Tools for Planning and Analyzing Randomized Controlled Trials and A/B Tests

Johann Gagnon-Bartsch, Adam Sales, Duy Pham,
Charlotte Mann, and Jaylin Lowe

Department of Statistics, University of Michigan &
Department of Mathematical Sciences, Worcester Polytechnic Institute

UNIVERSITY OF MICHIGAN

SREE 2024

September 18, 2024

WPI

While we are getting settled...

tinyurl.com/
sree-drct

Follow the steps at:

`https://tinyurl.com/sree-drct`

to get everything ready to follow along in *`RStudio`*!

Today's Plan

tinyurl.com/sree-drct

- 9:00–9:15 – Part I: Conceptual Overview
- 9:15–10:30 – Part II: Estimating Effects with RCT Data
- 10:30–11:00 – Part III: Incorporating Auxiliary Data
- 11:00–11:15 – Break 15 min
- 11:15–11:45 – Part IV: Treatment Effect Heterogeneity
- 11:45–12:15 – Part V: Planning Experiments

What You Need

tinyurl.com/sree-drct

- Tutorial website: *https://tinyurl.com/edmrct*
- RStudio
- Clone repo from Github: *https://github.com/mannczz/edm-rct-tutorial/*

Tutorial Structure

tinyurl.com/sree-drct

We will be alternating between:

- Conceptual descriptions of the methods
- Detailed walk-throughs of the software
- Opportunities for you to run analyses yourself, with our help

Please feel free to ask questions at any time!

- Calling out (unmute yourself if on Zoom)
- Zoom chat
- Any other way you can think of to get our attention

Conceptual Overview

Estimating Effects with RCT Data

Incorporating Auxiliary Data

Treatment Effect Heterogeneity

Planning Experiments

Experiments in Education Research

tinyurl.com/sree-drct

"Experiment" = "RCT" = "Randomized Controlled Trial"

- Randomize subjects (students? teachers? schools?) between condition
- Expose subjects to their randomized conditions
- Measure outcome(s) of interest

Experiments in Education Research

"Experiment" = "RCT" = "Randomized Controlled Trial"

tinyurl.com/sree-drct

- Randomize subjects (students? teachers? schools?) between condition
- Expose subjects to their randomized conditions
- Measure outcome(s) of interest
- Associations between condition and outcomes are causal

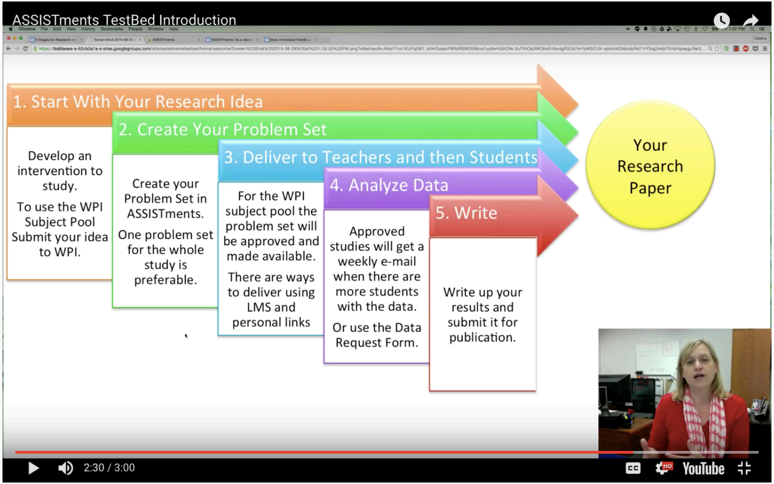
Experiments in Education Research

tinyurl.com/sree-drct

"Experiment" = "RCT" = "Randomized Controlled Trial"

- Randomize subjects (students? teachers? schools?) between condition
- Expose subjects to their randomized conditions
- Measure outcome(s) of interest
- Associations between condition and outcomes are causal
- Typical examples:
  - A/B tests in online learning
  - Field trials of (say) new curriculum vs. business as usual

Example 1: ASSISTments ETrials



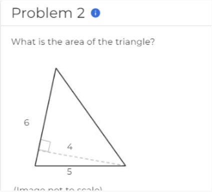
tinyurl.com/sree-drct

Integrated A/B test platform

Example 1: ASSISTments E-trials

tinyurl.com/sree-drct

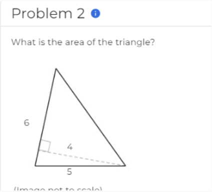
- Question: Text or video hints?
- Outcome: Complete skill builder?
- $n = 683$ middle school students



Example 1: ASSISTments E-trials

tinyurl.com/sree-drct

- Question: Text or video hints?
- Outcome: Complete skill builder?
- $n = 683$ middle school students



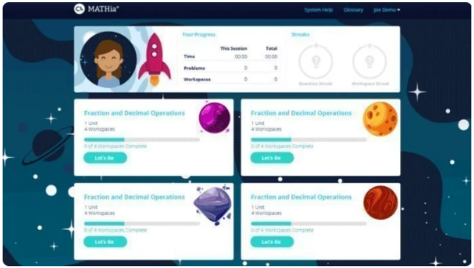
Results,

- Video: 205/337 (61%) completed
- Text: 193/346 (56%) completed

Example II: Cognitive Tutor Effectiveness Trial

tinyurl.com/sree-drct

- 73 High Schools & 74 Middle Schools in 7 states
- Similar schools paired
- In each pair, one school randomized to treatment, one to control
- Algebra I students in Trt school used CTAI, Control school used business as usual
- All students took a posttest at the end of the year



Example II: Cognitive Tutor Effectiveness Trial

tinyurl.com/sree-drct

Results

	Average Posttest			
	Middle		High	
	Year 1	Year 2	Year 1	Year 2
Control	17.4	16.9	10.3	9.7
Treatment	14.3	15.2	10.1	10.6

Scientific Goals

tinyurl.com/sree-drct

1. What is the average effect of [intervention] on [outcome]?

2. How Does the effect vary?

Scientific Goals

tinyurl.com/sree-drct

1. What is the average effect of [intervention] on [outcome]?
   - "Intervention" AKA "Treatment" (the thing you're randomizing)
     - Contrast between $2+$ conditions
     - E.g. access to ChatGPT hint vs teacher-written hint vs no hint
     - For today: focus on 2 conditions, "Treatment" vs "Control"
     - (those labels may be arbitrary)

2. How Does the effect vary?

Scientific Goals

1. What is the average effect of [intervention] on [outcome]?
   - "Intervention" AKA "Treatment" (the thing you're randomizing)
     - Contrast between $2+$ conditions
     - E.g. access to ChatGPT hint vs teacher-written hint vs no hint
     - For today: focus on 2 conditions, "Treatment" vs "Control"
     - (those labels may be arbitrary)
   - "Outcome"
     - Scalar quantity that the intervention might affect
     - E.g. student correctness on the next problem (0 or 1)

2. How Does the effect vary?

tinyurl.com/sree-drct

Scientific Goals

tinyurl.com/sree-drct

1. What is the average effect of [intervention] on [outcome]?
   - "Intervention" AKA "Treatment" (the thing you're randomizing)
     - Contrast between $2+$ conditions
     - E.g. access to ChatGPT hint vs teacher-written hint vs no hint
     - For today: focus on 2 conditions, "Treatment" vs "Control"
     - (those labels may be arbitrary)
   - "Outcome"
     - Scalar quantity that the intervention might affect
     - E.g. student correctness on the next problem (0 or 1)
   - "Average Effect" …to be defined soon!
2. How Does the effect vary?

