n, D_o, D_p \ldots) = Y_i(D_i)$$

3. SUTVA

- Stable Unit Treatment Value Assumption = **No Spillovers**
- Technically, treatment of unit j does not affect the potential outcomes for unit i

$$(Y_{1i}, Y_{0i}) \perp D_j$$

$$Y_i(D_i, D_j, D_k, D_l, D_m, D_n, D_o, D_p \ldots) = Y_i(D_i)$$

- But spillovers are common! If you get an award, I might feel more motivated or less motivated

3. SUTVA

- Stable Unit Treatment Value Assumption = **No Spillovers**
- Technically, treatment of unit j does not affect the potential outcomes for unit i

$$(Y_{1i}, Y_{0i}) \perp D_j$$

$$Y_i(D_i, D_j, D_k, D_l, D_m, D_n, D_o, D_p \ldots) = Y_i(D_i)$$

- But spillovers are common! If you get an award, I might feel more motivated or less motivated
- Why are spillovers a problem?

3. SUTVA

- Stable Unit Treatment Value Assumption = **No Spillovers**
- Technically, treatment of unit j does not affect the potential outcomes for unit i

$$(Y_{1i}, Y_{0i}) \perp D_j$$

$$Y_i(D_i, D_j, D_k, D_l, D_m, D_n, D_o, D_p \ldots) = Y_i(D_i)$$

- But spillovers are common! If you get an award, I might feel more motivated or less motivated
- Why are spillovers a problem?
  - **Design:** Limit risk of spillovers, eg. leave 20 miles between each unit
  - **Check:** Qualitative fieldwork
  - **Analysis:** Try to *measure* spillovers

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**
- Assignment to treatment causes a **’second’** treatment

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**
- Assignment to treatment causes a **’second’** treatment
- Eg. We share information about specific politicians on the radio, but the politicians then counter with their own broadcasts

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**
- Assignment to treatment causes a **'second'** treatment
- Eg. We share information about specific politicians on the radio, but the politicians then counter with their own broadcasts
- Our treatment effect is no longer *only* the effect of our information

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**
- Assignment to treatment causes a **’second’** treatment
- Eg. We share information about specific politicians on the radio, but the politicians then counter with their own broadcasts
- Our treatment effect is no longer *only* the effect of our information
- Or do we want to measure these additional effects?

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**
- Some responses to treatment we want to capture ('net effects')
  - Eg. One reason richer families change their attitudes to government is because they start paying taxes

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**
- Some responses to treatment we want to capture ('net effects')
  - Eg. One reason richer families change their attitudes to government is because they start paying taxes
- Others we don't want to capture
  - Eg. Measurement bias: Researchers treat treated units differently and record higher outcomes for them

4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**
- Some responses to treatment we want to capture ('net effects')
  - Eg. One reason richer families change their attitudes to government is because they start paying taxes
- Others we don't want to capture
  - Eg. Measurement bias: Researchers treat treated units differently and record higher outcomes for them
  - Or Hawthorne Effects arising from being studied, not treatment (more next week)
- *Design:* Careful specification of treatment and control

4. Excludability

- What if we find zero effect of government investment of \$1000 in healthcare on health outcomes, because households responded by reducing their spending by exactly \$1000?

4. Excludability

- What if we find zero effect of government investment of $1000 in healthcare on health outcomes, because households responded by reducing their spending by exactly $1000?
- Experimental treatment effects capture *all* net downstream effects

Section 4

Implementing Field Experiments

Implementing Field Experiments

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

Implementing Field Experiments

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

Implementing 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

Implementing 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

Implementing 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

Implementing 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!

Implementing Field Experiments

- How do we randomize?

Implementing Field Experiments

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

Implementing 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

Implementing 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

Implementing Field Experiments

- How do we randomize?

Implementing Field Experiments

- How do we randomize?
- Three options to assign treatment and control ’independent’ of potential outcomes:

Implementing Field Experiments

- How do we randomize?
- Three options to assign treatment and control ’independent’ of potential outcomes:
  - We have N units and want equal probability of treatment for each:

1. Flip a coin for every unit so every unit has probability 0.5 of treatment

Implementing Field Experiments

- How do we randomize?
- Three options to assign treatment and control 'independent' of potential outcomes:
  - We have N units and want equal probability of treatment for each:

1. Flip a coin for every unit so every unit has probability 0.5 of treatment
2. Randomize the order of the units and assign the first $\frac{N}{2}$ units to treatment

Implementing Field Experiments

- How do we randomize?
- Three options to assign treatment and control 'independent' of potential outcomes:
  - We have N units and want equal probability of treatment for each:

1. Flip a coin for every unit so every unit has probability 0.5 of treatment
2. Randomize the order of the units and assign the first $\frac{N}{2}$ units to treatment
3. Pair units and flip a coin to assign one to treatment so exactly $\frac{N}{2}$ get treatment

Implementing Field Experiments

- How do we randomize?
- Three options to assign treatment and control 'independent' of potential outcomes:
  - We have N units and want equal probability of treatment for each:

1. Flip a coin for every unit so every unit has probability 0.5 of treatment
2. Randomize the order of the units and assign the first $\frac{N}{2}$ units to treatment
3. Pair units and flip a coin to assign one to treatment so exactly $\frac{N}{2}$ get treatment

- What's the difference between these three options?

