# FLS 6441 - Methods III: Explanation and Causation

## Week 3 - Field Experiments

Jonathan Phillips

March 2020

# Rest of the Course

- The rest of the course is mostly about:
  - **Design-Based Solutions** to the Fundamental Problem of Causal Inference:

# Rest of the Course

- The rest of the course is mostly about:
  - **Design-Based Solutions** to the Fundamental Problem of Causal Inference:
    - Finding treatment assignment mechanisms that **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:
    - Finding treatment assignment mechanisms that **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:
    - Finding treatment assignment mechanisms that **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** | ✓ | ✓ |
| **Natural Experiments** | ✓ | |
| **Observational Studies** | | |

# 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:

$$E(Y_1) - E(Y_0) \quad (1)$$

## 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, $Y_1, Y_0 \perp D$:

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

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

## 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, $Y_1, Y_0 \perp D$:

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

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

$$E(Y_1|D=1) - E(Y_0|D=0) = E(Y_1) - E(Y_0) \tag{5}$$

## 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, $Y_1, Y_0 \perp D$:

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

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

$$E(Y_1|D=1) - E(Y_0|D=0) = E(Y_1) - E(Y_0) \tag{5}$$

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

# Randomization

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

## Randomization

- What is the treatment assignment mechanism under **randomization**?
  - *By definition* it has nothing to do with potential outcomes!

## Randomization

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

# Randomization

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

## Randomization

- What is the treatment assignment mechanism under **randomization**?
  - *By definition* 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'

# Randomization

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

# Randomization

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

# Randomization

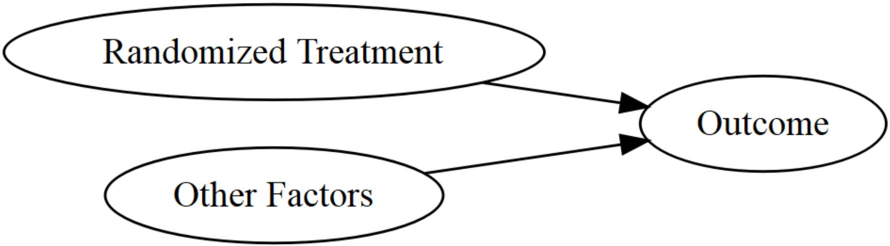
- What is the treatment assignment mechanism under **randomization**?
  - *By definition* it has nothing to do with potential outcomes!
  - So we get a representative sample of $Y_0$ and $Y_1$
    - Every unit has **exactly the same** probability of treatment
    - Potential outcomes are 'Completely Missing at Random'
    - No omitted variable bias is possible
    - No self-selection is possible
    - No reverse causation is possible

## Randomization

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

## Randomization

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



## Randomization

- Why does randomization remove selection bias?

## Randomization

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

## Randomization

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

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

## Randomization

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

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

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

## Randomization

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

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

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

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

## Randomization

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

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

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

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

- Now, use the Independence of Treatment Assignment:

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

## Randomization

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

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

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

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

- Now, use the Independence of Treatment Assignment:

