Regression with Randomized Experiments

Ryan T. Moore

2024-08-21

The Linear Regression Estimator

Estimating the ATE

In a policy experiment in India, Chattopadhyay and Duflo (2004) randomly assign reserved village council headship for women to villages.

Estimating the ATE

In a policy experiment in India, Chattopadhyay and Duflo (2004) randomly assign reserved village council headship for women to villages.

```r
library(readr)
seats <- read_csv("http://j.mp/2YfZdgv")
```

Estimating the ATE

In a policy experiment in India, Chattopadhyay and Duflo (2004) randomly assign reserved village council headship for women to villages.

```r
library(readr)
seats <- read_csv("http://j.mp/2YfZdgv")
```

```r
head(seats)
```

```
## # A tibble: 6 x 6
##      GP village reserved female irrigation water
##   <dbl>   <dbl>    <dbl>  <dbl>      <dbl> <dbl>
## 1     1       2        1      1          0    10
## 2     1       1        1      1          5     0
## 3     2       2        1      1          2     2
## 4     2       1        1      1          4    31
## 5     3       2        0      0          0     0
## 6     3       1        0      0          0     0
```

Estimating the ATE

Explore assignment:

```r
table(seats$reserved)
```

```
##
##   0   1
## 214 108
```

Estimating the ATE

Estimate treatment effect with difference in means:

```r
mean(seats$water[seats$reserved == 1]) -
  mean(seats$water[seats$reserved == 0])
```

```
## [1] 9.252423
```

Estimating the ATE

Estimate treatment effect with difference in means:

```
mean(seats$water[seats$reserved == 1]) -
  mean(seats$water[seats$reserved == 0])
```

```
## [1] 9.252423
```

In areas with reservations, about 9.3 more drinking water projects were undertaken.

With binary treatment as predictor, we can recover same result using least squares regression.

With binary treatment as predictor, we can recover same result using least squares regression.

Model:

$$\text{Water Projects}_i = \beta_0 + \beta_1 \left(\text{Women's Seats}_i\right) + \epsilon_i$$

With binary treatment as predictor, we can recover same result using least squares regression.

Model:

$$\text{Water Projects}_i = \beta_0 + \beta_1 \left(\text{Women's Seats}_i\right) + \epsilon_i$$

Estimate coefficients

```r
lm_out <- lm(water ~ reserved, data = seats)
lm_out
```

```
##
## Call:
## lm(formula = water ~ reserved, data = seats)
##
## Coefficients:
## (Intercept)     reserved
##      14.738        9.252
```

Estimated model

$$\widehat{\text{Water Projects}}_i = 14.74 + 9.25\,(\text{Women's Seats}_i)$$

Estimated model

$$\widehat{\text{Water Projects}}_i = 14.74 + 9.25\,(\text{Women's Seats}_i)$$

- Intercept: predicted value when Women's Seats are **not** reserved (`reserved == 0`)

Estimated model

$$\widehat{\text{Water Projects}}_i = 14.74 + 9.25\,(\text{Women's Seats}_i)$$

- Intercept: predicted value when Women's Seats are **not** reserved (`reserved == 0`)
- Slope: estimated average treatment effect (difference between `reserved == 0` and `reserved == 1`).

Estimated model

$$\widehat{\text{Water Projects}}_i = 14.74 + 9.25\,(\text{Women's Seats}_i)$$

- Intercept: predicted value when Women's Seats are **not** reserved (`reserved == 0`)
- Slope: estimated average treatment effect (difference between `reserved == 0` and `reserved == 1`).

Estimated model

$$\widehat{\text{Water Projects}}_i = 14.74 + 9.25\,(\text{Women's Seats}_i)$$

- Intercept: predicted value when Women's Seats are **not** reserved (`reserved == 0`)
- Slope: estimated average treatment effect (difference between `reserved == 0` and `reserved == 1`).

Causal interpretation of coefficient comes from the science of the policy experiment, **not** from regression.

Social Pressure Mail Experiment

Gerber, Green, and Larimer (2008): Social pressure mailer experiment; households randomly receive one of four messages

Social Pressure Mail Experiment

Gerber, Green, and Larimer (2008): Social pressure mailer experiment; households randomly receive one of four messages

Dear Registered Voter:

DO YOUR CIVIC DUTY AND VOTE!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

The whole point of democracy is that citizens are active participants in government; that we have a voice in government. Your voice starts with your vote. On August 8, remember your rights and responsibilities as a citizen. Remember to vote.

DO YOUR CIVIC DUTY — VOTE!

Social Pressure Mail Experiment

Gerber, Green, and Larimer (2008): Social pressure mailer experiment; households randomly receive one of four messages

Dear Registered Voter:

YOU ARE BEING STUDIED!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

This year, we're trying to figure out why people do or do not vote. We'll be studying voter turnout in the August 8 primary election.

Our analysis will be based on public records, so you will not be contacted again or disturbed in any way. Anything we learn about your voting or not voting will remain confidential and will not be disclosed to anyone else.

DO YOUR CIVIC DUTY — VOTE!

Social Pressure Mail Experiment

Gerber, Green, and Larimer (2008): Social pressure mailer experiment; households randomly receive one of four messages

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	
9995 JENNIFER KAY SMITH		Voted	
9997 RICHARD B JACKSON		Voted	
9999 KATHY MARIE JACKSON		Voted	
9999 BRIAN JOSEPH JACKSON		Voted	
9991 JENNIFER KAY THOMPSON		Voted	
9991 BOB R THOMPSON		Voted	
9993 BILL S SMITH			
9989 WILLIAM LUKE CASPER		Voted	
9989 JENNIFER SUE CASPER		Voted	
9987 MARIA S JOHNSON	Voted	Voted	
9987 TOM JACK JOHNSON	Voted	Voted	
9987 RICHARD TOM JOHNSON		Voted	

Multiple Regression: Many Predictors

- Gerber, Green, and Larimer (2008): Social pressure mailing experiment; households randomly receive one of four messages

Multiple Regression: Many Predictors

- Gerber, Green, and Larimer (2008): Social pressure mailing experiment; households randomly receive one of four messages
- *Four* predictors of turnout: assigned mailing with control, civic duty, Hawthorne, or social pressure message

Multiple Regression: Many Predictors

- Gerber, Green, and Larimer (2008): Social pressure mailing experiment; households randomly receive one of four messages
- *Four* predictors of turnout: assigned mailing with control, civic duty, Hawthorne, or social pressure message
- Use all four to predict turnout in 2006 primary

Multiple Regression: Many Predictors

- Gerber, Green, and Larimer (2008): Social pressure mailing experiment; households randomly receive one of four messages
- *Four* predictors of turnout: assigned mailing with control, civic duty, Hawthorne, or social pressure message
- Use all four to predict turnout in 2006 primary

Multiple Regression: Many Predictors

- Gerber, Green, and Larimer (2008): Social pressure mailing experiment; households randomly receive one of four messages
- *Four* predictors of turnout: assigned mailing with control, civic duty, Hawthorne, or social pressure message
- Use all four to predict turnout in 2006 primary

Model:

$$\begin{aligned} \text{Turnout 2006}_i &= \beta_0 + \beta_1 \cdot \text{Control}_i + \\ &\quad \beta_2 \cdot \text{Hawthorne}_i + \beta_3 \cdot \text{Neighbors}_i + \epsilon_i \end{aligned}$$

```r
social <- read_csv("http://j.mp/2YenEuU")

lm_out <- lm(primary2006 ~ messages, data = social)
```

```r
social <- read_csv("http://j.mp/2YenEuU")

lm_out <- lm(primary2006 ~ messages, data = social)
```

```
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.315      0.002 132.909    0.000
## messagesControl       -0.018      0.003  -6.905    0.000
## messagesHawthorne      0.008      0.003   2.341    0.019
## messagesNeighbors      0.063      0.003  18.944    0.000
```

Estimated model

$$\begin{aligned}\widehat{\text{Turnout 2006}}_i &= 0.31 + -0.02\,(\text{Control}_i) + 0.01\,(\text{Hawthorne}_i) + \\ &\quad 0.06\,(\text{Neighbors}_i)\end{aligned}$$

Estimated model

$$\widehat{\text{Turnout 2006}}_i = 0.31 + -0.02\,(\text{Control}_i) + 0.01\,(\text{Hawthorne}_i) + 0.06\,(\text{Neighbors}_i)$$

1. What is the predicted turnout probability under `Control`?
2. What is the predicted turnout probability under `Hawthorne`?
3. What is the predicted turnout probability under `Neighbors`?
4. What is the predicted turnout probability under `Civic Duty`?

Alternatively, estimate

$$\begin{aligned}\text{Turnout } 06_i &= \beta_0 \cdot \text{CivicDuty}_i + \beta_1 \cdot \text{Contrl}_i + \beta_2 \cdot \text{Hawthorne}_i + \\ &\quad \beta_3 \cdot \text{Neighbors}_i + \epsilon_i\end{aligned}$$

```r
lm_out <- lm(primary2006 ~ -1 + messages, data = social)
summary(lm_out)$coefficients
```

```
##                      Estimate  Std. Error  t value Pr(>|t
## messagesCivic Duty 0.3145377 0.002366570 132.9087
## messagesControl    0.2966383 0.001057939 280.3927
## messagesHawthorne  0.3223746 0.002367004 136.1952
## messagesNeighbors  0.3779482 0.002367097 159.6674
```

Estimated model:

$$\begin{aligned}\widehat{\text{Turnout 2006}}_i &= 0.31\,(\text{Civic Duty}_i) + 0.3\,(\text{Control}_i) + \\ &\quad 0.32\,(\text{Hawthorne}_i) + 0.38\,(\text{Neighbors}_i)\end{aligned}$$

Estimated model:

$$\begin{aligned}\widehat{\text{Turnout 2006}}_i &= 0.31\,(\text{Civic Duty}_i) + 0.3\,(\text{Control}_i) + \\ &\quad 0.32\,(\text{Hawthorne}_i) + 0.38\,(\text{Neighbors}_i)\end{aligned}$$

1. What is the predicted turnout probability under `Control`?
2. What is the predicted turnout probability under `Hawthorne`?
3. What is the predicted turnout probability under `Neighbors`?
4. What is the predicted turnout probability under `Civic Duty`?