Implementing Field Experiments

- How do we randomize?
- Three options to assign treatment and control 'independent' of potential outcomes:
  - We have N units and want equal probability of treatment for each:

1. Flip a coin for every unit so every unit has probability 0.5 of treatment
2. Randomize the order of the units and assign the first $\frac{N}{2}$ units to treatment
3. Pair units and flip a coin to assign one to treatment so exactly $\frac{N}{2}$ get treatment

- What's the difference between these three options?
- What % treated? 50:50 is usually most efficient

Implementing Field Experiments

- **Blocking**

Implementing Field Experiments

- **Blocking**
- Randomization is *inefficient* and risky

Implementing Field Experiments

- **Blocking**
- Randomization is *inefficient* and risky
- We know we need balance on key covariates, eg. gender, democracy
  - So why leave this to chance??

Implementing Field Experiments

- **Blocking**
- Randomization is *inefficient* and risky
- We know we need balance on key covariates, eg. gender, democracy
  - So why leave this to chance??
  - We can measure these variables and *enforce* balance (50% female in both treatment and control)

Implementing Field Experiments

- **Blocking**
- Randomization is *inefficient* and risky
- We know we need balance on key covariates, eg. gender, democracy
  - So why leave this to chance??
  - We can measure these variables and *enforce* balance (50% female in both treatment and control)
  - Blocking means randomizing *within* fixed groups

Implementing Field Experiments

- **Blocking**
- Randomization is *inefficient* and risky
- We know we need balance on key covariates, eg. gender, democracy
  - So why leave this to chance??
  - We can measure these variables and *enforce* balance (50% female in both treatment and control)
  - Blocking means randomizing *within* fixed groups
  - Eg. We have 10 states and a sample size of 5000 - so we fix 250 treated and 250 control in each state
- "Block what you can; randomize what you cannot"

Implementing Field Experiments

- Random treatment vs. Random samples

Implementing Field Experiments

- Random treatment vs. Random samples
- Representative potential outcomes vs. Sample representative of larger population

Implementing Field Experiments

- Random treatment vs. Random samples
- Representative potential outcomes vs. Sample representative of larger population
- Causal inference vs. Statistical inference

Implementing Field Experiments

- Random treatment vs. Random samples
- Representative potential outcomes vs. Sample representative of larger population
- Causal inference vs. Statistical inference
- Both work in the same way - randomization avoids selection (into the data/treatment)

Section 5

Critiquing Field Experiments

Critiquing Field Experiments

- Field experiments are easy to evaluate. What can go wrong??

1. Results are a Black Box

- We know that D causes Y in this population. So what?

1. Results are a Black Box

- We know that D causes Y in this population. So what?
- We know that giving people money causes them to vote more often. So what?

1. Results are a Black Box

- We know that D causes Y in this population. So what?
- We know that giving people money causes them to vote more often. So what?
- We want to learn about generalizable political processes.

1. Results are a Black Box

- We know that D causes Y in this population. So what?
- We know that giving people money causes them to vote more often. So what?
- We want to learn about generalizable political processes.
- What theory is this testing?

1. Results are a Black Box

- We know that D causes Y in this population. So what?
- We know that giving people money causes them to vote more often. So what?
- We want to learn about generalizable political processes.
- What theory is this testing? Does it reject any theory?

2. Generalizability of Context

- Our causal conclusions are **specific to the population we drew our sample from**

2. Generalizability of Context

- Our causal conclusions are **specific to the population we drew our sample from**
- Income makes attitudes to redistribution more negative in the USA

2. Generalizability of Context

- Our causal conclusions are **specific to the population we drew our sample from**
- Income makes attitudes to redistribution more negative in the USA
  - What is the effect in Angola?

2. Generalizability of Context

- Our causal conclusions are **specific to the population we drew our sample from**
- Income makes attitudes to redistribution more negative in the USA
  - What is the effect in Angola?
- Secondary school education leads to more conservative voting

2. Generalizability of Context

- Our causal conclusions are **specific to the population we drew our sample from**
- Income makes attitudes to redistribution more negative in the USA
  - What is the effect in Angola?
- Secondary school education leads to more conservative voting
  - What is the effect of university education?

