POLI666: Lab #1

Olivier Bergeron-Bouutin

January 19th, 2021

Moving from predictive to causal inference

Do we care about bias?

Remember this question from 618's 5th problem set?

- Is a biased estimate necessarily a sign of a useless model? Under what circumstances would you care if an estimate is biased, and under what circumstances would you not?

Do we care about bias?

Remember this question from 618's 5th problem set?

- Is a biased estimate necessarily a sign of a useless model? Under what circumstances would you care if an estimate is biased, and under what circumstances would you not?

How did we define bias?

Do we care about bias?

Remember this question from 618's 5th problem set?

- Is a biased estimate necessarily a sign of a useless model? Under what circumstances would you care if an estimate is biased, and under what circumstances would you not?

How did we define bias?

For regression coefficients: $E[\hat{\beta}] \neq \beta$

Do we care about bias?

Remember this question from 618's 5th problem set?

- Is a biased estimate necessarily a sign of a useless model? Under what circumstances would you care if an estimate is biased, and under what circumstances would you not?

How did we define bias?

For regression coefficients: $E[\hat{\beta}] \neq \beta$

More generally: $E[\hat{\theta}] \neq \theta$

Do we care about bias?

Remember this question from 618's 5th problem set?

- Is a biased estimate necessarily a sign of a useless model? Under what circumstances would you care if an estimate is biased, and under what circumstances would you not?

How did we define bias?

For regression coefficients: $E[\hat{\beta}] \neq \beta$

More generally: $E[\hat{\theta}] \neq \theta$

We didn't *necessarily* care about bias

Do we care about bias?

Remember this question from 618's 5th problem set?

- Is a biased estimate necessarily a sign of a useless model? Under what circumstances would you care if an estimate is biased, and under what circumstances would you not?

How did we define bias?

For regression coefficients: $E[\hat{\beta}] \neq \beta$

More generally: $E[\hat{\theta}] \neq \theta$

We didn't *necessarily* care about bias

- As long as our models could produce predictions that satisfied whichever inferential goal we had

Do we care about bias?

Remember this question from 618's 5th problem set?

- Is a biased estimate necessarily a sign of a useless model? Under what circumstances would you care if an estimate is biased, and under what circumstances would you not?

How did we define bias?

For regression coefficients: $E[\hat{\beta}] \neq \beta$

More generally: $E[\hat{\theta}] \neq \theta$

We didn't *necessarily* care about bias

- As long as our models could produce predictions that satisfied whichever inferential goal we had
- ...Any examples?

Intent of a model

Consider the following linear model from Gerber, Green and Larimer (2008):

$$\text{Turnout}_i = \beta_0 + \beta_1 \text{CivicDuty}_i + \beta_2 \text{Hawthorne}_i + \beta_3 \text{Self}_i + \beta_4 \text{Neighbors}_i + \epsilon_i \quad (1)$$

	Model 1
(Intercept)	0.297 [0.295, 0.299]
civicduty	0.018 [0.013, 0.023]
hawthorne	0.026 [0.021, 0.031]
self	0.049 [0.043, 0.054]
neighbors	0.081 [0.076, 0.086]
Num.Obs.	344084
R2	0.003
R2 Adj.	0.003

Intent of a model

What happens if we generate fitted values for our model?

Intent of a model

What happens if we generate fitted values for our model?

- Very poor predictive accuracy: the model can only generate 5 distinct fitted values: $E[\text{Turnout}|D_i]$

Intent of a model

What happens if we generate fitted values for our model?

- Very poor predictive accuracy: the model can only generate 5 distinct fitted values: $E[\text{Turnout}|D_i]$
- Yet our model is very useful because the treatments were randomly assigned: $\{Y_i(1), Y_i(0)\} \perp D_i$

Intent of a model

What happens if we generate fitted values for our model?

- Very poor predictive accuracy: the model can only generate 5 distinct fitted values: $E[\text{Turnout}|D_i]$
- Yet our model is very useful because the treatments were randomly assigned: $\{Y_i(1), Y_i(0)\} \perp D_i$
- Conversely, I can have a model riddled with biased coefficients but with high predictive power

Some food for thought

What can we do with this?

