PS207 Quantitative Causal Inference

Experiments

Matto Mildenberger

UC Santa Barbara

Special thanks to Chad Hazlett (UCLA) for select slides, used with permission

Randomization Solves the Selection Problem

Recall the selection bias formula for diff. in means ($\tilde{\tau}$)

$$\begin{aligned}
\tilde{\tau} &= \mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] \quad \text{(obs. diff. in means)} \\
&= \mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0] \\
&= \underbrace{\mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1]}_{\tau_{ATT}} + \underbrace{\mathbb{E}[Y_{0i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0]}_{\text{Bias}}
\end{aligned}$$

How can we eliminate the bias term?

Randomization Solves the Selection Problem

Recall the selection bias formula for diff. in means ($\tilde{\tau}$)

$$\begin{aligned}
\tilde{\tau} &= \mathbb{E}[Y_i|D_i=1] - \mathbb{E}[Y_i|D_i=0] \quad \text{(obs. diff. in means)} \\
&= \mathbb{E}[Y_{1i}|D_i=1] - \mathbb{E}[Y_{0i}|D_i=0] \\
&= \underbrace{\mathbb{E}[Y_{1i} - Y_{0i}|D_i=1]}_{\tau_{ATT}} + \underbrace{\mathbb{E}[Y_{0i}|D_i=1] - \mathbb{E}[Y_{0i}|D_i=0]}_{\text{Bias}}
\end{aligned}$$

How can we eliminate the bias term?

Random assignment of D_i will make the treated and untreated units identical on average, such that

$$\mathbb{E}[Y_{0i} \mid D_i = 1] = \mathbb{E}[Y_{0i} \mid D_i = 0]$$

This implies Bias $= 0$.

Are Experiments Feasible in Social Science?

- Large increase in the use of experiments in the social sciences: laboratory, survey, and field experiments

Are Experiments Feasible in Social Science?

- Large increase in the use of experiments in the social sciences: laboratory, survey, and field experiments
- *Abbreviated* list of examples (from Green 2008):
  - Program evaluation: development programs, education programs, SAT prep classes, weight loss programs, diversity training, deliberative polls, advertising campaigns, website designs...
  - Public policy evaluation: teacher pay, student incentives, class size, speed traps, vouchers, alternative sentencing, job training, health insurance subsidies, tax compliance, public housing
  - Behavioral research: persuasion, mobilization, education, income, interpersonal influence, conscientious health behaviors, media exposure, deliberation, discrimination
  - Research on institutions: transparency, corruption, electoral systems, information

Outline

1 Identification

2 Hypothesis Testing

3 Randomization Inference

4 Threats to Validity

5 Reviewing What We've Covered So Far

Identification vs. Estimation

Goal of causal inference: Learn about a counterfactual quantity of interest (QoI) using *finite, observed* data.

Identification vs. Estimation

Goal of causal inference: Learn about a counterfactual quantity of interest (QoI) using *finite, observed* data.

Causal inference thus involves two inferential hurdles:

1. Identification: If you can observe data from an entire *population*, can you learn about your QoI?
2. Estimation: Given your finite amount of data on a *sample*, how well can you learn about your QoI?

Identification vs. Estimation

Goal of causal inference: Learn about a counterfactual quantity of interest (QoI) using *finite, observed* data.

Causal inference thus involves two inferential hurdles:

1. Identification: If you can observe data from an entire *population*, can you learn about your QoI?
2. Estimation: Given your finite amount of data on a *sample*, how well can you learn about your QoI?

Golden rule of inference: IDENTIFICATION PRECEDES ESTIMATION

Identification vs. Estimation

Goal of causal inference: Learn about a counterfactual quantity of interest (QoI) using *finite, observed* data.

Causal inference thus involves two inferential hurdles:

1. Identification: If you can observe data from an entire *population*, can you learn about your QoI?
2. Estimation: Given your finite amount of data on a *sample*, how well can you learn about your QoI?

Golden rule of inference: IDENTIFICATION PRECEDES ESTIMATION

Example: $Y_i = X_i^\top \beta + \varepsilon_i$ (normal linear regression)

- When is β identified?

Identification vs. Estimation

Goal of causal inference: Learn about a counterfactual quantity of interest (QoI) using *finite, observed* data.

Causal inference thus involves two inferential hurdles:

1. Identification: If you can observe data from an entire *population*, can you learn about your QoI?
2. Estimation: Given your finite amount of data on a *sample*, how well can you learn about your QoI?

Golden rule of inference: IDENTIFICATION PRECEDES ESTIMATION

Example: $Y_i = X_i^\top \beta + \varepsilon_i$ (normal linear regression)

- When is β identified? No collinearity among X_i.
- When is β unbiasedly estimated?

Identification vs. Estimation

Goal of causal inference: Learn about a counterfactual quantity of interest (QoI) using *finite, observed* data.

Causal inference thus involves two inferential hurdles:

1. Identification: If you can observe data from an entire *population*, can you learn about your QoI?
2. Estimation: Given your finite amount of data on a *sample*, how well can you learn about your QoI?

Golden rule of inference: IDENTIFICATION PRECEDES ESTIMATION

Example: $Y_i = X_i^\top \beta + \varepsilon_i$ (normal linear regression)

- When is β identified? No collinearity among X_i.
- When is β unbiasedly estimated? $\mathbb{E}(\varepsilon_i \mid X_i) = 0$.
- What is an unbiased estimator for β?

Identification vs. Estimation

Goal of causal inference: Learn about a counterfactual quantity of interest (QoI) using *finite, observed* data.

Causal inference thus involves two inferential hurdles:

1. Identification: If you can observe data from an entire *population*, can you learn about your QoI?
2. Estimation: Given your finite amount of data on a *sample*, how well can you learn about your QoI?

Golden rule of inference: IDENTIFICATION PRECEDES ESTIMATION

Example: $Y_i = X_i^\top \beta + \varepsilon_i$ (normal linear regression)

- When is β identified? No collinearity among X_i.
- When is β unbiasedly estimated? $\mathbb{E}(\varepsilon_i \mid X_i) = 0$.
- What is an unbiased estimator for β? $\hat{\beta} = \left(\sum_{i=1}^n X_i X_i^\top\right)^{-1} \left(\sum_{i=1}^n X_i Y_i\right)$.

Classical Randomized Experiment

Setup:

- Units: $i = 1, ..., N$
- Treatment: $D_i \in \{0, 1\}$, randomly assigned
- Potential outcomes: Y_{0i}, Y_{1i}
- Observed outcome: $Y_i = Y_{D_i i}$
- Number of treated/untreated units: $N_1 = \sum_{i=1}^{N} D_i$ and $N_0 = N - N_1$

Classical Randomized Experiment

Setup:

- Units: $i = 1, ..., N$
- Treatment: $D_i \in \{0, 1\}$, randomly assigned
- Potential outcomes: Y_{0i}, Y_{1i}
- Observed outcome: $Y_i = Y_{D_i i}$
- Number of treated/untreated units: $N_1 = \sum_{i=1}^{N} D_i$ and $N_0 = N - N_1$

Notes:

- For now, we assume we have data on the entire population (N = sample size = population size)

Classical Randomized Experiment

Setup:

- Units: $i = 1, ..., N$
- Treatment: $D_i \in \{0, 1\}$, randomly assigned
- Potential outcomes: Y_{0i}, Y_{1i}
- Observed outcome: $Y_i = Y_{D_i i}$
- Number of treated/untreated units: $N_1 = \sum_{i=1}^{N} D_i$ and $N_0 = N - N_1$

Notes:

- For now, we assume we have data on the entire population (N = sample size = population size)
- Random assignment can take one of several forms:
  - Complete randomization: Exactly N_1 treated units
  - Simple (Bernoulli) randomization: Each unit independently assigned to treatment with probability p

Classical Randomized Experiment

Setup:

- Units: $i = 1, ..., N$
- Treatment: $D_i \in \{0, 1\}$, randomly assigned
- Potential outcomes: Y_{0i}, Y_{1i}
- Observed outcome: $Y_i = Y_{D_i i}$
- Number of treated/untreated units: $N_1 = \sum_{i=1}^{N} D_i$ and $N_0 = N - N_1$

Notes:

- For now, we assume we have data on the entire population (N = sample size = population size)
- Random assignment can take one of several forms:
  - Complete randomization: Exactly N_1 treated units
  - Simple (Bernoulli) randomization: Each unit independently assigned to treatment with probability p

Randomization (simple or complete) implies: $\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp D_i$

Identification of Average Treatment Effect

Identification assumption (guaranteed by random assignment):

$$\{Y_{1i},\ Y_{0i}\} \perp\!\!\!\perp D_i$$

Quantity of interest:

$$\tau_{ATE} \equiv \mathbb{E}[Y_{1i} - Y_{0i}] = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})$$

Identification of Average Treatment Effect

Identification assumption (guaranteed by random assignment):

$$\{Y_{1i},\ Y_{0i}\} \perp\!\!\!\perp D_i$$

Quantity of interest:

$$\tau_{ATE} \equiv \mathbb{E}[Y_{1i} - Y_{0i}] = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})$$

Is τ_{ATE} identified?

Identification of Average Treatment Effect

Identification assumption (guaranteed by random assignment):

$$\{Y_{1i},\ Y_{0i}\} \perp\!\!\!\perp D_i$$

Quantity of interest:

$$\tau_{ATE} \equiv \mathbb{E}[Y_{1i} - Y_{0i}] = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})$$

Is τ_{ATE} identified?

$$\begin{aligned}\mathbb{E}[Y_i|D_i = 1] &= \mathbb{E}[D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}|D_i = 1] \\ &= \mathbb{E}[Y_{1i}|D_i = 1]\end{aligned}$$

Identification of Average Treatment Effect

Identification assumption (guaranteed by random assignment):

$$\{Y_{1i},\ Y_{0i}\} \perp\!\!\!\perp D_i$$

Quantity of interest:

$$\tau_{ATE} \equiv \mathbb{E}[Y_{1i} - Y_{0i}] = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})$$

Is τ_{ATE} identified?

$$\begin{aligned}\mathbb{E}[Y_i|D_i = 1] &= \mathbb{E}[D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}|D_i = 1] \\ &= \mathbb{E}[Y_{1i}|D_i = 1] = \mathbb{E}[Y_{1i}] \quad \text{(by random assignment)}\end{aligned}$$

Similarly, $\mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{0i}]$

Identification of Average Treatment Effect

Identification assumption (guaranteed by random assignment):

$$\{Y_{1i},\ Y_{0i}\} \perp\!\!\!\perp D_i$$

Quantity of interest:

$$\tau_{ATE} \equiv \mathbb{E}[Y_{1i} - Y_{0i}] = \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i})$$

Is τ_{ATE} identified?

$$\begin{aligned}\mathbb{E}[Y_i|D_i = 1] &= \mathbb{E}[D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}|D_i = 1] \\ &= \mathbb{E}[Y_{1i}|D_i = 1] = \mathbb{E}[Y_{1i}] \quad \text{(by random assignment)}\end{aligned}$$

Similarly, $\mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[Y_{0i}]$
So it follows that

$$\tau_{ATE} = \mathbb{E}[Y_{1i}] - \mathbb{E}[Y_{0i}] = \underbrace{\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]}_{\text{observed difference in means}}$$

$$= \frac{1}{N_1}\sum_{i=1}^{N} D_i Y_i - \frac{1}{N_0}\sum_{i=1}^{N}(1 - D_i) Y_i$$

SATE and PATE

- Often we focus on estimating the average causal effect in a particular sample: **S**ample **A**verage **T**reatment **E**ffect (SATE)
  - uncertainty arises only from hypothetical randomizations.
  - inferences are limited to the sample in our study.

SATE and PATE

- Often we focus on estimating the average causal effect in a particular sample: **S**ample **A**verage **T**reatment **E**ffect (SATE)
  - uncertainty arises only from hypothetical randomizations.
  - inferences are limited to the sample in our study.
- Compare to the **P**opulation **A**verage **T**reatment **E**ffect (PATE)
  - requires knowledge about the sampling process that got you from population to sample
  - need to account for two sources of variation:
    - variation from the sampling process
    - variation from treatment assignment (randomness of D_i)

SATE and PATE

- Often we focus on estimating the average causal effect in a particular sample: **S**ample **A**verage **T**reatment **E**ffect (SATE)
  - uncertainty arises only from hypothetical randomizations.
  - inferences are limited to the sample in our study.
- Compare to the **P**opulation **A**verage **T**reatment **E**ffect (PATE)
  - requires knowledge about the sampling process that got you from population to sample
  - need to account for two sources of variation:
    - variation from the sampling process
    - variation from treatment assignment (randomness of D_i)
- Given additional uncertainty in PATE, you might (correctly) expect

$$\text{Var}(\widehat{PATE}) > \text{Var}(\widehat{SATE})$$

Uncertainty for SATE and PATE

A first attempt at standard errors,

- Treat $\overline{Y}_1$ as if computed from a random sampling of Y_{1i} in population
- Likewise, treat $\overline{Y}_0$ as if from a (separate) sampling of the Y_{0i}'s
- Chance variation in the two should be unrelated: $cov(\overline{Y}_1, \overline{Y}_1) = 0$

Uncertainty for SATE and PATE

A first attempt at standard errors,

- Treat $\overline{Y}_1$ as if computed from a random sampling of Y_{1i} in population
- Likewise, treat $\overline{Y}_0$ as if from a (separate) sampling of the Y_{0i}'s
- Chance variation in the two should be unrelated: $cov(\overline{Y}_1, \overline{Y}_1) = 0$

Get variance of the difference in means:

$$\begin{aligned}\mathbb{V}(\overline{Y}_1 - \overline{Y}_0) &= \mathbb{V}(\overline{Y}_1) + \mathbb{V}(\overline{Y}_0) + 2cov(\overline{Y}_1, \overline{Y}_0) \\ &= \frac{\sigma^2_{Y1}}{N_1} + \frac{\sigma^2_{Y0}}{N_0}\end{aligned}$$

$$SE_{ATE} = \sqrt{\frac{\sigma^2_{Y1}}{N_1} + \frac{\sigma^2_{Y0}}{N_0}}$$

where σ^2_{Y1} and σ^2_{Y0} are the variance of the Y_{1i} and Y_{0i} in the population

Uncertainty for SATE and PATE

A first attempt at standard errors,

- Treat $\overline{Y}_1$ as if computed from a random sampling of Y_{1i} in population
- Likewise, treat $\overline{Y}_0$ as if from a (separate) sampling of the Y_{0i}'s
- Chance variation in the two should be unrelated: $cov(\overline{Y}_1, \overline{Y}_1) = 0$

Get variance of the difference in means:

$$\begin{aligned}\mathbb{V}(\overline{Y}_1 - \overline{Y}_0) &= \mathbb{V}(\overline{Y}_1) + \mathbb{V}(\overline{Y}_0) + 2cov(\overline{Y}_1, \overline{Y}_0) \\ &= \frac{\sigma^2_{Y1}}{N_1} + \frac{\sigma^2_{Y0}}{N_0}\end{aligned}$$

$$SE_{ATE} = \sqrt{\frac{\sigma^2_{Y1}}{N_1} + \frac{\sigma^2_{Y0}}{N_0}}$$

where σ^2_{Y1} and σ^2_{Y0} are the variance of the Y_{1i} and Y_{0i} in the population

Uncertainty for SATE and PATE

A first attempt at standard errors,

- Treat $\overline{Y}_1$ as if computed from a random sampling of Y_{1i} in population
- Likewise, treat $\overline{Y}_0$ as if from a (separate) sampling of the Y_{0i}'s
- Chance variation in the two should be unrelated: $cov(\overline{Y}_1, \overline{Y}_1) = 0$

Get variance of the difference in means:

$$\begin{aligned}\mathbb{V}(\overline{Y}_1 - \overline{Y}_0) &= \mathbb{V}(\overline{Y}_1) + \mathbb{V}(\overline{Y}_0) + 2cov(\overline{Y}_1, \overline{Y}_0) \\ &= \frac{\sigma^2_{Y1}}{N_1} + \frac{\sigma^2_{Y0}}{N_0}\end{aligned}$$

$$SE_{ATE} = \sqrt{\frac{\sigma^2_{Y1}}{N_1} + \frac{\sigma^2_{Y0}}{N_0}}$$

where σ^2_{Y1} and σ^2_{Y0} are the variance of the Y_{1i} and Y_{0i} in the population

This is what we typically use, and it is

- the correct SE for the PATE
- what you get from t-tests (unequal variance), regression with robust SE
- conservative for the SATE, where the sample fixed by D_i is random

Standard Error for Sample ATE

In finite sample, sampling without replacement occurs; changes variance of group means, and induces (small) covariance

Standard Error for Sample ATE

Given complete randomization of N units with N_1 assigned to treatment and $N_0 = N - N_1$ to control, the true standard error of the estimated sample ATE is given by

$$SE_{\widehat{ATE}} = \sqrt{\left(\frac{N - N_1}{N - 1}\right)\frac{Var[Y_{1i}]}{N_1} + \left(\frac{N - N_0}{N - 1}\right)\frac{Var[Y_{0i}]}{N_0} + \left(\frac{1}{N - 1}\right)2Cov[Y_{1i}, Y_{0i}]}$$

with population variances and covariances $Var[Y_{di}]$, $Cov(Y_{1i}, Y_{0i})$.