Scientific Goals

tinyurl.com/sree-drct

1. What is the average effect of [intervention] on [outcome]?
   - “Intervention” AKA “Treatment” (the thing you’re randomizing)
     - Contrast between $2+$ conditions
     - E.g. access to ChatGPT hint vs teacher-written hint vs no hint
     - For today: focus on 2 conditions, “Treatment” vs “Control”
     - (those labels may be arbitrary)
   - “Outcome”
     - Scalar quantity that the intervention might affect
     - E.g. student correctness on the next problem (0 or 1)
   - “Average Effect” ...to be defined soon!
2. How Does the effect vary?
   - From one (type of) student to the next
   - From one context to the next

Statistical Goals

tinyurl.com/sree-drct

1. Get the most out of your data: more data $\rightarrow$ better estimation!!
2. ...Without making unnecessary assumptions
3. Easily
4. Design better experiments to start with

Statistical Goals

tinyurl.com/sree-drct

1. Get the most out of your data: more data $\rightarrow$ better estimation!!
    - Baseline covariate data
    - Historical user data
2. ...Without making unnecessary assumptions
3. Easily
4. Design better experiments to start with

Statistical Goals

tinyurl.com/sree-drct

1. Get the most out of your data: more data $\rightarrow$ better estimation!!
   - Baseline covariate data
   - Historical user data
2. ...Without making unnecessary assumptions
   - "Design-based" methods
   - NO assumptions about confounding, models, etc. etc.
3. Easily

4. Design better experiments to start with

Statistical Goals

tinyurl.com/sree-drct

1. Get the most out of your data: more data $\rightarrow$ better estimation!!
   - Baseline covariate data
   - Historical user data
2. ...Without making unnecessary assumptions
   - "Design-based" methods
   - NO assumptions about confounding, models, etc. etc.
3. Easily
   - i.e. without a PhD in statistics
   - Use our software package :)
4. Design better experiments to start with

Types of Variables: Baseline Covariates

tinyurl.com/sree-drct

- Fixed at baseline
- Unaffected by treatment

Types of Variables: Baseline Covariates

tinyurl.com/sree-drct

- Fixed at baseline
- Unaffected by treatment

Uses:

- More precise estimates
- Explore effect variation

Example: Covariates in ASSISTments

tinyurl.com/sree-drct

- Log data. For each previous skillbuilder,
  - Completed skill builder?
  - # problems attempted / completed?
  - Time to mastery
  - ...
- Demographic data

Example: Covariates in ASSISTments

tinyurl.com/sree-drct

- Log data. For each previous skillbuilder,
  - Completed skill builder?
  - # problems attempted / completed?
  - Time to mastery
  - . . .
- Demographic data

Don’t use post-treatment variables!

Auxiliary Data

- Covariate and outcome data from *other* subjects
- Often: historical data

tinyurl.com/sree-drct

Auxiliary Data

tinyurl.com/sree-drct

- Covariate and outcome data from *other* subjects
- Often: historical data
- Requirements
  - Separate sample from RCT
  - (some of the) same covariate data as for RCT subjects
  - similar outcome data as RCT

Auxiliary Data

tinyurl.com/
sree-drct

- Covariate and outcome data from *other* subjects
- Often: historical data
- Requirements
  - Separate sample from RCT
  - (some of the) same covariate data as for RCT subjects
  - similar outcome data as RCT

Uses:

- More precise estimates
- Planning experiments

Reiterate Goals

tinyurl.com/sree-drct

Estimate treatment effects

Reiterate Goals

tinyurl.com/
sree-drct

Estimate treatment effects

Using all our data

Reiterate Goals

tinyurl.com/sree-drct

Estimate treatment effects

Using all our data

- Covariates (even high-dimensional)
- Auxiliary/historical data

Reiterate Goals

tinyurl.com/sree-drct

Estimate treatment effects

Using all our data

- Covariates (even high-dimensional)
- Auxiliary/historical data

Without bias or extra assumptions

Conceptual Overview

Estimating Effects with RCT Data

Incorporating Auxiliary Data

Treatment Effect Heterogeneity

Planning Experiments

Potential Outcomes (Neyman-Rubin)

tinyurl.com/sree-drct

Consider a randomized experiment with:

- N participants
- One treatment group, one control group

Potential Outcomes (Neyman-Rubin)

tinyurl.com/
sree-drct

- The outcome depends on treatment.

Potential Outcomes (Neyman-Rubin)

tinyurl.com/sree-drct

- The outcome depends on treatment.
  If the coin had landed the other way, the outcome may have been different.

Potential Outcomes (Neyman-Rubin)

tinyurl.com/sree-drct

- The outcome depends on treatment.
  If the coin had landed the other way, the outcome may have been different.

- Each subject has two **potential outcomes**.

Potential Outcomes (Neyman-Rubin)

tinyurl.com/sree-drct

- The outcome depends on treatment.
  If the coin had landed the other way, the outcome may have been different.

- Each subject has two **potential outcomes**.
  One for treatment, one for control.

Potential Outcomes (Neyman-Rubin)

tinyurl.com/sree-drct

- The outcome depends on treatment.
  If the coin had landed the other way, the outcome may have been different.

- Each subject has two **potential outcomes**.
  One for treatment, one for control.
- We only ever observe **one** potential outcome.

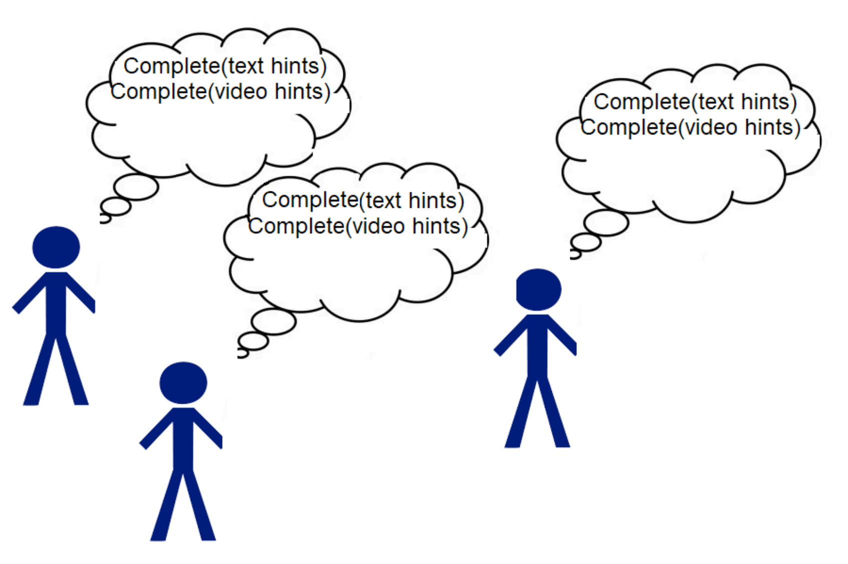
Potential Outcomes (Neyman-Rubin)

tinyurl.com/sree-drct

- The outcome depends on treatment.
  If the coin had landed the other way, the outcome may have been different.

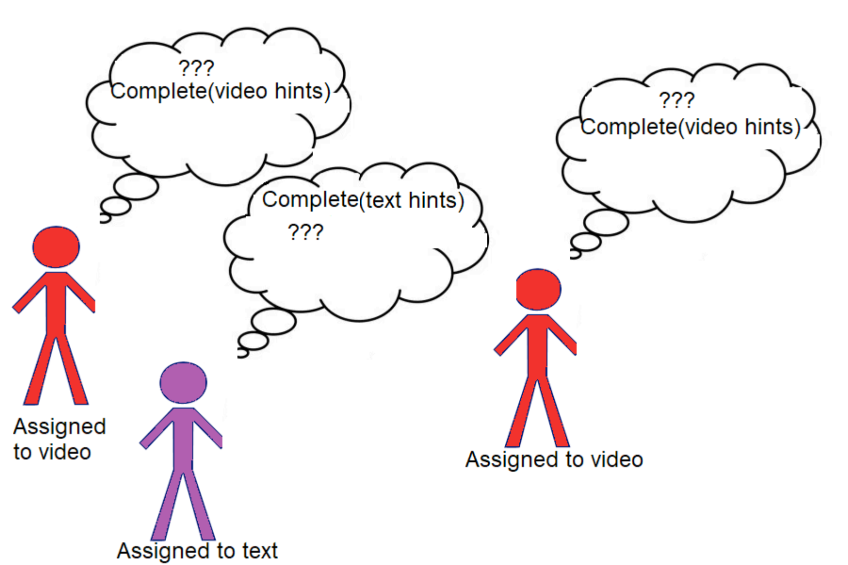
- Each subject has two **potential outcomes**.
  One for treatment, one for control.
- We only ever observe **one** potential outcome.
  The other is a counterfactual.

