

Lecture 2: Estimation and Inference for a Randomized Experiment

POL-GA 1251
Quantitative Political Analysis II
Prof. Cyrus Samii
NYU Politics

January 30, 2019

From identification to estimation

Random assignment **identifies** the ATE. We now turn to steps that follow identification: estimation and inference.

From identification to estimation

Random assignment **identifies** the ATE. We now turn to steps that follow identification: estimation and inference.

By **estimation** I mean using observable quantities to describe sample or population parameters. We are concerned with,

- ▶ Estimators and their bias, consistency, and efficiency.

From identification to estimation

Random assignment **identifies** the ATE. We now turn to steps that follow identification: estimation and inference.

By **estimation** I mean using observable quantities to describe sample or population parameters. We are concerned with,

- ▶ Estimators and their bias, consistency, and efficiency.

By **statistical inference** I mean making probabilistic statements about relationships between estimates and sample or population parameters.

We are concerned with,

- ▶ Intervals and coverage.
- ▶ Testing and Type I/II errors.

From identification to estimation

Random assignment **identifies** the ATE. We now turn to steps that follow identification: estimation and inference.

By **estimation** I mean using observable quantities to describe sample or population parameters. We are concerned with,

- ▶ Estimators and their bias, consistency, and efficiency.

By **statistical inference** I mean making probabilistic statements about relationships between estimates and sample or population parameters. We are concerned with,

- ▶ Intervals and coverage.
- ▶ Testing and Type I/II errors.

We will look at an **idealized randomized experiment** to illustrate key estimation and inference ideas.

The population and target

- ▶ A large population U of experimental units.

The population and target

- ▶ A large population U of experimental units.
- ▶ We will run a completely randomized experiment using a treatment, D_i , with support $\mathcal{D} = \{0, 1\}$, on a random sample, S .

The population and target

- ▶ A large population U of experimental units.
- ▶ We will run a completely randomized experiment using a treatment, D_i , with support $\mathcal{D} = \{0, 1\}$, on a random sample, S .
- ▶ Each unit in $j \in U$ has potential outcomes, $\{y_{dj}\}_{d \in \mathcal{D}}$.

The population and target

- ▶ A large population U of experimental units.
- ▶ We will run a completely randomized experiment using a treatment, D_i , with support $\mathcal{D} = \{0, 1\}$, on a random sample, S .
- ▶ Each unit in $j \in U$ has potential outcomes, $\{y_{dj}\}_{d \in \mathcal{D}}$.
- ▶ Each unit is also characterized by a continuous covariate, x_j .

The population and target

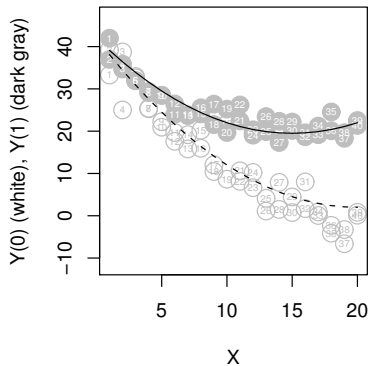
- ▶ A large population U of experimental units.
- ▶ We will run a completely randomized experiment using a treatment, D_i , with support $\mathcal{D} = \{0, 1\}$, on a random sample, S .
- ▶ Each unit in $j \in U$ has potential outcomes, $\{y_{dj}\}_{d \in \mathcal{D}}$.
- ▶ Each unit is also characterized by a continuous covariate, x_j .
- ▶ The population average treatment effect (PATE) is given by the expected value of difference in potential outcomes for an arbitrary unit drawn from U ,

$$PATE = \rho = E[Y_1 - Y_0].$$

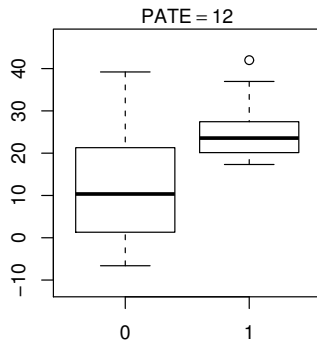
- ▶ The PATE is our **target** of estimation and inference.