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

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

## Randomization

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

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

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

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

- 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 influences

# Randomization

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

## Randomization

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

# Randomization

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

## Randomization

- 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

## Randomization

- 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

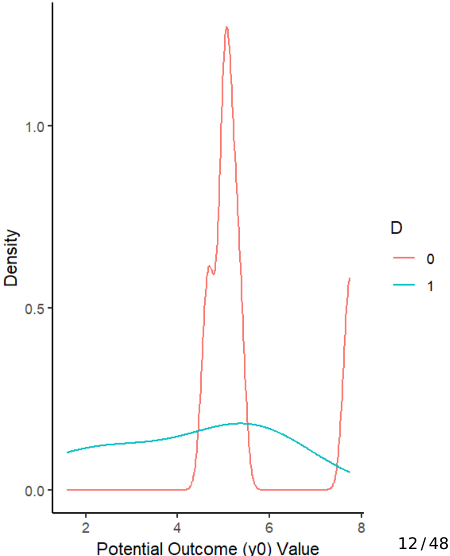
## Balance in Randomized Experiments

- Balance on potential outcomes is unlikely in small samples

## Balance in Randomized Experiments

- Balance on potential outcomes is unlikely in small samples

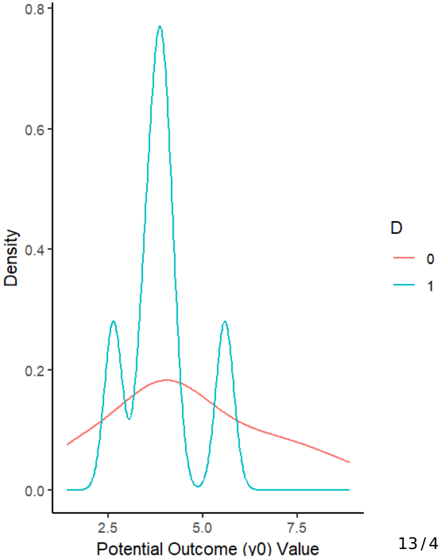
N=10



## Balance in Randomized Experiments

- Balance on potential outcomes is unlikely in small samples
- But the Law of Large Numbers helps us in large samples

N=20

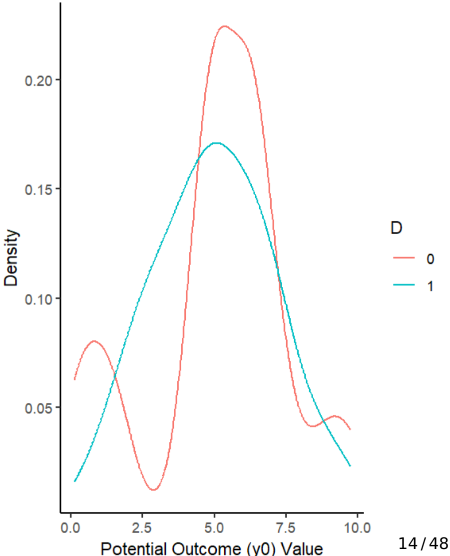


# Balance in Randomized Experiments

- Balance on potential outcomes is unlikely in small samples

- But the Law of Large Numbers helps us in large samples

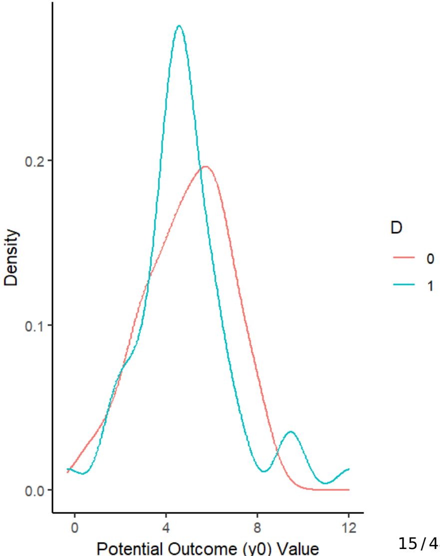
N=50



# Balance in Randomized Experiments

- Balance on potential outcomes is unlikely in small samples
- But the Law of Large Numbers helps us in large samples