Models with Covariates

Why add covariates to model?

We consider adding covariates to our linear model to

- adjust for chance differences

Why add covariates to model?

We consider adding covariates to our linear model to

- adjust for chance differences
    - reduce estimation error

Why add covariates to model?

We consider adding covariates to our linear model to

- adjust for chance differences
  - reduce estimation error
- improve precision of $\hat{\beta}_1$

Why add covariates to model?

We consider adding covariates to our linear model to

- adjust for chance differences
  - reduce estimation error
- improve precision of $\hat{\beta}_1$
- estimate effects for subgroups

Why add covariates to model?

We consider adding covariates to our linear model to

- adjust for chance differences
  - reduce estimation error
- improve precision of $\hat{\beta}_1$
- estimate effects for subgroups
- account for design features

Why add covariates to model?

We consider adding covariates to our linear model to

- adjust for chance differences
    - reduce estimation error
- improve precision of $\hat{\beta}_1$
- estimate effects for subgroups
- account for design features
- improve predictions of outcome (beyond causal inf)

Why add covariates to model?

We consider adding covariates to our linear model to

- adjust for chance differences
  - reduce estimation error
- improve precision of $\hat{\beta}_1$
- estimate effects for subgroups
- account for design features
- improve predictions of outcome (beyond causal inf)

Why add covariates to model?

We consider adding covariates to our linear model to

- adjust for chance differences
  - reduce estimation error
- improve precision of $\hat{\beta}_1$
- estimate effects for subgroups
- account for design features
- improve predictions of outcome (beyond causal inf)

- (not to reduce bias in estimator, if simple design)

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$ (diff in means, with indep from randomization)

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$ (diff in means, with indep from randomization)

Does adding covariates introduce bias?

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$ (diff in means, with indep from randomization)

Does adding covariates introduce bias?

- Freedman (2008): it can.

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$ (diff in means, with indep from randomization)

Does adding covariates introduce bias?

- Freedman (2008): it can.
- Green and Aronow (2011): Actually, not in practice.

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$ (diff in means, with indep from randomization)

Does adding covariates introduce bias?

- Freedman (2008): it can.
- Green and Aronow (2011): Actually, not in practice.

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$ (diff in means, with indep from randomization)

Does adding covariates introduce bias?

- Freedman (2008): it can.
- Green and Aronow (2011): Actually, not in practice. And, small relative to $SE(\widehat{TE})$.

Does adjusting for covariates risk bias?

$$Y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

- The bivariate regression $\hat{\beta}_1$ is unbiased for the ATE
- $E(\hat{\beta}_1) = \beta_1$ (diff in means, with indep from randomization)

Does adding covariates introduce bias?

- Freedman (2008): it can.
- Green and Aronow (2011): Actually, not in practice. And, small relative to $SE(\widehat{TE})$.
- "Included variable bias" (see https://t.ly/ttYpD)

Models with Covariates

- May want to adjust for or improve predictions based on several covariates

Models with Covariates

- May want to adjust for or improve predictions based on several covariates
- Covariates are *pre-treatment*!

Models with Covariates

- May want to adjust for or improve predictions based on several covariates
- Covariates are *pre-treatment*!
- E.g., experiment not quite balanced on age, prior turnout; adjust diff-in-means.

Models with Covariates

- May want to adjust for or improve predictions based on several covariates
- Covariates are *pre-treatment*!
- E.g., experiment not quite balanced on age, prior turnout; adjust diff-in-means.
- E.g., make better turnout predictions by including registrant's age and prior turnout, in addition to whether received `Neighbors` postcard

Models with Covariates

- May want to adjust for or improve predictions based on several covariates
- Covariates are *pre-treatment*!
- E.g., experiment not quite balanced on age, prior turnout; adjust diff-in-means.
- E.g., make better turnout predictions by including registrant's age and prior turnout, in addition to whether received `Neighbors` postcard
- Age and prior turnout are *causally prior* to treatment

Models with Covariates

- May want to adjust for or improve predictions based on several covariates
- Covariates are *pre-treatment*!
- E.g., experiment not quite balanced on age, prior turnout; adjust diff-in-means.
- E.g., make better turnout predictions by including registrant's age and prior turnout, in addition to whether received `Neighbors` postcard
- Age and prior turnout are *causally prior* to treatment
- Let's just compare `Neighbors` to `Control`.

Models with Covariates

```r
# Subset to two treatment conditions:
social.neighbor <- subset(social,
                          (messages == "Control") |
                            (messages == "Neighbors"))

# Calculate age:
social.neighbor <- social.neighbor |>
  mutate(age = 2006 - yearofbirth)
```

```r
# Multiple regression:
lm_out <- lm(primary2006 ~ age + primary2004 +
               messages, data = social.neighbor)

summary(lm_out)$coefficients |> round(3)
```

```
##                    Estimate Std. Error t value Pr(>|t
## (Intercept)           0.046      0.003  13.274
## age                   0.004      0.000  58.873
## primary2004           0.147      0.002  76.464
## messagesNeighbors     0.080      0.003  31.670
```

Now, make different prediction depending on registrant's age and prior turnout, not just treatment status.

Now, make different prediction depending on registrant's age and prior turnout, not just treatment status.

Estimated model:

$$\begin{aligned}\widehat{\text{Turnout 2006}}_i \quad &= \quad 0.05 + 0.004\,(\text{Age}_i) + 0.15\,(\text{Turnout 2004}_i) + \\ &\quad\; 0.08\,(\text{Neighbors}_i)\end{aligned}$$

Now, make different prediction depending on registrant's age and prior turnout, not just treatment status.

Estimated model:

$$\begin{aligned} \widehat{\text{Turnout 2006}}_i &= 0.05 + 0.004\,(\text{Age}_i) + 0.15\,(\text{Turnout 2004}_i) + \\ &\quad 0.08\,(\text{Neighbors}_i) \end{aligned}$$

Coefficients: how much 1-unit difference in predictor affects my prediction for turnout in 2006, *assuming the other predictors do not change.* (The *ceteris paribus* assumption.) Sometimes a valid assumption, sometimes not.

If compare two registrants, *ceteris paribus*, ages differ by 10 years, how much would turnout probability prediction differ?

If compare two registrants, *ceteris paribus*, ages differ by 10 years, how much would turnout probability prediction differ?

$$\begin{aligned}\widehat{\text{Turnout 2006}}_1 &= 0.05 + 0.004\,(\text{Age}_1) + 0.15\,(\text{Turnout 2004}_1) + \\ &\quad 0.08\,(\text{Neighbors}_1)\end{aligned}$$

If compare two registrants, *ceteris paribus*, ages differ by 10 years, how much would turnout probability prediction differ?

$$\begin{aligned}
\widehat{\text{Turnout } 2006}_1 &= 0.05 + 0.004\,(\text{Age}_1) + 0.15\,(\text{Turnout } 2004_1) + \\
&\quad 0.08\,(\text{Neighbors}_1) \\
\widehat{\text{Turnout } 2006}_2 &= 0.05 + 0.004\,(\text{Age}_1 + 10) + 0.15\,(\text{Turnout } 2004_1) + \\
&\quad 0.08\,(\text{Neighbors}_1)
\end{aligned}$$

If compare two registrants, *ceteris paribus*, ages differ by 10 years, how much would turnout probability prediction differ?

$$\begin{aligned}
\widehat{\text{Turnout 2006}}_1 &= 0.05 + 0.004\,(\text{Age}_1) + 0.15\,(\text{Turnout 2004}_1) + \\
&\quad 0.08\,(\text{Neighbors}_1) \\
\widehat{\text{Turnout 2006}}_2 &= 0.05 + 0.004\,(\text{Age}_1 + 10) + 0.15\,(\text{Turnout 2004}_1) + \\
&\quad 0.08\,(\text{Neighbors}_1)
\end{aligned}$$

$$\widehat{\text{Turnout 2006}}_2 - \widehat{\text{Turnout 2006}}_1 = 0.004\,(\text{Age}_1 + 10) - 0.004\,(\text{Age}_1)$$

If compare two registrants, *ceteris paribus*, ages differ by 10 years, how much would turnout probability prediction differ?

$$\begin{aligned}
\widehat{\text{Turnout 2006}}_1 &= 0.05 + 0.004\,(\text{Age}_1) + 0.15\,(\text{Turnout 2004}_1) + \\
&\quad 0.08\,(\text{Neighbors}_1) \\
\widehat{\text{Turnout 2006}}_2 &= 0.05 + 0.004\,(\text{Age}_1 + 10) + 0.15\,(\text{Turnout 2004}_1) + \\
&\quad 0.08\,(\text{Neighbors}_1)
\end{aligned}$$

$$\begin{aligned}
\widehat{\text{Turnout 2006}}_2 - \widehat{\text{Turnout 2006}}_1 &= 0.004\,(\text{Age}_1 + 10) - 0.004\,(\text{Age}_1) \\
&= 0.004 \times 10
\end{aligned}$$

If compare two registrants, *ceteris paribus*, ages differ by 10 years, how much would turnout probability prediction differ?

$$\begin{aligned}
\widehat{\text{Turnout 2006}}_1 &= 0.05 + 0.004\,(\text{Age}_1) + 0.15\,(\text{Turnout 2004}_1) + \\
&\quad 0.08\,(\text{Neighbors}_1) \\
\widehat{\text{Turnout 2006}}_2 &= 0.05 + 0.004\,(\text{Age}_1 + 10) + 0.15\,(\text{Turnout 2004}_1) + \\
&\quad 0.08\,(\text{Neighbors}_1)
\end{aligned}$$

$$\begin{aligned}
\widehat{\text{Turnout 2006}}_2 - \widehat{\text{Turnout 2006}}_1 &= 0.004\,(\text{Age}_1 + 10) - 0.004\,(\text{Age}_1) \\
&= 0.004 \times 10 \\
&\approx 0.038
\end{aligned}$$

If compare two registrants, *ceteris paribus*, ages differ by 10 years, how much would turnout probability prediction differ?

$$
\begin{aligned}
\widehat{\text{Turnout 2006}}_1 &= 0.05 + 0.004\,(\text{Age}_1) + 0.15\,(\text{Turnout 2004}_1) + \\
&\quad 0.08\,(\text{Neighbors}_1) \\
\widehat{\text{Turnout 2006}}_2 &= 0.05 + 0.004\,(\text{Age}_1 + 10) + 0.15\,(\text{Turnout 2004}_1) + \\
&\quad 0.08\,(\text{Neighbors}_1)
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\text{Turnout 2006}}_2 - \widehat{\text{Turnout 2006}}_1 &= 0.004\,(\text{Age}_1 + 10) - 0.004\,(\text{Age}_1) \\
&= 0.004 \times 10 \\
&\approx 0.038
\end{aligned}
$$

Not causal.