Standard Error for Sample ATE

In finite sample, sampling without replacement occurs; changes variance of group means, and induces (small) covariance

Standard Error for Sample ATE

Given complete randomization of N units with N_1 assigned to treatment and $N_0 = N - N_1$ to control, the true standard error of the estimated sample ATE is given by

$$SE_{\widehat{ATE}} = \sqrt{\left(\frac{N - N_1}{N - 1}\right)\frac{Var[Y_{1i}]}{N_1} + \left(\frac{N - N_0}{N - 1}\right)\frac{Var[Y_{0i}]}{N_0} + \left(\frac{1}{N - 1}\right)2Cov[Y_{1i}, Y_{0i}]}$$

with population variances and covariances $Var[Y_{di}]$, $Cov(Y_{1i}, Y_{0i})$.

Is this identifiable?

Standard Error for Sample ATE

In finite sample, sampling without replacement occurs; changes variance of group means, and induces (small) covariance

Standard Error for Sample ATE

Given complete randomization of N units with N_1 assigned to treatment and $N_0 = N - N_1$ to control, the true standard error of the estimated sample ATE is given by

$$SE_{\widehat{ATE}} = \sqrt{\left(\frac{N - N_1}{N - 1}\right)\frac{Var[Y_{1i}]}{N_1} + \left(\frac{N - N_0}{N - 1}\right)\frac{Var[Y_{0i}]}{N_0} + \left(\frac{1}{N - 1}\right)2Cov[Y_{1i}, Y_{0i}]}$$

with population variances and covariances $Var[Y_{di}]$, $Cov(Y_{1i}, Y_{0i})$.

Is this identifiable?

Standard error decreases if:

- N grows
- $Var[Y_1]$, $Var[Y_0]$ decrease
- $Cov[Y_1, Y_0]$ decreases

Proof: $SE_{\widehat{ATE}} \leq \hat{SE}_{\widehat{ATE}}$

Upper bound for standard error is when $Cor[Y_1, Y_0] = 1$:

$$Cor[Y_1, Y_0] = \frac{Cov[Y_1, Y_0]}{\sqrt{Var[Y_1]Var[Y_0]}} \leq 1 \iff Cov[Y_1, Y_0] \leq \sqrt{Var[Y_1]Var[Y_0]}$$

$$\begin{aligned}
SE_{\widehat{ATE}} &= \sqrt{\left(\frac{N-N_1}{N-1}\right)\frac{Var[Y_1]}{N_1} + \left(\frac{N-N_0}{N-1}\right)\frac{Var[Y_0]}{N_0} + \left(\frac{1}{N-1}\right)2Cov[Y_1, Y_0]} \\
&= \sqrt{\frac{1}{N-1}\left(\frac{N_0}{N_1}Var[Y_1] + \frac{N_1}{N_0}Var[Y_0] + 2Cov[Y_1, Y_0]\right)} \\
&\leq \sqrt{\frac{1}{N-1}\left(\frac{N_0}{N_1}Var[Y_1] + \frac{N_1}{N_0}Var[Y_0] + 2\sqrt{Var[Y_1]Var[Y_0]}\right)} \\
&\leq \sqrt{\frac{1}{N-1}\left(\frac{N_0}{N_1}Var[Y_1] + \frac{N_1}{N_0}Var[Y_0] + Var[Y_1] + Var[Y_0]\right)}
\end{aligned}$$

Last step follows from the following inequality

$$\begin{aligned}
(\sqrt{Var[Y_1]} - \sqrt{Var[Y_0]})^2 &\geq 0 \\
Var[Y_1] - 2\sqrt{Var[Y_1]Var[Y_0]} + Var[Y_0] &\geq 0 \iff Var[Y_1] + Var[Y_0] \geq 2\sqrt{Var[Y_1]Var[Y_0]}
\end{aligned}$$

Proof: $SE_{\widehat{ATE}} \leq \hat{SE}_{\widehat{ATE}}$

$$
\begin{aligned}
SE_{\widehat{ATE}} &\leq \sqrt{\frac{1}{N-1}\left(\frac{N_0}{N_1}Var[Y_1] + \frac{N_1}{N_0}Var[Y_0] + Var[Y_1] + Var[Y_0]\right)} \\
&\leq \sqrt{\frac{N_0^2 Var[Y_1] + N_1^2 Var[Y_0] + N_1 N_0 (Var[Y_1] + Var[Y_0])}{(N-1)N_1 N_0}} \\
&\leq \sqrt{\frac{(N_0^2 + N_1 N_0) Var[Y_1] + (N_1^2 + N_1 N_0) Var[Y_0]}{(N-1)N_1 N_0}} \\
&\leq \sqrt{\frac{(N_0 + N_1) N_0 Var[Y_1]}{(N-1)N_1 N_0} + \frac{(N_1 + N_0) N_1 Var[Y_0]}{(N-1)N_1 N_0}} \\
&\leq \sqrt{\frac{N\, Var[Y_1]}{(N-1)N_1} + \frac{N\, Var[Y_0]}{(N-1)N_0}} \\
&\leq \sqrt{\frac{N}{N-1}\left(\frac{Var[Y_1]}{N_1} + \frac{Var[Y_0]}{(N_0)}\right)} \\
&\leq \sqrt{\frac{N}{N-1}\left(\frac{\widehat{Var[Y_1]}}{N_1} + \frac{\widehat{Var[Y_0]}}{(N_0)}\right)}
\end{aligned}
$$

So the estimator for the standard error is conservative.

Seeing the Standard Error for Sample ATE

Want to know how ATE would differ under J randomizations of treatment, same sample: $SE_{\widehat{\theta}} \equiv \sqrt{\frac{1}{J}\sum_1^J(\widehat{\theta}_j - \bar{\widehat{\theta}})^2}$

i	Y_{1i}	Y_{0i}	Y_i	D_i	$P(D_i = 1)$
1	3	0	3	1	2/4
2	1	1	1	1	2/4
3	2	0	0	0	2/4
4	2	1	1	0	2/4

Seeing the Standard Error for Sample ATE

Want to know how ATE would differ under J randomizations of treatment, same sample: $SE_{\widehat{\theta}} \equiv \sqrt{\frac{1}{J}\sum_1^J(\widehat{\theta}_j - \bar{\widehat{\theta}})^2}$

i	Y_{1i}	Y_{0i}	Y_i	D_i	$P(D_i = 1)$
1	3	0	3	1	2/4
2	1	1	1	1	2/4
3	2	0	0	0	2/4
4	2	1	1	0	2/4

Seeing the Standard Error for Sample ATE

Want to know how ATE would differ under J randomizations of treatment, same sample: $SE_{\widehat{\theta}} \equiv \sqrt{\frac{1}{J}\sum_1^J(\widehat{\theta}_j - \bar{\widehat{\theta}})^2}$

i	Y_{1i}	Y_{0i}	Y_i	D_i	$P(D_i = 1)$
1	3	0	3	1	2/4
2	1	1	1	1	2/4
3	2	0	0	0	2/4
4	2	1	1	0	2/4

ATE estimates given all possible random assignments with two treated units:

Treated Units:	1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
$\widehat{ATE}$:	1.5	1.5	2	1	1.5	1.5

Seeing the Standard Error for Sample ATE

Want to know how ATE would differ under J randomizations of treatment, same sample: $SE_{\widehat{\theta}} \equiv \sqrt{\frac{1}{J}\sum_1^J(\widehat{\theta}_j - \bar{\widehat{\theta}})^2}$

i	Y_{1i}	Y_{0i}	Y_i	D_i	$P(D_i = 1)$
1	3	0	3	1	2/4
2	1	1	1	1	2/4
3	2	0	0	0	2/4
4	2	1	1	0	2/4

ATE estimates given all possible random assignments with two treated units:

Treated Units:	1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
$\widehat{ATE}$:	1.5	1.5	2	1	1.5	1.5

Average $\widehat{ATE}$ is 1.5 and true standard error is $SE_{\widehat{ATE}} =$

$$\sqrt{\tfrac{1}{6}[(1.5-1.5)^2 + (1.5-1.5)^2 + (2-1.5)^2 + (1-1.5)^2 + (1.5-1.5)^2 + (1.5-1.5)^2]} \approx .28$$

Standard Error for Sample ATE

To convince yourself the "true" formula is correct, try it using the (partial unobservable) quantities in the prior table. You would get:

$$
\begin{aligned}
SE_{\widehat{ATE}} &= \sqrt{\left(\frac{N-N_1}{N-1}\right)\frac{Var[Y_{1i}]}{N_1} + \left(\frac{N-N_0}{N-1}\right)\frac{Var[Y_{0i}]}{N_0} + \left(\frac{1}{N-1}\right)2Cov[Y_{1i}, Y_{0i}]} \\
&= \sqrt{\left(\frac{4-2}{4-1}\right)\frac{.25}{2} + \left(\frac{4-2}{4-1}\right)\frac{.5}{2} + \left(\frac{1}{4-1}\right)2(-.25)} \\
&\approx .28
\end{aligned}
$$

Standard Error for Sample ATE

To convince yourself the "true" formula is correct, try it using the (partial unobservable) quantities in the prior table. You would get:

$$
\begin{aligned}
SE_{\widehat{ATE}} &= \sqrt{\left(\frac{N-N_1}{N-1}\right)\frac{Var[Y_{1i}]}{N_1} + \left(\frac{N-N_0}{N-1}\right)\frac{Var[Y_{0i}]}{N_0} + \left(\frac{1}{N-1}\right)2Cov[Y_{1i}, Y_{0i}]} \\
&= \sqrt{\left(\frac{4-2}{4-1}\right)\frac{.25}{2} + \left(\frac{4-2}{4-1}\right)\frac{.5}{2} + \left(\frac{1}{4-1}\right)2(-.25)} \\
&\approx .28
\end{aligned}
$$

But of course both ways we got to this number involved unobservables.

Standard Error for Sample ATE

How does this compare to what we'll actually compute, on average?

i	Y_{1i}	Y_{0i}	Y_i
1	3	0	3
2	1	1	1
3	2	0	0
4	2	1	1

$\widehat{SE}_{\widehat{ATE}}$ estimates given all possible assignments with two treated units:

Treated Units:	1 & 2	1 & 3	1 & 4	2 & 3	2 & 4	3 & 4
$\widehat{ATE}$:	1.5	1.5	2	1	1.5	1.5
$\widehat{SE}_{\widehat{ATE}}$:	1.11	.5	.71	.71	.5	.5

The average $\widehat{SE}_{\widehat{ATE}}$ is $\approx .67$ compared to true $SE_{\widehat{ATE}} \approx .28$

Outline

1 Identification

2 Hypothesis Testing

3 Randomization Inference

4 Threats to Validity

5 Reviewing What We've Covered So Far

Example: Effect of Training on Earnings

- Treatment Group:
  - $N_1 = 7,487$
  - Estimated Average Earnings $\bar{Y}_1$: $\$16,199$
  - Estimated Sample Standard deviation $\widehat{\sigma}_{Y|D_i=1}$: $\$17,038$
- Control Group :
  - $N_0 = 3,717$
  - Estimated Average Earnings $\bar{Y}_0$: $\$15,040$
  - Estimated Sample deviation $\widehat{\sigma}_{Y|D_i=0}$: $\$16,180$
- Estimated average effect of training:

Example: Effect of Training on Earnings

- Treatment Group:
  - $N_1 = 7,487$
  - Estimated Average Earnings $\bar{Y}_1$: $\$16,199$
  - Estimated Sample Standard deviation $\widehat{\sigma}_{Y|D_i=1}$: $\$17,038$
- Control Group :
  - $N_0 = 3,717$
  - Estimated Average Earnings $\bar{Y}_0$: $\$15,040$
  - Estimated Sample deviation $\widehat{\sigma}_{Y|D_i=0}$: $\$16,180$
- Estimated average effect of training:
  - $\widehat{\tau}_{ATE} = \bar{Y}_1 - \bar{Y}_0 = 16,199 - 15,040 = \$1,159$
- Estimated standard error for effect of training:

Example: Effect of Training on Earnings

- Treatment Group:
  - $N_1 = 7,487$
  - Estimated Average Earnings $\bar{Y}_1$: \$16,199
  - Estimated Sample Standard deviation $\widehat{\sigma}_{Y|D_i=1}$: \$17,038
- Control Group :
  - $N_0 = 3,717$
  - Estimated Average Earnings $\bar{Y}_0$: \$15,040
  - Estimated Sample deviation $\widehat{\sigma}_{Y|D_i=0}$: \$16,180
- Estimated average effect of training:
  - $\widehat{\tau}_{ATE} = \bar{Y}_1 - \bar{Y}_0 = 16,199 - 15,040 = \$1,159$
- Estimated standard error for effect of training:

$$\widehat{SE}_{\widehat{ATE}} = \sqrt{\frac{\widehat{\sigma}^2_{Y|D_i=1}}{N_1} + \frac{\widehat{\sigma}^2_{Y|D_i=0}}{(N_0)}} = \sqrt{\frac{17,038^2}{7,487} + \frac{16,180^2}{3,717}} \approx \$330$$

- Is this consistent with a zero average treatment effect $\alpha_{ATE} = 0$?

Testing the Null Hypothesis of Zero Average Effect

Null hypothesis H_0: $\tau_{ATE} = 0$, the average potential outcomes in the population are the same for treatment and control: $\mathbb{E}[Y_1] = \mathbb{E}[Y_0]$.

- However, we observe a difference in mean earnings of $\hat{\tau}_{ATE} = 1,159$
- What is the probability of observing a difference this large if the true average effect of the training were zero (i.e. the null hypothesis were true)?

Testing the Null Hypothesis of Zero Average Effect

- Use a two-sample t-test with unequal variances:

$$t = \frac{\widehat{\tau}}{\sqrt{\frac{\widehat{\sigma}^2_{Y_i|D_i=1}}{N_1} + \frac{\widehat{\sigma}^2_{Y_i|D_i=0}}{N_0}}} = \frac{\$1,159}{\sqrt{\frac{\$17,038^2}{7,487} + \frac{\$16,180^2}{3,717}}} \approx 3.5$$

  - We know that $t_N \xrightarrow{d} \mathcal{N}(0,1)$
  - And for a standard normal distribution, the probability of observing a value of t that is larger than $|t| > 1.96$ is $< .05$
  - So obtaining a value as high as $t = 3.5$ is very unlikely under the null hypothesis of a zero average effect
  - We reject the null hypothesis H_0: $\tau_0 = 0$ against the alternative H_1: $\tau_0 \neq 0$ at asymptotic 5% significance level whenever $|t| > 1.96$.
  - Inverting the test statistic we can construct a 95% confidence interval

$$\widehat{\tau}_{ATE} \pm 1.96 \cdot \widehat{SE}_{\widehat{ATE}}$$

  - What assumptions did we need along the way?

Testing the Null Hypothesis of Zero Average Effect

```
> d <- read.dta("jtpa.dta")
> t.test(earnings~assignmt,data=d,var.equal=FALSE)

        Welch Two Sample t-test

data:  earnings by assignmt
t = -3.5084, df = 7765.599, p-value = 0.0004533
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1807.2427  -511.6239
sample estimates:
mean in group 0 mean in group 1
       15040.50        16199.94
```

Regression to Estimate the Average Treatment Effect

Estimator (Regression)

The ATE can be expressed as a regression equation:

$$
\begin{aligned}
Y_i &= D_i\, Y_{1i} + (1 - D_i)\, Y_{0i} \\
&= Y_{0i} + (Y_{1i} - Y_{0i})\, D_i \\
&= \underbrace{\bar{Y}_0}_{\alpha} + \underbrace{(\bar{Y}_1 - \bar{Y}_0)}_{\tau_{Reg}} D_i + \underbrace{\{(Y_{i0} - \bar{Y}_0) + D_i \cdot [(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)]\}}_{\varepsilon} \\
&= \alpha + \tau_{Reg} D_i + \varepsilon_i
\end{aligned}
$$

Regression to Estimate the Average Treatment Effect

Estimator (Regression)

The ATE can be expressed as a regression equation:

$$
\begin{aligned}
Y_i &= D_i\, Y_{1i} + (1 - D_i)\, Y_{0i} \\
&= Y_{0i} + (Y_{1i} - Y_{0i})\, D_i \\
&= \underbrace{\bar{Y}_0}_{\alpha} + \underbrace{(\bar{Y}_1 - \bar{Y}_0)}_{\tau_{Reg}} D_i + \underbrace{\{(Y_{i0} - \bar{Y}_0) + D_i \cdot [(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)]\}}_{\varepsilon} \\
&= \alpha + \tau_{Reg} D_i + \varepsilon_i
\end{aligned}
$$

- Does this assume constant treatment effects?