The population and target

**Potential outcomes
w/ covariate**



Diffs. in pot. outcomes



From population to sample

- ▶ We take a random sample S indexed by $i = 1, \dots, n$.
- ▶ An arbitrary member of S is characterized by potential outcomes, (Y_{1i}, Y_{0i}) , and a covariate, X_i .

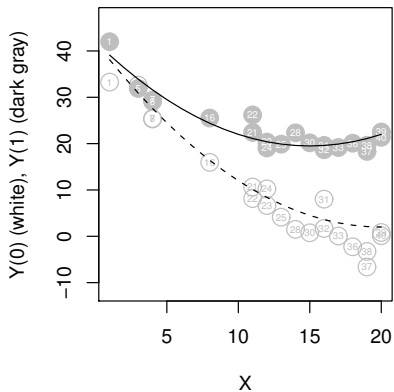
From population to sample

- ▶ We take a random sample S indexed by $i = 1, \dots, n$.
- ▶ An arbitrary member of S is characterized by potential outcomes, (Y_{1i}, Y_{0i}) , and a covariate, X_i .
- ▶ Another possible target of inference is thus the sample average treatment effect (SATE), given by the average difference in potential outcomes over members of the sample,

$$SATE = \rho_S = \frac{1}{n} \sum_{i \in S} Y_{1i} - Y_{0i}$$

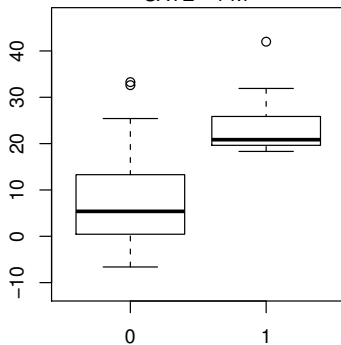
- ▶ The SATE is a random quantity whose distribution is characterized by the sampling design.
- ▶ For any given sample, the SATE may not equal the PATE, introducing one **source of variation** for which we may want to account.

**Potential outcomes
w/ covariate**



Diffs. in pot. outcomes

SATE = 14.7



Simple randomized experiment

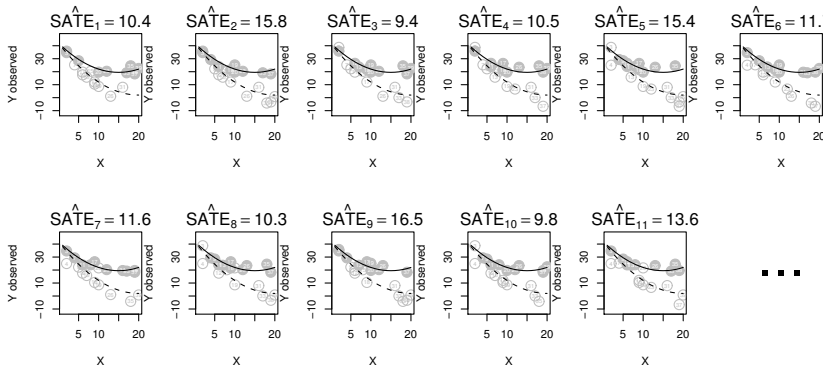
- ▶ We randomly assign $1 < n_1 < n - 1$ units to treatment ($D_i = 1$), in which case $n_0 = n - n_1$ are assigned to control ($D_i = 0$).
- ▶ We observe $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$.

Simple randomized experiment

- ▶ We randomly assign $1 < n_1 < n - 1$ units to treatment ($D_i = 1$), in which case $n_0 = n - n_1$ are assigned to control ($D_i = 0$).
- ▶ We observe $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$.
- ▶ Consider an intuitive **estimator** for the SATE,

$$S\hat{ATE} = \frac{1}{n_1} \sum_{i:D_i=1} Y_i - \frac{1}{n_0} \sum_{i:D_i=0} Y_i = \bar{Y}_1 - \bar{Y}_0.$$

- ▶ For any given experiment, this quantity will not equal SATE exactly. Assignment presents another **source of variation** for which we may want to account.