Heterogeneous Treatment Effects

Heterogeneous Treatment Effects

Often, believe τ_i vary systematically by some X.

Heterogeneous Treatment Effects

Often, believe τ_i vary systematically by some X.

- Reserved seats for women may only produce more water projects in *larger* villages, e.g.

Heterogeneous Treatment Effects

Often, believe τ_i vary systematically by some X.

- Reserved seats for women may only produce more water projects in *larger* villages, e.g.
- Reserved seats may produce *fewer* water projects in small villages

A Test for Heterogeneity

- "Heterogeneity" is $\mathrm{Var}(\tau_i) \neq 0$

A Test for Heterogeneity

- "Heterogeneity" is $\text{Var}(\tau_i) \neq 0$
- If no heterogeneity, then $\text{Var}(Y_i(1)) = \text{Var}(Y_i(0))$

A Test for Heterogeneity

- "Heterogeneity" is $\mathrm{Var}(\tau_i) \neq 0$
- If no heterogeneity, then $\mathrm{Var}(Y_i(1)) = \mathrm{Var}(Y_i(0))$

A Test for Heterogeneity

- "Heterogeneity" is $\mathrm{Var}(\tau_i) \neq 0$
- If no heterogeneity, then $\mathrm{Var}(Y_i(1)) = \mathrm{Var}(Y_i(0)) \rightsquigarrow$ null hypothesis

A Test for Heterogeneity

- "Heterogeneity" is $\text{Var}(\tau_i) \neq 0$
- If no heterogeneity, then $\text{Var}(Y_i(1)) = \text{Var}(Y_i(0)) \rightsquigarrow$ null hypothesis

RI test for heterogeneity:

- Calculate observed diff-in-variances

A Test for Heterogeneity

- "Heterogeneity" is $\text{Var}(\tau_i) \neq 0$
- If no heterogeneity, then $\text{Var}(Y_i(1)) = \text{Var}(Y_i(0)) \rightsquigarrow$ null hypothesis

RI test for heterogeneity:

- Calculate observed diff-in-variances
- Assume $\tau_i = \hat{\tau}$

A Test for Heterogeneity

- "Heterogeneity" is $\text{Var}(\tau_i) \neq 0$
- If no heterogeneity, then $\text{Var}(Y_i(1)) = \text{Var}(Y_i(0)) \rightsquigarrow$ null hypothesis

RI test for heterogeneity:

- Calculate observed diff-in-variances
- Assume $\tau_i = \hat{\tau}$
- Impute missing potential outcomes

A Test for Heterogeneity

- "Heterogeneity" is $\text{Var}(\tau_i) \neq 0$
- If no heterogeneity, then $\text{Var}(Y_i(1)) = \text{Var}(Y_i(0)) \rightsquigarrow$ null hypothesis

RI test for heterogeneity:

- Calculate observed diff-in-variances
- Assume $\tau_i = \hat{\tau}$
- Impute missing potential outcomes
- Create reference dist'n of 1000 diff-in-variances

A Test for Heterogeneity

- "Heterogeneity" is $\mathrm{Var}(\tau_i) \neq 0$
- If no heterogeneity, then $\mathrm{Var}(Y_i(1)) = \mathrm{Var}(Y_i(0)) \rightsquigarrow$ null hypothesis

RI test for heterogeneity:

- Calculate observed diff-in-variances
- Assume $\tau_i = \widehat{\tau}$
- Impute missing potential outcomes
- Create reference dist'n of 1000 diff-in-variances
- p-value = prop of ref dist'n as extreme or more extreme than observed diff-in-variances.

Testing for Heterogeneity

```
obs_diff_in_vars <- var(seats$water[seats$reserved == 1])
  var(seats$water[seats$reserved == 0])
```

Testing for Heterogeneity

```
obs_diff_in_vars <- var(seats$water[seats$reserved == 1]) -
  var(seats$water[seats$reserved == 0])
```

```
obs_diff_in_vars
```

```
## [1] 2266.654
```

Testing for Heterogeneity

```
n_perms <- 2000                     # how many permutations
store_dvs <- rep(NA, n_perms) # storage
base_assg <- c(rep(0, 214), rep(1, 108))
```

Testing for Heterogeneity

```r
n_perms <- 2000                     # how many permutations
store_dvs <- rep(NA, n_perms) # storage
base_assg <- c(rep(0, 214), rep(1, 108))
```

```r
for(i in 1:n_perms){
  perm_tr <- sample(base_assg)
  diff_in_var <- var(seats$water[perm_tr == 1]) -
    var(seats$water[perm_tr == 0])
  store_dvs[i] <- diff_in_var
}
```

Testing for Heterogeneity

```r
n_perms <- 2000                   # how many permutations
store_dvs <- rep(NA, n_perms) # storage
base_assg <- c(rep(0, 214), rep(1, 108))
```

```r
for(i in 1:n_perms){
  perm_tr <- sample(base_assg)
  diff_in_var <- var(seats$water[perm_tr == 1]) -
    var(seats$water[perm_tr == 0])
  store_dvs[i] <- diff_in_var
}
```

```r
p_val <- mean(abs(store_dvs) >= abs(obs_diff_in_vars))

p_val
```

```
## [1] 0.005
```

Testing for Heterogeneity

```r
n_perms <- 2000                    # how many permutations
store_dvs <- rep(NA, n_perms) # storage
base_assg <- c(rep(0, 214), rep(1, 108))
```

```r
for(i in 1:n_perms){
  perm_tr <- sample(base_assg)
  diff_in_var <- var(seats$water[perm_tr == 1]) -
    var(seats$water[perm_tr == 0])
  store_dvs[i] <- diff_in_var
}
```

```r
p_val <- mean(abs(store_dvs) >= abs(obs_diff_in_vars))

p_val
```

```
## [1] 0.005
```

Since $p < \alpha = 0.05$, reject null of "no heterogeneity".

Heterogeneous Effects

If evidence or theory for heterogeneous effects, estimate them!

Heterogeneous Effects

If evidence or theory for heterogeneous effects, estimate them!

Estimand: Conditional Average Treatment Effect (CATE)

Heterogeneous Effects

If evidence or theory for heterogeneous effects, estimate them!

Estimand: Conditional Average Treatment Effect (CATE)

$$\begin{aligned} CATE &= E(Y_1 - Y_0 | X = x) \\ &= E(Y_1 | X = x) - E(Y_0 | X = x) \end{aligned}$$

Estimating Heterogeneous Treatment Effects

- Subset on covariate X, estimate ATE within subsets

Estimating Heterogeneous Treatment Effects

- Subset on covariate X, estimate ATE within subsets
- Estimate with interaction term in single model

Interaction Terms for Heterogeneous Effects

- Add predictor variable representing subgroups in multiple regression equation

Interaction Terms for Heterogeneous Effects

- Add predictor variable representing subgroups in multiple regression equation
- "Interact it with" (i.e., multiply it by) treatment to get us treatment effect in subgroups

Interaction Terms for Heterogeneous Effects

- Add predictor variable representing subgroups in multiple regression equation
- "Interact it with" (i.e., multiply it by) treatment to get us treatment effect in subgroups

Interaction Terms for Heterogeneous Effects

- Add predictor variable representing subgroups in multiple regression equation
- "Interact it with" (i.e., multiply it by) treatment to get us treatment effect in subgroups

```
lm_out <- lm(primary2006 ~ primary2004 +
               messages + primary2004:messages,
             data = social.neighbor)
summary(lm_out)$coef |> round(3)
```

```
##                                Estimate Std. Error t valu
## (Intercept)                       0.237      0.001 176.32
## primary2004                       0.149      0.002  69.96
## messagesNeighbors                 0.069      0.003  20.93
## primary2004:messagesNeighbors     0.027      0.005   5.23
```

Estimated model:

$$\widehat{\text{Turnout 2006}}_i = 0.24 + 0.15\,(\text{Turnout 2004}_i) + 0.07\,(\text{Neighbors}_i) + 0.03\,(\text{Turnout 2004}_i)\,(\text{Neighbors}_i)$$

Estimated model:

$$
\begin{aligned}
\widehat{\text{Turnout 2006}}_i &= 0.24 + 0.15\,(\text{Turnout 2004}_i) + 0.07\,(\text{Neighbors}_i) + \\
&\quad 0.03\,(\text{Turnout 2004}_i)\,(\text{Neighbors}_i)
\end{aligned}
$$

Calculate

1. Predicted turnout prob under `Control`, if no 2004 vote?

Estimated model:

$$\begin{aligned}\widehat{\text{Turnout 2006}}_i &= 0.24 + 0.15\,(\text{Turnout 2004}_i) + 0.07\,(\text{Neighbors}_i) + \\ &\quad 0.03\,(\text{Turnout 2004}_i)\,(\text{Neighbors}_i)\end{aligned}$$

Calculate

1. Predicted turnout prob under `Control`, if no 2004 vote?
2. Predicted turnout prob under `Neighbors`, if no 2004 vote?