Regression to Estimate the Average Treatment Effect

Estimator (Regression)

The ATE can be expressed as a regression equation:

$$
\begin{aligned}
Y_i &= D_i\, Y_{1i} + (1 - D_i)\, Y_{0i} \\
&= Y_{0i} + (Y_{1i} - Y_{0i})\, D_i \\
&= \underbrace{\bar{Y}_0}_{\alpha} + \underbrace{(\bar{Y}_1 - \bar{Y}_0)}_{\tau_{Reg}} D_i + \underbrace{\{(Y_{i0} - \bar{Y}_0) + D_i \cdot [(Y_{i1} - \bar{Y}_1) - (Y_{i0} - \bar{Y}_0)]\}}_{\varepsilon} \\
&= \alpha + \tau_{Reg} D_i + \varepsilon_i
\end{aligned}
$$

- Does this assume constant treatment effects?
- Our SE estimator allows different variance for $D = 1$ and $D = 0$.
  - Implies heteroskedasticity
  - Use "HC2" heteroskedasticity-robust variance:

$$\hat{\sigma}^2_{HC2} = \frac{S_1^2}{N_1} + \frac{S_0^2}{N_0} = \tilde{\mathbb{V}}(\tilde{\tau})$$

Regression to Estimate the Average Treatment Effect

```
> library(sandwich)
> library(lmtest)
>
> lout <- lm(earnings~assignmt,data=d)
> coeftest(lout,vcov = vcovHC(lout, type = "HC2"))

t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept) 15040.50     265.38 56.6752 < 2.2e-16 ***
assignmt     1159.43     330.46  3.5085 0.0004524 ***
--
```

Outline

1 Identification

2 Hypothesis Testing

3 Randomization Inference

4 Threats to Validity

5 Reviewing What We've Covered So Far

Testing in Small Samples: Fisher's Exact Test

- Test of differences in means with large N:

$$H_0 : \mathbb{E}[Y_1] = \mathbb{E}[Y_0], \quad H_1 : \mathbb{E}[Y_1] \neq \mathbb{E}[Y_0] \text{ (weak null)}$$

Testing in Small Samples: Fisher's Exact Test

- Test of differences in means with large N:

$$H_0 : \mathbb{E}[Y_1] = \mathbb{E}[Y_0], \quad H_1 : \mathbb{E}[Y_1] \neq \mathbb{E}[Y_0] \text{ (weak null)}$$

- Fisher's Exact Test:

$$H_0 : Y_{1i} = Y_{0i},$$

Testing in Small Samples: Fisher's Exact Test

- Test of differences in means with large N:

$$H_0 : \mathbb{E}[Y_1] = \mathbb{E}[Y_0], \quad H_1 : \mathbb{E}[Y_1] \neq \mathbb{E}[Y_0] \text{ (weak null)}$$

- Fisher's Exact Test:

$$H_0 : Y_{1i} = Y_{0i}, \quad H_1 : Y_1 \neq Y_0 \ \ \forall i, \qquad \text{(sharp null of no effect)}$$

Testing in Small Samples: Fisher's Exact Test

- Test of differences in means with large N:

$$H_0 : \mathbb{E}[Y_1] = \mathbb{E}[Y_0], \quad H_1 : \mathbb{E}[Y_1] \neq \mathbb{E}[Y_0] \text{ (weak null)}$$

- Fisher's Exact Test:

$$H_0 : Y_{1i} = Y_{0i}, \quad H_1 : Y_1 \neq Y_0 \ \ \forall i, \qquad \text{(sharp null of no effect)}$$

Key idea: Under the sharp null, we "observe" all potential outcomes – can compute what ATE would be under alternate randomizations

Testing in Small Samples: Fisher's Exact Test

- Test of differences in means with large N:

$$H_0 : \mathbb{E}[Y_1] = \mathbb{E}[Y_0], \quad H_1 : \mathbb{E}[Y_1] \neq \mathbb{E}[Y_0] \text{ (weak null)}$$

- Fisher's Exact Test:

$$H_0 : Y_{1i} = Y_{0i}, \quad H_1 : Y_1 \neq Y_0 \ \ \forall i, \qquad \text{(sharp null of no effect)}$$

Key idea: Under the sharp null, we "observe" all potential outcomes – can compute what ATE would be under alternate randomizations

Let Ω be the set of all possible ways to assign treatments.

Testing in Small Samples: Fisher's Exact Test

- Test of differences in means with large N:

$$H_0 : \mathbb{E}[Y_1] = \mathbb{E}[Y_0], \quad H_1 : \mathbb{E}[Y_1] \neq \mathbb{E}[Y_0] \text{ (weak null)}$$

- Fisher's Exact Test:

$$H_0 : Y_{1i} = Y_{0i}, \quad H_1 : Y_1 \neq Y_0 \ \ \forall i, \qquad \text{(sharp null of no effect)}$$

Key idea: Under the sharp null, we "observe" all potential outcomes – can compute what ATE would be under alternate randomizations

Let Ω be the set of all possible ways to assign treatments.

Fisher's exact test procedure:

1. Calculate a statistic $\hat{\theta}_{true}$ (e.g. difference in means) from original treatment assignment data
2. Obtain the null distribution of the statistic by calculating the same statistic $\hat{\theta}(\omega)$ under the sharp null for every possible (or many) ω in Ω (or many)
3. Compare $\hat{\theta}_{true}$ to the null distribution of $\hat{\theta}(\omega)$'s to see how "extreme" it is

Testing in Small Samples: Fisher's Exact Test

i	Y_{1i}	Y_{0i}	D_i
1	3	?	1
2	1	?	1
3	?	0	0
4	?	1	0
$\widehat{\tau}_{ATE}$			1.5

What do we know given the sharp null $H_0 : Y_1 = Y_0$?

Testing in Small Samples: Fisher's Exact Test

i	Y_{1i}	Y_{0i}	D_i
1	3	3	1
2	1	1	1
3	0	0	0
4	1	1	0
$\widehat{\tau}_{ATE}$			1.5
$\hat{\tau}(\omega)$			1.5

Given the full schedule of potential outcomes under the sharp null, we can compute the null distribution of ATE_{H_0} across all possible randomization.

Testing in Small Samples: Fisher's Exact Test

i	Y_{1i}	Y_{0i}	D_i	ω_1
1	3	3	1	1
2	1	1	1	0
3	0	0	0	1
4	1	1	0	0
$\widehat{\tau}_{ATE}$			1.5	
$\hat{\tau}(\omega)$			1.5	0.5

Testing in Small Samples: Fisher's Exact Test

i	Y_{1i}	Y_{0i}	D_i	ω_1	ω_2
1	3	3	1	1	1
2	1	1	1	0	0
3	0	0	0	1	0
4	1	1	0	0	1
$\widehat{\tau}_{ATE}$			1.5		
$\hat{\tau}(\omega)$			1.5	0.5	1.5

Testing in Small Samples: Fisher's Exact Test

i	Y_{1i}	Y_{0i}	D_i	ω_1	ω_2	ω_3
1	3	3	1	1	1	0
2	1	1	1	0	0	1
3	0	0	0	1	0	1
4	1	1	0	0	1	0
$\widehat{\tau}_{ATE}$			1.5			
$\hat{\tau}(\omega)$			1.5	0.5	1.5	-1.5

Testing in Small Samples: Fisher's Exact Test

i	Y_{1i}	Y_{0i}	D_i	ω_1	ω_2	ω_3	ω_4
1	3	3	1	1	1	0	0
2	1	1	1	0	0	1	1
3	0	0	0	1	0	1	0
4	1	1	0	0	1	0	1
$\widehat{\tau}_{ATE}$			1.5				
$\hat{\tau}(\omega)$			1.5	0.5	1.5	-1.5	-.5

Testing in Small Samples: Fisher's Exact Test

i	Y_{1i}	Y_{0i}	D_i	ω_1	ω_2	ω_3	ω_4	ω_5
1	3	3	1	1	1	0	0	0
2	1	1	1	0	0	1	1	0
3	0	0	0	1	0	1	0	1
4	1	1	0	0	1	0	1	1
$\widehat{\tau}_{ATE}$			1.5					
$\hat{\tau}(\omega)$			1.5	0.5	1.5	-1.5	-.5	-1.5

So $\Pr(\hat{\tau}(\omega) \geq \widehat{\tau}_{ATE}) = 2/6 \approx .33$.

Which assumptions are needed?

Testing in Small Samples: Fisher's Exact Test

i	Y_{1i}	Y_{0i}	D_i	ω_1	ω_2	ω_3	ω_4	ω_5
1	3	3	1	1	1	0	0	0
2	1	1	1	0	0	1	1	0
3	0	0	0	1	0	1	0	1
4	1	1	0	0	1	0	1	1
$\widehat{\tau}_{ATE}$			1.5					
$\hat{\tau}(\omega)$			1.5	0.5	1.5	-1.5	-.5	-1.5

So $\Pr(\hat{\tau}(\omega) \geq \widehat{\tau}_{ATE}) = 2/6 \approx .33$.

Which assumptions are needed? None! Randomization as "reasoned basis for causal inference" (Fisher 1935)

Outline

1 Identification

2 Hypothesis Testing

3 Randomization Inference

4 Threats to Validity

5 Reviewing What We've Covered So Far

Threats to Internal and External Validity

- Internal validity: can we estimate the treatment effect for our particular sample?
  - Fails when there are differences between treated and controls (other than the treatment itself) that affect the outcome
  - In other words, how sure is the randomization?

Threats to Internal and External Validity

- Internal validity: can we estimate the treatment effect for our particular sample?
  - Fails when there are differences between treated and controls (other than the treatment itself) that affect the outcome
  - In other words, how sure is the randomization?
- External validity: can we extrapolate our estimates to other populations?
  - Might Y_{1i} and Y_{0i} have looked different in a different part of the population?

Most Common Threats to Internal Validity

- Failure of randomization
  - e.g. implementing partners assign their favorites to treatment group; imbalance due to small sample size

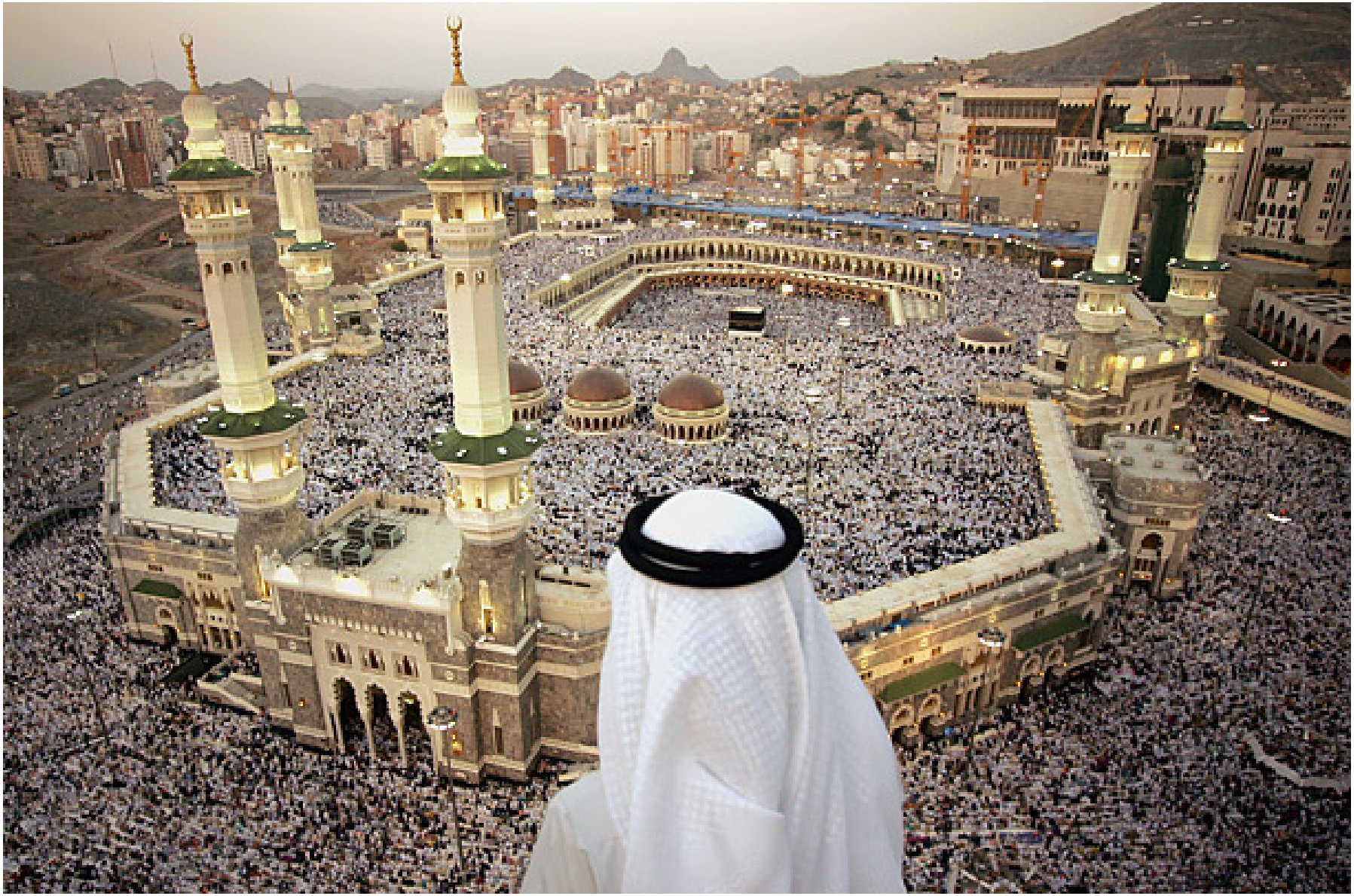
Most Common Threats to Internal Validity

- Failure of randomization
  - e.g. implementing partners assign their favorites to treatment group; imbalance due to small sample size
- Noncompliance with experimental protocol
  - e.g. failure to treat or “crossover”. Some members of the control group receive the treatment and some members of the treatment group go untreated
  - e.g. in JTPA: only 62% of women and 66% of men assigned for treatment actually enrolled in JTPA training (on the other side, compliance was almost perfect in the control group).

Most Common Threats to Internal Validity

- Failure of randomization
  - e.g. implementing partners assign their favorites to treatment group; imbalance due to small sample size
- Noncompliance with experimental protocol
  - e.g. failure to treat or “crossover”. Some members of the control group receive the treatment and some members of the treatment group go untreated
  - e.g. in JTPA: only 62% of women and 66% of men assigned for treatment actually enrolled in JTPA training (on the other side, compliance was almost perfect in the control group).
- Differential attrition
  - if probability of attrition depends on treatment status and potential outcomes, you break the randomization.
  - Two scenarios: (i) people attrite from treatment but you still get to measure outcome; (ii) people attrite and you can’t measure anything.

Example: Clingingsmith, Khwaja, Kremer



Example: Natural Experiment in Pakistan

- Pakistan allocated about 135,000 visas to Saudi Arabia for the Hajj via a randomized lottery.

Example: Natural Experiment in Pakistan

- Pakistan allocated about 135,000 visas to Saudi Arabia for the Hajj via a randomized lottery.
- Wealthier Pakistanis tend to use private Hajj tour operators rather than the lottery.

Example: Natural Experiment in Pakistan

- Pakistan allocated about 135,000 visas to Saudi Arabia for the Hajj via a randomized lottery.
- Wealthier Pakistanis tend to use private Hajj tour operators rather than the lottery.
- Randomization occurs among individuals grouped into “parties”, where parties are stratified by sect, region, and accommodation.

Example: Natural Experiment in Pakistan

- Pakistan allocated about 135,000 visas to Saudi Arabia for the Hajj via a randomized lottery.
- Wealthier Pakistanis tend to use private Hajj tour operators rather than the lottery.
- Randomization occurs among individuals grouped into “parties”, where parties are stratified by sect, region, and accommodation.
- Compliance with the experiment is imperfect:
  - 99% who win lottery attend the Hajj.
  - 11% who lose lottery still attend the Hajj (via private tours).

Example: Natural Experiment in Pakistan

- Pakistan allocated about 135,000 visas to Saudi Arabia for the Hajj via a randomized lottery.
- Wealthier Pakistanis tend to use private Hajj tour operators rather than the lottery.
- Randomization occurs among individuals grouped into “parties”, where parties are stratified by sect, region, and accommodation.
- Compliance with the experiment is imperfect:
  - 99% who win lottery attend the Hajj.
  - 11% who lose lottery still attend the Hajj (via private tours).
- Since randomization is not controlled by researcher, balance checks and qualitative checks are crucial!