Simple randomized experiment

- ▶ By random assignment,

$$\begin{aligned} E_D[S\hat{ATE}|S] &= \frac{1}{n_1}n_1E_D[Y_{1i}|S] - \frac{1}{n_0}n_0E_D[Y_{0i}|S] \\ &= \sum_{y_1 \in \{Y_{1i}:i \in S\}} y_1 \Pr[Y_{1i} = y_1] - \sum_{y_0 \in \{Y_{0i}:i \in S\}} y_0 \Pr[Y_{0i} = y_0] \\ &= \frac{1}{n} \sum_{i \in S} Y_{1i} - Y_{0i} = SATE, \end{aligned}$$

implying **unbiasedness** with respect to the SATE.

Simple randomized experiment

- ▶ By random assignment,

$$\begin{aligned} E_D[S\hat{ATE}|S] &= \frac{1}{n_1}n_1E_D[Y_{1i}|S] - \frac{1}{n_0}n_0E_D[Y_{0i}|S] \\ &= \sum_{y_1 \in \{Y_{1i}:i \in S\}} y_1 \Pr[Y_{1i} = y_1] - \sum_{y_0 \in \{Y_{0i}:i \in S\}} y_0 \Pr[Y_{0i} = y_0] \\ &= \frac{1}{n} \sum_{i \in S} Y_{1i} - Y_{0i} = SATE, \end{aligned}$$

implying **unbiasedness** with respect to the SATE.

- ▶ Over samples, similar calculations show,

$$E[E_D[S\hat{ATE}|S]] = E[SATE] = PATE,$$

in which case $S\hat{ATE}$ is also unbiased for the PATE.

Simple randomized experiment

- ▶ By standard sample theoretic results and a bit of algebra,

$$\begin{aligned}\text{Var}_D[\widehat{SATE}|S] &= \text{Var}_D[\bar{Y}_1] + \text{Var}_D[\bar{Y}_0] - 2\text{Cov}_D[\bar{Y}_1, \bar{Y}_0] \\ &= \frac{s_{Y_1}^2}{n_1} \left(\frac{n-n_1}{n} \right) + \frac{s_{Y_0}^2}{n_0} \left(\frac{n-n_0}{n} \right) - 2 \left[-\frac{s_{Y_1, Y_0}}{n} \right] \\ &= \frac{s_{Y_1}^2}{n_1} + \frac{s_{Y_0}^2}{n_0} - \underbrace{\frac{s_{Y_1}^2 + s_{Y_0}^2 - 2s_{Y_1, Y_0}}{n}}_{\text{numerator is } \text{Var}[Y_1 - Y_0] = \text{Var}[\rho_i]} \\ &= \frac{s_{Y_1}^2}{n_1} + \frac{s_{Y_0}^2}{n_0} - \frac{s_{\rho}^2}{n},\end{aligned}$$

(s_W^2 is sample var. for W_i , and $s_{W,V}$ is sample cov. for W_i, V_i).

Simple randomized experiment

- ▶ By standard sample theoretic results and a bit of algebra,

$$\begin{aligned}\text{Var}_D[\widehat{SATE}|S] &= \text{Var}_D[\bar{Y}_1] + \text{Var}_D[\bar{Y}_0] - 2\text{Cov}_D[\bar{Y}_1, \bar{Y}_0] \\ &= \frac{s_{Y_1}^2}{n_1} \left(\frac{n-n_1}{n} \right) + \frac{s_{Y_0}^2}{n_0} \left(\frac{n-n_0}{n} \right) - 2 \left[-\frac{s_{Y_1, Y_0}}{n} \right] \\ &= \frac{s_{Y_1}^2}{n_1} + \frac{s_{Y_0}^2}{n_0} - \underbrace{\frac{s_{Y_1}^2 + s_{Y_0}^2 - 2s_{Y_1, Y_0}}{n}}_{\text{numerator is } \text{Var}[Y_1 - Y_0] = \text{Var}[\rho_i]} \\ &= \frac{s_{Y_1}^2}{n_1} + \frac{s_{Y_0}^2}{n_0} - \frac{s_{\rho}^2}{n},\end{aligned}$$

(s_W^2 is sample var. for W_i , and $s_{W,V}$ is sample cov. for W_i, V_i).