Estimated model:

$$\begin{aligned}\widehat{\text{Turnout 2006}}_i &= 0.24 + 0.15\,(\text{Turnout 2004}_i) + 0.07\,(\text{Neighbors}_i) + \\ &\quad 0.03\,(\text{Turnout 2004}_i)\,(\text{Neighbors}_i)\end{aligned}$$

Calculate

1. Predicted turnout prob under `Control`, if no 2004 vote?
2. Predicted turnout prob under `Neighbors`, if no 2004 vote?
3. Treatment effect for those who didn't vote in 2004?

Estimated model:

$$\widehat{\text{Turnout 2006}}_i = 0.24 + 0.15\,(\text{Turnout 2004}_i) + 0.07\,(\text{Neighbors}_i) + 0.03\,(\text{Turnout 2004}_i)\,(\text{Neighbors}_i)$$

Calculate

1. Predicted turnout prob under `Control`, if no 2004 vote?
2. Predicted turnout prob under `Neighbors`, if no 2004 vote?
3. Treatment effect for those who didn't vote in 2004?
4. Predicted turnout prob under `Control`, if voted in 2004?
5. Predicted turnout prob under `Neighbors`, if voted in 2004?
6. Treatment effect for those who voted in 2004?

Interaction Terms for Heterogeneous Effects

$$\widehat{\text{Turnout 2006}}_i \quad = \quad \beta_0 + \beta_1 X_i + \beta_2 T_i + \beta_3 X_i T_i$$

Interaction Terms for Heterogeneous Effects

$$\begin{aligned}
\widehat{\text{Turnout 2006}}_i &= \beta_0 + \beta_1 X_i + \beta_2 T_i + \beta_3 X_i T_i \\
&= \underbrace{\beta_0}_{\bar{Y} \text{ if } (0,0)} + \underbrace{\beta_1}_{\bar{Y} \text{ if } X_i=1} X_i + \underbrace{\beta_2}_{\text{ATE if } X_i=0} T_i + \underbrace{\beta_3}_{\text{HTE}} X_i T_i
\end{aligned}$$

Interaction Terms for Heterogeneous Effects

$$\begin{aligned}
\widehat{\text{Turnout 2006}}_i &= \beta_0 + \beta_1 X_i + \beta_2 T_i + \beta_3 X_i T_i \\
&= \underbrace{\beta_0}_{\bar{Y}\text{ if }(0,0)} + \underbrace{\beta_1}_{\bar{Y}\text{ if }X_i=1} X_i + \underbrace{\beta_2}_{\text{ATE if }X_i=0} T_i + \underbrace{\beta_3}_{\text{HTE}} X_i T_i \\
\widehat{\text{Turnout 2006}}_i &= \underbrace{0.24}_{\bar{Y}\text{ for (0,0), Prior Nonvoters, Control}} + \\
&\quad \underbrace{0.15}_{\text{Additional }\bar{Y}\text{ for Prior Voters}} (\text{Turnout 2004}_i) + \\
&\quad \underbrace{0.07}_{\text{ATE for Prior Nonvoters; part of ATE for Prior Voters}} (\text{Neigh} \\
&\quad \underbrace{0.03}_{\text{Additional ATE for Prior Voters}} (\text{Turnout 2004}_i)(\text{Neighb}
\end{aligned}$$

Exercise:

Consider the data on UK MP candidates at
http://j.mp/32PHfFd

1. Read the data into R and name it `mps`
2. Create a new variable called `winner`. Set it equal to `1` if `margin` is greater than 0; set it to `0` if `margin` is less than zero.
3. Consider `winner` to be a randomly assigned treatment. Estimate the causal effect of `winner` on `ln.net`, net worth at death of these candidates.
4. Investigate whether the effect of `winner` varies by `party`.

Quantile Average Treatment Effects

A *quantile* is a cut-point, or a position, in a statistical distribution.

Quantile Average Treatment Effects

A *quantile* is a cut-point, or a position, in a statistical distribution.

What is the name for the value that has

- half the values below it?

Quantile Average Treatment Effects

A *quantile* is a cut-point, or a position, in a statistical distribution.

What is the name for the value that has

- half the values below it?
- $\frac{3}{4}$ of the values below it?

Quantile Average Treatment Effects

A *quantile* is a cut-point, or a position, in a statistical distribution.

What is the name for the value that has

- half the values below it?
- $\frac{3}{4}$ of the values below it?
- $\frac{1}{5}$ of the values below it?

Quantile Average Treatment Effects

A *quantile* is a cut-point, or a position, in a statistical distribution.

What is the name for the value that has

- half the values below it?
- $\frac{3}{4}$ of the values below it?
- $\frac{1}{5}$ of the values below it?
- $\ldots$

Quantile Average Treatment Effects

A *quantile* is a cut-point, or a position, in a statistical distribution.

What is the name for the value that has

- half the values below it?
- $\frac{3}{4}$ of the values below it?
- $\frac{1}{5}$ of the values below it?
- …

Quantile Average Treatment Effects

A *quantile* is a cut-point, or a position, in a statistical distribution.

What is the name for the value that has

- half the values below it?
- $\frac{3}{4}$ of the values below it?
- $\frac{1}{5}$ of the values below it?
- $\dots$

These are *quantiles*.

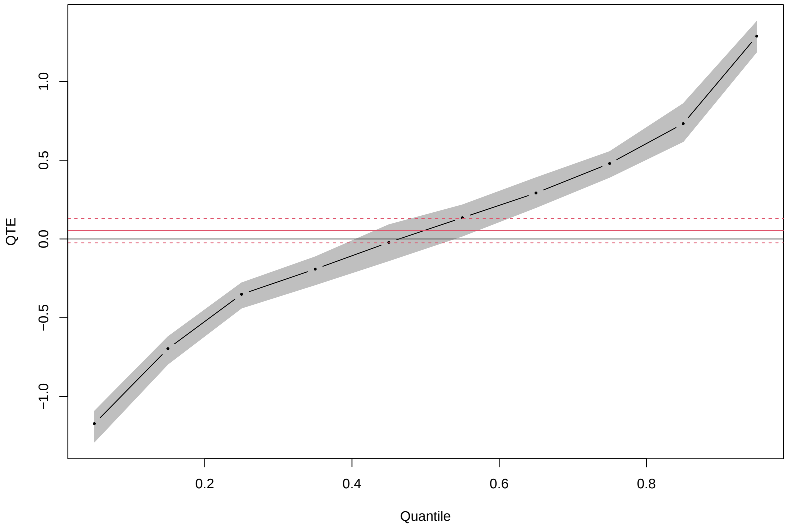
Quantile Average Treatment Effects (data)

```
set.seed(326370675)
n <- 1000
Y1 <- Y0 <- runif(n)
# If low baseline, negative TE:
Y1[Y0 <.5] <- Y0[Y0 <.5] - rnorm(length(Y0[Y0 <.5]))
# If high baseline, positive TE:
Y1[Y0 >.5] <- Y0[Y0 >.5] + rnorm(length(Y0[Y0 >.5]))
D <- sample((1:n) %% 2)  # Assign 0/1
Y <- D * Y1 + (1 - D) * Y0  # Y_obs
samp <- tibble(D, Y)
library(quantreg)
ate <- coef(lm(Y ~ D, data = samp))[2]
qtes <- rq(Y ~ D,
           tau = seq(.05, .95, length.out = 10), # At what
           data = samp, method = "fn")
```

Quantile Average Treatment Effects (data)

```r
set.seed(326370675)
n <- 1000
Y1 <- Y0 <- runif(n)
# If low baseline, negative TE:
Y1[Y0 <.5] <- Y0[Y0 <.5] - rnorm(length(Y0[Y0 <.5]))
# If high baseline, positive TE:
Y1[Y0 >.5] <- Y0[Y0 >.5] + rnorm(length(Y0[Y0 >.5]))
D <- sample((1:n) %% 2)  # Assign 0/1
Y <- D * Y1 + (1 - D) * Y0  # Y_obs
samp <- tibble(D, Y)
library(quantreg)
ate <- coef(lm(Y ~ D, data = samp))[2]
qtes <- rq(Y ~ D,
           tau = seq(.05, .95, length.out = 10), # At what p
           data = samp, method = "fn")
```

(See https://bit.ly/3dnNhGP)