Example: Natural Experiment in Pakistan

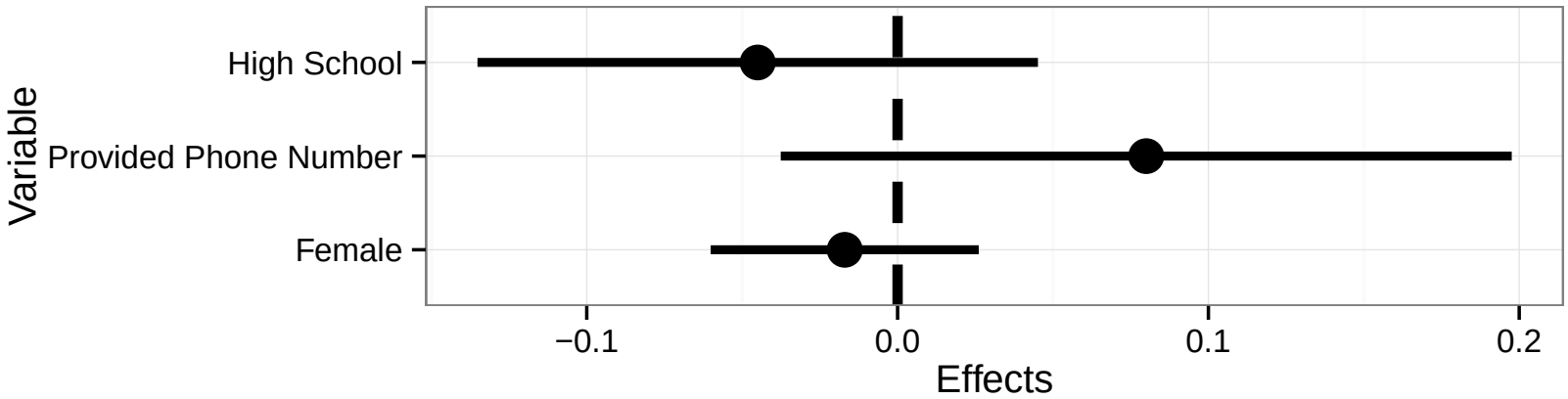
Two pieces of information to bolster the randomization assumption:

- Qualitative information:
  - The lottery selection algorithm was designed and implemented by an independent and reputable third party, and there were no reports of lottery manipulation.

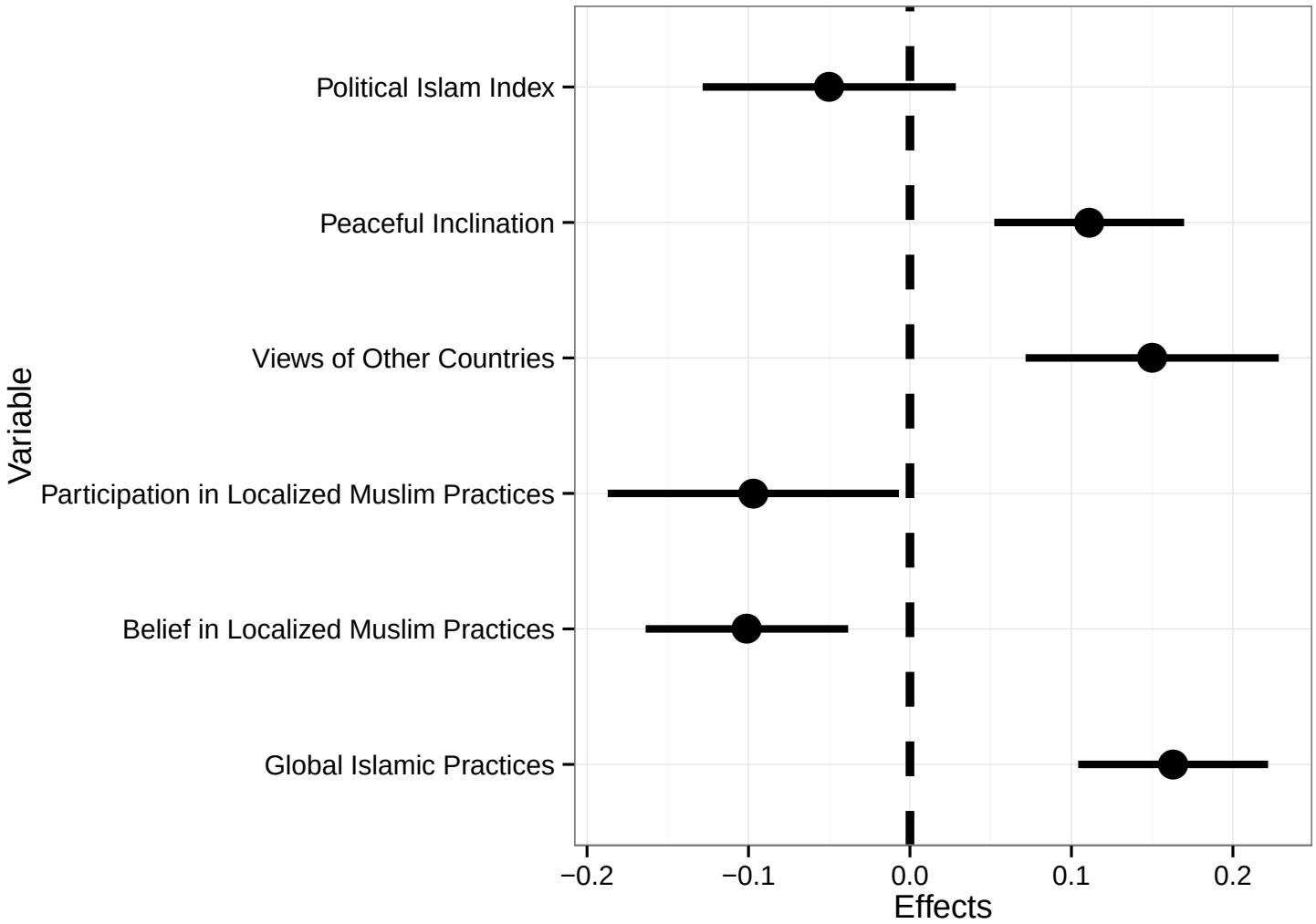
Example: Natural Experiment in Pakistan

Two pieces of information to bolster the randomization assumption:

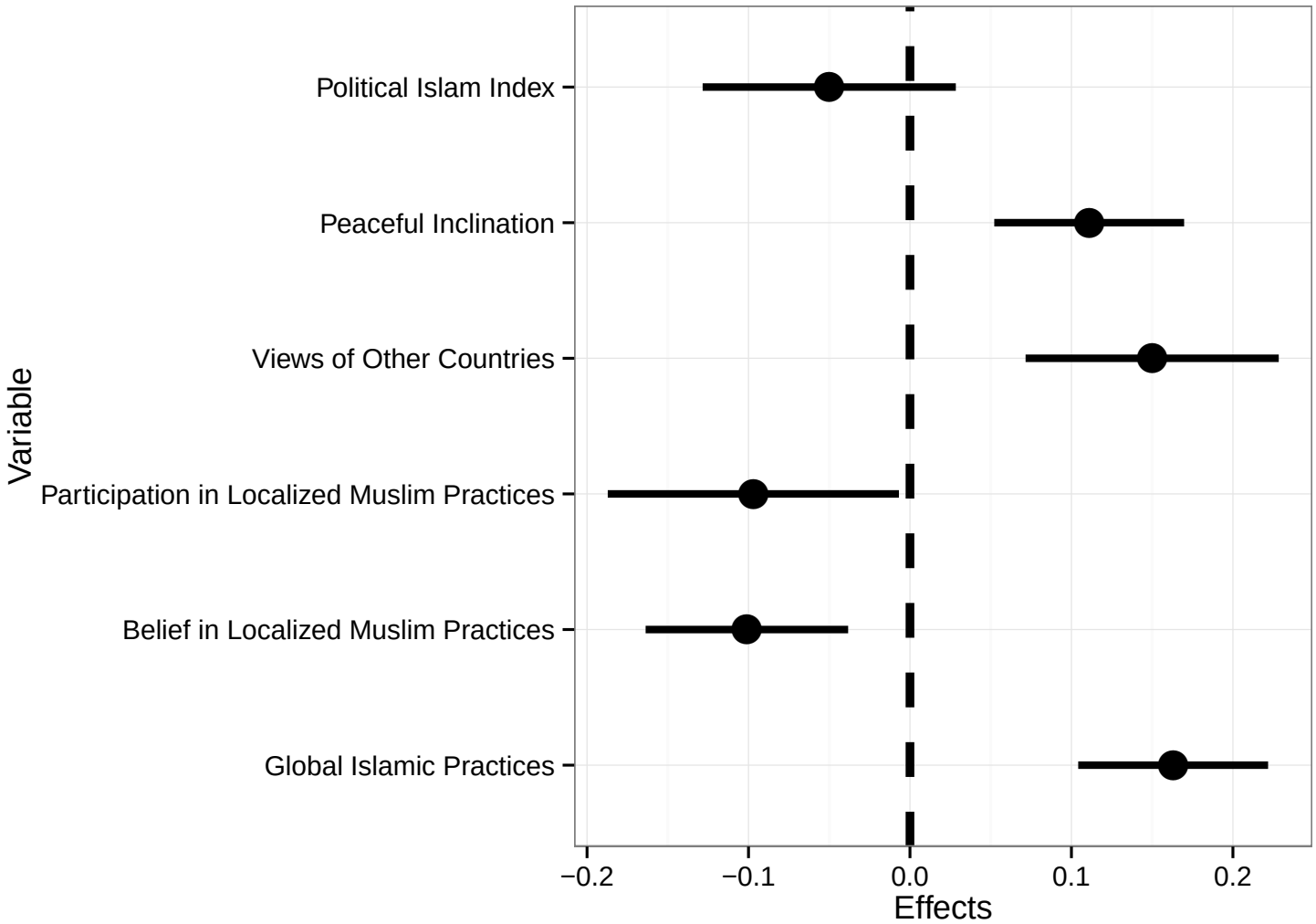
- Qualitative information:
  - The lottery selection algorithm was designed and implemented by an independent and reputable third party, and there were no reports of lottery manipulation.
- Balance tests:



Effects of winning the Hajj lottery



Effects of winning the Hajj lottery



Would you say these are the “effects of going on the Hajj”?

Most Common Threats to External Validity

- Non-representative sample
  - e.g. laboratory experiment using a convenience sample
  - units are randomly assigned, but not from the pop of interest
- Non-representative treatment
  - the treatment differs in actual implementations
  - e.g. survey experiment about the effect of media priming on voting
  - scale effects
  - actual treatments may be bundled

Internal vs. External Validity

Which one is more important?

Internal vs. External Validity

Which one is more important?

"One common view is that internal validity comes first. If you do not know the effects of the treatment on the units in your study, you are not well-positioned to infer the effects on units you did not study who live in circumstances you did not study." (Rosenbaum 2010, p. 56)

- Randomization ensures internal validity
- External validity may be partially addressed by comparing the results of several internally valid studies conducted in different circumstances and at different times
- Note that the same external validity issues often apply in observation studies

Internal vs. External Validity

Which one is more important?

"One common view is that internal validity comes first. If you do not know the effects of the treatment on the units in your study, you are not well-positioned to infer the effects on units you did not study who live in circumstances you did not study." (Rosenbaum 2010, p. 56)

- Randomization ensures internal validity
- External validity may be partially addressed by comparing the results of several internally valid studies conducted in different circumstances and at different times
- Note that the same external validity issues often apply in observation studies

That said, it is a balancing act

Internal vs. External Validity

Which one is more important?

"One common view is that internal validity comes first. If you do not know the effects of the treatment on the units in your study, you are not well-positioned to infer the effects on units you did not study who live in circumstances you did not study." (Rosenbaum 2010, p. 56)

- Randomization ensures internal validity
- External validity may be partially addressed by comparing the results of several internally valid studies conducted in different circumstances and at different times
- Note that the same external validity issues often apply in observation studies

That said, it is a balancing act

Often a series of studies with lower internal validity gives us the inspiration and qualitative knowledge needed for an experiment

Outline

1 Identification

2 Hypothesis Testing

3 Randomization Inference

4 Threats to Validity

5 Reviewing What We've Covered So Far

Review

- Random assignment solves the identification problem for causal inference based on minimal assumptions that the researcher can control

Review

- Random assignment solves the identification problem for causal inference based on minimal assumptions that the researcher can control
- Put differently, random assignment "balances observed and unobserved confounders", or "makes treated and control units comparable".

Review

- Random assignment solves the identification problem for causal inference based on minimal assumptions that the researcher can control
- Put differently, random assignment "balances observed and unobserved confounders", or "makes treated and control units comparable".
- Regression *can* be used for analyzing experiments: simple regression with robust SE yields unbiased estimates with conservative confidence intervals. Will be more useful when we add covariates.

Review

- Random assignment solves the identification problem for causal inference based on minimal assumptions that the researcher can control
- Put differently, random assignment "balances observed and unobserved confounders", or "makes treated and control units comparable".
- Regression *can* be used for analyzing experiments: simple regression with robust SE yields unbiased estimates with conservative confidence intervals. Will be more useful when we add covariates.
- Possible tradeoff between internal validity and external validity

Review

- Random assignment solves the identification problem for causal inference based on minimal assumptions that the researcher can control
- Put differently, random assignment "balances observed and unobserved confounders", or "makes treated and control units comparable".
- Regression *can* be used for analyzing experiments: simple regression with robust SE yields unbiased estimates with conservative confidence intervals. Will be more useful when we add covariates.
- Possible tradeoff between internal validity and external validity
- What's Next: Advanced Experimental Design and Analysis
  - covariate adjustment
  - cluster randomization
  - block randomization

Some Important Topics in Experimental Design and Analysis

A brief introduction to some common topics:

Some Important Topics in Experimental Design and Analysis

A brief introduction to some common topics:

- Covariate adjustment

Some Important Topics in Experimental Design and Analysis

A brief introduction to some common topics:

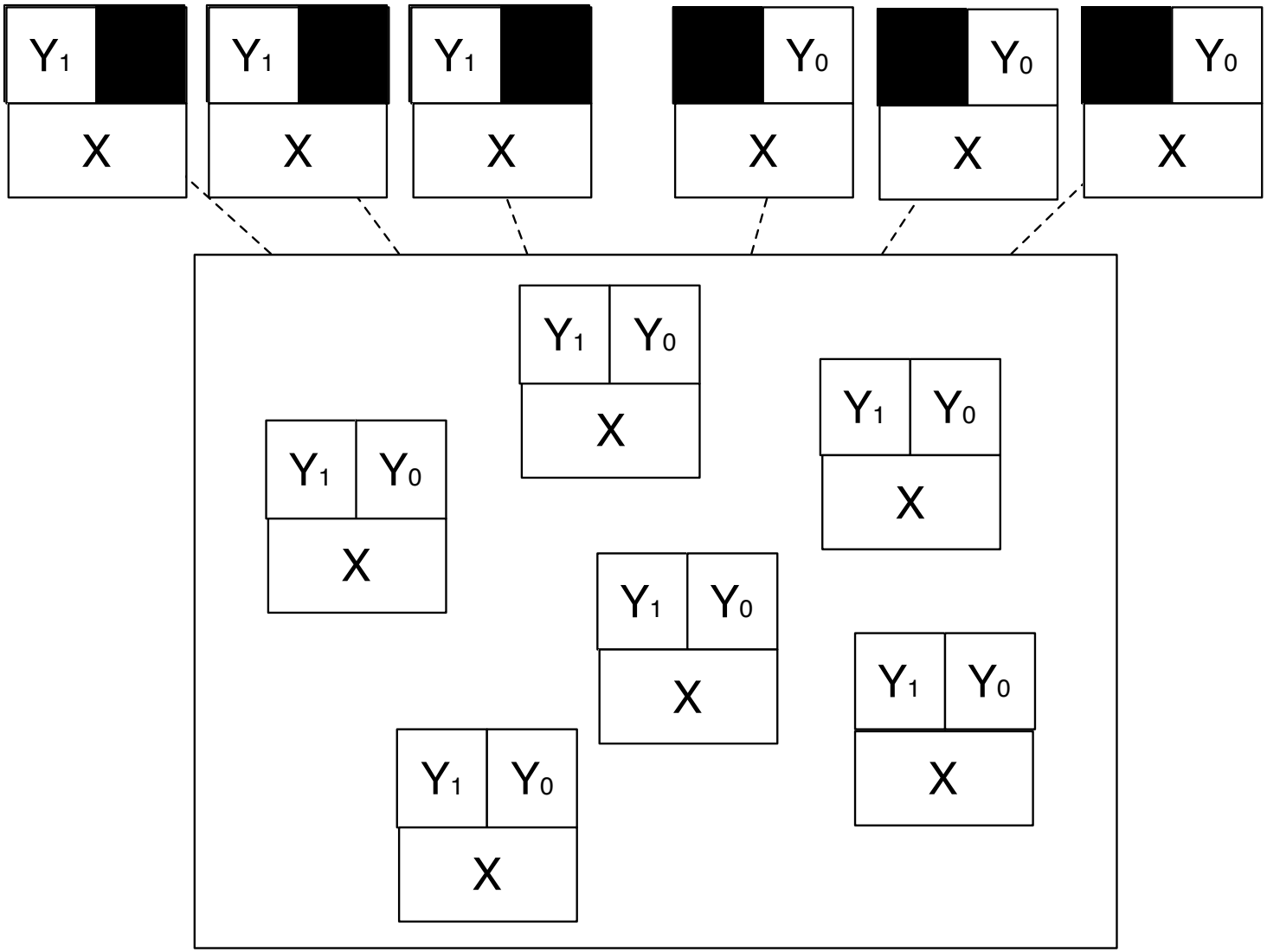
- Covariate adjustment
- Blocking

Some Important Topics in Experimental Design and Analysis

A brief introduction to some common topics:

- Covariate adjustment
- Blocking
- Ethical Considerations

Covariates and Experiments



Covariates

- Randomization is gold standard for causal inference because in expectation it balances observed but also unobserved characteristics between treatment and control group.