Potential Outcomes (Neyman-Rubin)



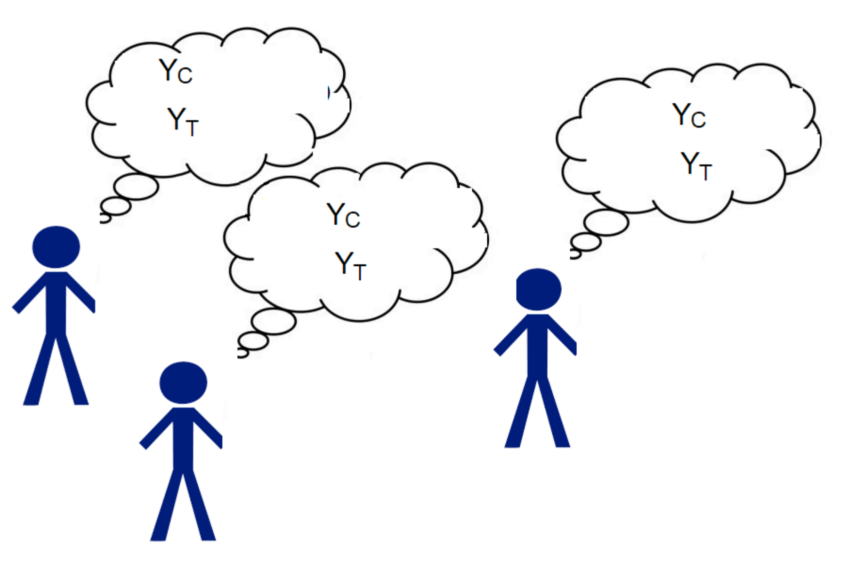
tinyurl.com/sree-drct

Potential Outcomes (Neyman-Rubin)



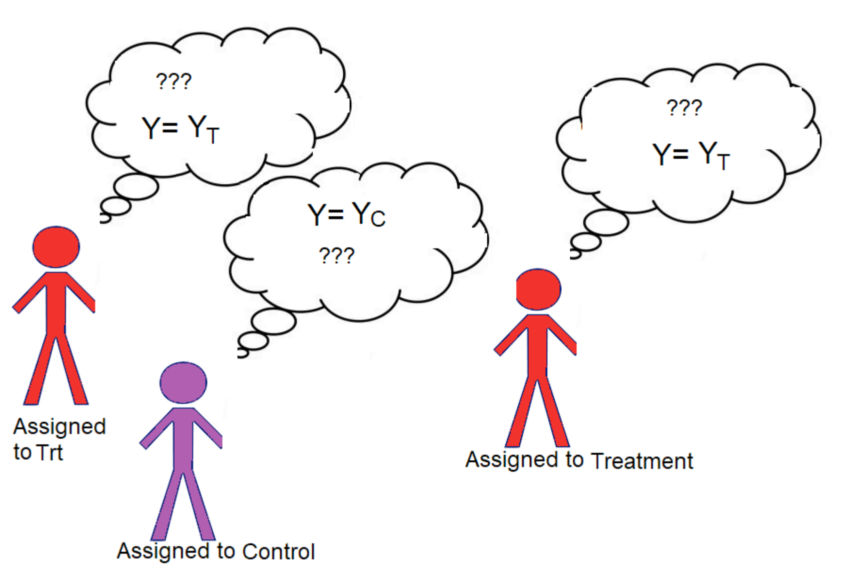
tinyurl.com/
sree-drct

Potential Outcomes (Neyman-Rubin)



tinyurl.com/
sree-drct

Potential Outcomes (Neyman-Rubin)



tinyurl.com/sree-drct

Potential Outcomes (Neyman-Rubin)

- For each participant i there are two potential outcomes, y_i^t and y_i^c

tinyurl.com/
sree-drct

Potential Outcomes (Neyman-Rubin)

- For each participant i there are two potential outcomes, y_i^t and y_i^c
- Potential outcomes are **fixed** values, not random

tinyurl.com/sree-drct

Potential Outcomes (Neyman-Rubin)

- For each participant i there are two potential outcomes, y_i^t and y_i^c
- Potential outcomes are **fixed** values, not random
- Let T_i be the treatment assignment of unit i

$$T_i = \begin{cases} 1, & \text{Unit } i \text{ is assigned to treatment} \\ 0, & \text{Unit } i \text{ is assigned to control} \end{cases}$$

tinyurl.com/sree-drct

Potential Outcomes (Neyman-Rubin)

tinyurl.com/sree-drct

- For each participant i there are two potential outcomes, y_i^t and y_i^c
- Potential outcomes are **fixed** values, not random
- Let T_i be the treatment assignment of unit i

$$T_i = \begin{cases} 1, & \text{Unit } i \text{ is assigned to treatment} \\ 0, & \text{Unit } i \text{ is assigned to control} \end{cases}$$

- Let Y_i be the observed outcome for unit i. If unit i is assigned to treatment, we observe y_i^t; otherwise, we observe y_i^c:

$$Y_i = \begin{cases} y_i^c & \text{if } T_i = 0 \\ y_i^t & \text{if } T_i = 1 \end{cases}$$

Individual and Average Treatment Effects

- The individual treatment effect is

$$\tau_i = y_i^t - y_i^c$$

tinyurl.com/sree-drct

Individual and Average Treatment Effects

tinyurl.com/sree-drct

- The individual treatment effect is

$$\tau_i = y_i^t - y_i^c$$

- The individual treatment effect is **never observed**.

Individual and Average Treatment Effects

tinyurl.com/sree-drct

- The individual treatment effect is

$$\tau_i = y_i^t - y_i^c$$

- The individual treatment effect is **never observed**.

- The average treatment effect (ATE) is

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} \tau_i$$

Individual and Average Treatment Effects

tinyurl.com/sree-drct

- The individual treatment effect is

$$\tau_i = y_i^t - y_i^c$$

- The individual treatment effect is **never observed**.

- The average treatment effect (ATE) is

$$\bar{\tau} = \frac{1}{N}\sum_{i=1}^{N} \tau_i$$

- The average treatment effect can be estimated.

Individual and Average Treatment Effects

tinyurl.com/sree-drct

- The individual treatment effect is

$$\tau_i = y_i^t - y_i^c$$

- The individual treatment effect is **never observed**.

- The average treatment effect (ATE) is

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} \tau_i$$

- The average treatment effect can be estimated.
- Also: average effects for subgroups of subjects (more later)

Estimating Average Treatment Effects

tinyurl.com/sree-drct

"The fundamental problem of causal inference"

- We only observe one potential outcome for each subject
  - For treatment subjects y^t
  - For control, y^c

Estimating Average Treatment Effects

tinyurl.com/sree-drct

"The fundamental problem of causal inference"

- We only observe one potential outcome for each subject
  - For treatment subjects y^t
  - For control, y^c
- One potential outcome is always missing
- We need to *impute* the missing potential outcome

Estimating Average Treatment Effects

tinyurl.com/sree-drct

"The fundamental problem of causal inference"

- We only observe one potential outcome for each subject
  - For treatment subjects y^t
  - For control, y^c
- One potential outcome is always missing
- We need to *impute* the missing potential outcome
- Two approaches to imputation:
  1. Use randomization: unbiased, but imprecise

Estimating Average Treatment Effects

tinyurl.com/sree-drct

“The fundamental problem of causal inference”

- We only observe one potential outcome for each subject
    - For treatment subjects y^t
    - For control, y^c
- One potential outcome is always missing
- We need to *impute* the missing potential outcome
- Two approaches to imputation:
    1. Use randomization: unbiased, but imprecise
    2. Use covariates & and model: biased, but precise

Estimating Average Treatment Effects

tinyurl.com/sree-drct

"The fundamental problem of causal inference"

- We only observe one potential outcome for each subject
  - For treatment subjects y^t
  - For control, y^c
- One potential outcome is always missing
- We need to *impute* the missing potential outcome
- Two approaches to imputation:
  1. Use randomization: unbiased, but imprecise
  2. Use covariates & and model: biased, but precise
  3. Our approach: use both!