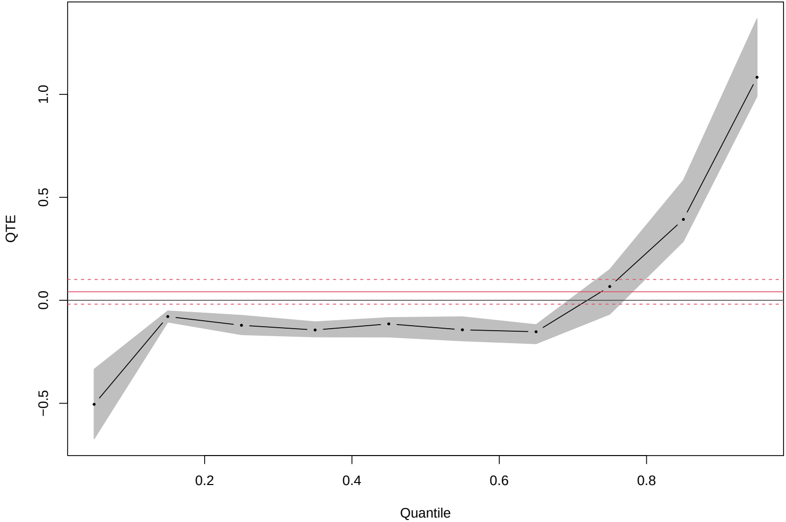


QTE
1.0
0.5
0.0
−0.5
−1.0
0.2
0.4
0.6
0.8
Quantile

Quantile Average Treatment Effects (new data)

```r
set.seed(21578100)
n <- 1000
Y1 <- Y0 <- runif(n)
# If low baseline, zero TE:
Y1[Y0 <.5] <- Y0[Y0 <.5]
# If high baseline, positive TE:
Y1[Y0 >.5] <- Y0[Y0 >.5] + rnorm(length(Y0[Y0 >.5]))
D <- sample((1:n) %% 2)  # Assign 0/1
Y <- D * Y1 + (1 - D) * Y0  # Y_obs
samp <- tibble(D, Y)

ate <- coef(lm(Y ~ D, data = samp))[2]
qtes <- rq(Y ~ D,
           tau = seq(.05, .95, length.out = 10),
           data = samp, method = "fn")
```



1.0
0.5
0.0
−0.5
QTE
0.2
0.4
0.6
0.8
Quantile

Inference for Experiments using Regression Adjustment

Inference for Experiments using Regression Adjustment

Lin (2013) compares several estimates of SE for linear models with experimental data, adjusting for covariates x.

Tailored for fixed-sample, variation-from-assignment situations.

Inference for Experiments using Regression Adjustment

Lin (2013) compares several estimates of SE for linear models with experimental data, adjusting for covariates x.

Tailored for fixed-sample, variation-from-assignment situations.

“Sandwich” estimators of variance:

Inference for Experiments using Regression Adjustment

Lin (2013) compares several estimates of SE for linear models with experimental data, adjusting for covariates x.

Tailored for fixed-sample, variation-from-assignment situations.

"Sandwich" estimators of variance:

$$Var\left(\widehat{\beta_{\text{OLS}}}\right) = (X'X)^{-1}(X'\text{diag}(\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n)X)(X'X)^{-1}$$

Inference for Experiments using Regression Adjustment

Model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (x_i - \bar{x}) + \beta_3 T_i (x_i - \bar{x}) + \epsilon_i$$

Inference for Experiments using Regression Adjustment

Model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (x_i - \bar{x}) + \beta_3 T_i (x_i - \bar{x}) + \epsilon_i$$

"Full set of treatment by (demeaned) covariate interactions"

Inference for Experiments using Regression Adjustment

Model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (x_i - \bar{x}) + \beta_3 T_i (x_i - \bar{x}) + \epsilon_i$$

"Full set of treatment by (demeaned) covariate interactions"

Then, use "sandwich" HC2 standard errors.

Inference for Experiments using Regression Adjustment

Model:

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (x_i - \bar{x}) + \beta_3 T_i (x_i - \bar{x}) + \epsilon_i$$

"Full set of treatment by (demeaned) covariate interactions"

Then, use "sandwich" HC2 standard errors.

Causal estimate:

$$\widehat{ATE}_{\text{interact}} = \hat{\beta}_1$$

Inference for Experiments using Regression Adjustment

Result: SE's that are

Inference for Experiments using Regression Adjustment

Result: SE's that are

- Consistent under heteroskedasticity

Inference for Experiments using Regression Adjustment

Result: SE's that are

- Consistent under heteroskedasticity
- At least as efficient as usual $\widehat{ATE}_{\text{OLS}}$

Inference for Experiments using Regression Adjustment

Result: SE's that are

- Consistent under heteroskedasticity
- At least as efficient as usual $\widehat{ATE}_{\text{OLS}}$
- More efficient if $n_T \neq n_C$

Inference for Experiments using Regression Adjustment

Result: SE's that are

- Consistent under heteroskedasticity
- At least as efficient as usual $\widehat{ATE}_{\text{OLS}}$
- More efficient if $n_T \neq n_C$
- Doesn't *hurt* precision
  ($\widehat{ATE}_{\text{OLS}}$ does if x varies more with τ_i than Y_1, Y_0, e.g.)

Inference for Experiments using Regression Adjustment

Recovering the ATE

Suppose categorical G with 2 levels. Let G_i be indicator for first level. Expect heterogeneous treatment effects.

Inference for Experiments using Regression Adjustment

Recovering the ATE

Suppose categorical G with 2 levels. Let G_i be indicator for first level. Expect heterogeneous treatment effects.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 G_i + \beta_3 T_i G_i + \epsilon_i \tag{1}$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

Suppose categorical G with 2 levels. Let G_i be indicator for first level. Expect heterogeneous treatment effects.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 G_i + \beta_3 T_i G_i + \epsilon_i \tag{1}$$

What's $\widehat{ATE}$?

Inference for Experiments using Regression Adjustment

Recovering the ATE

Suppose categorical G with 2 levels. Let G_i be indicator for first level. Expect heterogeneous treatment effects.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 G_i + \beta_3 T_i G_i + \epsilon_i \tag{1}$$

What's $\widehat{ATE}$? Not $\hat{\beta}_1$.

Inference for Experiments using Regression Adjustment

Recovering the ATE

Suppose categorical G with 2 levels. Let G_i be indicator for first level. Expect heterogeneous treatment effects.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 G_i + \beta_3 T_i G_i + \epsilon_i \tag{1}$$

What's $\widehat{ATE}$? Not $\hat{\beta}_1$. Not $\hat{\beta}_1 + \hat{\beta}_3$.

Inference for Experiments using Regression Adjustment

Recovering the ATE

Suppose categorical G with 2 levels. Let G_i be indicator for first level. Expect heterogeneous treatment effects.

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 G_i + \beta_3 T_i G_i + \epsilon_i \tag{1}$$

What's $\widehat{ATE}$? Not $\hat{\beta}_1$. Not $\hat{\beta}_1 + \hat{\beta}_3$.

(Need to weight by how many of each G type we have.)

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$?

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$?

$$\hat{\beta}_1$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$?

$$\hat{\beta}_1$$

What's $\widehat{ATE}$ for $G_i = 1$ group?

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$?

$$\hat{\beta}_1$$

What's $\widehat{ATE}$ for $G_i = 1$ group?

$$\widehat{ATE}_{G1} \quad = \quad \hat{\beta}_1 + \hat{\beta}_3 (1 - \bar{G})$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$?

$$\hat{\beta}_1$$

What's $\widehat{ATE}$ for $G_i = 1$ group?

$$\begin{aligned} \widehat{ATE}_{G1} &= \hat{\beta}_1 + \hat{\beta}_3 (1 - \bar{G}) \\ &= \hat{\beta}_1 + \hat{\beta}_3 \times (\text{prop. of } G = 0) \end{aligned}$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2(G_i - \bar{G}) + \beta_3 T_i(G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$ for $G_i = 0$ group?

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$ for $G_i = 0$ group?

$$\widehat{ATE}_{G0} \quad = \quad \hat{\beta}_1 + \hat{\beta}_3 (0 - \bar{G})$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$ for $G_i = 0$ group?

$$\begin{aligned} \widehat{ATE}_{G0} &= \hat{\beta}_1 + \hat{\beta}_3 (0 - \bar{G}) \\ &= \hat{\beta}_1 - \hat{\beta}_3 \bar{G} \end{aligned}$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$ for $G_i = 0$ group?

$$\begin{aligned} \widehat{ATE}_{G0} &= \hat{\beta}_1 + \hat{\beta}_3 (0 - \bar{G}) \\ &= \hat{\beta}_1 - \hat{\beta}_3 \bar{G} \\ &= \hat{\beta}_1 - \hat{\beta}_3 \times (\text{prop. of } G = 1) \end{aligned}$$

Inference for Experiments using Regression Adjustment

Recovering the ATE

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2 (G_i - \bar{G}) + \beta_3 T_i (G_i - \bar{G}) + \epsilon_i$$

What's $\widehat{ATE}$ for $G_i = 0$ group?

$$\begin{aligned}
\widehat{ATE}_{G0} &= \hat{\beta}_1 + \hat{\beta}_3 (0 - \bar{G}) \\
&= \hat{\beta}_1 - \hat{\beta}_3 \bar{G} \\
&= \hat{\beta}_1 - \hat{\beta}_3 \times (\text{prop. of } G = 1)
\end{aligned}$$

Inference for Experiments using Regression Adjustment

So, to adjust for covariates in experimental data,

1. Estimate

$$Y_i = \beta_0 + \beta_1 T_i + \beta_2(x_i - \bar{x}) + \beta_3 T_i(x_i - \bar{x}) + \epsilon_i$$

2. Use HC2 standard errors (or analogous CR2 for clustered designs)

```r
library(estimatr)
lm_robust(y ~ t + I(x - mean(x)) + t * I(x - mean(x)),
          data = df)
```

Inference for Experiments using Regression Adjustment

So, to adjust for covariates x in experimental data,

```
lm_lin(y ~ t, covariates = ~ x, data = df)
```

Inference for Experiments using Regression Adjustment

With two levels of treatment,

$$Y_i = \beta_0 + \beta_1 T1_i + \beta_2 T2_i + \beta_3 (x_i - \bar{x}) + \beta_4 T1_i (x_i - \bar{x}) + \beta_5 T2_i (x_i - \bar{x}) + \epsilon_i$$

Inference for Experiments using Regression Adjustment

With two levels of treatment,

$$Y_i = \beta_0 + \beta_1 T1_i + \beta_2 T2_i + \beta_3 (x_i - \bar{x}) + \beta_4 T1_i (x_i - \bar{x}) + \beta_5 T2_i (x_i - \bar{x}) + \epsilon_i$$

etc.