N=100

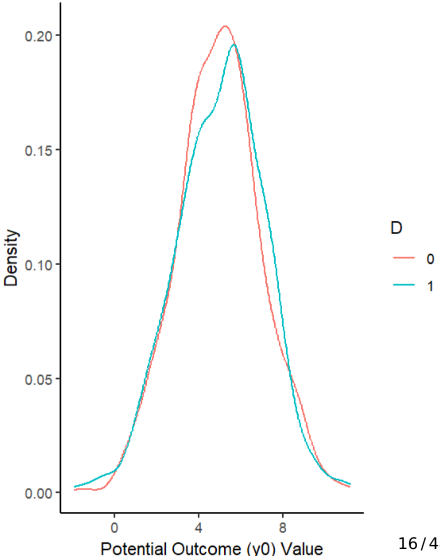


# Balance in Randomized Experiments

- Balance on potential outcomes is unlikely in small samples

- But the Law of Large Numbers helps us in large samples

N=1000

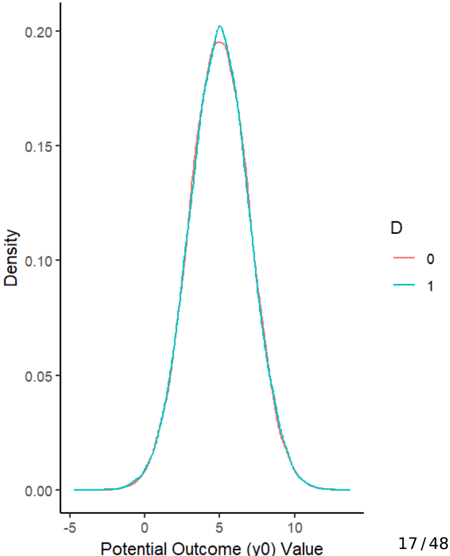


# Balance in Randomized Experiments

- Balance on potential outcomes is unlikely in small samples

- But the Law of Large Numbers helps us in large samples

N=100000



# Section 2

# Analysis

## Analyzing Field Experiments

- If treatment is random we know that:

$$\hat{ATE} = E(Y_1|D=1) - E(Y_0|D=0) \tag{6}$$
$$= E(Y_1) - E(Y_0) \tag{7}$$
$$= \text{Real ATE} \tag{8}$$

## Analyzing Field Experiments

- If treatment is random we know that:

$$\hat{ATE} = E(Y_1|D=1) - E(Y_0|D=0) \quad (6)$$
$$= E(Y_1) - E(Y_0) \quad (7)$$
$$= \text{Real ATE} \quad (8)$$

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

## Analyzing Field Experiments

- If treatment is random we know that:

$$\hat{ATE} = E(Y_1|D=1) - E(Y_0|D=0) \quad (6)$$
$$= E(Y_1) - E(Y_0) \quad (7)$$
$$= \text{Real ATE} \quad (8)$$

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

## Analyzing Field Experiments

- If treatment is random we know that:

$$\hat{ATE} = E(Y_1|D=1) - E(Y_0|D=0) \quad (6)$$
$$= E(Y_1) - E(Y_0) \quad (7)$$
$$= \text{Real ATE} \quad (8)$$

- 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:

$$\hat{ATE} = E(Y_1|D=1) - E(Y_0|D=0) \quad (6)$$
$$= E(Y_1) - E(Y_0) \quad (7)$$
$$= \text{Real ATE} \quad (8)$$

- 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:

$$\hat{ATE} = E(Y_1|D=1) - E(Y_0|D=0) \quad (6)$$
$$= E(Y_1) - E(Y_0) \quad (7)$$
$$= \text{Real ATE} \quad (8)$$

- 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:

$$\hat{ATE} = E(Y_1|D=1) - E(Y_0|D=0) \quad (6)$$
$$= E(Y_1) - E(Y_0) \quad (7)$$
$$= \text{Real ATE} \quad (8)$$

- 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
- By definition:

$$Y_i^{obs} = Y_{0i}(1 - D_i) + Y_{1i}D_i$$

$$Y_i^{obs} = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

## Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test
- By definition:

$$Y_i^{obs} = Y_{0i}(1-D_i) + Y_{1i}D_i$$

$$Y_i^{obs} = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

- We can estimate:

$$Y_i^{obs} = \alpha + \beta D_i + \epsilon_i$$

## Analyzing Field Experiments

- Simple Regression = Difference-in-means T-test
- By definition:

$$Y_i^{obs} = Y_{0i}(1 - D_i) + Y_{1i}D_i$$

$$Y_i^{obs} = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

- We can estimate:

$$Y_i^{obs} = \alpha + \beta D_i + \epsilon_i$$

- So:

$$\hat{\beta} = E(Y_{1i} - Y_{0i})$$

# Analyzing Field Experiments

- Simple Regression is **identical** to a Difference-in-means T-test
- T-test Results:

| | estimate | statistic | p.value |
|---|---|---|---|
| 1 | 0.27065 | 2.69475 | 0.00706 |