Our Method, in a Nutshell

Step 1:

Train algorithms to predict y^c, y^t as a function of covariates

$$f^c : \mathbf{X} \to y^c \text{ (use data from ctl group)}$$

$$f^t : \mathbf{X} \to y^t \text{ (use data from trt group)}$$

tinyurl.com/sree-drct

Our Method, in a Nutshell

tinyurl.com/sree-drct

Step 1:

Train algorithms to predict y^c, y^t as a function of covariates

$$f^c : \mathbf{X} \rightarrow y^c \text{ (use data from ctl group)}$$

$$f^t : \mathbf{X} \rightarrow y^t \text{ (use data from trt group)}$$

Step 2:

Use algorithms to get imputations:

$$\hat{y}_i^c = f^c(X_i)$$

$$\hat{y}_i^t = f^t(X_i)$$

Our Method, in a Nutshell

tinyurl.com/sree-drct

Step 1:

Train algorithms to predict y^c, y^t as a function of covariates

$$f^c : \mathbf{X} \rightarrow y^c \text{ (use data from ctl group)}$$

$$f^t : \mathbf{X} \rightarrow y^t \text{ (use data from trt group)}$$

Step 2:

Use algorithms to get imputations:

$$\hat{y}_i^c = f^c(X_i)$$

$$\hat{y}_i^t = f^t(X_i)$$

Step 3: Calculate $\hat{m}_i = Pr(Z_i = 0)\hat{y}_i^t + Pr(Z_i = 1)\hat{y}_i^c$

Our Method, in a Nutshell

tinyurl.com/sree-drct

Step 1:

Train algorithms to predict y^c, y^t as a function of covariates

$$f^c : \mathbf{X} \rightarrow y^c \text{ (use data from ctl group)}$$
$$f^t : \mathbf{X} \rightarrow y^t \text{ (use data from trt group)}$$

Step 2:

Use algorithms to get imputations:

$$\hat{y}_i^c = f^c(X_i)$$
$$\hat{y}_i^t = f^t(X_i)$$

Step 3: Calculate $\hat{m}_i = Pr(Z_i = 0)\hat{y}_i^t + Pr(Z_i = 1)\hat{y}_i^c$

Step 4:

Use randomization-based method to estimate effects on $Y - \hat{m}$ instead of Y

Important Caveat

tinyurl.com/sree-drct

For this to be *strictly* unbiased, we need:

$$\hat{m}_i \text{ independent of } T_i$$

Important Caveat

tinyurl.com/sree-drct

For this to be *strictly* unbiased, we need:

$$\hat{m}_i \text{ independent of } T_i$$

Since Y_i is a function of T_i, that means we need:

$$\hat{y}^c \text{ and } \hat{y}^t \text{ independent of } Y_i$$

Important Caveat

For this to be *strictly* unbiased, we need:

$$\hat{m}_i \text{ independent of } T_i$$

Since Y_i is a function of T_i, that means we need:

$$\hat{y}^c \text{ and } \hat{y}^t \text{ independent of } Y_i$$

We can't use i's data to train f^c and f^t!

tinyurl.com/sree-drct

Important Caveat

tinyurl.com/sree-drct

For this to be *strictly* unbiased, we need:

$$\hat{m}_i \text{ independent of } T_i$$

Since Y_i is a function of T_i, that means we need:

$$\hat{y}^c \text{ and } \hat{y}^t \text{ independent of } Y_i$$

We can't use i's data to train f^c and f^t!

Solution: re-train f^c and f^t for each subject i, leaving out i's data

Important Caveat

tinyurl.com/sree-drct

For this to be *strictly* unbiased, we need:

$\hat{m}_i$ independent of T_i

Since Y_i is a function of T_i, that means we need:

$\hat{y}^c$ and $\hat{y}^t$ independent of Y_i

We can't use i's data to train f^c and f^t!

Solution: re-train f^c and f^t for each subject i, leaving out i's data

"Leave-One-Out Potential Outcomes" or "LOOP"

Sticks and Stones May Break my Bones, but Bad Models Won't Hurt Me

tinyurl.com/sree-drct

- What if f^c and f^t are totally wrong and bad??

Sticks and Stones May Break my Bones, but Bad Models Won't Hurt Me

tinyurl.com/sree-drct

- What if f^c and f^t are totally wrong and bad??
- Estimate will still be unbiased!

Sticks and Stones May Break my Bones, but Bad Models Won't Hurt Me

tinyurl.com/sree-drct

- What if f^c and f^t are totally wrong and bad??
- Estimate will still be unbiased!
- Standard errors, p-values, and confidence intervals will still be valid!

Sticks and Stones May Break my Bones, but Bad Models Won't Hurt Me

tinyurl.com/sree-drct

- What if f^c and f^t are totally wrong and bad??
- Estimate will still be unbiased!
- Standard errors, p-values, and confidence intervals will still be valid!
- (core of inference is based on randomization)

Sticks and Stones May Break my Bones, but Bad Models Won't Hurt Me

tinyurl.com/
sree-drct

- What if f^c and f^t are totally wrong and bad??
- Estimate will still be unbiased!
- Standard errors, p-values, and confidence intervals will still be valid!
- (core of inference is based on randomization)
- Covariate adjustment won't help much

Sticks and Stones May Break my Bones, but Bad Models Won't Hurt Me

tinyurl.com/sree-drct

- What if f^c and f^t are totally wrong and bad??
- Estimate will still be unbiased!
- Standard errors, p-values, and confidence intervals will still be valid!
- (core of inference is based on randomization)
- Covariate adjustment won't help much
- In moderate/large samples, it won't hurt either!

Digression: What about old-fashioned regression?

tinyurl.com/sree-drct

Regression method:

Fit model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_{1i} + \beta_3 X_{2i} + \ldots$$

Estimated effect: $\hat{\beta}_2$

Digression: What about old-fashioned regression?

tinyurl.com/sree-drct

Regression method:

Fit model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_{1i} + \beta_3 X_{2i} + \dots$$

Estimated effect: $\hat{\beta}_2$

Problem: What if the model is false?

- E.g. Y isn't linear in covariates
- E.g. What if there should be interactions?

Digression: What about old-fashioned regression?

tinyurl.com/sree-drct

Regression method:

Fit model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 X_{1i} + \beta_3 X_{2i} + \dots$$

Estimated effect: $\hat{\beta}_2$

Problem: What if the model is false?

- E.g. Y isn't linear in covariates
- E.g. What if there should be interactions?

Good news: $\hat{\beta}$ is *approximately* unbiased *in large samples*

Digression: What about old-fashioned regression?

tinyurl.com/sree-drct

Why our method?

1. *Exactly* unbiased *in any sample*
2. Use *any* algorithm for f^c, f^t
   - High dimensional covariates
   - Flexible for non-linearity, interactions

Digression: What about old-fashioned regression?

tinyurl.com/sree-drct

Why our method?

1. *Exactly* unbiased *in any sample*
2. Use *any* algorithm for f^c, f^t
   - High dimensional covariates
   - Flexible for non-linearity, interactions
   - $\Rightarrow$ better imputations
   - $\Rightarrow$ better effect estimates

Digression: What about old-fashioned regression?

tinyurl.com/sree-drct

Why our method?

1. *Exactly* unbiased *in any sample*
2. Use *any* algorithm for f^c, f^t
   - High dimensional covariates
   - Flexible for non-linearity, interactions
   - $\Rightarrow$ better imputations
   - $\Rightarrow$ better effect estimates
   - We recommend random forest

What You Need for Our Method

tinyurl.com/sree-drct

1. Randomized treatment variable
2. Outcome variable
3. Covariates
4. What is the experimental design?

What You Need for Our Method

tinyurl.com/sree-drct

1. Randomized treatment variable
2. Outcome variable
3. Covariates
4. What is the experimental design?

One last digression[1]: experimental designs

What You Need for Our Method

tinyurl.com/sree-drct

1. Randomized treatment variable
2. Outcome variable
3. Covariates
4. What is the experimental design?

One last digression[1]: experimental designs

[1] This is not a promise.

The Two Questions of Experimental Design

tinyurl.com/
sree-drct

1. Who or What is being randomized?
2. How are they being randomized?

Who/What is being randomized?

tinyurl.com/sree-drct

- Individual randomization
- Cluster or Group randomization

How are they being randomized?

tinyurl.com/sree-drct

- What's the probability each unit is assigned to treatment?
- How does one unit's assignment affect other units?