2. Generalizability of Context

- Our causal conclusions are **specific to the population we drew our sample from**
- Income makes attitudes to redistribution more negative in the USA
  - What is the effect in Angola?
- Secondary school education leads to more conservative voting
  - What is the effect of university education?
- Yes, you randomly sampled and randomly assigned treatment, but the population you drew from was not the full population

2. Generalizability of Context

- Our causal conclusions are **specific to the population we drew our sample from**
- Income makes attitudes to redistribution more negative in the USA
  - What is the effect in Angola?
- Secondary school education leads to more conservative voting
  - What is the effect of university education?
- Yes, you randomly sampled and randomly assigned treatment, but the population you drew from was not the full population
  - Experiments are only possible where people agree, and those places are not representatie

2. Generalizability of Context

- Our causal conclusions are **specific to the population we drew our sample from**
- Income makes attitudes to redistribution more negative in the USA
  - What is the effect in Angola?
- Secondary school education leads to more conservative voting
  - What is the effect of university education?
- Yes, you randomly sampled and randomly assigned treatment, but the population you drew from was not the full population
  - Experiments are only possible where people agree, and those places are not representatie
  - Selection bias has come back!

3. Generalizability of Treatment

- The effect of an education intervention in an experiment in Butantã raised test scores by 20%, and was evaluated and verified by USP

3. Generalizability of Treatment

- The effect of an education intervention in an experiment in Butantã raised test scores by 20%, and was evaluated and verified by USP
- The government expands the program nationwide. Do Brazilian students' scores improve on average by 20%?

3. Generalizability of Treatment

- The effect of an education intervention in an experiment in Butantã raised test scores by 20%, and was evaluated and verified by USP
- The government expands the program nationwide. Do Brazilian students' scores improve on average by 20%?
- Three problems:
    1. **Implementation Varies:** Implementing at scale is **hard**, costly and requires delegation to less motivated and skilled actors. Oversight pressure is also reduced.

3. Generalizability of Treatment

- The effect of an education intervention in an experiment in Butantã raised test scores by 20%, and was evaluated and verified by USP
- The government expands the program nationwide. Do Brazilian students' scores improve on average by 20%?
- Three problems:
    1. **Implementation Varies:** Implementing at scale is **hard**, costly and requires delegation to less motivated and skilled actors. Oversight pressure is also reduced.
    2. **Ownership and Excludability:** Telling someone to implement an intervention is different from working with a self-motivated actor who designed the intervention. Knowing you were randomly assigned to treatment rather than choosing treatment changes political ownership, perceptions and motivation.

3. Generalizability of Treatment

- The effect of an education intervention in an experiment in Butantã raised test scores by 20%, and was evaluated and verified by USP
- The government expands the program nationwide. Do Brazilian students' scores improve on average by 20%?
- Three problems:
  1. **Implementation Varies:** Implementing at scale is **hard**, costly and requires delegation to less motivated and skilled actors. Oversight pressure is also reduced.
  2. **Ownership and Excludability:** Telling someone to implement an intervention is different from working with a self-motivated actor who designed the intervention. Knowing you were randomly assigned to treatment rather than choosing treatment changes political ownership, perceptions and motivation.
  3. **General Equilibrium Effects:** Average test scores went from 75% to 95%, so the exam board readjusted the test and made it harder.

3. Generalizability of Treatment

- Eg. The Millennium Villages Project

3. Generalizability of Treatment

- Eg. The Millennium Villages Project
- WB/UN/Columbia University tried to invest USD$120 per person in 14 African villages

3. Generalizability of Treatment

- Eg. The Millennium Villages Project
- WB/UN/Columbia University tried to invest USD$120 per person in 14 African villages
- Mixed but positive results: crop yields increased 85-350%, malaria reduced 50%

3. Generalizability of Treatment

- Eg. The Millennium Villages Project
- WB/UN/Columbia University tried to invest USD$120 per person in 14 African villages
- Mixed but positive results: crop yields increased 85-350%, malaria reduced 50%
- But sites were not representative (close to main roads and cities so they're easy to visit)

3. Generalizability of Treatment

- Eg. The Millennium Villages Project
- WB/UN/Columbia University tried to invest USD$120 per person in 14 African villages
- Mixed but positive results: crop yields increased 85-350%, malaria reduced 50%
- But sites were not representative (close to main roads and cities so they're easy to visit)
- Treatment could not be scaled (Every village cannot get visits from Columbia professors twice a year)

3. Generalizability of Treatment

- Eg. The Millennium Villages Project
- WB/UN/Columbia University tried to invest USD$120 per person in 14 African villages
- Mixed but positive results: crop yields increased 85-350%, malaria reduced 50%
- But sites were not representative (close to main roads and cities so they're easy to visit)
- Treatment could not be scaled (Every village cannot get visits from Columbia professors twice a year)
- And politics was ignored (No implementation unless you give them responsibility, but lose control)

4. Skewed Learning

- Research focuses on where experiments are most possible, not where it is most needed

4. Skewed Learning

- Research focuses on where experiments are most possible, not where it is most needed
- Selection bias in research findings

4. Skewed Learning

- Research focuses on where experiments are most possible, not where it is most needed
- Selection bias in research findings
-