- ▶ The first two terms are identified, however the third term is not.
- ▶ Ignoring third term will overestimate randomization variance.
- ▶ Largest when ρ_i is constant over i or, equivalently, when (Y_{1i}, Y_{0i}) are perfectly correlated (linear shift).
(perfect correlation means high treated values implies low control values.)

Simple randomized experiment

- ▶ The variance goes to zero in n_1, n_0 .
- ▶ As such, the distribution of $S\hat{A}TE$ zeroes in on SATE as n_1, n_0 become large.
- ▶ Because of this, we say that “ $S\hat{A}TE$ is **consistent** for SATE in n_1, n_0 .”

Simple randomized experiment

- ▶ The variance goes to zero in n_1, n_0 .
- ▶ As such, the distribution of \widehat{SATE} zeroes in on SATE as n_1, n_0 become large.
- ▶ Because of this, we say that “ \widehat{SATE} is **consistent** for SATE in n_1, n_0 .”
- ▶ Unbiasedness and consistency are two distinct properties:
 - ▶ Both are desirable.
 - ▶ Consistency is often more important in applied settings, e.g. in cases where a consistent but biased estimator has lower variance than any unbiased estimators. (An example is multiple regression with experimental data.)

Back to the population

- ▶ Over samples, the ANOVA theorem yields (MHE, p. 33),

$$\begin{aligned}\text{Var}[\widehat{SATE}] &= \text{E}[\text{Var}_D[\widehat{SATE}|S]] + \text{Var}[\text{E}_D[\widehat{SATE}|S]] \\ &= \text{E}\left[\frac{s_{Y_1}^2}{n_1} + \frac{s_{Y_0}^2}{n_0} - \frac{s_\rho^2}{n}\right] + \frac{\sigma_\rho^2}{n} \\ &= \frac{\sigma_{Y_1}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_0} - \frac{\sigma_\rho^2}{n} + \frac{\sigma_\rho^2}{n}. \\ &= \frac{\sigma_{Y_1}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_0}.\end{aligned}$$

Back to the population

- ▶ Over samples, the ANOVA theorem yields (MHE, p. 33),

$$\begin{aligned}\text{Var}[\widehat{SATE}] &= \text{E}[\text{Var}_D[\widehat{SATE}|S]] + \text{Var}[\text{E}_D[\widehat{SATE}|S]] \\ &= \text{E}\left[\frac{s_{Y_1}^2}{n_1} + \frac{s_{Y_0}^2}{n_0} - \frac{s_\rho^2}{n}\right] + \frac{\sigma_\rho^2}{n} \\ &= \frac{\sigma_{Y_1}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_0} - \frac{\sigma_\rho^2}{n} + \frac{\sigma_\rho^2}{n}. \\ &= \frac{\sigma_{Y_1}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_0}.\end{aligned}$$

- ▶ Shows that \widehat{SATE} is also consistent for PATE in n_1, n_0 .

Back to the population

- ▶ Over samples, the ANOVA theorem yields (MHE, p. 33),

$$\begin{aligned}\text{Var}[\widehat{SATE}] &= \text{E}[\text{Var}_D[\widehat{SATE}|S]] + \text{Var}[\text{E}_D[\widehat{SATE}|S]] \\ &= \text{E}\left[\frac{s_{Y_1}^2}{n_1} + \frac{s_{Y_0}^2}{n_0} - \frac{s_\rho^2}{n}\right] + \frac{\sigma_\rho^2}{n} \\ &= \frac{\sigma_{Y_1}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_0} - \frac{\sigma_\rho^2}{n} + \frac{\sigma_\rho^2}{n}. \\ &= \frac{\sigma_{Y_1}^2}{n_1} + \frac{\sigma_{Y_0}^2}{n_0}.\end{aligned}$$

- ▶ Shows that \widehat{SATE} is also consistent for PATE in n_1, n_0 .
- ▶ By sample theoretic results, conventional sample variance estimators are unbiased for these population variances, so

$$\widehat{V} = \frac{\widehat{s}_{Y_1}^2}{n_1} + \frac{\widehat{s}_{Y_0}^2}{n_0},$$

is unbiased for $\text{Var}[\widehat{SATE}]$ (different than $\text{Var}_D[\widehat{SATE}|S]$).