Examples We'll Cover

tinyurl.com/sree-drct

- Individual randomization
  - Bernoulli
  - Paired
- Cluster randomization
  - Paired

Examples

tinyurl.com/sree-drct

- ASSISTments E-Trials A/B test
  - Students are randomized individually
  - Students are randomized independently
  - $\Rightarrow$ Bernoullli

Examples

tinyurl.com/sree-drct

- ASSISTments E-Trials A/B test
  - Students are randomized individually
  - Students are randomized independently
  - $\Rightarrow$ Bernoullli
- Cognitive Tutor Effectiveness Study
  - *Schools* are randomized
  - Randomization is within pairs
  - (if your school is randomized to treatment, its pair *must* be randomized to control)
  - $\Rightarrow$ paired cluster design

Other Designs

tinyurl.com/sree-drct

To be implemented (hopefully) soon:

- "Completely randomized design"
    - At the outset, fix # randomized to treatment, # randomized to control
    - Now T_i and T_j are dependent!
- Block-randomized design
    - e.g. a separate completely randomized experiment in each classroom
    - Paired designs are a special case

Other Designs

tinyurl.com/sree-drct

To be implemented (hopefully) soon:

- “Completely randomized design”
  - At the outset, fix # randomized to treatment, # randomized to control
  - Now T_i and T_j are dependent!
- Block-randomized design
  - e.g. a separate completely randomized experiment in each classroom
  - Paired designs are a special case

Probably won’t get to for a while:

- Bandit designs
  - Probability i is assigned to treatment depends on previous subjects’ outcomes

tinyurl.com/
sree-drct

Estimating Effects in Practice

dRCT Package Overview

tinyurl.com/sree-drct

Installation:

- You will need to install the package from Github using the `devtools` package in *R*
- e.g. `install_github("manncz/dRCT")`

Primary Functions:

```
loop(Y, Tr, Z, pred, p, ...)
p_loop(Y, Tr, Z, pred, P, n, ...)
```

Covariate Adjustment with Bernoulli Randomized Trails (LOOP)

tinyurl.com/sree-drct

`loop(Y, Tr, Z, pred, p, ...)`

- `Y`: outcome vector
- `Tr`: treatment assignment vector
- `Z`: matrix of covariates
- `pred`: interpolation algorithm
- p: probability of treatment
- `...`: optional inputs for interpolation algorithm

LOOP interpolation algorithms

tinyurl.com/sree-drct

pred

- *loop_rf*
- *loop_ols*
- *loop_glm*

Covariate Adjustment with Paired Trails (P-LOOP)

tinyurl.com/sree-drct

```
p_loop(Y, Tr, Z, pred, P, n, ...)
```

- `Y`: outcome vector
- `Tr`: treatment assignment vector
- `Z`: matrix of covariates
- `pred`: interpolation algorithm
- P: vector of pair assignments
- n: optional vector of cluster sizes
- `...`: optional inputs for interpolation algorithm

P-LOOP interpolation algorithms

tinyurl.com/sree-drct

pred

- *p_ols_po*
- *p_ols_v12*
- **p_ols_interp**
- *p_rf_po*
- *p_rf_v12*
- **p_rf_interp**

Real Data Example: Texas School Data

tinyurl.com/sree-drct

- AEIS: School-level data from Texas Education Agency from 2003-2011
- > 3,000 schools
- TAKS (standardized test) passing rates
- Thousands of additional possible covariates

TEA Texas Education Agency

Real Data Example: Synthetic School-Level RCT

tinyurl.com/sree-drct

- Inspired by the Cognitive Tutor Algebra I study (Pane et al. 2014)

Real Data Example: Synthetic School-Level RCT

tinyurl.com/sree-drct

- Inspired by the Cognitive Tutor Algebra I study (Pane et al. 2014)
- **RCT Sample:** 50 Texas middle schools
- **Treatment:** Alternative 8th grade mathematics curriculum
- **Design:** Schools randomly assigned to implement new curriculum or continue standard in the 2007/8 school year

Real Data Example: Synthetic School-Level RCT

tinyurl.com/sree-drct

- Inspired by the Cognitive Tutor Algebra I study (Pane et al. 2014)
- **RCT Sample:** 50 Texas middle schools
- **Treatment:** Alternative 8th grade mathematics curriculum
- **Design:** Schools randomly assigned to implement new curriculum or continue standard in the 2007/8 school year
- **Outcome:** 2008 8th grade math TAKS passing rate

Real Data Example: Synthetic School-Level RCT

tinyurl.com/sree-drct

- Inspired by the Cognitive Tutor Algebra I study (Pane et al. 2014)
- **RCT Sample:** 50 Texas middle schools
- **Treatment:** Alternative 8th grade mathematics curriculum
- **Design:** Schools randomly assigned to implement new curriculum or continue standard in the 2007/8 school year
- **Outcome:** 2008 8th grade math TAKS passing rate
- **Pretest:** 2007 8th grade math TAKS passing rate

Your Turn!

tinyurl.com/sree-drct

1. Follow along while we talk through *01-explore-aeis-data.Rmd*
2. Work through *02-effect-est.Rmd*
   - Effect estimate for Bernoilli randomized trial
   - Effect estimate for paired randomed trial
   - Effect esitmate for paired cluster randomed trial
3. Flag any of us down as you have questions!

Auxiliary Data

By “Auxiliary Data” we mean a dataset that meets these criteria:

1. Doesn’t include data from RCT participants
2. Includes covariate data
3. Includes outcome data

tinyurl.com/sree-drct

Auxiliary Data

tinyurl.com/sree-drct

By "Auxiliary Data" we mean a dataset that meets these criteria:

1. Doesn't include data from RCT participants
2. Includes covariate data
3. Includes outcome data

Examples:

- A/B test: historical log data from users who worked on similar modules before the experiment started
- Field trial: Administrative (e.g. SLDS) data from students in schools that were not part of the RCT

Auxiliary Data

tinyurl.com/sree-drct

By "Auxiliary Data" we mean a dataset that meets these criteria:

1. Doesn't include data from RCT participants
2. Includes covariate data
3. Includes outcome data

Examples:

- A/B test: historical log data from users who worked on similar modules before the experiment started
- Field trial: Administrative (e.g. SLDS) data from students in schools that were not part of the RCT

Note: we have sometimes called this the "remnant"

What use is more data??

tinyurl.com/sree-drct

- Already imputing potential outcomes with f^c and f^t in LOOP
- f^c and f^t can be flexible, high dimensional
- They are fit to representative data

What use is more data??

tinyurl.com/sree-drct

- Already imputing potential outcomes with f^c and f^t in LOOP
- f^c and f^t can be flexible, high dimensional
- They are fit to representative data

Limits on f^c and f^t

- RCT sample size might be too small to fit *really* good models
- Human-adaptive modeling: no good!

Example 1: ASSISTments

Covariates:

- Log data. For each previous skillbuilder,
  - Completed skill builder?
  - # problems attempted / completed?
  - Time to mastery
- Demographic data

tinyurl.com/sree-drct

Example 1: ASSISTments

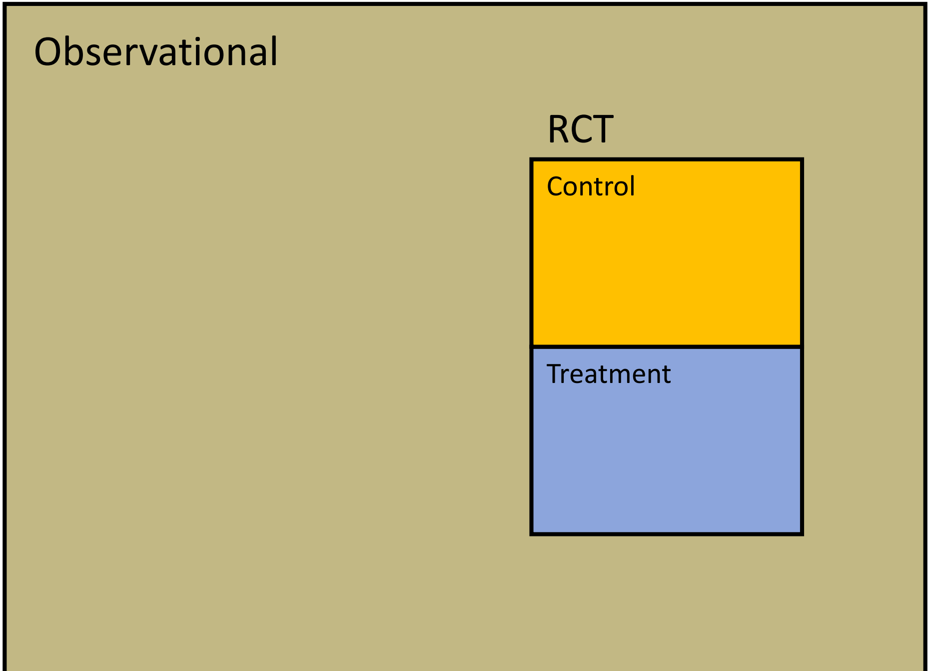
tinyurl.com/sree-drct

Covariates:

- Log data. For each previous skillbuilder,
    - Completed skill builder?
    - # problems attempted / completed?
    - Time to mastery
- Demographic data

Auxiliary Data:

- Observational
- Students who were *not* randomzied
    - Previous users
    - Current users not assigned to that skillbuilder
- Same covariates available



tinyurl.com/sree-drct

Observational

Step 1:

Train Model $\hat{y}(\cdot) : \boldsymbol{x} \to Y$

With auxiliary data

RCT

Control

Treatment

tinyurl.com/sree-drct

Observational

Step 1:
Train Model $\hat{y}(\cdot) : \boldsymbol{x} \rightarrow Y$
With auxiliary data

Step 2:
Extrapolate
With fitted model & RCT data

RCT

Control $\hat{y}(\boldsymbol{x}_i)$
Treatment $\hat{y}(\boldsymbol{x}_j)$

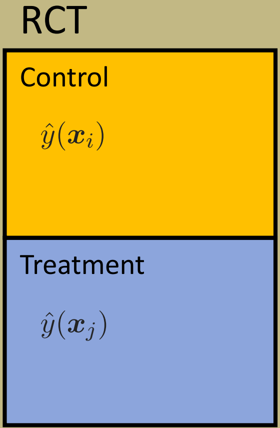
tinyurl.com/sree-drct

Observational

Step 1:
Train Model $\hat{y}(\cdot) : \boldsymbol{x} \to Y$
With auxiliary data

Step 2:
Extrapolate
With fitted model & RCT data

Step 3:
Use $\hat{y}(\boldsymbol{x})$ as a "super-covariate"



tinyurl.com/sree-drct

Auxiliary Data

tinyurl.com/sree-drct

Basic idea: Use auxiliary-based predictions $\hat{y}(\boldsymbol{x}_i)$ as a *covariate* in the RCT.

Auxiliary Data

tinyurl.com/sree-drct

Basic idea: Use auxiliary-based predictions $\hat{y}(\boldsymbol{x}_i)$ as a *covariate* in the RCT.

- The function $\hat{y}(\cdot)$ is fit on auxiliary data
- The covariates $\boldsymbol{x}$ are pre-treatment
- $\Rightarrow \hat{y}(\boldsymbol{x})$ is invariant to treatment assignment

Auxiliary Data

tinyurl.com/sree-drct

Basic idea: Use auxiliary-based predictions $\hat{y}(\boldsymbol{x}_i)$ as a *covariate* in the RCT.

- The function $\hat{y}(\cdot)$ is fit on auxiliary data
- The covariates $\boldsymbol{x}$ are pre-treatment
- $\Rightarrow \hat{y}(\boldsymbol{x})$ is invariant to treatment assignment
- $\hat{y}(\boldsymbol{x})$ might be an amazing covariate

Auxiliary Data

tinyurl.com/sree-drct

Basic idea: Use auxiliary-based predictions $\hat{y}(\boldsymbol{x}_i)$ as a *covariate* in the RCT.

- The function $\hat{y}(\cdot)$ is fit on auxiliary data
- The covariates $\boldsymbol{x}$ are pre-treatment
- $\Rightarrow \hat{y}(\boldsymbol{x})$ is invariant to treatment assignment
- $\hat{y}(\boldsymbol{x})$ might be an amazing covariate
- ...or it might not

Special Prediction Algorithm for LOOP

tinyurl.com/sree-drct

- If $\hat{y}(x)$ predicts Y really well, we would expect a linear relationship
  - $\Rightarrow$ fit OLS models within LOOP

Special Prediction Algorithm for LOOP

tinyurl.com/sree-drct

- If $\hat{y}(x)$ predicts Y really well, we would expect a linear relationship
  - $\Rightarrow$ fit OLS models within LOOP
- Maybe $\hat{y}(x)$ isn't that much better than other covariates (or, maybe it's useless)
  - $\Rightarrow$ use random forest within LOOP

Special Prediction Algorithm for LOOP

tinyurl.com/sree-drct

- If $\hat{y}(x)$ predicts Y really well, we would expect a linear relationship
  - $\Rightarrow$ fit OLS models within LOOP
- Maybe $\hat{y}(x)$ isn't that much better than other covariates (or, maybe it's useless)
  - $\Rightarrow$ use random forest within LOOP
- Let the data decide!
  - `pred=reloop`

tinyurl.com/sree-drct

Incorporating Auxiliary Data in Practice

Incorporating Auxiliary Information

`loop( Y , Tr , Z , pred = reloop , p, yhat , ...)`

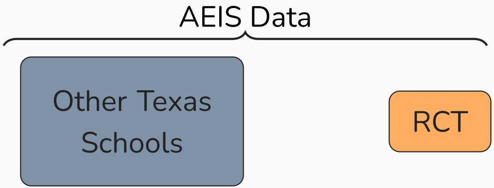
- `Y`: outcome vector
- `Tr`: treatment assignment vector
- `Z`: matrix of covariates
- `pred = reloop`: specify auxiliary data interpolation algorithm
- *p*: probability of treatment
- `yhat`: vector of auxiliary predictions
- `...`: optional inputs for interpolation algorithm

tinyurl.com/sree-drct

Real Data Example: Texas Schools

tinyurl.com/sree-drct

- AEIS data includes thousands of schools not in our RCT
- A great setting for integrating auxiliary and RCT data



Your Turn!

tinyurl.com/sree-drct

1. Work through *03-integrate-aux.Rmd*
   - We fit an auxiliary model and generate predictions to input as *yhat*
2. Apply what you learned in *04-effect-estABtest.Rmd*
3. Flag any of us down as you have questions!

Conceptual Overview

Estimating Effects with RCT Data

Incorporating Auxiliary Data

Treatment Effect Heterogeneity

Planning Experiments

Heterogeneous Treatment Effects

tinyurl.com/sree-drct

Recall: The individual treatment effect is $\tau_i = y_i^t - y_i^c$

- Until now, our goal has been the average treatment effect (ATE)

$$\bar{\tau} = \frac{1}{N} \sum_{i=1}^{N} \tau_i$$

Heterogeneous Treatment Effects

tinyurl.com/sree-drct

Recall: The individual treatment effect is $\tau_i = y_i^t - y_i^c$

- Until now, our goal has been the average treatment effect (ATE)

$$\bar{\tau} = \frac{1}{N}\sum_{i=1}^{N} \tau_i$$

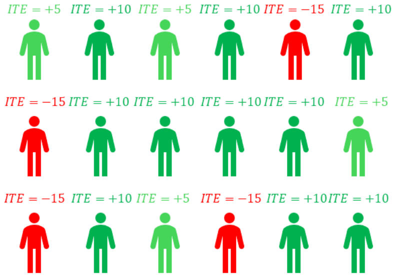
- We can use the same tools for other models of τ_i:
  - Averages for subgroups (subgroup effects)
  - Moderation: look for patterns in effects$\leftrightarrow$covariates $\tau_i|\boldsymbol{x}_i$
  - Predict an individual's treatment effect $\hat{\tau}_i$

Example

tinyurl.com/sree-drct

Example

tinyurl.com/sree-drct



Conditional Average Treatment Effect

tinyurl.com/sree-drct

- The conditional average treatment effect (CATE) is

$$\tau(x) = \mathbb{E}[\tau_i | \boldsymbol{X}_i = \boldsymbol{x}] = \mathbb{E}[y_i^t - y_i^c | \boldsymbol{X}_i = \boldsymbol{x}]$$

- The expected treatment effect **conditional on having a specific set of covariate values**.