# Analyzing Field Experiments

- Simple Regression is **identical** to a 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
- Estimated treatment effects are still *probabilistic* (random variables) so we may get the wrong answer by chance

## Repeated Experiments

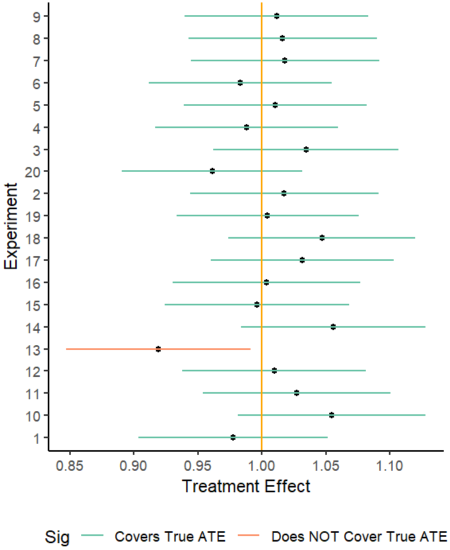
- The results from one experiment are not perfect
- Estimated treatment effects are still *probabilistic* (random variables) so we may get the wrong answer by chance
- In repeated experiments, 95% of confidence intervals will cross the true treatment effect

## Repeated Experiments

- The results from one experiment are not perfect
- Estimated 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
- Try repeated experiments in an App

## Repeated Experiments

- The results from one experiment are not perfect
- Estimated 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
- Try repeated experiments in an App



## 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/sampling** (eg. villages)

## 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/sampling** (eg. villages)
- In general, causal 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/sampling** (eg. villages)
- In general, causal 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:
  1. **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:
  1. **Small sample**, but note causal inference is now model-dependent
  2. **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:
  1. **Small sample**, but note causal inference is now model-dependent
  2. **Chance/residual imbalance** on a specific variable which we want to adjust for
  3. **To improve precision**, i.e. reduce the standard errors on $\beta$
     - The more variation in $Y$ we can explain with covariates, the more certain we can be on the effect of $D$

# Other Quantities of Interest

- Average Treatment Effects are just one summary statistic
  - Treatment effects are not normally constant

## Other Quantities of Interest

- Average Treatment Effects are just one summary statistic
    - Treatment effects are not normally constant
    - Averages can be influenced by outliers

# Other Quantities of Interest

- Average Treatment Effects are just one summary statistic
  - Treatment effects are not normally constant
  - Averages 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
  - Treatment effects are not normally constant
  - Averages 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

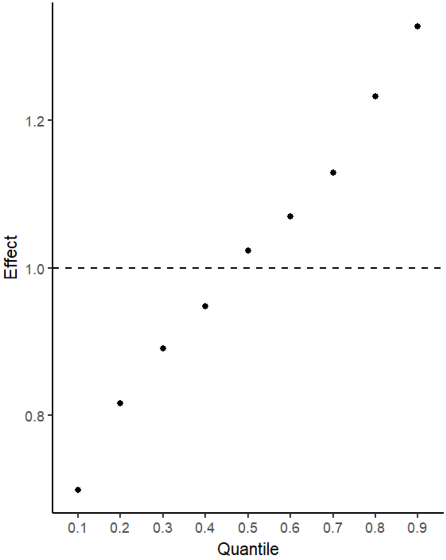
- Average Treatment Effects are just one summary statistic
  - Treatment effects are not normally constant
  - Averages can be influenced by outliers
- What if an average effect of +5% income leaves half the population hugely rich and half very poor?
- Average treatment effects are easiest (difference-in-means equals mean-difference)
- But we can also estimate Quantile treatment effects, eg. the effect of treatment on the bottom 10% of the distribution