Inference for SATE or PATE

The quantity \hat{V} thus offers two interpretations when we have a randomized experiment with a random sample from the population:

Inference for SATE or PATE

The quantity \hat{V} thus offers two interpretations when we have a randomized experiment with a random sample from the population:

- ▶ A conservative approximation of the **randomization distribution variance**, thus providing a way to make inferences about $\bar{Y}_1 - \bar{Y}_0$ relative to SATE.

Inference for SATE or PATE

The quantity \hat{V} thus offers two interpretations when we have a randomized experiment with a random sample from the population:

- ▶ A conservative approximation of the **randomization distribution variance**, thus providing a way to make inferences about $\bar{Y}_1 - \bar{Y}_0$ relative to SATE.
- ▶ An unbiased estimator for the **sampling distribution plus randomization distribution variance**, thus providing a way to make inferences about $\bar{Y}_1 - \bar{Y}_0$ relative to PATE.

Inference for SATE or PATE: intervals

- ▶ Freedman (2008a) Theorem 1 shows that by a central limit theorem for non-independent random variables, $\bar{Y}_1 - \bar{Y}_0$ is asymptotically normal under well-behaved higher order moments.
- ▶ Thus, for large n_0, n_1 , $(1 - \alpha)100\%$ confidence intervals for SATE or PATE (under interpretations above) are given by,

$$\left((\bar{Y}_1 - \bar{Y}_0) - z_{\alpha/2} \sqrt{\hat{V}}, \quad (\bar{Y}_1 - \bar{Y}_0) + z_{\alpha/2} \sqrt{\hat{V}} \right).$$

Inference for SATE or PATE: intervals

- ▶ Freedman (2008a) Theorem 1 shows that by a central limit theorem for non-independent random variables, $\bar{Y}_1 - \bar{Y}_0$ is asymptotically normal under well-behaved higher order moments.
- ▶ Thus, for large n_0, n_1 , $(1 - \alpha)100\%$ confidence intervals for SATE or PATE (under interpretations above) are given by,

$$\left((\bar{Y}_1 - \bar{Y}_0) - z_{\alpha/2} \sqrt{\hat{V}}, \quad (\bar{Y}_1 - \bar{Y}_0) + z_{\alpha/2} \sqrt{\hat{V}} \right).$$

- ▶ This interval yields an asymptotic coverage rate of $(1 - \alpha)100\%$.

Inference for SATE or PATE: intervals

- ▶ Freedman (2008a) Theorem 1 shows that by a central limit theorem for non-independent random variables, $\bar{Y}_1 - \bar{Y}_0$ is asymptotically normal under well-behaved higher order moments.
- ▶ Thus, for large n_0, n_1 , $(1 - \alpha)100\%$ confidence intervals for SATE or PATE (under interpretations above) are given by,

$$\left((\bar{Y}_1 - \bar{Y}_0) - z_{\alpha/2} \sqrt{\hat{V}}, \quad (\bar{Y}_1 - \bar{Y}_0) + z_{\alpha/2} \sqrt{\hat{V}} \right).$$

- ▶ This interval yields an asymptotic coverage rate of $(1 - \alpha)100\%$.
- ▶ For not-so-large n_0, n_1 , the coverage rate may be improved by using critical values from the t -distribution, which has “fatter tails.” The t approximation is motivated by the exact finite sample distribution of normally distributed outcomes.

Inference for SATE or PATE: testing

- ▶ We first consider testing under the Neyman-Pearson framework, which aims to control error rates in a binary decision problem.
- ▶ The “average null” (or “weak null”) hypothesis stipulates,

$$H_0^{av} : \rho = 0 \quad \text{versus} \quad H_a^{av} : \rho \neq 0 \text{ (two-sided).}$$

(Could also state in terms of SATE.)