Cool causal inference work that can get you thinking:

What can we do with this?

Cool causal inference work that can get you thinking:

Eggers and Hainmueller 2009 (RDD)

- Exploit randomness of close elections to estimate “returns to office”: political candidates who barely won died with substantially larger wealth than candidates who barely lost

What can we do with this?

Cool causal inference work that can get you thinking:

Eggers and Hainmueller 2009 (RDD)

- Exploit randomness of close elections to estimate “returns to office”: political candidates who barely won died with substantially larger wealth than candidates who barely lost

Kalla and Broockman 2016 (RCT)

- Letters sent to congressmembers by a political organization randomly mention/don’t mention campaign donations made by the organization to the legislator; mentioning donations leads to more access

What can we do with this?

Cool causal inference work that can get you thinking:

Eggers and Hainmueller 2009 (RDD)

- Exploit randomness of close elections to estimate “returns to office”: political candidates who barely won died with substantially larger wealth than candidates who barely lost

Kalla and Broockman 2016 (RCT)

- Letters sent to congressmembers by a political organization randomly mention/don’t mention campaign donations made by the organization to the legislator; mentioning donations leads to more access

What can we do with this?

Garz and Martin 2020 (RDD)

- How do we differentiate the effect of economic conditions from the effect of media coverage of the economy? Crossing an unemployment “milestone” (i.e. crossing a round number) is quasi-random and produces an exogenous change in media coverage.

What can we do with this?

Garz and Martin 2020 (RDD)

- How do we differentiate the effect of economic conditions from the effect of media coverage of the economy? Crossing an unemployment “milestone” (i.e. crossing a round number) is quasi-random and produces an exogenous change in media coverage.

Fournaies and Hall 2018 (DiD)

- Use data data from the 99 state legislatures on legislators’ committee assignments; assignment to legislative committees causes an increase in campaign donations from relevant interest groups.

What can we do with this?

Garz and Martin 2020 (RDD)

- How do we differentiate the effect of economic conditions from the effect of media coverage of the economy? Crossing an unemployment "milestone" (i.e. crossing a round number) is quasi-random and produces an exogenous change in media coverage.

Fournaies and Hall 2018 (DiD)

- Use data data from the 99 state legislatures on legislators' committee assignments; assignment to legislative committees causes an increase in campaign donations from relevant interest groups.

Acharya, Blackwell and Sen 2016 (IV)

- Cotton suitability as an instrument for % of Southern counties' population that is black before the abolition of slavery.

Epistemic modesty?

Causal inference is hard! We must keep in mind several questions/elements:

Epistemic modesty?

Causal inference is hard! We must keep in mind several questions/elements:

- In the social sciences, treatment effects are typically small!

Epistemic modesty?

Causal inference is hard! We must keep in mind several questions/elements:

- In the social sciences, treatment effects are typically small!
- Is my research question substantively interesting?

Epistemic modesty?

Causal inference is hard! We must keep in mind several questions/elements:

- In the social sciences, treatment effects are typically small!
- Is my research question substantively interesting?
- Is it non-trivial?

Epistemic modesty?

Causal inference is hard! We must keep in mind several questions/elements:

- In the social sciences, treatment effects are typically small!
- Is my research question substantively interesting?
- Is it non-trivial?
- Do I have a plausible identification strategy?

Epistemic modesty?

Causal inference is hard! We must keep in mind several questions/elements:

- In the social sciences, treatment effects are typically small!
- Is my research question substantively interesting?
- Is it non-trivial?
- Do I have a plausible identification strategy?
- Do I have the statistical power to detect anticipated τ?