“Controlling for Blocks”

Beware LSDV: "controlling for blocks"

A first-thought for estimating ATE for blocked design ...

Beware LSDV: "controlling for blocks"

A first-thought for estimating ATE for blocked design ...

Estimate linear model with indicators for each block:

$$Y_i = \beta_0 + \beta_1 T_i + \gamma_1 B_{i1} + \gamma_2 B_{i2} + \ldots + \epsilon_i$$

Beware LSDV: "controlling for blocks"

However, when

1. TE varies by block
2. $P(T = 1)$ varies by block

Beware LSDV: "controlling for blocks"

However, when

1. TE varies by block
2. $P(T = 1)$ varies by block

Using "block fixed effects" or "controlling for block IDs" is

Beware LSDV: "controlling for blocks"

However, when

1. TE varies by block
2. $P(T = 1)$ varies by block

Using "block fixed effects" or "controlling for block IDs" is

- biased for the estimate (ATE)

Beware LSDV: "controlling for blocks"

However, when

1. TE varies by block
2. $P(T = 1)$ varies by block

Using "block fixed effects" or "controlling for block IDs" is

- biased for the estimate (ATE)
- biased for the standard error ($SE(ATE)$)

Beware LSDV: "controlling for blocks"

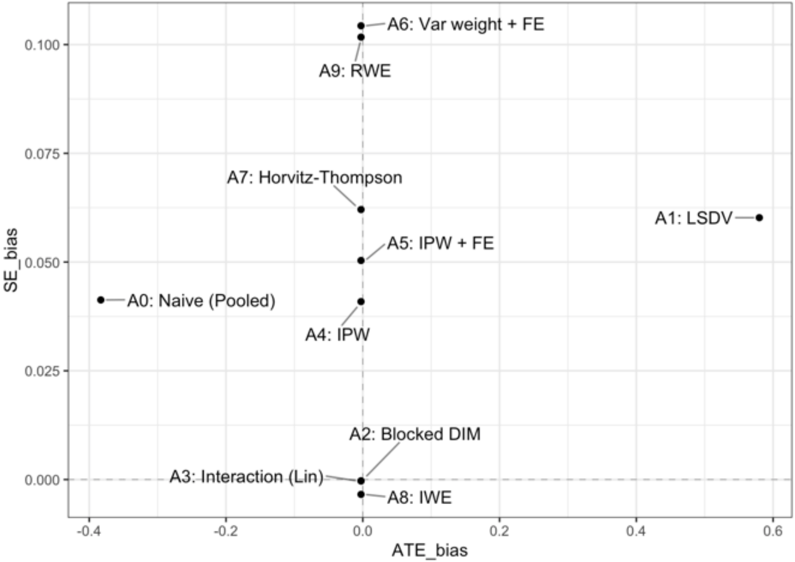
However, when

1. TE varies by block
2. $P(T = 1)$ varies by block

Using "block fixed effects" or "controlling for block IDs" is

- biased for the estimate (ATE)
- biased for the standard error ($SE(ATE)$)
- worse than most everything else!

Beware LSDV: "controlling for blocks"



Beware LSDV: "controlling for blocks"

The intuition: regression over-weights high-variance blocks

block	block_ATE	block_ATE_est	n_j	p_j	sample_weight	fe_weight
1	4	4.06	100	0.5	0.33	0.45
2	2	1.66	100	0.7	0.33	0.38
3	0	0.02	100	0.9	0.33	0.16

Beware LSDV: "controlling for blocks"

- Should I adjust for continuous x, or block IDs, where blocks come from coarsened x?

Beware LSDV: "controlling for blocks"

- Should I adjust for continuous x, or block IDs, where blocks come from coarsened x?
- If both TE and Tr Prob vary across blocks, use **block IDs** to remove bias
  (since those determine assignment)

Beware LSDV: "controlling for blocks"

- Should I adjust for continuous x, or block IDs, where blocks come from coarsened x?
- If both TE and Tr Prob vary across blocks, use **block IDs** to remove bias
  (since those determine assignment)
- In all 4 of the "probabilities vary/constant" by "treatment effects vary/constant" possibilities, using **both** the IDs and the variables themselves feels like overkill, but performs well.

(Related note: regression weights the cases)

FIGURE 1 Example of nominal and effective samples from Jensen (2003)

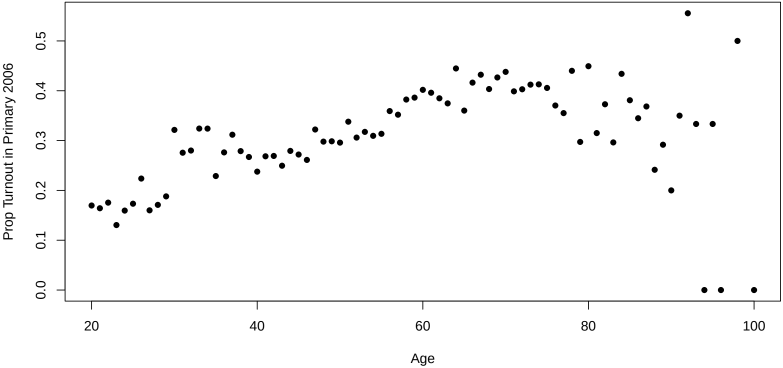
- Aronow and Samii (2016) shows how multiple regression weights its cases
- Interp: for causal inf., need causal methods, not just regression + representative sample

Nonlinear Terms

Terms to Capture Nonlinear Relationships

Let's take a step back from causal inference, and just think about trying to model the relationship between turnout probability and age.

```r
# Calculate turnout prop for every year of age:
prop_turnout <- tapply(social.neighbor$primary2006,
                       social.neighbor$age, mean)
# Plot:
plot(prop_turnout ~ names(prop_turnout), xlab = "Age",
     ylab = "Prop Turnout in Primary 2006", pch = 16)
```



This looks like a nonlinear relationship. Let's include an age-squared term to try to model this nonlinearity:

```
lm_out <- lm(primary2006 ~ age + I(age^2),
             data = social.neighbor)
lm_out
```

```
##
## Call:
## lm(formula = primary2006 ~ age + I(age^2), data = social
##
## Coefficients:
## (Intercept)          age     I(age^2)
##  -2.868e-02    9.818e-03   -5.572e-05
```

The estimated model is

$$\widehat{\text{Turnout 2006}}_i = -0.02868 + 0.00982\,(\text{Age}_i) + -6 \times 10^{-5}\left(\text{Age}_i^2\right)$$

1. What is the predicted turnout probability for a 30-year-old?
2. What is the predicted turnout probability for a 60-year-old?
3. What is the predicted turnout probability for a 80-year-old?
4. What is the predicted turnout probability for a 90-year-old?

```r
predict(lm_out, data.frame(age = 40))
```

```
##         1
## 0.2748905
```

```r
predict(lm_out, data.frame(age = 60))
```

```
##         1
## 0.3598197
```

```r
predict(lm_out, data.frame(age = 80))
```

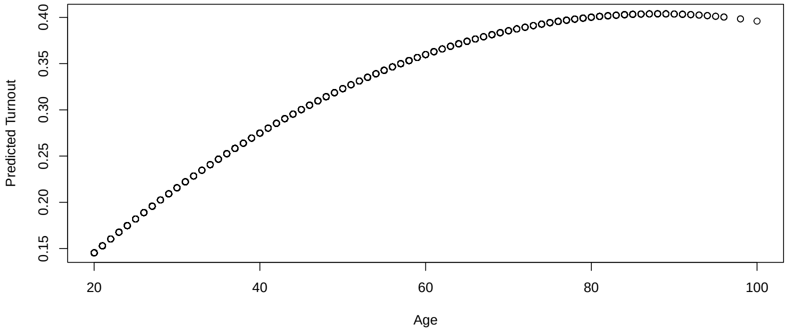
```
##         1
## 0.4001768
```

```r
predict(lm_out, data.frame(age = 100))
```

```
##         1
## 0.3959618
```

We can also look at all the predicted values:

```
plot(lm_out$fitted.values ~ social.neighbor$age,
     xlab = "Age", ylab = "Predicted Turnout")
```



How Causal Inference is Different

Data Science Approaches

Three tasks of data science:

Data Science Approaches

Three tasks of data science:

- Description

Three tasks of data science:

- Description
- Prediction

Three tasks of data science:

- Description
- Prediction
- Causal Inference

Data Science Approaches

Three tasks of data science:

- Description
- Prediction
- Causal Inference

Data Science Approaches

Three tasks of data science:

- Description
- Prediction
- Causal Inference

Models/algorithms central to all three.

Data Science Approaches

Three tasks of data science:

- Description
- Prediction
- Causal Inference

Models/algorithms central to all three.

Hernán, Hsu, and Healy (2019)

Description

- Identifying patterns, etc.

Description

- Identifying patterns, etc.
- E.g., clustering to discover groups

Data Science Approaches

Prediction

- Components

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)
  - Mapping from inputs to outputs
    (linear model, decision tree, ...)

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)
  - Mapping from inputs to outputs
    (linear model, decision tree, ...)
  - Metric for evaluating mapping

Prediction

- Components
  - Inputs/outputs
    (predictors/outcomes, features/responses, ...)
  - Mapping from inputs to outputs
    (linear model, decision tree, ...)
  - Metric for evaluating mapping
- With these, model/machine learning algorithm does the work

Prediction

- Components
  - Inputs/outputs (predictors/outcomes, features/responses, ...)
  - Mapping from inputs to outputs (linear model, decision tree, ...)
  - Metric for evaluating mapping
- With these, model/machine learning algorithm does the work
- E.g., regression, random forests, neural networks, ...

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same
  - T v. $\mathbf{X}$ – very different!

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same
  - T v. $\mathbf{X}$ – very different!
  - (the more knowledge, the better!)