Conditional Average Treatment Effect

tinyurl.com/sree-drct

- The conditional average treatment effect (CATE) is

$$\tau(x) = \mathbb{E}[\tau_i | \boldsymbol{X}_i = \boldsymbol{x}] = \mathbb{E}[y_i^t - y_i^c | \boldsymbol{X}_i = \boldsymbol{x}]$$

- The expected treatment effect **conditional on having a specific set of covariate values**.
- “iCATE”: expected effect based on i’s covariates,

$$\tau(\boldsymbol{x}_i) = \mathbb{E}[\tau_i | \boldsymbol{X}_i = \boldsymbol{x}_i]$$

We're already almost there

tinyurl.com/sree-drct

In order to get the ATE, we already have imputations:

$$\hat{y}_i^c = f^c(X_i)$$
$$\hat{y}_i^t = f^t(X_i)$$

And weighted average: $\hat{m}_i = Pr(Z_i = 0)\hat{y}_i^t + Pr(Z_i = 1)\hat{y}_i^c$
(For each i, we use everyone *but* i to estimate functions $f^c(\cdot)$ and $f^t(\cdot)$.)

We're already almost there

tinyurl.com/sree-drct

Also: an unbiased estimator for τ_i (!):

$$\hat{\tau}_i = U_i \left(Y_i - \hat{m}_i \right)$$

where

$$U_i = \begin{cases} \frac{1}{p_i} \text{ if } T_i = 1 \\ \frac{-1}{1-p_i} \text{ if } T_i = 0 \end{cases}$$

We're already almost there

Also: an unbiased estimator for τ_i (!):

$$\hat{\tau}_i = U_i \left(Y_i - \hat{m}_i \right)$$

where

$$U_i = \begin{cases} \frac{1}{p_i} \text{ if } T_i = 1 \\ \frac{-1}{1-p_i} \text{ if } T_i = 0 \end{cases}$$

If

$$T_i \hat{m}_i$$

Then

$$\mathbb{E}[\hat{\tau}_i] = \tau_i$$

tinyurl.com/sree-drct

Modeling with $\hat{\tau}$

tinyurl.com/sree-drct

$\hat{\tau}$ will typically be too noisy to be of much use by itself

$\ldots$ but it can be used for modeling

Modeling with $\hat{\tau}$

tinyurl.com/sree-drct

$\hat{\tau}$ will typically be too noisy to be of much use by itself

... but it can be used for modeling

- Estimating subgroup effects

Modeling with $\hat{\tau}$

tinyurl.com/sree-drct

$\hat{\tau}$ will typically be too noisy to be of much use by itself

... but it can be used for modeling

- Estimating subgroup effects
- Parametric moderation modeling

Modeling with $\hat{\tau}$

tinyurl.com/sree-drct

$\hat{\tau}$ will typically be too noisy to be of much use by itself

... but it can be used for modeling

- Estimating subgroup effects
- Parametric moderation modeling
- Non-parametric (or ML) modeling for the iCATE

Subgroup Effects

tinyurl.com/sree-drct

The sample mean of $\hat{\tau}$ for a subgroup is unbiased for the CATE (conditional average treatment effect) in that subgroup.

Moderation Analysis

tinyurl.com/sree-drct

Example: OLS

Fit model:

$$\hat{\tau}_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i$$

Moderation Analysis

tinyurl.com/sree-drct

Example: OLS

Fit model:

$$\hat{\tau}_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i$$

- If τ is linear in $X_1, \ldots, X_k$ then estimated slopes $\hat{\beta}$ are unbiased for true slopes
- If not, estimated slopes $\hat{\beta}$ are unbiased for "populaiton regression"—slopes you would estimate if you had true τ instead of estimates

Moderation Analysis

tinyurl.com/sree-drct

Example: OLS

Fit model:

$$\hat{\tau}_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i$$

- If τ is linear in $X_1, \ldots, X_k$ then estimated slopes $\hat{\beta}$ are unbiased for true slopes
- If not, estimated slopes $\hat{\beta}$ are unbiased for "populaiton regression"—slopes you would estimate if you had true τ instead of estimates

Use heteroskedasticity-robust SEs

Moderation Analysis

tinyurl.com/sree-drct

Example: OLS

Fit model:

$$\hat{\tau}_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \epsilon_i$$

- If τ is linear in $X_1, \ldots, X_k$ then estimated slopes $\hat{\beta}$ are unbiased for true slopes
- If not, estimated slopes $\hat{\beta}$ are unbiased for "populaiton regression"—slopes you would estimate if you had true τ instead of estimates

Use heteroskedasticity-robust SEs

OLS is only one example

Estimating iCATEs

tinyurl.com/sree-drct

iCATE: average treatment effect for subgroup with $\boldsymbol{x} = \boldsymbol{x}_i$

Use any old model for $\hat{\tau}(\boldsymbol{x})$, as long as it fits well

Advantages of This Approach

tinyurl.com/sree-drct

- Can model y^C with any model–including ML
  - Regardless of heterogeneity research question
  - Model need not be correct
- Flexible with regards to model for τ as a function of $\boldsymbol{x}$
- Built off of unbiased $\hat{\tau}$
  - If model for $\tau|\boldsymbol{x}$ is wrong, may still get biased estimators
  - ... but probably less biased than methods built on *biased* $\hat{\tau}$

In the Package

tinyurl.com/sree-drct

You can use the function $getITE$ to retrieve the ITE estimates from a LOOP estimator.

In the Package

tinyurl.com/sree-drct

You can use the function $getITE$ to retrieve the ITE estimates from a LOOP estimator.

- This function will work for an estimator built with or without auxiliary data, which allows us to improve precision further.

In the Package

tinyurl.com/sree-drct

You can use the function *getITE* to retrieve the ITE estimates from a LOOP estimator.

- This function will work for an estimator built with or without auxiliary data, which allows us to improve precision further.
- However, it is currently only for Bernoulli-randomized experiments.

In the Package

tinyurl.com/sree-drct

You can use the function $getITE$ to retrieve the ITE estimates from a LOOP estimator.

- This function will work for an estimator built with or without auxiliary data, which allows us to improve precision further.
- However, it is currently only for Bernoulli-randomized experiments.
- Once you have retrieve the estimates, choose your favorite model and do some regressing!

Your Turn!

tinyurl.com/sree-drct

1. Work through *`05-heterogeneousEffects.Rmd`*
    - We fit retrieve ITE estimates from the model in *`04-effect-estABtest.Rmd`*.
    - We then estimate the CATE by regressing these estimates on the covariates.
2. Flag any of us down as you have questions!

What You Need

tinyurl.com/sree-drct

- We'll be using the *dRCTpower* package to plan experiments
- Main function is *run_app*
- You can download the package in R using the following commands:

```
install.packages("devtools")
devtools::install_github("jaylinlowe/dRCTpower")
```

- We will be using the *aux_dat_small.csv* file from the Github repo

Main Question

tinyurl.com/sree-drct

How to choose a sample size for our experiment, particularly if auxiliary data will be incorporated?

Overview

tinyurl.com/
sree-drct

- Incorporating auxiliary data in our analysis can improve precision, meaning we can have a smaller sample size with the same power

Overview

tinyurl.com/sree-drct

- Incorporating auxiliary data in our analysis can improve precision, meaning we can have a smaller sample size with the same power
- Gain in precision is determined by how predictive a model fit on the auxiliary data is for the RCT

Overview

tinyurl.com/sree-drct

- Incorporating auxiliary data in our analysis can improve precision, meaning we can have a smaller sample size with the same power
- Gain in precision is determined by how predictive a model fit on the auxiliary data is for the RCT
- But....we don't have the RCT data!