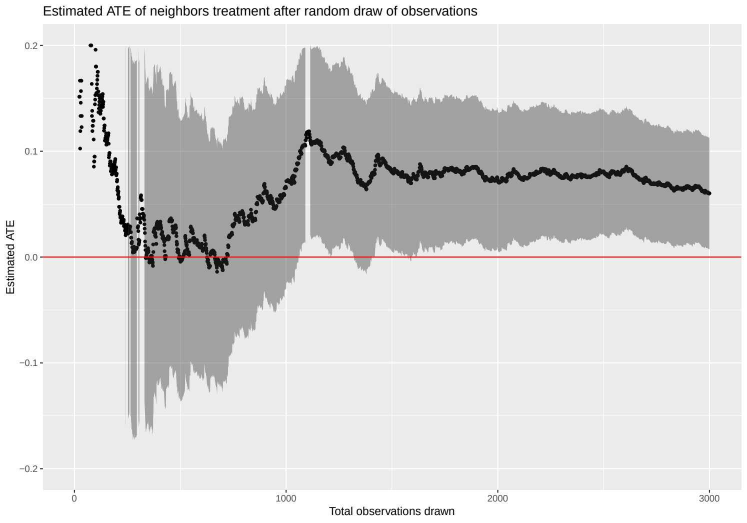
Epistemic modesty?

Causal inference is hard! We must keep in mind several questions/elements:

- In the social sciences, treatment effects are typically small!
- Is my research question substantively interesting?
- Is it non-trivial?
- Do I have a plausible identification strategy?
- Do I have the statistical power to detect anticipated τ?
- Does the treatment effect identified in my sample generalize?

Effect size in Gerber et al. 2008

Rolling ATE as we draw random observations from the Gerber et al. sample:



What can we NOT do?

Perhaps even more important is recognizing the limits of our methods

What can we NOT do?

Perhaps even more important is recognizing the limits of our methods

- In terms of isolating exogenous change

What can we NOT do?

Perhaps even more important is recognizing the limits of our methods

- In terms of isolating exogenous change
- And in terms of reliably estimating effects of fairly small magnitude

What can we NOT do?

Perhaps even more important is recognizing the limits of our methods

- In terms of isolating exogenous change
- And in terms of reliably estimating effects of fairly small magnitude

We do *not* want to approach this atheoretically! Given your substantive knowledge, the estimated effect should pass the “smell test”

What can we NOT do?

Perhaps even more important is recognizing the limits of our methods

- In terms of isolating exogenous change
- And in terms of reliably estimating effects of fairly small magnitude

We do *not* want to approach this atheoretically! Given your substantive knowledge, the estimated effect should pass the “smell test”

- John Stewart paper in *Electoral Studies*: the claim that John Stewart leaving The Daily Show increased county-level vote for Trump by 1.1% just does not sound right! (paper was retracted due to a coding error)

What can we NOT do?

Perhaps even more important is recognizing the limits of our methods

- In terms of isolating exogenous change
- And in terms of reliably estimating effects of fairly small magnitude

We do *not* want to approach this atheoretically! Given your substantive knowledge, the estimated effect should pass the "smell test"

- John Stewart paper in *Electoral Studies*: the claim that John Stewart leaving The Daily Show increased county-level vote for Trump by 1.1% just does not sound right! (paper was retracted due to a coding error)
- Hannity vs Carlson paper: "a one standard deviation increase in relative viewership of *Hannity* relative to *Tucker Carlson Tonight* is associated with approximately 34 percent more COVID-19 cases on March 14 and approximately 24 percent more COVID-19 deaths on March 28."

Notation and terminology

Three conceptual levels

We have to differentiate three terms:

Three conceptual levels

We have to differentiate three terms:

Estimand

An unknown value that describes a population relationship. It's the value that we want to estimate. We'll call this θ.

Three conceptual levels

We have to differentiate three terms:

Estimand

An unknown value that describes a population relationship. It's the value that we want to estimate. We'll call this θ.

↓

Estimator

A rule to produce a numerical value that represents the estimand.

↓

Three conceptual levels

We have to differentiate three terms:

Estimand

An unknown value that describes a population relationship. It's the value that we want to estimate. We'll call this θ.

↓

Estimator

A rule to produce a numerical value that represents the estimand.

↓

Estimate

The numerical value taken on by an estimator for a particular sample of data.

Three conceptual levels

Some things to keep in mind:

Three conceptual levels

Some things to keep in mind:

1. Estimators produce estimates.

Three conceptual levels

Some things to keep in mind:

1. Estimators produce estimates.
2. Under repeated sampling, estimators produce many estimates.

Three conceptual levels

Some things to keep in mind:

1. Estimators produce estimates.
2. Under repeated sampling, estimators produce many estimates.
3. Estimates are a random variable, i.e. they have a stochastic/random component.