Covariates

- Randomization is gold standard for causal inference because in expectation it balances observed but also unobserved characteristics between treatment and control group.
- Unlike potential outcomes, you observe baseline covariates for all units. Covariate values are predetermined with respect to the treatment and do not depend on D_i.

Covariates

- Randomization is gold standard for causal inference because in expectation it balances observed but also unobserved characteristics between treatment and control group.
- Unlike potential outcomes, you observe baseline covariates for all units. Covariate values are predetermined with respect to the treatment and do not depend on D_i.
- Randomization implies *covariate balance*,

$$p(X|D = 1) = p(X|D = 0)$$

Covariates

- Randomization is gold standard for causal inference because in expectation it balances observed but also unobserved characteristics between treatment and control group.
- Unlike potential outcomes, you observe baseline covariates for all units. Covariate values are predetermined with respect to the treatment and do not depend on D_i.
- Randomization implies *covariate balance*,

$$p(X|D = 1) = p(X|D = 0)$$

- If this is not the case, then one of two possibilities:
  - randomization was compromised.
  - unlucky sampling

Covariates

- Randomization is gold standard for causal inference because in expectation it balances observed but also unobserved characteristics between treatment and control group.
- Unlike potential outcomes, you observe baseline covariates for all units. Covariate values are predetermined with respect to the treatment and do not depend on D_i.
- Randomization implies *covariate balance*,

$$p(X|D = 1) = p(X|D = 0)$$

- If this is not the case, then one of two possibilities:
  - randomization was compromised.
  - unlucky sampling
- Good idea to test for covariate balance on important covariates, using "balance tests" (eg. t-tests, F-tests, etc.)

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

- But why include X_i when experiments "control" for covariates by design?

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

- But why include X_i when experiments "control" for covariates by design?
  - Correct chance imbalances that may push $\hat{\tau}$ far from τ_{ATE}.

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

- But why include X_i when experiments "control" for covariates by design?
  - Correct chance imbalances that may push $\hat{\tau}$ far from τ_{ATE}.
  - Increase precision: How does this happen?

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

- But why include X_i when experiments "control" for covariates by design?
  - Correct chance imbalances that may push $\hat{\tau}$ far from τ_{ATE}.
  - Increase precision: How does this happen?
- ATE estimates are robust to model specification (with sufficient N).

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

- But why include X_i when experiments "control" for covariates by design?
  - Correct chance imbalances that may push $\hat{\tau}$ far from τ_{ATE}.
  - Increase precision: How does this happen?
- ATE estimates are robust to model specification (with sufficient N).
- But, do not control for **post-treatment** covariates!

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

- But why include X_i when experiments "control" for covariates by design?
  - Correct chance imbalances that may push $\hat{\tau}$ far from τ_{ATE}.
  - Increase precision: How does this happen?
- ATE estimates are robust to model specification (with sufficient N).
- But, do not control for **post-treatment** covariates!
  One intuition: consider post-treatment variable Z

$$\begin{aligned}\tau(z) &= \mathbb{E}[Y_i|D_i = 1, Z_i = z] - \mathbb{E}[Y_i|D_i = 0, Z_i = z] \\ &= \mathbb{E}[Y_{1i}|D_i = 1, Z_{1i} = z] - \mathbb{E}[Y_{0i}|D_i = 0, Z_{0i} = z]\end{aligned}$$

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

- But why include X_i when experiments "control" for covariates by design?
  - Correct chance imbalances that may push $\hat{\tau}$ far from τ_{ATE}.
  - Increase precision: How does this happen?

- ATE estimates are robust to model specification (with sufficient N).
- But, do not control for **post-treatment** covariates!
  One intuition: consider post-treatment variable Z

$$\begin{aligned}\tau(z) &= \mathbb{E}[Y_i|D_i = 1, Z_i = z] - \mathbb{E}[Y_i|D_i = 0, Z_i = z] \\ &= \mathbb{E}[Y_{1i}|D_i = 1, Z_{1i} = z] - \mathbb{E}[Y_{0i}|D_i = 0, Z_{0i} = z]\end{aligned}$$

  - there is good chance $Z_{1i} - Z_{0i}$ correlated with $Y_{1i} - Y_{0i}$...

Regression with Covariates

- A common model with experimental data:

$$Y_i = \alpha + \tau D_i + X_i\beta + \varepsilon_i$$

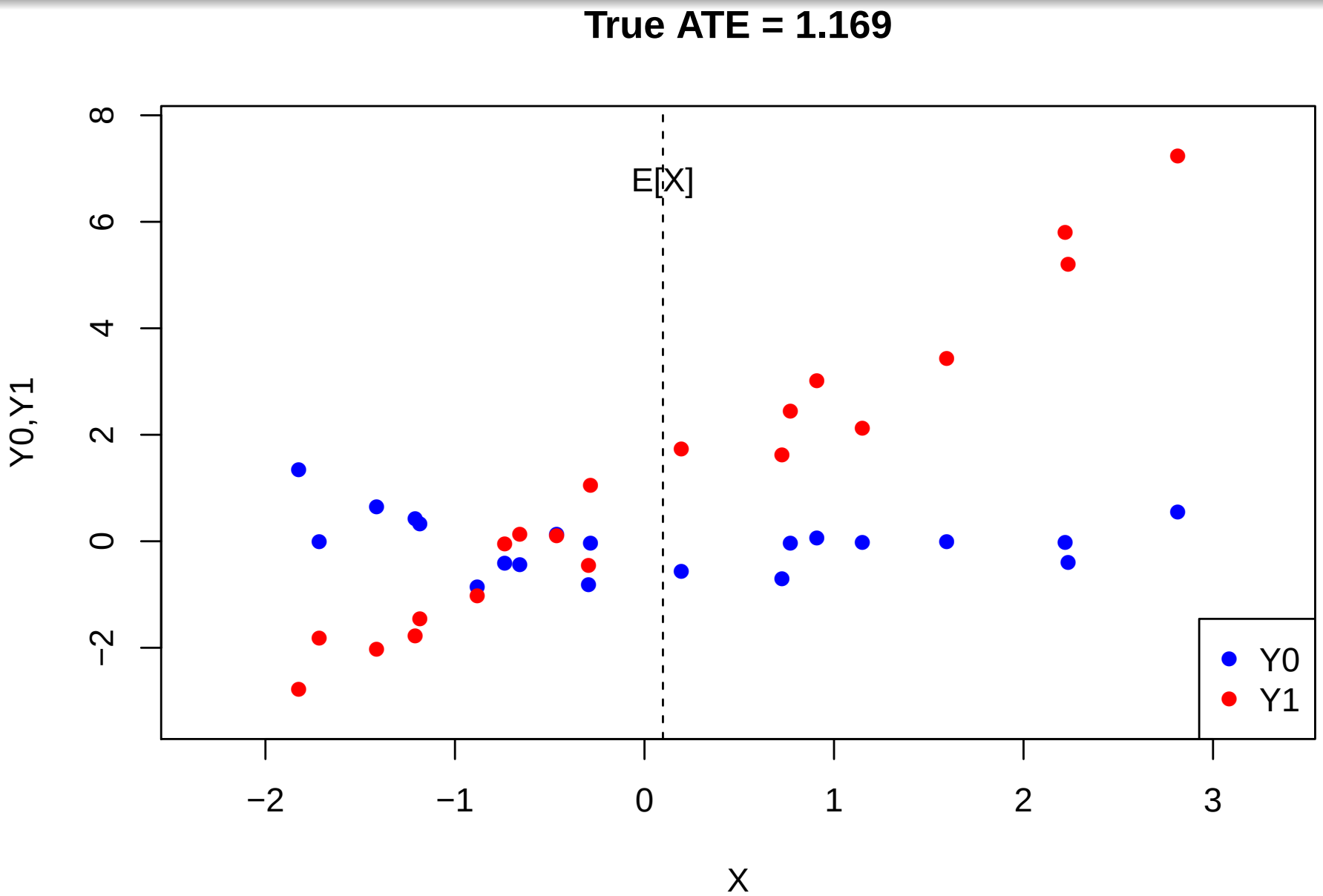
- But why include X_i when experiments "control" for covariates by design?
  - Correct chance imbalances that may push $\hat{\tau}$ far from τ_{ATE}.
  - Increase precision: How does this happen?

- ATE estimates are robust to model specification (with sufficient N).
- But, do not control for **post-treatment** covariates!
  One intuition: consider post-treatment variable Z

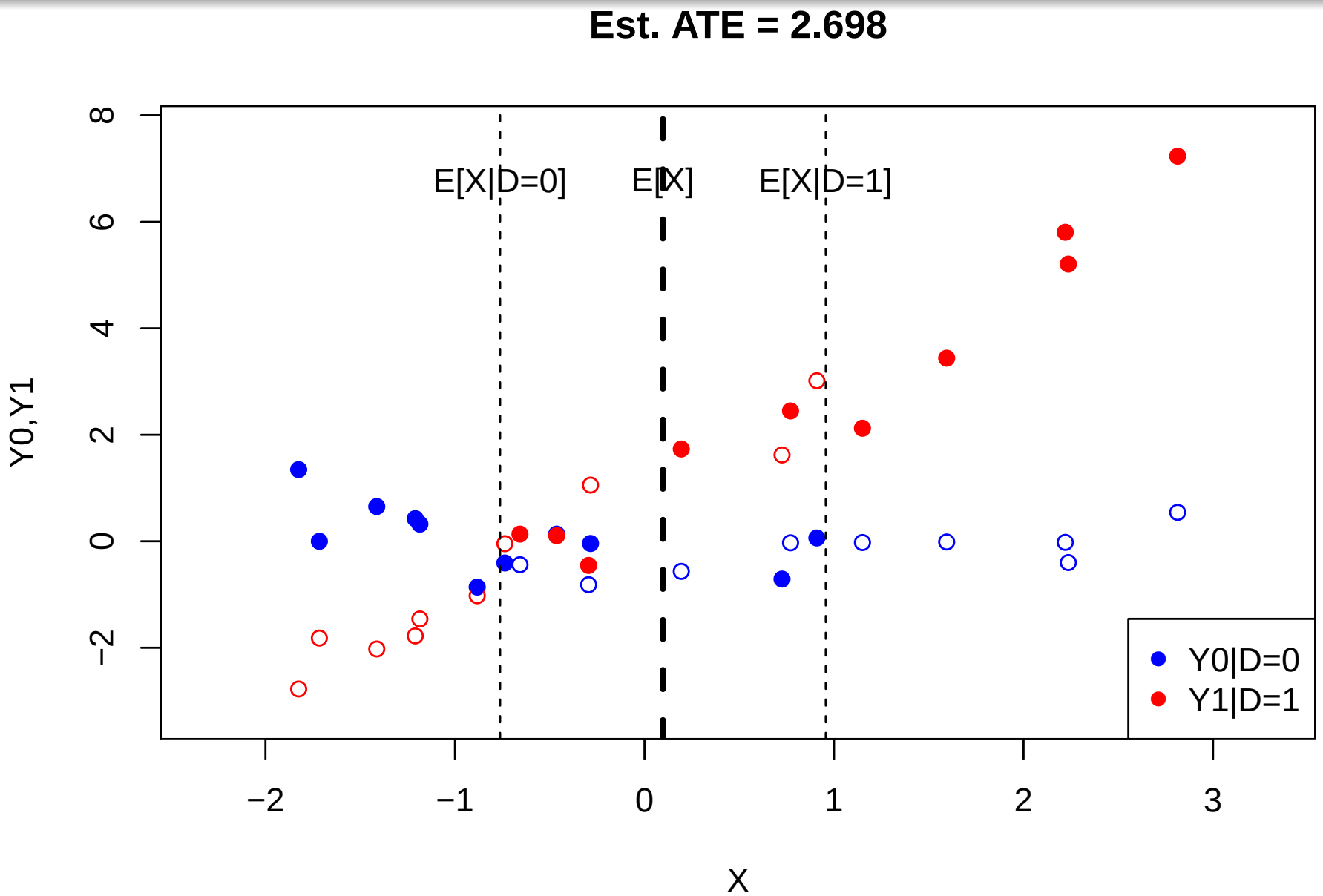
$$\begin{aligned}\tau(z) &= \mathbb{E}[Y_i|D_i = 1, Z_i = z] - \mathbb{E}[Y_i|D_i = 0, Z_i = z] \\ &= \mathbb{E}[Y_{1i}|D_i = 1, Z_{1i} = z] - \mathbb{E}[Y_{0i}|D_i = 0, Z_{0i} = z]\end{aligned}$$

  - there is good chance $Z_{1i} - Z_{0i}$ correlated with $Y_{1i} - Y_{0i}$...
  - so only comparing cases with $Z_{1i} \approx Z_{0i}$ makes problematic comparison!