Method

tinyurl.com/sree-drct

1. Break auxiliary dataset into subgroups

Method

tinyurl.com/sree-drct

1. Break auxiliary dataset into subgroups
2. For each subgroup, treat it as the RCT and the rest of the data as the auxiliary data

Method

tinyurl.com/sree-drct

1. Break auxiliary dataset into subgroups
2. For each subgroup, treat it as the RCT and the rest of the data as the auxiliary data
3. Calculate the required sample size under this framework

Requirements

tinyurl.com/
sree-drct

Large auxiliary dataset that:

- is substantially larger than the RCT will be

Requirements

tinyurl.com/sree-drct

Large auxiliary dataset that:

- is substantially larger than the RCT will be
- has covariates

Requirements

tinyurl.com/sree-drct

Large auxiliary dataset that:

- is substantially larger than the RCT will be
- has covariates
- has the same outcome of interest as the RCT

Limitations

tinyurl.com/sree-drct

- Method is only plausible if it's reasonable to assume the RCT looks like some subgroup of the auxiliary data, even if we don't know what subgroup that is

Limitations

tinyurl.com/sree-drct

- Method is only plausible if it's reasonable to assume the RCT looks like some subgroup of the auxiliary data, even if we don't know what subgroup that is
- Dangerous to assume RCT looks like any one subgroup

Limitations

tinyurl.com/sree-drct

- Method is only plausible if it's reasonable to assume the RCT looks like some subgroup of the auxiliary data, even if we don't know what subgroup that is
- Dangerous to assume RCT looks like any one subgroup
- Dangerous to choose most optimistic option

General Power Calculations

$$n = 2\sigma^2 \frac{(\xi_{1-\alpha/2} + \xi_{1-\beta})^2}{\Delta_A^2}$$

tinyurl.com/sree-drct

General Power Calculations

tinyurl.com/sree-drct

$$n = 2\sigma^2 \frac{(\xi_{1-\alpha/2} + \xi_{1-\beta})^2}{\Delta_A^2}$$

- $\xi_{1-\alpha/2}$ is the critical value obtained from a normal distribution for Type I error equal to α.

General Power Calculations

tinyurl.com/sree-drct

$$n = 2\sigma^2 \frac{(\xi_{1-\alpha/2} + \xi_{1-\beta})^2}{\Delta_A^2}$$

- $\xi_{1-\alpha/2}$ is the critical value obtained from a normal distribution for Type I error equal to α.
- $\xi_{1-\beta}$ is the critical value from a normal distribution for Type II error rate β.

General Power Calculations

tinyurl.com/sree-drct

$$n = 2\sigma^2 \frac{(\xi_{1-\alpha/2} + \xi_{1-\beta})^2}{\Delta_A^2}$$

- $\xi_{1-\alpha/2}$ is the critical value obtained from a normal distribution for Type I error equal to α.
- $\xi_{1-\beta}$ is the critical value from a normal distribution for Type II error rate β.
- Δ_A is the effect size, typically 20% of the standard deviation of the outcome in the population

General Power Calculations

tinyurl.com/sree-drct

$$n = 2\sigma^2 \frac{(\xi_{1-\alpha/2} + \xi_{1-\beta})^2}{\Delta_A^2}$$

- $\xi_{1-\alpha/2}$ is the critical value obtained from a normal distribution for Type I error equal to α.
- $\xi_{1-\beta}$ is the critical value from a normal distribution for Type II error rate β.
- Δ_A is the effect size, typically 20% of the standard deviation of the outcome in the population
- σ^2 is the true variance of the outcome in the population, typically replaced with an estimate from a sample

Our Modification

tinyurl.com/sree-drct

- We replace σ^2 with an estimate from each subgroup

Our Modification

tinyurl.com/sree-drct

- We replace σ^2 with an estimate from each subgroup
- Shiny app gives two estimates, one if you were to use auxiliary data in analysis, and one without

Our Modification

tinyurl.com/sree-drct

- We replace σ^2 with an estimate from each subgroup
- Shiny app gives two estimates, one if you were to use auxiliary data in analysis, and one without
- ”Without auxiliary data” estimate is variance of outcome for that subgroup

Our Modification

tinyurl.com/sree-drct

- We replace σ^2 with an estimate from each subgroup
- Shiny app gives two estimates, one if you were to use auxiliary data in analysis, and one without
- "Without auxiliary data" estimate is variance of outcome for that subgroup
- "With auxiliary data" estimate is variance of the residuals, $(y_i - \hat{y}_i)$, where $\hat{y}_i$ are out-of-bag predictions from model

Defining Subgroups

tinyurl.com/sree-drct

Three options:

1. Categorical Variable
    - Divide based on levels of categorical variable
    - Can create your own categorical variables

Defining Subgroups

tinyurl.com/sree-drct

Three options:

1. Categorical Variable
   - Divide based on levels of categorical variable
   - Can create your own categorical variables
2. Numerical Variable
   - Divide into 10 (adjustable) equally sized groups
   - May need to divide into fewer if there isn't enough variation

Defining Subgroups

tinyurl.com/sree-drct

Three options:

1. Categorical Variable
   - Divide based on levels of categorical variable
   - Can create your own categorical variables
2. Numerical Variable
   - Divide into 10 (adjustable) equally sized groups
   - May need to divide into fewer if there isn't enough variation
3. Best-Worst Case Scenario
   - Divide based on how predictive we expect the auxiliary model to be for that group
   - Good starting point

tinyurl.com/sree-drct

Shiny App Demo

References

tinyurl.com/
sree-drct

tinyurl.com/sree-drct

Gagnon-Bartsch, Johann A., Adam C. Sales, Edward Wu, Anthony F. Botelho, John A. Erickson, Luke W. Miratrix and Neil T. Heffernan. 2023. “Precise unbiased estimation in randomized experiments using auxiliary observational data.” *Journal of Causal Inference* 11(1):20220011. **URL:** *https://www.degruyter.com/document/doi/10.1515/jci-2022-0011/html*

Künzel, Sören R, Jasjeet S Sekhon, Peter J Bickel and Bin Yu. 2019. “Metalearners for estimating heterogeneous treatment effects using machine learning.” *Proceedings of the National Academy of Sciences* 116(10):4156–4165.

Lowe, Jaylin, Charlotte Mann, Jiaying Wang, Adam Sales and Johann Gagnon-Bartsch. Forthcoming. “Power Calculations for Randomized Controlled Trials with Auxiliary Observational Data.” *EDM 2024* .

Mann, Charlotte, Jiaying Wang, Adam Sales and Johann Gagnon-Bartsch. Forthcoming. “Using Publicly Available Auxiliary Data to Improve Precision of Treatment Effect Estimation in a Randomized Efficacy Trial.” *EDM 2024* .

Pane, John F., Beth Ann Griffin, Daniel F. McCaffrey and Rita Karam. 2014. “Effectiveness of Cognitive Tutor Algebra I at Scale.” *Educational Evaluation and Policy Analysis* 36(2):127–144.
URL: *https://doi.org/10.3102/0162373713507480*

Pham, Duy, Kirk Vanacore, Adam Sales and Johann Gagnon-Bartsch. Forthcoming. “LOOL: Towards Personalization with Flexible Robust Estimation of Heterogeneous Treatment Effects.” *EDM 2024* .

Sales, Adam C, Ethan B Prihar, Johann A Gagnon-Bartsch and Neil T Heffernan. 2023. “Using Auxiliary Data to Boost Precision in the Analysis of A/B Tests on an Online Educational Platform: New Data and New Results.” *arXiv preprint arXiv:2306.06273* .

tinyurl.com/sree-drct

Wager, Stefan and Susan Athey. 2018. “Estimation and Inference of Heterogeneous Treatment Effects using Random Forests.” *Journal of the American Statistical Association* 113(523):1228–1242.

Wu, Edward and Johann A. Gagnon-Bartsch. 2018. “The LOOP Estimator: Adjusting for Covariates in Randomized Experiments.” *Evaluation Review* 42(4):458–488. Publisher: SAGE Publications Inc.
URL: *https://doi.org/10.1177/0193841X18808003*

Wu, Edward and Johann A. Gagnon-Bartsch. 2021. “Design-Based Covariate Adjustments in Paired Experiments.” *Journal of Educational and Behavioral Statistics* 46(1):109–132. Publisher: American Educational Research Association.
URL: *https://doi.org/10.3102/1076998620941469*

tinyurl.com/
sree-drct