Three conceptual levels

Some things to keep in mind:

1. Estimators produce estimates.
2. Under repeated sampling, estimators produce many estimates.
3. Estimates are a random variable, i.e. they have a stochastic/random component.
4. Under repeated sampling, estimates have a distribution, which we call a sampling distribution.

Three conceptual levels

Some things to keep in mind:

1. Estimators produce estimates.
2. Under repeated sampling, estimators produce many estimates.
3. Estimates are a random variable, i.e. they have a stochastic/random component.
4. Under repeated sampling, estimates have a distribution, which we call a sampling distribution.
5. Estimators have different properties.

Potential outcomes framework

Potential outcomes

For each unit i, we imagine two states of the world: under treatment and under control

Potential outcomes

For each unit i, we imagine two states of the world: under treatment and under control

Realized outcome

$$Y_i(d) = \begin{cases} Y_i(1), & \text{Potential outcome for unit } i \text{ under treatment} \\ Y_i(0), & \text{Potential outcome for unit } i \text{ under control} \end{cases} \tag{2}$$

Potential outcomes

For each unit i, we imagine two states of the world: under treatment and under control

Realized outcome

$$Y_i(d) = \begin{cases} Y_i(1), & \text{Potential outcome for unit } i \text{ under treatment} \\ Y_i(0), & \text{Potential outcome for unit } i \text{ under control} \end{cases} \qquad (2)$$

If my treatment is going to the hospital:

- $Y_i(1)$ is the value that outcome Y would take if unit i went to the hospital

Potential outcomes

For each unit i, we imagine two states of the world: under treatment and under control

Realized outcome

$$Y_i(d) = \begin{cases} Y_i(1), & \text{Potential outcome for unit } i \text{ under treatment} \\ Y_i(0), & \text{Potential outcome for unit } i \text{ under control} \end{cases} \tag{2}$$

If my treatment is going to the hospital:

- $Y_i(1)$ is the value that outcome Y would take if unit i went to the hospital
- $Y_i(0)$ is the value that outcome Y would take if unit i did not go to the hospital

Estimand: Individual-level treatment effect

We then define the individual-level treatment effect as the difference in potential outcomes for unit i between treatment and control:

Estimand: Individual-level treatment effect

We then define the individual-level treatment effect as the difference in potential outcomes for unit i between treatment and control:

$$\tau_i = Y_i(1) - Y_i(0)$$

Estimand: Individual-level treatment effect

We then define the individual-level treatment effect as the difference in potential outcomes for unit i between treatment and control:

$$\tau_i = Y_i(1) - Y_i(0)$$

Will I ever observe this quantity?

Estimand: Individual-level treatment effect

We then define the individual-level treatment effect as the difference in potential outcomes for unit i between treatment and control:

$$\tau_i = Y_i(1) - Y_i(0)$$

Will I ever observe this quantity? **NO!** That's the fundamental problem of causal inference

- And why we need to learn all of this complicated stuff in the first place...

Fundamental problem of causal inference:

Only one of the potential outcomes is realized:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) = \begin{cases} Y_i(1), & \text{if } D_i = 1 \\ Y_i(0), & \text{if } D_i = 0 \end{cases} \tag{3}$$

Fundamental problem of causal inference:

Only one of the potential outcomes is realized:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) = \begin{cases} Y_i(1), & \text{if } D_i = 1 \\ Y_i(0), & \text{if } D_i = 0 \end{cases} \tag{3}$$

FPCI: this parameter τ_i can never be observed!

Fundamental problem of causal inference:

Only one of the potential outcomes is realized:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) = \begin{cases} Y_i(1), & \text{if } D_i = 1 \\ Y_i(0), & \text{if } D_i = 0 \end{cases} \tag{3}$$

FPCI: this parameter τ_i can never be observed!

- Unit i is either treated or it is not treated

Fundamental problem of causal inference:

Only one of the potential outcomes is realized:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) = \begin{cases} Y_i(1), & \text{if } D_i = 1 \\ Y_i(0), & \text{if } D_i = 0 \end{cases} \tag{3}$$

FPCI: this parameter τ_i can never be observed!