Regression Adjustment For Experiments

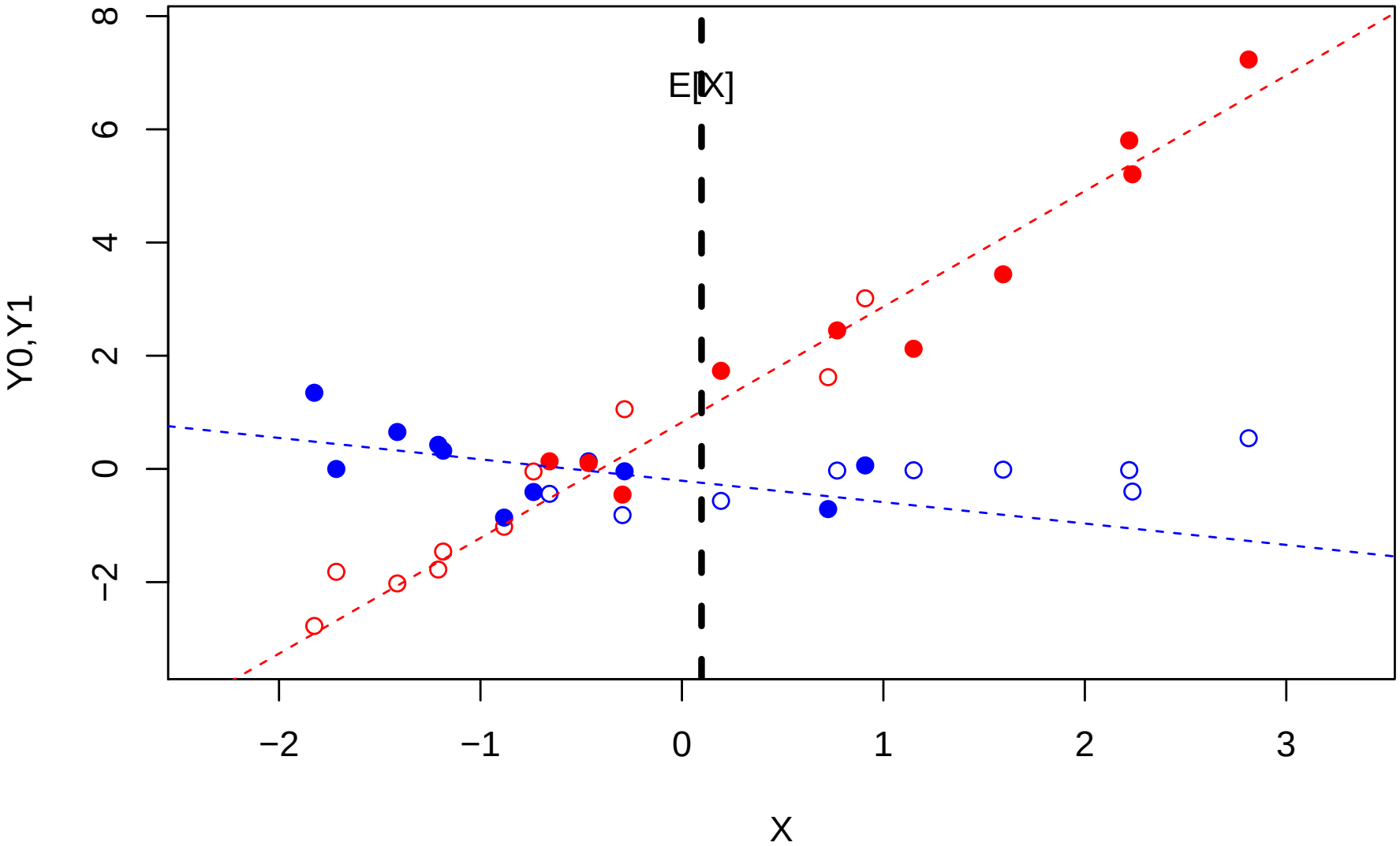


Regression Adjustment For Experiments

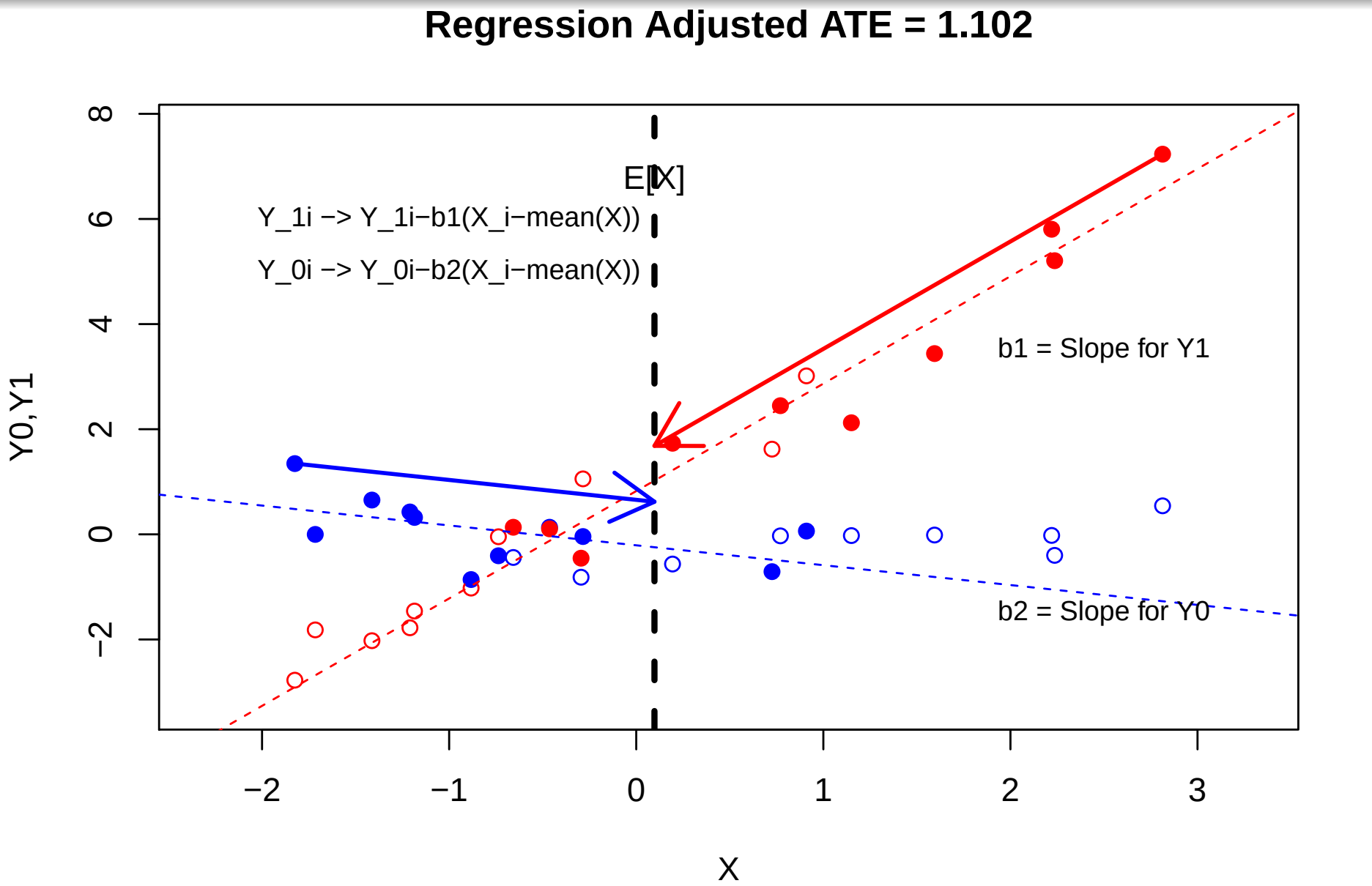


Regression Adjustment For Experiments

Consider regression lines (interacting with D)



Regression Adjustment For Experiments



Covariate Adjustment with Regression

Freedman (2008) shows that regression of the form:

$$Y_i = \alpha + \tau_{reg} D_i + \beta_1 X_i + \varepsilon_i$$

- $\hat{\tau}_{reg}$; small bias on order of $1/N$

Covariate Adjustment with Regression

Freedman (2008) shows that regression of the form:

$$Y_i = \alpha + \tau_{reg} D_i + \beta_1 X_i + \varepsilon_i$$

- $\hat{\tau}_{reg}$; small bias on order of $1/N$
- $\hat{\tau}_{reg}$ will not necessarily improve precision, confidence intervals wrong.

Covariate Adjustment with Regression

Freedman (2008) shows that regression of the form:

$$Y_i = \alpha + \tau_{reg} D_i + \beta_1 X_i + \varepsilon_i$$

- $\hat{\tau}_{reg}$; small bias on order of $1/N$
- $\hat{\tau}_{reg}$ will not necessarily improve precision, confidence intervals wrong.

Lin (2013) replies, showing:

- in sufficient samples, these problems are minimal
- robust SEs give asymptotically consistent or conservative CIs
- if you interact covariates with treatment, it never hurts asymptotic precision, likely increases precision:

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

- interactions not needed in large sample, but might as well

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Interpreting $\tau_{interact}$:

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Interpreting $\tau_{interact}$:

- What is $\mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}]$?

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Interpreting $\tau_{interact}$:

- What is $\mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}] - \mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Interpreting $\tau_{interact}$:

- What is $\mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}] - \mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = x^*] - \mathbb{E}[Y_i | D_i = 0, X_i = x^*]$ for $x^* \neq \overline{X}$?

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Interpreting $\tau_{interact}$:

- What is $\mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}] - \mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = x^*] - \mathbb{E}[Y_i | D_i = 0, X_i = x^*]$ for $x^* \neq \overline{X}$?

So,

- $\tau_{interact}$ tells you comparison after "moving" all points to $X_i = \overline{X}$

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Interpreting $\tau_{interact}$:

- What is $\mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}] - \mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = x^*] - \mathbb{E}[Y_i | D_i = 0, X_i = x^*]$ for $x^* \neq \overline{X}$?

So,

- $\tau_{interact}$ tells you comparison after "moving" all points to $X_i = \overline{X}$
- same as $\mathbb{E}_X[\hat{\tau} | X = x]$, so convenient

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Interpreting $\tau_{interact}$:

- What is $\mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}] - \mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = x^*] - \mathbb{E}[Y_i | D_i = 0, X_i = x^*]$ for $x^* \neq \overline{X}$?

So,

- $\tau_{interact}$ tells you comparison after "moving" all points to $X_i = \overline{X}$
- same as $\mathbb{E}_X[\hat{\tau} | X = x]$, so convenient
- model also gives you $ATE(X)$ for any $x \in X$

Why specify the interactive model this way?

Consider our model,

$$Y_i = \alpha + \tau_{interact} D_i + \beta_1 \cdot (X_i - \bar{X}) + \beta_2 \cdot D_i \cdot (X_i - \bar{X}) + \varepsilon_i$$

Review:

- What is $\frac{\partial Y_i}{\partial X_i}$? $\frac{\partial Y_i}{\partial D_i}$?

Interpreting $\tau_{interact}$:

- What is $\mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = \overline{X}] - \mathbb{E}[Y_i | D_i = 0, X_i = \overline{X}]$?
- What is $\mathbb{E}[Y_i | D_i = 1, X_i = x^*] - \mathbb{E}[Y_i | D_i = 0, X_i = x^*]$ for $x^* \neq \overline{X}$?

So,

- $\tau_{interact}$ tells you comparison after "moving" all points to $X_i = \overline{X}$
- same as $\mathbb{E}_X[\hat{\tau} | X = x]$, so convenient
- model also gives you $ATE(X)$ for any $x \in X$
- allowing the interaction is a good idea, though not yet standard practice

Why are Experimental Findings Robust to Alternative Specifications?

Two facts:

1. *any* bivariate regression of some Y on X has $\hat{\beta} = \frac{\widehat{Cov}(Y,X)}{\widehat{\text{Var}}(X)}$.

Why are Experimental Findings Robust to Alternative Specifications?

Two facts:

1. *any* bivariate regression of some Y on X has $\hat{\beta} = \frac{\widehat{Cov}(Y,X)}{\widehat{\text{Var}}(X)}$.
2. the coefficient on the kth variable, X_k, from a multivariate regression can be obtained by (i) regress X_k on other X's, call residual $\tilde{X}_k$.(ii) regress Y on $\tilde{X}_k$. (Or $\tilde{Y}$ on $\tilde{X}_k$ for similarly defined $\tilde{Y}$)

Why are Experimental Findings Robust to Alternative Specifications?

Two facts:

1. *any* bivariate regression of some Y on X has $\hat{\beta} = \frac{\widehat{Cov}(Y,X)}{\widehat{\text{Var}}(X)}$.
2. the coefficient on the kth variable, X_k, from a multivariate regression can be obtained by (i) regress X_k on other X's, call residual $\tilde{X}_k$.(ii) regress Y on $\tilde{X}_k$. (Or $\tilde{Y}$ on $\tilde{X}_k$ for similarly defined $\tilde{Y}$)

All together, this gives the Frisch-Waugh-Lovell (FWL) theorem:

$$\beta_k = \frac{Cov(\tilde{Y}_i, \tilde{X}_{ki})}{Var(\tilde{X}_{ki})}$$

Where,

- $\tilde{X}_{ki}$ is the residual from a regression of X_{ki} on all other covariates
- $\tilde{Y}_i$ may have all other X partialled out as well or you can use the original Y_i.

Why are Experimental Findings Robust to Alternative Specifications?

If our "X_k" of interest is the D:

$$\hat{\beta}_D = \frac{\widehat{Cov}(\tilde{Y}_i, \tilde{D}_i)}{\widehat{\text{Var}}(\tilde{D}_i)}$$

where $\tilde{D}_i$ is the residual part of D_i after regressing on X_i

Why are Experimental Findings Robust to Alternative Specifications?

If our "X_k" of interest is the D:

$$\hat{\beta}_D = \frac{\widehat{Cov}(\tilde{Y}_i, \tilde{D}_i)}{\widehat{\text{Var}}(\tilde{D}_i)}$$

where $\tilde{D}_i$ is the residual part of D_i after regressing on X_i

- For experimental data, on average, what will $\tilde{D}_i$ look like?

Why are Experimental Findings Robust to Alternative Specifications?

If our "X_k" of interest is the D:

$$\hat{\beta}_D = \frac{\widehat{Cov}(\tilde{Y}_i, \tilde{D}_i)}{\widehat{\text{Var}}(\tilde{D}_i)}$$

where $\tilde{D}_i$ is the residual part of D_i after regressing on X_i

- For experimental data, on average, what will $\tilde{D}_i$ look like?
- What does this suggest about the effect of adding covariates to the simple regression of Y on D?

Summary: Covariate Adjustment with Regression

- "Agnostic" view of regression: one does not need to believe in the classical linear model (linearity and constant treatment effects) to advocate OLS covariate adjustment in randomized experiments

Summary: Covariate Adjustment with Regression

- "Agnostic" view of regression: one does not need to believe in the classical linear model (linearity and constant treatment effects) to advocate OLS covariate adjustment in randomized experiments
- Helpful both to protect against chance imbalance and improve efficiency.

Summary: Covariate Adjustment with Regression

- "Agnostic" view of regression: one does not need to believe in the classical linear model (linearity and constant treatment effects) to advocate OLS covariate adjustment in randomized experiments
- Helpful both to protect against chance imbalance and improve efficiency.

Summary: Covariate Adjustment with Regression

- "Agnostic" view of regression: one does not need to believe in the classical linear model (linearity and constant treatment effects) to advocate OLS covariate adjustment in randomized experiments
- Helpful both to protect against chance imbalance and improve efficiency.
- Since covariates are balanced by design, results typically not model dependent

Summary: Covariate Adjustment with Regression

- "Agnostic" view of regression: one does not need to believe in the classical linear model (linearity and constant treatment effects) to advocate OLS covariate adjustment in randomized experiments
- Helpful both to protect against chance imbalance and improve efficiency.
- Since covariates are balanced by design, results typically not model dependent
- Best reason to be sceptical of adjustment: fishing.
  - always report unadjusted result first
  - best if adjustment strategy pre-specified

Blocking

- Prior to randomization, collect background data on each unit.

Blocking

- Prior to randomization, collect background data on each unit.
- Why leave it to chance to balance on these, especially in small sample?

Blocking

- Prior to randomization, collect background data on each unit.
- Why leave it to chance to balance on these, especially in small sample?
- Blocking: pre-stratify sample, randomize within each stratum to ensure groups start out with identical distribution of the blocked factors.

Blocking

- Prior to randomization, collect background data on each unit.
- Why leave it to chance to balance on these, especially in small sample?
- Blocking: pre-stratify sample, randomize within each stratum to ensure groups start out with identical distribution of the blocked factors.
- You effectively run a separate experiment within each stratum; randomization will balance unobserved attributes

Blocking

- Prior to randomization, collect background data on each unit.
- Why leave it to chance to balance on these, especially in small sample?
- Blocking: pre-stratify sample, randomize within each stratum to ensure groups start out with identical distribution of the blocked factors.
- You effectively run a separate experiment within each stratum; randomization will balance unobserved attributes
- Why is this helpful?
  - Four subjects with pre-tx outcomes of $\{2, 2, 8, 8\}$. (Say Y_{i0} are similar).

Blocking

- Prior to randomization, collect background data on each unit.
- Why leave it to chance to balance on these, especially in small sample?
- Blocking: pre-stratify sample, randomize within each stratum to ensure groups start out with identical distribution of the blocked factors.
- You effectively run a separate experiment within each stratum; randomization will balance unobserved attributes
- Why is this helpful?
  - Four subjects with pre-tx outcomes of $\{2, 2, 8, 8\}$. (Say Y_{i0} are similar).
  - Simple random assignment into two treated/control will place $\{2, 2\}$ and $\{8, 8\}$ together in the same group one-third of the time

Blocking

Imagine you run an experiment where you block on gender.

Blocking

Imagine you run an experiment where you block on gender.

Obtain block specific ATEs, $\tau_f = \mathbb{E}[\tau|f]$, $\tau_m = \mathbb{E}[\tau|m]$.

Blocking

Imagine you run an experiment where you block on gender.

Obtain block specific ATEs, $\tau_f = \mathbb{E}[\tau|f]$, $\tau_m = \mathbb{E}[\tau|m]$.

(Let's review law of iterated expectations)

Blocking

Imagine you run an experiment where you block on gender.

Obtain block specific ATEs, $\tau_f = \mathbb{E}[\tau|f]$, $\tau_m = \mathbb{E}[\tau|m]$.

(Let's review law of iterated expectations)

Applying the LIE,

$$\mathbb{E}[\tau] = p[f]\mathbb{E}[\tau|f] + p[m]\mathbb{E}[\tau|m]$$

Blocking

Imagine you run an experiment where you block on gender.

Obtain block specific ATEs, $\tau_f = \mathbb{E}[\tau|f]$, $\tau_m = \mathbb{E}[\tau|m]$.

(Let's review law of iterated expectations)

Applying the LIE,

$$\mathbb{E}[\tau] = p[f]\mathbb{E}[\tau|f] + p[m]\mathbb{E}[\tau|m]$$

The sample analog is unbiased for this:

$$\hat{\tau}_B = \frac{N_f}{N_f + N_m} \cdot \hat{\tau}_f + \frac{N_m}{N_f + N_m} \cdot \hat{\tau}_m$$

Blocking

Imagine you run an experiment where you block on gender.

Obtain block specific ATEs, $\tau_f = \mathbb{E}[\tau|f]$, $\tau_m = \mathbb{E}[\tau|m]$.

(Let's review law of iterated expectations)

Applying the LIE,

$$\mathbb{E}[\tau] = p[f]\mathbb{E}[\tau|f] + p[m]\mathbb{E}[\tau|m]$$

The sample analog is unbiased for this:

$$\hat{\tau}_B = \frac{N_f}{N_f + N_m} \cdot \hat{\tau}_f + \frac{N_m}{N_f + N_m} \cdot \hat{\tau}_m$$

or more generally, if there are J strata or blocks, then

$$\hat{\tau}_B = \sum_{j=1}^{J} \frac{N_j}{N} \hat{\tau}_j$$

Blocking

Taking the number of units per block as fixed (i.e. conditional on the experimental design), the variance is:

Blocking

Taking the number of units per block as fixed (i.e. conditional on the experimental design), the variance is:

$$\text{Var}(\hat{\tau}_B) = \left(\frac{N_f}{N_f + N_m}\right)^2 \text{Var}(\hat{\tau}_f) + \left(\frac{N_m}{N_f + N_m}\right)^2 \text{Var}(\hat{\tau}_m)$$

Blocking

Taking the number of units per block as fixed (i.e. conditional on the experimental design), the variance is:

$$\text{Var}(\hat{\tau}_B) = \left(\frac{N_f}{N_f + N_m}\right)^2 \text{Var}(\hat{\tau}_f) + \left(\frac{N_m}{N_f + N_m}\right)^2 \text{Var}(\hat{\tau}_m)$$

or more generally

$$Var(\hat{\tau}_B) = \sum_{j=1}^{J} \left(\frac{N_j}{N}\right)^2 \text{Var}(\hat{\tau}_j)$$

Blocking with Regression

Can we treat these blocks as regressors in OLS?

Blocking with Regression

Can we treat these blocks as regressors in OLS? Yes, with caution

Blocking with Regression

Can we treat these blocks as regressors in OLS? Yes, with caution

If probability of treatment is equal across the blocks, then OLS with block "fixed effects" will estimate the ATE:

$$Y_i = \beta_0 + \tau D_i + \sum_{j=2}^{J} \beta_j \cdot B_{ij} + \varepsilon_i$$

where B_j is a dummy for the j-th block (one omitted as reference category).