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same
  - T v. $\mathbf{X}$ – very different!
  - (the more knowledge, the better!)
  - (alternative: solve fundamental problem of causal inference! ☺)

Data Science Approaches

Causal Inference

- Potential outcomes/counterfactual/interventionist perspective
- Requires *expertise* different to description/prediction
- Requires more than summary statistics, metrics, etc.
- Requires some knowledge of causal structure
  - Not all inputs treated same
  - T v. $\mathbf{X}$ – very different!
  - (the more knowledge, the better!)
  - (alternative: solve fundamental problem of causal inference! ☺)
- E.g., experiments, observational causal designs, …

Causal Inference with Models

Consider two loaded datasets:

Causal Inference with Models

Consider two loaded datasets:

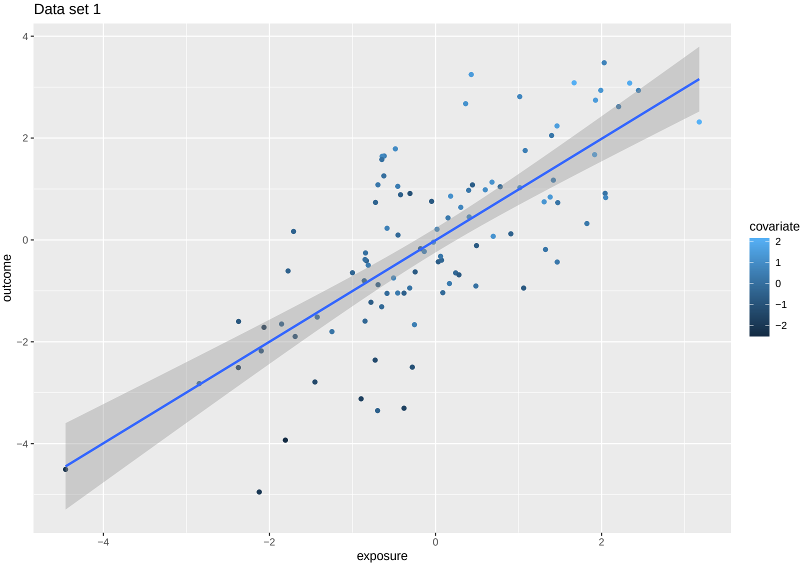
```r
str(df1)
```

```
## tibble [100 x 3] (S3: tbl_df/tbl/data.frame)
##  $ covariate: num [1:100] -0.622 1.137 -0.238 1.529 -0.1
##  $ exposure : num [1:100] 0.0332 0.3627 0.2422 1.4633 0.
##  $ outcome  : num [1:100] -0.429 2.675 -0.647 2.238 1.04
```

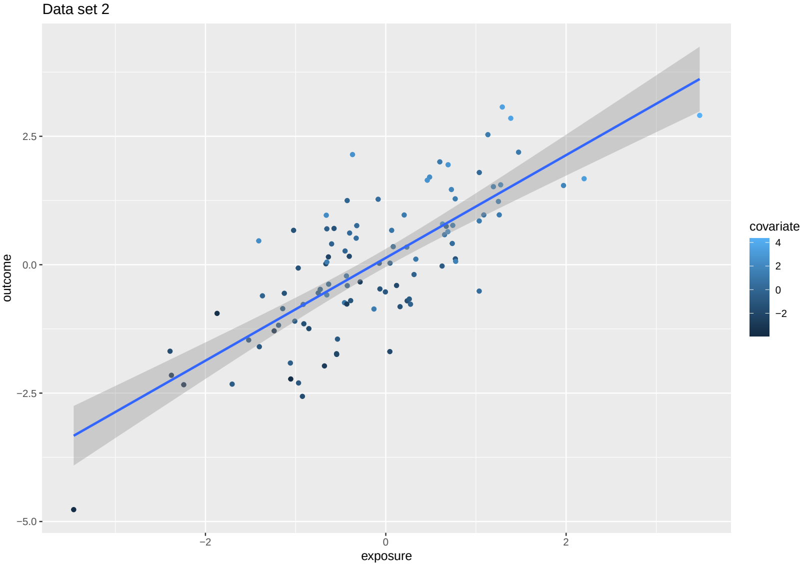
```r
str(df2)
```

```
## tibble [100 x 3] (S3: tbl_df/tbl/data.frame)
##  $ exposure : num [1:100] 0.4862 0.0653 -1.4021 -0.546 -
##  $ outcome  : num [1:100] 1.706 0.669 -1.597 -1.733 0.61
##  $ covariate: num [1:100] 2.24 0.924 -0.999 -2.343 0.207
```

Causal Inference with Models



Causal Inference with Models



Causal Inference with Models

Model each

```
lm_df1 <- lm(outcome ~ exposure, data = df1)
lm_df2 <- lm(outcome ~ exposure, data = df2)
```

```
## # A tibble: 4 x 4
##   data  term        estimate std.error
##   <chr> <chr>          <dbl>     <dbl>
## 1 df1   (Intercept) -0.00671    0.120
## 2 df1   exposure     0.996      0.0927
## 3 df2   (Intercept)  0.133      0.0890
## 4 df2   exposure     1.00       0.0841
```

Causal Inference with Models

Model each

```r
lm_df1 <- lm(outcome ~ exposure, data = df1)
lm_df2 <- lm(outcome ~ exposure, data = df2)
```

```
## # A tibble: 4 x 4
##   data  term        estimate std.error
##   <chr> <chr>          <dbl>     <dbl>
## 1 df1   (Intercept) -0.00671    0.120
## 2 df1   exposure     0.996      0.0927
## 3 df2   (Intercept)  0.133      0.0890
## 4 df2   exposure     1.00       0.0841
```

- Both cases: effect of exposure ≈ 1.

Causal Inference with Models

Model each

```r
lm_df1 <- lm(outcome ~ exposure, data = df1)
lm_df2 <- lm(outcome ~ exposure, data = df2)
```

```
## # A tibble: 4 x 4
##   data  term        estimate std.error
##   <chr> <chr>          <dbl>     <dbl>
## 1 df1   (Intercept) -0.00671    0.120
## 2 df1   exposure     0.996      0.0927
## 3 df2   (Intercept)  0.133      0.0890
## 4 df2   exposure     1.00       0.0841
```

- Both cases: effect of exposure ≈ 1.
- Is this good? Is it correct?

Causal Inference with Models

Model each

```
lm_df1 <- lm(outcome ~ exposure, data = df1)
lm_df2 <- lm(outcome ~ exposure, data = df2)
```

```
## # A tibble: 4 x 4
##   data  term        estimate std.error
##   <chr> <chr>          <dbl>     <dbl>
## 1 df1   (Intercept) -0.00671    0.120
## 2 df1   exposure     0.996      0.0927
## 3 df2   (Intercept)  0.133      0.0890
## 4 df2   exposure     1.00       0.0841
```

- Both cases: effect of exposure ≈ 1.
- Is this good? Is it correct?
- What if we adjust for covariate?

Causal Inference with Models

```r
lm_df1_adj <- lm(outcome ~ exposure + covariate, data = df1
lm_df2_adj <- lm(outcome ~ exposure + covariate, data = df2
```

```
## # A tibble: 4 x 4
##   data  term      estimate std.error
##   <chr> <chr>        <dbl>     <dbl>
## 1 df1   exposure     0.501    0.108
## 2 df1   covariate    0.970    0.147
## 3 df2   exposure     0.554    0.0990
## 4 df2   covariate    0.385    0.0598
```

- Both cases: effect of exposure ≈ 0.5.

Causal Inference with Models

```
lm_df1_adj <- lm(outcome ~ exposure + covariate, data = df1
lm_df2_adj <- lm(outcome ~ exposure + covariate, data = df2
```

```
## # A tibble: 4 x 4
##   data  term      estimate std.error
##   <chr> <chr>        <dbl>     <dbl>
## 1 df1   exposure     0.501    0.108
## 2 df1   covariate    0.970    0.147
## 3 df2   exposure     0.554    0.0990
## 4 df2   covariate    0.385    0.0598
```

- Both cases: effect of exposure ≈ 0.5.
- Is this good? Is it correct?

Causal Inference with Models

```
lm_df1_adj <- lm(outcome ~ exposure + covariate, data = df1
lm_df2_adj <- lm(outcome ~ exposure + covariate, data = df2
```

```
## # A tibble: 4 x 4
##   data  term      estimate std.error
##   <chr> <chr>        <dbl>     <dbl>
## 1 df1   exposure     0.501    0.108
## 2 df1   covariate    0.970    0.147
## 3 df2   exposure     0.554    0.0990
## 4 df2   covariate    0.385    0.0598
```

- Both cases: effect of exposure ≈ 0.5.
- Is this good? Is it correct?
- Which is correct? $\beta = 1$? $\beta = 0.5$?

Causal Inference with Models

```r
lm_df1_adj <- lm(outcome ~ exposure + covariate, data = df1
lm_df2_adj <- lm(outcome ~ exposure + covariate, data = df2
```

```
## # A tibble: 4 x 4
##   data  term      estimate std.error
##   <chr> <chr>        <dbl>     <dbl>
## 1 df1   exposure     0.501    0.108
## 2 df1   covariate    0.970    0.147
## 3 df2   exposure     0.554    0.0990
## 4 df2   covariate    0.385    0.0598
```

- Both cases: effect of exposure ≈ 0.5.
- Is this good? Is it correct?
- Which is correct? $\beta = 1$? $\beta = 0.5$?
- *Should* we adjust for covariate?

Causal Inference with Models

There is nothing in the data that tells us.