- Unit i is either treated or it is not treated
- I observe either $Y_i(1)$ OR $Y_i(0)$, but never both

Fundamental problem of causal inference:

Only one of the potential outcomes is realized:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0) = \begin{cases} Y_i(1), & \text{if } D_i = 1 \\ Y_i(0), & \text{if } D_i = 0 \end{cases} \tag{3}$$

FPCI: this parameter τ_i can never be observed!

- Unit i is either treated or it is not treated
- I observe either $Y_i(1)$ OR $Y_i(0)$, but never both
- Causal inference as a problem of missing data

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i
1	2	1	?
1	3	3	?
0	5	4	?
1	3	1	?
0	2	4	?

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i
1	2	1	2
1	3	3	3
0	5	4	4
1	3	1	3
0	2	4	4

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	?
1	3	3	3	?
0	5	4	4	?
1	3	1	3	?
0	2	4	4	?

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

Ideal data

Without the fundamental problem of causal inference, this is what our data would look like:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	1	2	1
1	3	3	3	0
0	5	4	4	1
1	3	1	3	2
0	2	4	4	-2

I can compute τ_i since for each unit i, I have access to both potential outcomes

Real data

With the fundamental problem of causal inference, we only see one realized outcome Y_i for each unit i:

Real data

With the fundamental problem of causal inference, we only see one realized outcome Y_i for each unit i:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	?	2	?
1	3	?	3	?
0	?	4	4	?
1	3	?	3	?
0	?	4	4	?

Real data

With the fundamental problem of causal inference, we only see one realized outcome Y_i for each unit i:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	?	2	?
1	3	?	3	?
0	?	4	4	?
1	3	?	3	?
0	?	4	4	?

As a result, we cannot observe τ_i

- And it is unidentified unless we make strong assumptions

Real data

With the fundamental problem of causal inference, we only see one realized outcome Y_i for each unit i:

D_i	$Y_i(1)$	$Y_i(0)$	Y_i	τ_i
1	2	?	2	?
1	3	?	3	?
0	?	4	4	?
1	3	?	3	?
0	?	4	4	?

As a result, we cannot observe τ_i

- And it is unidentified unless we make strong assumptions
- Namely: $\tau_i = \tau \; \forall i$

Other estimands

We will never observe τ_i and can (in practice) never plausibly claim to be able to identify it

Other estimands

We will never observe τ_i and can (in practice) never plausibly claim to be able to identify it

- So we will have to focus on other quantities of interest

Other estimands

We will never observe τ_i and can (in practice) never plausibly claim to be able to identify it

- So we will have to focus on other quantities of interest

Average treatment effect (ATE)

$$\tau = \mathbb{E}[\tau_i] = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$$

Other estimands

We will never observe τ_i and can (in practice) never plausibly claim to be able to identify it

- So we will have to focus on other quantities of interest

Average treatment effect (ATE)

$\tau = \mathbb{E}[\tau_i] = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$

We can further disaggregate the above into Population ATE (PATE) and Sample ATE (SATE)

Other estimands

We will never observe τ_i and can (in practice) never plausibly claim to be able to identify it

- So we will have to focus on other quantities of interest

Average treatment effect (ATE)

$\tau = \mathbb{E}[\tau_i] = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$

We can further disaggregate the above into Population ATE (PATE) and Sample ATE (SATE)

Conditional Average treatment effect (CATE)

$\tau(x) = \mathbb{E}[\tau_i | X_i = x]$

Other estimands

We will never observe τ_i and can (in practice) never plausibly claim to be able to identify it

- So we will have to focus on other quantities of interest

Average treatment effect (ATE)

$$\tau = \mathbb{E}[\tau_i] = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$$

We can further disaggregate the above into Population ATE (PATE) and Sample ATE (SATE)

Conditional Average treatment effect (CATE)

$$\tau(x) = \mathbb{E}[\tau_i | X_i = x]$$

ATE for some subpopulation of interest (heterogenous effects)

Other estimands

We will never observe τ_i and can (in practice) never plausibly claim to be able to identify it