Blocking with Regression

Can we treat these blocks as regressors in OLS? Yes, with caution

If probability of treatment is equal across the blocks, then OLS with block "fixed effects" will estimate the ATE:

$$Y_i = \beta_0 + \tau D_i + \sum_{j=2}^{J} \beta_j \cdot B_{ij} + \varepsilon_i$$

where B_j is a dummy for the j-th block (one omitted as reference category).

If (empirical) $Pr(D = 1)$ changes by block, then OLS will put different weight on different blocks.

Blocking with Regression

Can we treat these blocks as regressors in OLS? Yes, with caution

If probability of treatment is equal across the blocks, then OLS with block "fixed effects" will estimate the ATE:

$$Y_i = \beta_0 + \tau D_i + \sum_{j=2}^{J} \beta_j \cdot B_{ij} + \varepsilon_i$$

where B_j is a dummy for the j-th block (one omitted as reference category).

If (empirical) $Pr(D = 1)$ changes by block, then OLS will put different weight on different blocks.

To correct for this, weight each observation by something that makes probabilities the same. Let p_{ij} be $Pr(D_i = 1 |$ block for unit i):

$$w_{ij} = \left(\frac{1}{p_{ij}}\right) D_i + \left(\frac{1}{1 - p_{ij}}\right) (1 - D_i)$$

When Does Blocking Help?

Imagine models for complete and block-randomized designs:

$$Y_i = \alpha + \tau_{CR} D_i + \varepsilon_i$$

When Does Blocking Help?

Imagine models for complete and block-randomized designs:

$$Y_i = \alpha + \tau_{CR} D_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha + \tau_{BR} D_i + \sum_{j=2}^{J} \beta_j B_{ij} + \varepsilon_i^* \quad (2)$$

where B_j is a dummy for the j-th block. Then given iid sampling:

When Does Blocking Help?

Imagine models for complete and block-randomized designs:

$$Y_i = \alpha + \tau_{CR} D_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha + \tau_{BR} D_i + \sum_{j=2}^{J} \beta_j B_{ij} + \varepsilon_i^* \quad (2)$$

where B_j is a dummy for the j-th block. Then given iid sampling:

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \qquad \text{with } \widehat{\sigma}_\varepsilon^2 =$$

When Does Blocking Help?

Imagine models for complete and block-randomized designs:

$$Y_i = \alpha + \tau_{CR} D_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha + \tau_{BR} D_i + \sum_{j=2}^{J} \beta_j B_{ij} + \varepsilon_i^* \quad (2)$$

where B_j is a dummy for the j-th block. Then given iid sampling:

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \qquad \text{with } \widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{n} \widehat{\varepsilon}_i^2}{n-2} = \frac{SSR_{\widehat{\varepsilon}}}{n-2}$$

When Does Blocking Help?

Imagine models for complete and block-randomized designs:

$$Y_i = \alpha + \tau_{CR} D_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha + \tau_{BR} D_i + \sum_{j=2}^{J} \beta_j B_{ij} + \varepsilon_i^* \quad (2)$$

where B_j is a dummy for the j-th block. Then given iid sampling:

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \quad \text{with } \widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{n} \widehat{\varepsilon}_i^2}{n-2} = \frac{SSR_{\widehat{\varepsilon}}}{n-2}$$

$$Var[\widehat{\tau}_{BR}] = \frac{\sigma_{\varepsilon^*}^2}{\sum_{i=1}^{n}(D_i - \bar{D})^2(1-R_j^2)} \quad \text{with } \widehat{\sigma}_{\varepsilon^*}{}^2 = \frac{\sum_{i=1}^{n} \widehat{\varepsilon}_i^{*2}}{n-k-1} = \frac{SSR_{\widehat{\varepsilon^*}}}{n-k-1}$$

When Does Blocking Help?

Imagine models for complete and block-randomized designs:

$$Y_i = \alpha + \tau_{CR} D_i + \varepsilon_i \quad (1)$$

$$Y_i = \alpha + \tau_{BR} D_i + \sum_{j=2}^{J} \beta_j B_{ij} + \varepsilon_i^* \quad (2)$$

where B_j is a dummy for the j-th block. Then given iid sampling:

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma_\varepsilon^2}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \quad \text{with } \widehat{\sigma}_\varepsilon^2 = \frac{\sum_{i=1}^{n} \widehat{\varepsilon}_i^2}{n-2} = \frac{SSR_{\widehat{\varepsilon}}}{n-2}$$

$$Var[\widehat{\tau}_{BR}] = \frac{\sigma_{\varepsilon^*}^2}{\sum_{i=1}^{n}(D_i - \bar{D})^2(1-R_j^2)} \quad \text{with } \widehat{\sigma}_{\varepsilon^*}{}^2 = \frac{\sum_{i=1}^{n} \widehat{\varepsilon}_i^{*2}}{n-k-1} = \frac{SSR_{\widehat{\varepsilon}^*}}{n-k-1}$$

where R_j^2 is R^2 from regression of D on all B_j variables and a constant.

When Does Blocking Help?

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma^2_{\varepsilon}}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \quad \text{with } \widehat{\sigma}^2_{\varepsilon} = \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}{n-2} = \frac{SSR_{\widehat{\varepsilon}}}{n-2}$$

$$Var[\widehat{\tau}_{BR}] = \frac{\sigma^2_{\varepsilon *}}{\sum_{i=1}^{n}(D_i - \bar{D})^2(1-R_j^2)} \text{ with } \widehat{\sigma}_{\varepsilon^*}{}^2 = \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^{*2}}{n-k-1} = \frac{SSR_{\widehat{\varepsilon}^*}}{n-k-1}$$

where R_j^2 is R^2 from regression of D on all B_j variables and a constant.

When Does Blocking Help?

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma^2_{\varepsilon}}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \quad \text{with } \widehat{\sigma}^2_{\varepsilon} = \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}{n-2} = \frac{SSR_{\widehat{\varepsilon}}}{n-2}$$

$$Var[\widehat{\tau}_{BR}] = \frac{\sigma^2_{\varepsilon *}}{\sum_{i=1}^{n}(D_i - \bar{D})^2(1-R_j^2)} \text{ with } \widehat{\sigma}_{\varepsilon^*}{}^2 = \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^{*2}}{n-k-1} = \frac{SSR_{\widehat{\varepsilon}^*}}{n-k-1}$$

where R_j^2 is R^2 from regression of D on all B_j variables and a constant.

When Does Blocking Help?

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma^2_{\varepsilon}}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \quad \text{with } \widehat{\sigma}^2_{\varepsilon} = \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}{n-2} = \frac{SSR_{\widehat{\varepsilon}}}{n-2}$$

$$Var[\widehat{\tau}_{BR}] = \frac{\sigma^2_{\varepsilon *}}{\sum_{i=1}^{n}(D_i - \bar{D})^2(1-R_j^2)} \text{ with } \widehat{\sigma}_{\varepsilon^*}{}^2 = \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^{*2}}{n-k-1} = \frac{SSR_{\widehat{\varepsilon}^*}}{n-k-1}$$

where R_j^2 is R^2 from regression of D on all B_j variables and a constant.

- If R_j^2 is high, variance inflated. But what can we say about R_{ij}^2 when we're talking about blocking?

When Does Blocking Help?

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma^2_{\varepsilon}}{\sum_{i=1}^{n}(D_i - \bar{D})^2} \quad \text{with } \widehat{\sigma}^2_{\varepsilon} = \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^2}{n-2} = \frac{SSR_{\widehat{\varepsilon}}}{n-2}$$

$$Var[\widehat{\tau}_{BR}] = \frac{\sigma^2_{\varepsilon*}}{\sum_{i=1}^{n}(D_i - \bar{D})^2(1-R_j^2)} \text{ with } \widehat{\sigma}_{\varepsilon^*}{}^2 = \frac{\sum_{i=1}^{n}\widehat{\varepsilon}_i^{*2}}{n-k-1} = \frac{SSR_{\widehat{\varepsilon}^*}}{n-k-1}$$

where R_j^2 is R^2 from regression of D on all B_j variables and a constant.

- If R_j^2 is high, variance inflated. But what can we say about R_{ij}^2 when we're talking about blocking?
- If blocks explain outcomes strongly, how will $\widehat{\sigma}_{\varepsilon^*}{}^2$ compare to $\widehat{\sigma}^2_{\varepsilon}$?

When Does Blocking Help?

$$Var[\widehat{\tau}_{CR}] = \frac{\sigma^2_\varepsilon}{\sum_{i=1}^n (D_i - \bar{D})^2} \quad \text{with } \widehat{\sigma}^2_\varepsilon = \frac{\sum_{i=1}^n \widehat{\varepsilon}_i^2}{n-2} = \frac{SSR_{\widehat{\varepsilon}}}{n-2}$$

$$Var[\widehat{\tau}_{BR}] = \frac{\sigma^2_{\varepsilon *}}{\sum_{i=1}^n (D_i - \bar{D})^2 (1 - R_j^2)} \text{ with } \widehat{\sigma}_{\varepsilon^*}{}^2 = \frac{\sum_{i=1}^n \widehat{\varepsilon}_i^{*2}}{n-k-1} = \frac{SSR_{\widehat{\varepsilon}^*}}{n-k-1}$$

where R_j^2 is R^2 from regression of D on all B_j variables and a constant.

- If R_j^2 is high, variance inflated. But what can we say about R_{ij}^2 when we're talking about blocking?
- If blocks explain outcomes strongly, how will $\widehat{\sigma}_{\varepsilon^*}{}^2$ compare to $\widehat{\sigma}^2_\varepsilon$?
- Adding DOF adjustment, we end up asking whether $\frac{SSR_{\widehat{\varepsilon}^*}}{n-k-1} < \frac{SSR_{\widehat{\varepsilon}}}{n-2}$

Fair Trade Labeling Experiment (Hainmueller et al, 2012)

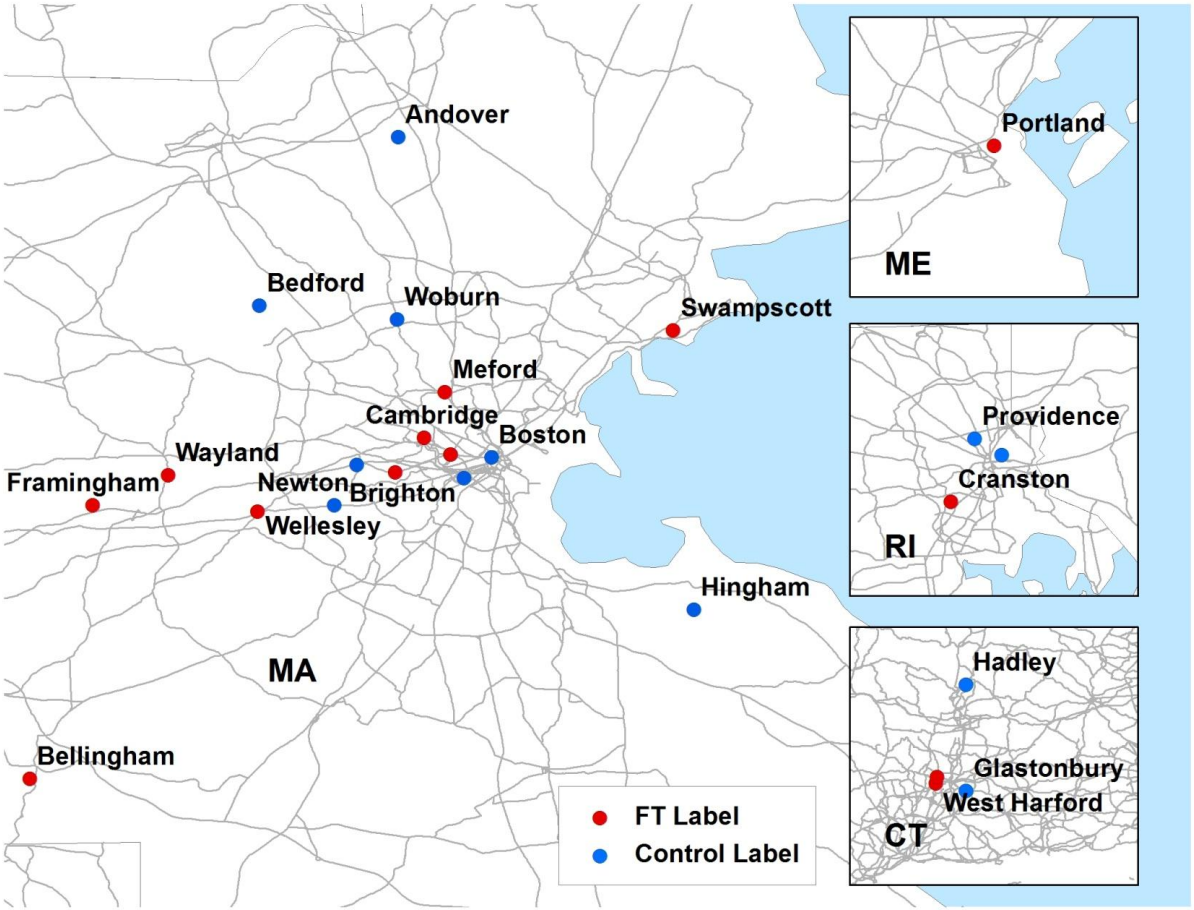
Label Experiment

Treatment

Control

Example: Fair Trade Labeling Experiment

Matched Pairs: Phase 1



Example: Fair Trade Labeling Experiment

```
> d <- read.dta("FTdata.dta")
> head(d)
  store pair FTweek lnsalesd
1     1    1      1     3.20
2     4    1      0     2.77
3     6    2      1     4.18
4     9    2      0     4.04
5    21    3      1     4.30
6    24    3      0     3.93
```

Example: Fair Trade Labeling Experiment

```
> cr.out <- lm(lnsalesd~FTweek,data=d)
> coeftest(cr.out,vcov = vcovHC(cr.out, type = "HC1"))

t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.35000    0.16079 27.0537   <2e-16 ***
FTweek       0.12385    0.21424  0.5781   0.5686
--
```

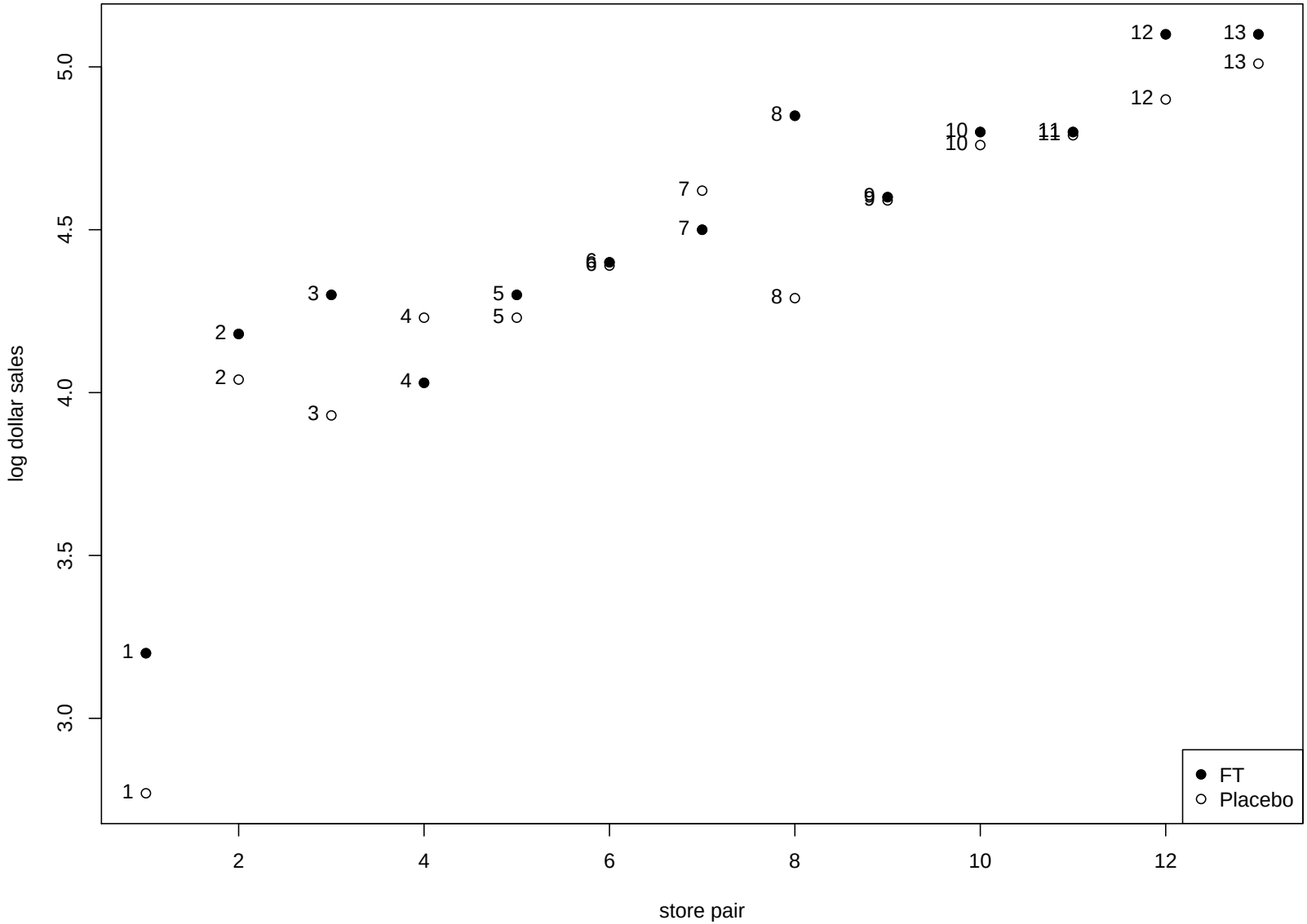
Example: Fair Trade Labeling Experiment

```
> br.out <- lm(lnsalesd~FTweek+as.factor(pair),data=d)
> coeftest(br.out,vcov = vcovHC(br.out, type = "HC1"))

t test of coefficients:

                   Estimate Std. Error t value  Pr(>|t|)
(Intercept)        2.923077   0.162144 18.0277 4.671e-10 ***
FTweek             0.123846   0.060176  2.0581 0.0619840
as.factor(pair)2   1.125000   0.159549  7.0511 1.335e-05 ***
as.factor(pair)3   1.130000   0.204440  5.5273 0.0001304 ***
as.factor(pair)4   1.145000   0.231925  4.9369 0.0003439 ***
as.factor(pair)5   1.280000   0.161773  7.9123 4.208e-06 ***
as.factor(pair)6   1.410000   0.169987  8.2948 2.591e-06 ***
as.factor(pair)7   1.575000   0.203689  7.7324 5.317e-06 ***
as.factor(pair)8   1.585000   0.277319  5.7154 9.675e-05 ***
as.factor(pair)9   1.610000   0.169987  9.4713 6.420e-07 ***
as.factor(pair)10 1.795000   0.165195 10.8660 1.450e-07 ***
as.factor(pair)11 1.810000   0.169987 10.6479 1.810e-07 ***
as.factor(pair)12 2.015000   0.164183 12.2729 3.763e-08 ***
as.factor(pair)13 2.070000   0.160298 12.9134 2.127e-08 ***
--
```