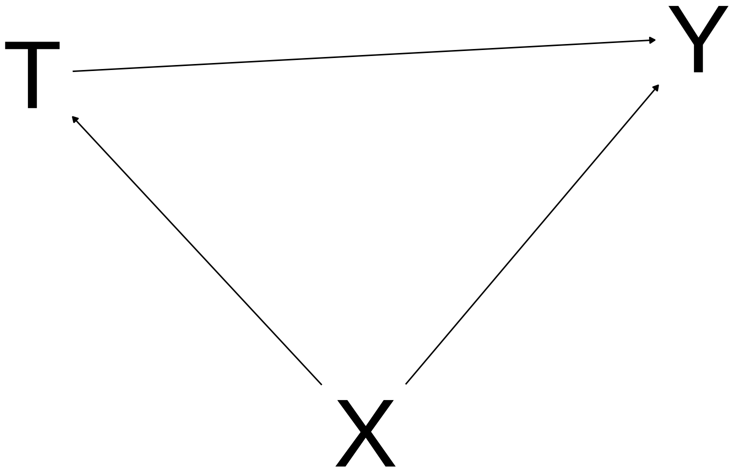
Causal Inference with Models

There is nothing in the data that tells us. ☹

Causal Inference with Models

There is nothing in the data that tells us. ☹

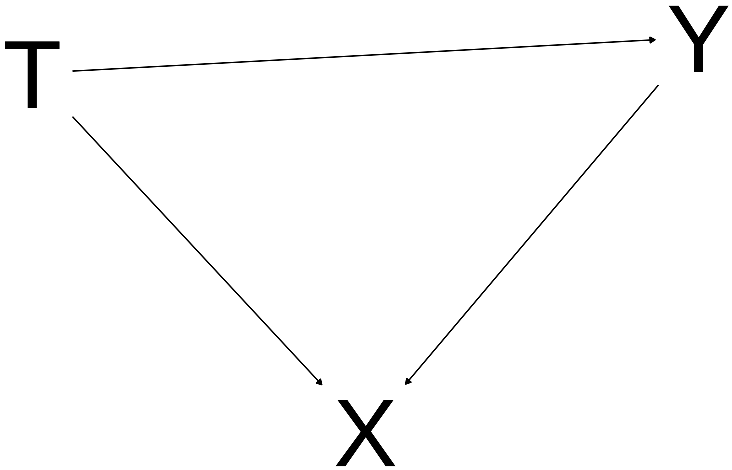
Here are the true structures: First



Causal Inference with Models

There is nothing in the data that tells us. ☹

Here are the true structures: Second



Causal Inference with Models

When know structures, adjustment sets for unbiasedness differ:

Causal Inference with Models

When know structures, adjustment sets for unbiasedness differ:

- `df1`: confounding $\Rightarrow$ **adjust for** X

Causal Inference with Models

When know structures, adjustment sets for unbiasedness differ:

- `df1`: confounding $\Rightarrow$ **adjust for** X
- `df2`: collider $\Rightarrow$ **do not adjust for** X

Causal Inference with Models

When know structures, adjustment sets for unbiasedness differ:

- `df1`: confounding $\Rightarrow$ **adjust for** X
- `df2`: collider $\Rightarrow$ **do not adjust for** X

Causal Inference with Models

When know structures, adjustment sets for unbiasedness differ:

- `df1`: confounding $\Rightarrow$ **adjust for** X
- `df2`: collider $\Rightarrow$ **do not adjust for** X

```
g_conf <- dagitty("dag{ x -> y ; x <- c -> y }")
g_coll <- dagitty("dag{ x -> y ; x -> c <- y }")
```

Causal Inference with Models

When know structures, adjustment sets for unbiasedness differ:

- `df1`: confounding $\Rightarrow$ **adjust for** X
- `df2`: collider $\Rightarrow$ **do not adjust for** X

```
g_conf <- dagitty("dag{ x -> y ; x <- c -> y }")
g_coll <- dagitty("dag{ x -> y ; x -> c <- y }")
```

```
adjustmentSets(g_conf, "x", "y")
```

```
## { c }
```

Causal Inference with Models

When know structures, adjustment sets for unbiasedness differ:

- `df1`: confounding $\Rightarrow$ **adjust for** X
- `df2`: collider $\Rightarrow$ **do not adjust for** X

```
g_conf <- dagitty("dag{ x -> y ; x <- c -> y }")
g_coll <- dagitty("dag{ x -> y ; x -> c <- y }")
```

```
adjustmentSets(g_conf, "x", "y")
```

```
## { c }
```

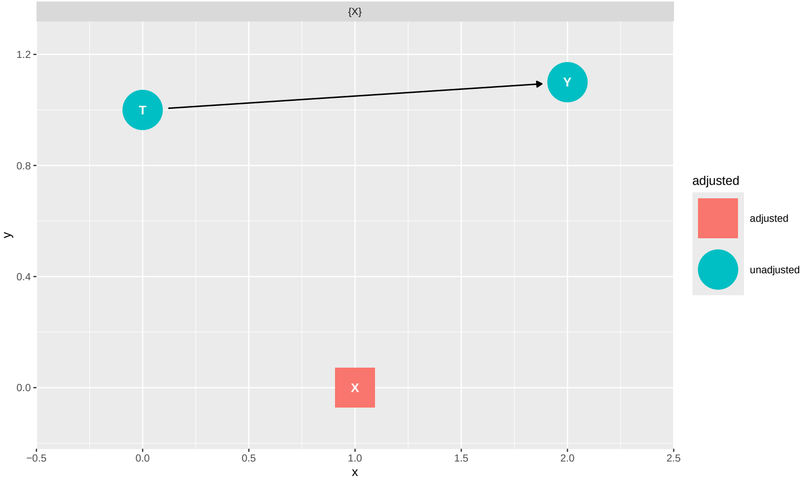
```
adjustmentSets(g_coll, "x", "y")
```

```
##  {}
```

Causal Inference with Models

When know structures, adjustment sets for unbiasedness differ:

- `df1`: confounding ⇒ **adjust for** X
- `df2`: collider ⇒ **do not adjust for** X



Causal Inference with Models

So, correct adjustments to reveal causal effect of $T \rightsquigarrow Y$:

Causal Inference with Models

So, correct adjustments to reveal causal effect of $T \rightsquigarrow Y$:

`df1`, **adjust** for X, $\beta = 0.5$:

```
## # A tibble: 1 x 4
##   data  term     estimate std.error
##   <chr> <chr>       <dbl>     <dbl>
## 1 df1   exposure    0.501     0.108
```

`df2`, **do not adjust** for X, $\beta = 1$:

```
## # A tibble: 1 x 4
##   data  term     estimate std.error
##   <chr> <chr>       <dbl>     <dbl>
## 1 df2   exposure     1.00    0.0841
```

Causal Inference with Models

So, correct adjustments to reveal causal effect of $T \rightsquigarrow Y$:

`df1`, **adjust** for X, $\beta = 0.5$:

```
## # A tibble: 1 x 4
##   data  term     estimate std.error
##   <chr> <chr>       <dbl>     <dbl>
## 1 df1   exposure    0.501     0.108
```

`df2`, **do not adjust** for X, $\beta = 1$:

```
## # A tibble: 1 x 4
##   data  term     estimate std.error
##   <chr> <chr>       <dbl>     <dbl>
## 1 df2   exposure     1.00    0.0841
```

(Data from D'Agostino McGowan (2023))

Causal Inference with Models

- Importance of identifying "pre-treatment covariates", "proper covariates"; doing "design before analysis"

Causal Inference with Models

- Importance of identifying "pre-treatment covariates", "proper covariates"; doing "design before analysis"
- Importance of experiments: strong knowledge about (part of) causal structure – assgn mechanism

Causal Inference with Models

- Importance of identifying "pre-treatment covariates", "proper covariates"; doing "design before analysis"
- Importance of experiments: strong knowledge about (part of) causal structure – assgn mechanism
- Causal inference is critical to scientific questions, and separate from prediction

Causal Inference with Models

- Importance of identifying "pre-treatment covariates", "proper covariates"; doing "design before analysis"
- Importance of experiments: strong knowledge about (part of) causal structure – assgn mechanism
- Causal inference is critical to scientific questions, and separate from prediction
- Though, methods from prediction can aid causal inference (see, especially, *causal forests*)

Causal Inference with Models

- Importance of identifying "pre-treatment covariates", "proper covariates"; doing "design before analysis"
- Importance of experiments: strong knowledge about (part of) causal structure – assgn mechanism
- Causal inference is critical to scientific questions, and separate from prediction
- Though, methods from prediction can aid causal inference (see, especially, *causal forests*)
- "Causal euphimisms" don't help (Hernán 2018)

References I

Aronow, Peter M., and Cyrus Samii. 2016. "Does Regression Produce Representative Estimates of Causal Effects?" *American Journal of Political Science* 60 (1): 250–67.

Chattopadhyay, Raghabendra, and Esther Duflo. 2004. "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica* 72 (5): 1409–43.

D'Agostino McGowan, Lucy. 2023. *quartets: Datasets to Help Teach Statistics*. https://r-causal.github.io/quartets/.

Freedman, David A. 2008. "On Regression Adjustments to Experimental Data." *Advances in Applied Mathematics* 40: 180–93.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review* 102 (1): 33–48.

Green, Donald P., and Peter M. Aronow. 2011. "Analyzing Experimental Data Using Regression: When Is Bias a Practical Concern?" *SSRN Working Paper*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1466886.

References II

Hernán, Miguel A. 2018. "The c-Word: Scientific Euphemisms Do Not Improve Causal Inference from Observational Data." *American Journal of Public Health* 108 (5): 616–19.

Hernán, Miguel A., John Hsu, and Brian Healy. 2019. "A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks." *CHANCE* 32 (1): 42–49. https://doi.org/10.1080/09332480.2019.1579578.

Lin, Winston. 2013. "Agnostic Notes on Regression Adjustments to Experimental Data: Reexaminig Freedman's Critique." *Annals of Applied Statistics*, 295–318. https://doi.org/10.1214/12-AOAS583.