Inference for SATE or PATE: testing

- ▶ We first consider testing under the Neyman-Pearson framework, which aims to control error rates in a binary decision problem.
- ▶ The “average null” (or “weak null”) hypothesis stipulates,

$$H_0^{av} : \rho = 0 \quad \text{versus} \quad H_a^{av} : \rho \neq 0 \text{ (two-sided).}$$

(Could also state in terms of SATE.)

- ▶ Under H_0^{av} , the t-statistic ($t = \text{“test”}$),

$$t = \frac{\bar{Y}_1 - \bar{Y}_0}{\sqrt{\hat{V}}}$$

is distributed approximately $N(0, 1)$ (or, under the t -distribution refinement, t with $n - 2$ degrees of freedom). If t exceeds the relevant critical value ($t_{\alpha/2}$ for a two-sided test), we reject H_0^{av} .

Inference for SATE or PATE: testing

- ▶ We first consider testing under the Neyman-Pearson framework, which aims to control error rates in a binary decision problem.
- ▶ The “average null” (or “weak null”) hypothesis stipulates,

$$H_0^{av} : \rho = 0 \quad \text{versus} \quad H_a^{av} : \rho \neq 0 \text{ (two-sided).}$$

(Could also state in terms of SATE.)

- ▶ Under H_0^{av} , the t-statistic ($t = \text{“test”}$),

$$t = \frac{\bar{Y}_1 - \bar{Y}_0}{\sqrt{\hat{V}}}$$

is distributed approximately $N(0, 1)$ (or, under the t -distribution refinement, t with $n - 2$ degrees of freedom). If t exceeds the relevant critical value ($t_{\alpha/2}$ for a two-sided test), we reject H_0^{av} .

- ▶ The asymptotic type I error rate is α (if the null is true, you nonetheless reject $\alpha\%$ of the time).
- ▶ (The type II error rate is $1 - \text{power}$, which depends on the standing of H_a .)
- ▶ The t distribution is used to improve error rates in finite samples.

Inference for SATE or PATE: testing

We may also wish to test the “sharp null” hypothesis,

$$H_0^{sh} : Y_{1i} = Y_{0i} \text{ for all } i \in S \quad \text{versus} \quad H_a^{sh} : Y_{1i} = Y_{0i} \text{ for some } i \in S.$$

This hypothesis is a departure from the estimation and inference framework examined thus far, and gets us into the world of Fisher exact tests (next slide).

Note that $H_0^{sh} \Rightarrow H_0^{av}$, and so rejection of H_0^{av} implies rejection of H_0^{sh} (but not the other way around).

An aside on exact tests

- ▶ If H_0^{sh} is true, we actually know all potential outcomes!
- ▶ Thus, for those with $D_i = 1$, we impute $Y_{0i} = Y_i$, and for those with $D_i = 0$, we impute $Y_{1i} = Y_i$.

An aside on exact tests

- ▶ If H_0^{sh} is true, we actually know all potential outcomes!
- ▶ Thus, for those with $D_i = 1$, we impute $Y_{0i} = Y_i$, and for those with $D_i = 0$, we impute $Y_{1i} = Y_i$.
- ▶ We can examine how “unusual” is our estimate, $\bar{Y}_1 - \bar{Y}_0$ relative to all the possible values it could take under H_0^{sh} .
- ▶ With our imputed potential outcomes, we compute $\bar{Y}_1 - \bar{Y}_0$ for all of the $\binom{n}{n_1}$ treatment assignment possibilities, and see what proportion of those estimates are larger than what we observed.
- ▶ This provides an exact one-sided p -value for H_0^{sh} . That is, if we reject against an α threshold, in *finite samples*, the exact type I error rate is α .