Example: Fair Trade Labeling Experiment



How much is explained by blocks?

```
> summary(lm(lnsalesd~as.factor(pair),data=d))
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.9850     0.1212  24.621 2.72e-12 ***
as.factor(pair)2     1.1250     0.1715   6.562 1.82e-05 ***
as.factor(pair)3     1.1300     0.1715   6.591 1.74e-05 ***
as.factor(pair)4     1.1450     0.1715   6.678 1.52e-05 ***
as.factor(pair)5     1.2800     0.1715   7.466 4.73e-06 ***
as.factor(pair)6     1.4100     0.1715   8.224 1.65e-06 ***
as.factor(pair)7     1.5750     0.1715   9.186 4.77e-07 ***
as.factor(pair)8     1.5850     0.1715   9.245 4.44e-07 ***
as.factor(pair)9     1.6100     0.1715   9.390 3.71e-07 ***
as.factor(pair)10    1.7950     0.1715  10.469 1.05e-07 ***
as.factor(pair)11    1.8100     0.1715  10.557 9.56e-08 ***
as.factor(pair)12    2.0150     0.1715  11.752 2.68e-08 ***
as.factor(pair)13    2.0700     0.1715  12.073 1.94e-08 ***
--
Residual standard error: 0.1715 on 13 degrees of freedom
Multiple R-squared:  0.9474,        Adjusted R-squared:  0.8988
F-statistic:  19.5 on 12 and 13 DF,  p-value: 2.356e-06
```

Conclusions: Blocking

How does blocking help?

- Increases efficiency if the blocking variables predict outcomes
- Can help with small sample bias due to “bad” randomization
- Powerful, especially in small to medium sized samples. Can save your experiment!

Conclusions: Blocking

How does blocking help?

- Increases efficiency if the blocking variables predict outcomes
- Can help with small sample bias due to “bad” randomization
- Powerful, especially in small to medium sized samples. Can save your experiment!

What to block on?

- Two constraints: data you can get, number of units to work with.
- Irrelevant variables: burn up DOFs, worsen matches on good ones
- Most important: predictors of potential outcomes!
- Pre-Tx outcomes is great; add other main predictors if available/helpful
- Another consideration: variables desired for subgroup analysis

Conclusions: Blocking

How does blocking help?

- Increases efficiency if the blocking variables predict outcomes
- Can help with small sample bias due to “bad” randomization
- Powerful, especially in small to medium sized samples. Can save your experiment!

What to block on?

- Two constraints: data you can get, number of units to work with.
- Irrelevant variables: burn up DOFs, worsen matches on good ones
- Most important: predictors of potential outcomes!
- Pre-Tx outcomes is great; add other main predictors if available/helpful
- Another consideration: variables desired for subgroup analysis

How to block?

- Stratification
- Pair or two pair-matching
- Check: `blockTools` library.

Analysis with Blocking

"**As ye randomize, so shall ye analyze**" (Senn 2004)

Analysis with Blocking

"**As ye randomize, so shall ye analyze**" (Senn 2004)

- Account for the method of randomization when doing analysis

Analysis with Blocking

"**As ye randomize, so shall ye analyze**" (Senn 2004)

- Account for the method of randomization when doing analysis
- If using OLS,
  - include block dummies
  - typically use clustered SEs
  - if probability of treatment assignment varies across blocks, reweight

Analysis with Blocking

"**As ye randomize, so shall ye analyze**" (Senn 2004)

- Account for the method of randomization when doing analysis
- If using OLS,
  - include block dummies
  - typically use clustered SEs
  - if probability of treatment assignment varies across blocks, reweight
- If doing randomization inference, block design influences construction of Ω: re-randomization is within blocks.

Analysis with Blocking

"**As ye randomize, so shall ye analyze**" (Senn 2004)

- Account for the method of randomization when doing analysis
- If using OLS,
  - include block dummies
  - typically use clustered SEs
  - if probability of treatment assignment varies across blocks, reweight
- If doing randomization inference, block design influences construction of Ω: re-randomization is within blocks.
- If doings bootstrap, do block bootstrap.

Analysis with Blocking

"**As ye randomize, so shall ye analyze**" (Senn 2004)

- Account for the method of randomization when doing analysis
- If using OLS,
  - include block dummies
  - typically use clustered SEs
  - if probability of treatment assignment varies across blocks, reweight
- If doing randomization inference, block design influences construction of Ω: re-randomization is within blocks.
- If doings bootstrap, do block bootstrap.
- Failure to account for the method of randomization in these ways can result in incorrect test size.

Hardwork in the Design and Implementation

Statistics are often easy; challenge is in implementation and design

Hardwork in the Design and Implementation

Statistics are often easy; challenge is in implementation and design

- Finding partners, manage relationships will be common failure point

Hardwork in the Design and Implementation

Statistics are often easy; challenge is in implementation and design

- Finding partners, manage relationships will be common failure point
- Possible benefits to implementers
  - free “consulting” to assist with monitoring and evaluation
  - allocating limited resources (dealing with large target group)
  - encouragement designs increase uptake
  - ...but if implementers are happy with a much cheaper (non-causal) evaluation, may not be willing to do what it takes

Hardwork in the Design and Implementation

Statistics are often easy; challenge is in implementation and design

- Finding partners, manage relationships will be common failure point
- Possible benefits to implementers
  - free “consulting” to assist with monitoring and evaluation
  - allocating limited resources (dealing with large target group)
  - encouragement designs increase uptake
  - ...but if implementers are happy with a much cheaper (non-causal) evaluation, may not be willing to do what it takes
- Iron law of field experiments: keep it simple. Many ways to fail.

Hardwork in the Design and Implementation

Statistics are often easy; challenge is in implementation and design

- Finding partners, manage relationships will be common failure point
- Possible benefits to implementers
  - free “consulting” to assist with monitoring and evaluation
  - allocating limited resources (dealing with large target group)
  - encouragement designs increase uptake
  - ...but if implementers are happy with a much cheaper (non-causal) evaluation, may not be willing to do what it takes
- Iron law of field experiments: keep it simple. Many ways to fail.
- Getting difficult to publish studies that only examine the efficacy of some intervention: look for substantively interesting treatment and outcome.

Ethics and Experimentation

- Fearon, Humphreys, and Weinstein (2009) used a field experiment to examine if community-driven reconstruction programs foster social reconciliation in post-conflict Liberian villages.
- Outcome: funding raised for collective projects in public goods game played with 24 villagers. Total payout to village is publicly announced.

Ethics and Experimentation

- Fearon, Humphreys, and Weinstein (2009) used a field experiment to examine if community-driven reconstruction programs foster social reconciliation in post-conflict Liberian villages.
- Outcome: funding raised for collective projects in public goods game played with 24 villagers. Total payout to village is publicly announced.

We received a report that leaders in one community had gathered villagers together after we left and asked people to report how much they had contributed. We moved quickly to prevent any retribution in that village, but also decided to alter the protocol for subsequent games to ensure greater protection for game participants.

Ethics and Experimentation

- Fearon, Humphreys, and Weinstein (2009) used a field experiment to examine if community-driven reconstruction programs foster social reconciliation in post-conflict Liberian villages.
- Outcome: funding raised for collective projects in public goods game played with 24 villagers. Total payout to village is publicly announced.

We received a report that leaders in one community had gathered villagers together after we left and asked people to report how much they had contributed. We moved quickly to prevent any retribution in that village, but also decided to alter the protocol for subsequent games to ensure greater protection for game participants.

These changes included stronger language about the importance of protecting anonymity, random audits of community behavior, facilitation of anonymous reporting of violations of game protocol by participants, and a new opportunity to receive supplemental funds in a postproject lottery if no reports of harassment were received.

Ethics and Experimentation

IRB approval typically required.

Ethics and Experimentation

IRB approval typically required.

Key principals (Belmont Report, 1978):

Ethics and Experimentation

IRB approval typically required.

Key principals (Belmont Report, 1978):

- **Respect for persons**: informed consent.
  - Informed consent often done as part of the baseline survey.
  - If risks are minimal and consent will undermine the study, then informed consent rules sometimes waived.

Ethics and Experimentation

IRB approval typically required.

Key principals (Belmont Report, 1978):

- **Respect for persons**: informed consent.
  - Informed consent often done as part of the baseline survey.
  - If risks are minimal and consent will undermine the study, then informed consent rules sometimes waived.
- **Benefice**: Avoid knowingly doing harm. Does not mean that all risk can be eliminated, but possible risks must be balanced against overall benefits to society of the research.

Ethics and Experimentation

IRB approval typically required.

Key principals (Belmont Report, 1978):

- **Respect for persons**: informed consent.
  - Informed consent often done as part of the baseline survey.
  - If risks are minimal and consent will undermine the study, then informed consent rules sometimes waived.
- **Benefice**: Avoid knowingly doing harm. Does not mean that all risk can be eliminated, but possible risks must be balanced against overall benefits to society of the research.
  - Control group is often difficult: denying access to some benefit.

Ethics and Experimentation

IRB approval typically required.

Key principals (Belmont Report, 1978):

- **Respect for persons**: informed consent.
  - Informed consent often done as part of the baseline survey.
  - If risks are minimal and consent will undermine the study, then informed consent rules sometimes waived.
- **Benefice**: Avoid knowingly doing harm. Does not mean that all risk can be eliminated, but possible risks must be balanced against overall benefits to society of the research.
  - Control group is often difficult: denying access to some benefit.
  - however, project benefits often finite anyway, so why not distribute them by randomization...but that won't always prevent possible envy problems

Ethics and Experimentation

IRB approval typically required.

Key principals (Belmont Report, 1978):

- **Respect for persons**: informed consent.
  - Informed consent often done as part of the baseline survey.
  - If risks are minimal and consent will undermine the study, then informed consent rules sometimes waived.
- **Benefice**: Avoid knowingly doing harm. Does not mean that all risk can be eliminated, but possible risks must be balanced against overall benefits to society of the research.
  - Control group is often difficult: denying access to some benefit.
  - however, project benefits often finite anyway, so why not distribute them by randomization...but that won't always prevent possible envy problems
  - treatments that are occurring already in the wild pose minimal additional risk, but still may be criticized when used for research

Ethics and Experimentation

IRB approval typically required.

Key principals (Belmont Report, 1978):

- **Respect for persons**: informed consent.
  - Informed consent often done as part of the baseline survey.
  - If risks are minimal and consent will undermine the study, then informed consent rules sometimes waived.
- **Benefice**: Avoid knowingly doing harm. Does not mean that all risk can be eliminated, but possible risks must be balanced against overall benefits to society of the research.
  - Control group is often difficult: denying access to some benefit.
  - however, project benefits often finite anyway, so why not distribute them by randomization...but that won't always prevent possible envy problems
  - treatments that are occurring already in the wild pose minimal additional risk, but still may be criticized when used for research
- **Justice**: Benefits and risks should accrue to same group or individuals

Ethics and Experimentation

IRB approval typically required.

Key principals (Belmont Report, 1978):

- **Respect for persons**: informed consent.
  - Informed consent often done as part of the baseline survey.
  - If risks are minimal and consent will undermine the study, then informed consent rules sometimes waived.
- **Benefice**: Avoid knowingly doing harm. Does not mean that all risk can be eliminated, but possible risks must be balanced against overall benefits to society of the research.
  - Control group is often difficult: denying access to some benefit.
  - however, project benefits often finite anyway, so why not distribute them by randomization...but that won't always prevent possible envy problems
  - treatments that are occurring already in the wild pose minimal additional risk, but still may be criticized when used for research
- **Justice**: Benefits and risks should accrue to same group or individuals
  - evaluate interventions that are relevant to the subject population


FACEBOOK SHOULDN'T CHOOSE WHAT STUFF THEY SHOW US TO CONDUCT UNETHICAL PSYCHOLOGICAL RESEARCH.
THEY SHOULD ONLY MAKE THOSE DECISIONS BASED ON, UH...
HOWEVER THEY WERE DOING IT BEFORE.
WHICH WAS PROBABLY ETHICAL, RIGHT?

Conclusion: Experiments

- Random assignment is great, but can use a helping hand sometimes...

Conclusion: Experiments

- Random assignment is great, but can use a helping hand sometimes...
- Design: blocking to improve balance
  - improves balance
  - improves efficiency
  - can save your experiments in small samples

Conclusion: Experiments

- Random assignment is great, but can use a helping hand sometimes...
- Design: blocking to improve balance
  - improves balance
  - improves efficiency
  - can save your experiments in small samples
- Analysis: covariate adjustment

Conclusion: Experiments

- Random assignment is great, but can use a helping hand sometimes...
- Design: blocking to improve balance
  - improves balance
  - improves efficiency
  - can save your experiments in small samples
- Analysis: covariate adjustment
  - can improve efficiency and finite sample bias
  - relatively safe, use interactive model (Lin 2013)
  - main risk is user-defined specification (fishing)

Conclusion: Experiments

- Random assignment is great, but can use a helping hand sometimes...
- Design: blocking to improve balance
  - improves balance
  - improves efficiency
  - can save your experiments in small samples
- Analysis: covariate adjustment
  - can improve efficiency and finite sample bias
  - relatively safe, use interactive model (Lin 2013)
  - main risk is user-defined specification (fishing)
    - report unadjusted results first
    - commit to adjustment plan, use non-arbitrary choices
    - typically results will stable though

Conclusion: Experiments

- Random assignment is great, but can use a helping hand sometimes...
- Design: blocking to improve balance
  - improves balance
  - improves efficiency
  - can save your experiments in small samples
- Analysis: covariate adjustment
  - can improve efficiency and finite sample bias
  - relatively safe, use interactive model (Lin 2013)
  - main risk is user-defined specification (fishing)
    - report unadjusted results first
    - commit to adjustment plan, use non-arbitrary choices
    - typically results will stable though

Conclusion: Experiments

- Random assignment is great, but can use a helping hand sometimes...
- Design: blocking to improve balance
  - improves balance
  - improves efficiency
  - can save your experiments in small samples
- Analysis: covariate adjustment
  - can improve efficiency and finite sample bias
  - relatively safe, use interactive model (Lin 2013)
  - main risk is user-defined specification (fishing)
    - report unadjusted results first
    - commit to adjustment plan, use non-arbitrary choices
    - typically results will stable though
- Beyond methodology, remember to consider:

Conclusion: Experiments

- Random assignment is great, but can use a helping hand sometimes...
- Design: blocking to improve balance
  - improves balance
  - improves efficiency
  - can save your experiments in small samples
- Analysis: covariate adjustment
  - can improve efficiency and finite sample bias
  - relatively safe, use interactive model (Lin 2013)
  - main risk is user-defined specification (fishing)
    - report unadjusted results first
    - commit to adjustment plan, use non-arbitrary choices
    - typically results will stable though
- Beyond methodology, remember to consider:
  - can your experiment reveal an interesting effect?
  - will it have external validity
  - ethical concerns