## Other Quantities of Interest

- Assume the treatment effect is normally-distributed: $N(\mu = 1, \sigma^2 = 1)$

# Other Quantities of Interest

- Assume the treatment effect is normally-distributed: $N(\mu = 1, \sigma^2 = 1)$



## Heterogeneous Effects

- **Experiment:** We place a new health centre in half of all communities at random, and want to measure whether the health centre has a bigger effect in poor or rich neighbourhoods

## Heterogeneous Effects

- **Experiment:** We place a new health centre in half of all communities at random, and want to measure whether the health centre has a bigger effect in poor or rich neighbourhoods
- **Analysis:** Run a single regression with an interaction between treatment and neighbourhood income

## Heterogeneous Effects

- **Experiment:** We place a new health centre in half of all communities at random, and want to measure whether the health centre has a bigger effect in poor or rich neighbourhoods
- **Analysis:** Run a single regression with an interaction between treatment and neighbourhood income
- **Result:** The health centre boosts health by 20% in rich neighbourhoods and reduces health in poor neighbourhoods by 20%

## Heterogeneous Effects

- **Experiment:** We place a new health centre in half of all communities at random, and want to measure whether the health centre has a bigger effect in poor or rich neighbourhoods
- **Analysis:** Run a single regression with an interaction between treatment and neighbourhood income
- **Result:** The health centre boosts health by 20% in rich neighbourhoods and reduces health in poor neighbourhoods by 20%
- **Interpretation:** Does neighbourhood poverty cause health centres to have a negative impact?

## Heterogeneous Effects

- **Experiment:** We place a new health centre in half of all communities at random, and want to measure whether the health centre has a bigger effect in poor or rich neighbourhoods
- **Analysis:** Run a single regression with an interaction between treatment and neighbourhood income
- **Result:** The health centre boosts health by 20% in rich neighbourhoods and reduces health in poor neighbourhoods by 20%
- **Interpretation:** Does neighbourhood poverty cause health centres to have a negative impact?
  - We cannot interpret the 'moderator' variable as having a causal effect, the different treatment effects could be due to omitted variables or selection

## Heterogeneous Effects

- **Experiment:** We place a new health centre in half of all communities at random, and want to measure whether the health centre has a bigger effect in poor or rich neighbourhoods
- **Analysis:** Run a single regression with an interaction between treatment and neighbourhood income
- **Result:** The health centre boosts health by 20% in rich neighbourhoods and reduces health in poor neighbourhoods by 20%
- **Interpretation:** Does neighbourhood poverty cause health centres to have a negative impact?
  - We cannot interpret the 'moderator' variable as having a causal effect, the different treatment effects could be due to omitted variables or selection
  - Only the health centre was randomly assigned, not neighbourhood income!

# Section 3

# Assumptions

# Assumptions

1. Compliance with Randomization procedure

# Assumptions

1. Compliance with Randomization procedure
2. Randomization produced balance on potential outcomes

## Assumptions

1. Compliance with Randomization procedure
2. Randomization produced balance on potential outcomes
3. No Spillovers (SUTVA)

# Assumptions

1. Compliance with Randomization procedure
2. Randomization produced balance on potential outcomes
3. No Spillovers (SUTVA)
4. Excludability

# Assumptions

1. Compliance with Randomization procedure
2. Randomization produced balance on potential outcomes
3. No Spillovers (SUTVA)
4. Excludability

# 1. Compliance with Randomization procedure

- Randomization is unpopular, political, and sometimes resisted

# 1. Compliance with Randomization procedure

- Randomization is unpopular, political, and sometimes resisted
- Need to verify treatment allocation
  - Transparency, documentation