An aside on exact tests

- ▶ If H_0^{sh} is true, we actually know all potential outcomes!
- ▶ Thus, for those with $D_i = 1$, we impute $Y_{0i} = Y_i$, and for those with $D_i = 0$, we impute $Y_{1i} = Y_i$.
- ▶ We can examine how “unusual” is our estimate, $\bar{Y}_1 - \bar{Y}_0$ relative to all the possible values it could take under H_0^{sh} .
- ▶ With our imputed potential outcomes, we compute $\bar{Y}_1 - \bar{Y}_0$ for all of the $\binom{n}{n_1}$ treatment assignment possibilities, and see what proportion of those estimates are larger than what we observed.
- ▶ This provides an exact one-sided p -value for H_0^{sh} . That is, if we reject against an α threshold, in *finite samples*, the exact type I error rate is α .
- ▶ Note to test H_0^{sh} we don't have to use $\bar{Y}_1 - \bar{Y}_0$. Imbens and Rubin (CIS, Ch. 5) show that the difference in average *ranks* weakly dominates $\bar{Y}_1 - \bar{Y}_0$ in terms of power.
- ▶ This approach to testing is due to Fisher (1935).
- ▶ “Inverted test” intervals can be constructed (Rosenbaum, 2002).

Covariates and efficiency

- ▶ In our examination of estimation and inference concepts thus far, we have not made any use of our **covariates**.

Covariates and efficiency

- ▶ In our examination of estimation and inference concepts thus far, we have not made any use of our **covariates**.
- ▶ Because of random assignment and random sampling, there has been no need to use covariates to obtain unbiased and consistent estimates.

Covariates and efficiency

- ▶ In our examination of estimation and inference concepts thus far, we have not made any use of our **covariates**.
- ▶ Because of random assignment and random sampling, there has been no need to use covariates to obtain unbiased and consistent estimates.
- ▶ However, we can use covariate information to improve **efficiency** (that is, reduce the randomization and sampling distribution variance of our estimate of ρ) while maintaining consistency (though not unbiasedness).

Covariates and efficiency

Consider the interacted regression, allowing X_i to be a vector,

$$Y_i = \alpha + \rho_{reg}D_i + X_i'\beta_0 + D_i(X_i - \bar{X}_D)'\beta_1 + \varepsilon_i$$

where \bar{X}_D is covariate means for the treatment groups.

Covariates and efficiency

Consider the interacted regression, allowing X_i to be a vector,

$$Y_i = \alpha + \rho_{reg}D_i + X_i'\beta_0 + D_i(X_i - \bar{X}_D)'\beta_1 + \varepsilon_i$$

where \bar{X}_D is covariate means for the treatment groups.

- ▶ OLS estimation yields $\hat{\rho}_{reg} = \bar{Y}_{1adj} - \bar{Y}_{0adj}$, where

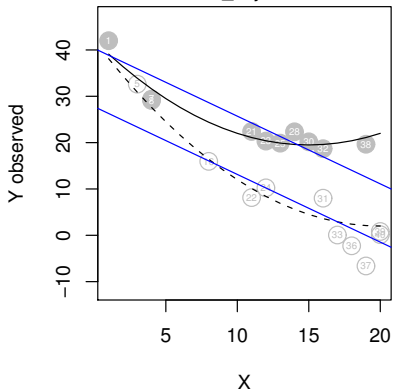
$$\bar{Y}_{1adj} = \bar{Y}_1 + (\bar{X} - \bar{X}_1)'(\hat{\beta}_0 + \hat{\beta}_1)$$

$$\bar{Y}_{0adj} = \bar{Y}_0 + (\bar{X} - \bar{X}_0)'\hat{\beta}_0,$$

- ▶ If we exclude the interaction, we would set $\beta_1 = 0$
- ▶ No presumption that the regression specification is “correct.”

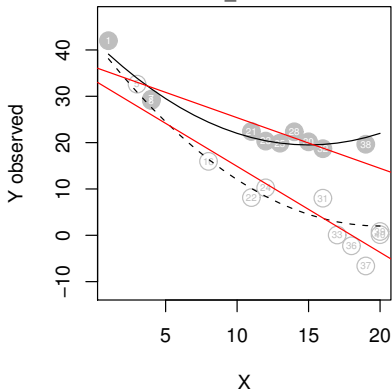
Simple covariance adj.