- So we will have to focus on other quantities of interest

Average treatment effect (ATE)

$$\tau = \mathbb{E}[\tau_i] = \mathbb{E}[Y_i(1) - Y_i(0)] = \mathbb{E}[Y_i(1)] - \mathbb{E}[Y_i(0)]$$

We can further disaggregate the above into Population ATE (PATE) and Sample ATE (SATE)

Conditional Average treatment effect (CATE)

$$\tau(x) = \mathbb{E}[\tau_i | X_i = x]$$

ATE for some subpopulation of interest (heterogenous effects)

But, again, we cannot actually compute $\mathbb{E}[\tau_i]$ because we never observe τ_i

This is where causal **inference** comes in

This is where causal **inference** comes in

- We have "missing data" that prevents us from observing τ_i, and therefore τ

This is where causal **inference** comes in

- We have "missing data" that prevents us from observing τ_i, and therefore τ
- We have to use the observed (incomplete) data to *infer* the estimand

The role of randomization

Why do we need randomization?

MHE: randomization solves the selection problem

Why do we need randomization?

MHE: randomization solves the selection problem

In practice, what does that mean?

Why do we need randomization?

MHE: randomization solves the selection problem

In practice, what does that mean?

- What is the selection problem?

Why do we need randomization?

MHE: randomization solves the selection problem

In practice, what does that mean?

- What is the selection problem?
- How would we state the role of randomization in mathematical notation?

Why do we need randomization?

MHE: randomization solves the selection problem

In practice, what does that mean?

- What is the selection problem?
- How would we state the role of randomization in mathematical notation?

$$\{Y_i(1), Y_i(0)\} \perp D_i$$

MHE: randomization solves the selection problem

In practice, what does that mean?

- What is the selection problem?
- How would we state the role of randomization in mathematical notation?

$$\{Y_i(1), Y_i(0)\} \perp D_i$$

$$\mathbb{E}[Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(0)|D_i = 0]$$

Why do we need randomization?

MHE: randomization solves the selection problem

In practice, what does that mean?

- What is the selection problem?
- How would we state the role of randomization in mathematical notation?

$$\{Y_i(1), Y_i(0)\} \perp D_i$$

$$\mathbb{E}[Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(0)|D_i = 0]$$

Is this something we can empirically test?

Why do we need randomization?

MHE: randomization solves the selection problem

In practice, what does that mean?

- What is the selection problem?
- How would we state the role of randomization in mathematical notation?

$$\{Y_i(1), Y_i(0)\} \perp D_i$$

$$\mathbb{E}[Y_i(0)|D_i = 1] = \mathbb{E}[Y_i(0)|D_i = 0]$$

Is this something we can empirically test?

No: I still can't observe the potential outcomes.

Balance checks in R

We'll use Dunning and Nilekani (2013), who investigate the effect of ethnic quotas on redistribution in India. The unit of analysis is the village council.

```
dunning <- read_dta("data/dunning_bal.dta")
```

Balance checks in R

We'll use Dunning and Nilekani (2013), who investigate the effect of ethnic quotas on redistribution in India. The unit of analysis is the village council.

```
dunning <- read_dta("data/dunning_bal.dta")
```

Our **pre-treatment** (emphasis important!) covariates are:

- *P_ILL*: Mean number of illiterates
- *MARGWORK_P*: Mean number of marginal workers
- *No_HH*: Number of households
- *MAIN_AL_P*: Mean agricultural laborers
- *MAIN_CL_P*: Mean cultivators
- *NON_WORK_F*: Mean female nonworkers

Balance checks in R

We'll use Dunning and Nilekani (2013), who investigate the effect of ethnic quotas on redistribution in India. The unit of analysis is the village council.

```r
dunning <- read_dta("data/dunning_bal.dta")
```

Our **pre-treatment** (emphasis important!) covariates are:

- *P_ILL*: Mean number of illiterates
- *MARGWORK_P*: Mean number of marginal workers
- *No_HH*: Number of households
- *MAIN_AL_P*: Mean agricultural laborers
- *MAIN_CL_P*: Mean cultivators
- *NON_WORK_F*: Mean female nonworkers

We want to show the control and treatment groups are not systematically different. There are many ways of going about this! Ideas?