# 1. Compliance with Randomization procedure

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

# 1. Compliance with Randomization procedure

- Randomization is unpopular, political, and sometimes resisted
- Need to verify treatment allocation
  - Transparency, documentation
- And treatment compliance
  - Did anyone assigned to control manage to get treatment?
  - Did anyone assigned to treatment refuse?
- **Design:** Double-blind assignment

# 1. Compliance with Randomization procedure

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

## 1. Compliance with Randomization procedure

- Randomization is unpopular, political, and sometimes resisted
- Need to verify treatment allocation
  - Transparency, documentation
- And treatment compliance
  - Did anyone assigned to control manage to get treatment?
  - Did anyone assigned to treatment refuse?
- **Design:** Double-blind assignment
- **Checks:** Qualitative fieldwork
- **Analysis:** More on how to respond to non-compliance next week

## 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
- If lots of variables are balanced, it's likely potential outcomes are too

# 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
- If lots of variables are balanced, it's likely potential outcomes are too
- **Check:** Normally a difference in means T-test of covariates between treatment and control groups

## 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
- If lots of variables are balanced, it's likely potential outcomes are too
- **Check:** Normally a difference in means T-test of covariates between treatment and control groups
- **Check:** Or a Kolmogorov-Smirnov (KS) 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 so much 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 so much 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 so much 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 so much 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*

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

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

- Spillovers interfere with our control group, so the comparison does not measure the direct effect of a treatment on person *i*

## 3. SUTVA

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

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

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

- Spillovers interfere with our control group, so the comparison does not measure the direct effect of a treatment on person $i$
- 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 $i$

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

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

- Spillovers interfere with our control group, so the comparison does not measure the direct effect of a treatment on person $i$
- But spillovers are common! If you get an award, I might feel more motivated or less motivated
- What should we do?

# 3. SUTVA

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

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

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

- Spillovers interfere with our control group, so the comparison does not measure the direct effect of a treatment on person *i*
- But spillovers are common! If you get an award, I might feel more motivated or less motivated
- What should we do?
  - **Design:** Limit risk of spillovers, eg. leave 20 miles between each unit in sampling
  - **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 **'parallel'** treatment

# 4. Excludability

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

# 4. Excludability

- **Nothing else correlated with treatment affects potential outcomes**
- Assignment to treatment causes a **’parallel’** treatment
  - Eg. We decide to share information about specific politicians on the radio, but the politicians find out and counter with their own broadcasts
- Our treatment effect is no longer *only* the effect of our information intervention

# 4. Excludability

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

# 4. Excludability

- Distinguish between the downstream consequences of treatment and ’parallel’ treatments

## 4. Excludability

- Distinguish between the downstream consequences of treatment and 'parallel' treatments

**Downstream ('net') Consequences**

# 4. Excludability

- Distinguish between the downstream consequences of treatment and ’parallel’ treatments

**Downstream (’net’) Consequences**

- Eg. We give a cash handout to families, and then they also start paying taxes; which explains their changing attitudes to government?

# 4. Excludability

- Distinguish between the downstream consequences of treatment and ’parallel’ treatments

**Downstream (’net’) Consequences**

- Eg. We give a cash handout to families, and then they also start paying taxes; which explains their changing attitudes to government?
- We find zero effect of government investing $1000 in healthcare on health outcomes, because households responded by reducing their spending by exactly $1000

# 4. Excludability

- Distinguish between the downstream consequences of treatment and ’parallel’ treatments

**Downstream (’net’) Consequences**

- Eg. We give a cash handout to families, and then they also start paying taxes; which explains their changing attitudes to government?
- We find zero effect of government investing $1000 in healthcare on health outcomes, because households responded by reducing their spending by exactly $1000

**Parallel Treatments**

# 4. Excludability

- Distinguish between the downstream consequences of treatment and ’parallel’ treatments

**Downstream (’net’) Consequences**

- Eg. We give a cash handout to families, and then they also start paying taxes; which explains their changing attitudes to government?
- We find zero effect of government investing $1000 in healthcare on health outcomes, because households responded by reducing their spending by exactly $1000