$\hat{SATE}_{adj} = 12.6$



Interacted covariance adj.

$\hat{SATE}_{int} = 12.5$



Covariates and efficiency

$$Y_i = \alpha + \rho_{reg}D_i + X_i'\beta_0 + D_i(X_i - \bar{X}_D)'\beta_1 + \varepsilon_i$$

- ▶ $\hat{\rho}_{reg}$ is biased but consistent for SATE.
- ▶ $\hat{\rho}_{reg}$ has lower asymptotic variance than $\hat{\rho} = \bar{Y}_1 - \bar{Y}_0$ for PATE.
- ▶ As such, $\hat{\rho}_{reg}$ is **consistent and more efficient** than $\hat{\rho}$, albeit biased. The bias diminishes quickly in sample size ($O(1/n)$).
- ▶ Simple adjustment (i.e., setting $\beta_1 = 0$) improves efficiency if experiment is not strongly imbalanced ($\min[n_1/n, n_0/n] < .25$) and $\text{Cov}(X_i, Y_i)$ is large relative to $\text{Cov}(X_i, Y_i(1) - Y_i(0))$ (Lin, 2013).

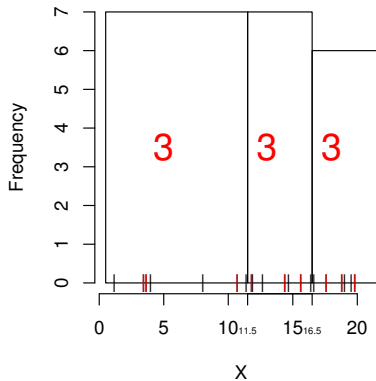
Covariates and efficiency

- ▶ Regression in previous slides motivated by efficiency.
- ▶ It may be used to address “incidental confounds,” with consistency following from usual regression assumptions.

Covariates and efficiency

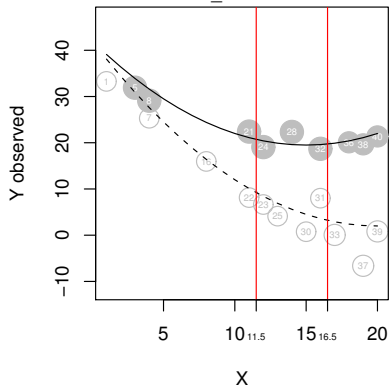
- ▶ A **design-based approach** (i.e., ex ante approach) to boost efficiency and address covariate imbalance is to incorporate covariates into the randomization.
- ▶ One approach is “block randomization.”
- ▶ Observations are divided into blocks, $b = 1, \dots, B$, typically by prognostic covariates.
- ▶ Random assignment occurs within blocks.

Covariate histogram



Outcomes w/ block wghts.

$\hat{SATE}_{block} = 14.5$



Covariates and efficiency

- ▶ By principles of stratified sampling, an unbiased estimator for SATE is, $\widehat{SATE}_{block} = \sum_b (N_b/N) \widehat{SATE}_b$.
- ▶ \widehat{SATE}_{block} has lower variance than \widehat{SATE} if outcome variation is reduced within strata.
- ▶ \widehat{SATE}_{block} is algebraically equivalent to coefficient from regression with block FEs and inverse propensity score weights.

Conclusion

- ▶ The analysis here is “design-based”: we evaluated bias, consistency, coverage rates, and error rates with reference to the *sampling and randomization* distributions.

Conclusion

- ▶ The analysis here is “design-based”: we evaluated bias, consistency, coverage rates, and error rates with reference to the *sampling and randomization* distributions.
- ▶ For the idealized experiment, these distributions and associated error rates are *created* by the research design.
- ▶ They are under the *researcher's control*.
- ▶ It is in this sense that design-based inference for experiments is a “principled basis for inference” (Fisher 1935).
- ▶ For observational studies, we can apply this framework in an “as-if” sense (Rubin 2008; Rosenbaum 2002), although it is *nature*, not the researcher, who controls the sampling, randomization, and therefore error rates.