**Parallel Treatments**

- Eg. Measurement bias: Researchers give treated units ’the benefit of the doubt’ and record higher outcomes for them

# 4. Excludability

- Distinguish between the downstream consequences of treatment and 'parallel' treatments

**Downstream ('net') Consequences**

- Eg. We give a cash handout to families, and then they also start paying taxes; which explains their changing attitudes to government?
- We find zero effect of government investing $1000 in healthcare on health outcomes, because households responded by reducing their spending by exactly $1000

**Parallel Treatments**

- Eg. Measurement bias: Researchers give treated units 'the benefit of the doubt' and record higher outcomes for them
- Or Hawthorne Effects: Participants respond to being studied, not treatment (more next week)

# 4. Excludability

- Distinguish between the downstream consequences of treatment and 'parallel' treatments

**Downstream ('net') Consequences**

- Eg. We give a cash handout to families, and then they also start paying taxes; which explains their changing attitudes to government?
- We find zero effect of government investing $1000 in healthcare on health outcomes, because households responded by reducing their spending by exactly $1000

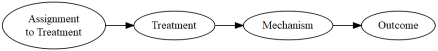
**Parallel Treatments**

- Eg. Measurement bias: Researchers give treated units 'the benefit of the doubt' and record higher outcomes for them
- Or Hawthorne Effects: Participants respond to being studied, not treatment (more next week)

- **Design:** Careful specification of treatment and control

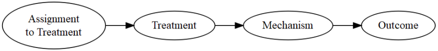
# 4. Excludability

## Downstream Consequence of Treatment

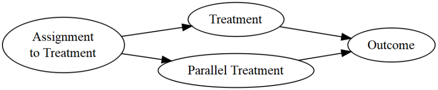


# 4. Excludability

## Downstream Consequence of Treatment




## Parallel Treatment



# Section 4

# Implementation

## 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?
- 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 similar units and flip a coin to assign one from each pair 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 similar units and flip a coin to assign one from each pair to 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 similar units and flip a coin to assign one from each pair to treatment
- What's the difference between these three options?
- What % treated? 50:50 is usually most efficient

# 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 similar units and flip a coin to assign one from each pair to treatment
- What's the difference between these three options?
- What % treated? 50:50 is usually most efficient
- To actually randomize, use the 'randomizr' package

# 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, 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, 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, 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, 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 a sample size of 4000, half male, half female

## Implementing Field Experiments

- **Blocking**
- Randomization is *inefficient* and risky
- We know we need balance on key covariates, eg. gender, 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 a sample size of 4000, half male, half female

Without Blocking:

| | M | F |
|---|---|---|
| Treated | 1042 | 958 |
| Control | 972 | 1028 |

With Blocking:

| | M | F |
|---|---|---|
| Treated | 1000 | 1000 |
| Control | 1000 | 1000 |

## Implementing Field Experiments

- **Blocking**
- Randomization is *inefficient* and risky
- We know we need balance on key covariates, eg. gender, 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 a sample size of 4000, half male, half female

Without Blocking:

| | M | F |
|---|---|---|
| Treated | 1042 | 958 |
| Control | 972 | 1028 |

With Blocking:

| | M | F |
|---|---|---|
| Treated | 1000 | 1000 |
| Control | 1000 | 1000 |

- "Block what you can; randomize what you cannot"

# Implementing Field Experiments

- **Blocking**
- Randomization is *inefficient* and risky
- We know we need balance on key covariates, eg. gender, 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 a sample size of 4000, half male, half female

Without Blocking:

| | M | F |
|---|---|---|
| Treated | 1042 | 958 |
| Control | 972 | 1028 |

With Blocking:

| | M | F |
|---|---|---|
| Treated | 1000 | 1000 |
| Control | 1000 | 1000 |

- "Block what you can; randomize what you cannot"
- We focus on within-block variation: $Y_i = \alpha + D_i + B_i + \epsilon_i$

# Implementing Field Experiments

- Random treatment vs. Random samples

**Random Treatment**

# Implementing Field Experiments

- Random treatment vs. Random samples

**Random Treatment**

- Representative potential outcomes

# Implementing Field Experiments

- Random treatment vs. Random samples

**Random Treatment**

- Representative potential outcomes
- *Causal* Inference

**Random Samples**

# Implementing Field Experiments

- Random treatment vs. Random samples

**Random Treatment**

- Representative potential outcomes
- *Causal* Inference

**Random Samples**

- Sample representative of larger population

# Implementing Field Experiments

- Random treatment vs. Random samples

**Random Treatment**

- Representative potential outcomes
- *Causal* Inference

**Random Samples**

- Sample representative of larger population
- *Statistical* Inference

# Implementing Field Experiments

- Random treatment vs. Random samples

**Random Treatment**

- Representative potential outcomes
- *Causal* Inference

**Random Samples**

- Sample representative of larger population
- *Statistical* Inference

- Both work in the same way - randomization avoids selection (into the data/treatment)

# Section 5

# Critiquing

# 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.

# 1. Results are a Black Box

- We know that $D$ causes $Y$ in this population. So what? What did we learn about political science?

# 1. Results are a Black Box

- We know that $D$ causes $Y$ in this population. So what? What did we learn about political science?
  - We know that giving citizens health insurance makes them more likely to vote.

# 1. Results are a Black Box

- We know that $D$ causes $Y$ in this population. So what? What did we learn about political science?
  - We know that giving citizens health insurance makes them more likely to vote. Why?? How?? What is the mechanism?

# 1. Results are a Black Box

- We know that $D$ causes $Y$ in this population. So what? What did we learn about political science?
  - We know that giving citizens health insurance makes them more likely to vote. Why?? How?? What is the mechanism?
  - Due to increased wealth? Increased trust in government? More mobility?

## 1. Results are a Black Box

- We know that $D$ causes $Y$ in this population. So what? What did we learn about political science?
  - We know that giving citizens health insurance makes them more likely to vote. Why?? How?? What is the mechanism?
  - Due to increased wealth? Increased trust in government? More mobility?
- What theory is this testing?

## 1. Results are a Black Box

- We know that $D$ causes $Y$ in this population. So what? What did we learn about political science?
  - We know that giving citizens health insurance makes them more likely to vote. Why?? How?? What is the mechanism?
  - Due to increased wealth? Increased trust in government? More mobility?
- What theory is this testing? Does it reject any theory?

# 1. Results are a Black Box

- We know that $D$ causes $Y$ in this population. So what? What did we learn about political science?
  - We know that giving citizens health insurance makes them more likely to vote. Why?? How?? What is the mechanism?
  - Due to increased wealth? Increased trust in government? More mobility?
- What theory is this testing? Does it reject any theory?
- We want to test theories, not treatments

## 2. Generalizability of Context

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

# 2. Generalizability of Context

- Our causal conclusions are **restricted 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 **restricted 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 **restricted 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 **restricted 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 **restricted 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 not in the full population we want to learn about

# 2. Generalizability of Context

- Our causal conclusions are **restricted 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 not in the full population we want to learn about
  - The places that agree to field experiments are not representative

## 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.

## 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.
  2. **Ownership and Excludability:**
     - Telling someone to implement an intervention is different from working with a self-motivated actor who designed the intervention.

# 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.
  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.
  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 70% to 90%, 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% compared to controls

# 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% compared to controls
- But:
  1. 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% compared to controls
- But:
  1. Sites were not representative (close to main roads and cities so they're easy to visit)
  2. 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% compared to controls
- But:
  1. Sites were not representative (close to main roads and cities so they're easy to visit)
  2. Treatment could not be scaled (Every village cannot get visits from Columbia professors twice a year)
  3. And politics was ignored (No implementation unless you give locals responsibility, but then you 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