GOV 1000/ 2000e/ 2000/ E-2000 Section 11

Solé Prillaman

Harvard University

November 14, 2013

Logistics

Problem Set 9 - Due by 1pm on Tuesday to course dropbox.

Problem Set 8 Corrections - Due by 1pm on Tuesday to course dropbox.

Reading Quiz - Due by 9pm on Sunday on course website.

Randomized Experiment Example

- We want to study how information on unemployment insurance programs affects the likelihood of applying for benefits

  Treatment: receive information on unemployment insurance program (how it works)

  $T = 1$ if receive information, $T = 0$ if do not receive information

  Outcome: likelihood of applying for benefits (measured from 1 to 5)

  $Y \in \{1, 2, 3, 4, 5\}$, where 5 indicates very likely to apply

POTENTIAL OUTCOMES TABLE

i	Name	T_i	Y_i	$Y_i(1)$	$Y_i(0)$
1	Benjamin Franklin	?	?	?	?
2	Thomas Jefferson	?	?	?	?
3	George Washington	?	?	?	?
4	Alexander Hamilton	?	?	?	?
5	James Madison	?	?	?	?
6	John Jay	?	?	?	?
7	Roger Sherman	?	?	?	?
8	Charles Pinckney	?	?	?	?
9	John Hancock	?	?	?	?
10	Robert Morris	?	?	?	?

Assumptions of Causal Inference

SUTVA

1. No interference – the treatment value of unit i does not affect the *potential outcomes* of unit j for all $i \neq j$
2. No Multiple Versions of Treatment *or* Treatment Variation Irrelevance (multiple versions of treatment may exist but all result in same outcome $Y_i(a)$)

In our example, when would this not hold?

Note: This assumption is implicit in Hernan and Robins

Assumptions of Causal Inference

Consistency

$$Y_i(1) = Y_i | T_i = 1 \text{ and } Y_i(0) = Y_i | T_i = 0$$

$$\Rightarrow \quad Y_i = T_i \cdot Y_i(1) + (1 - T_i) \cdot Y_i(0)$$

Note: In Rubin framework this is included in SUTVA

Assumptions of Causal Inference

Exchangeability (Unconfoundedness)

- The observed treatment is independent of the *potential outcomes*
- $Y_i(0), Y_i(1) \perp\!\!\!\perp T_i$ (Hernan and Robins)
- $P(T_i|Y_i(0), Y_i(1)) = P(T_i)$ (Rubin)

In our example, does this hold?

Yes - because treatment was randomly assigned!

When would this not hold?

Potential Outcomes Table

i	Name	T_i	Y_i	$Y_i(1)$	$Y_i(0)$
1	Benjamin Franklin	?	?	3	2
2	Thomas Jefferson	?	?	3	3
3	George Washington	?	?	5	2
4	Alexander Hamilton	?	?	1	2
5	James Madison	?	?	2	1
6	John Jay	?	?	4	3
7	Roger Sherman	?	?	3	4
8	Charles Pinckney	?	?	4	5
9	John Hancock	?	?	4	3
10	Robert Morris	?	?	1	3

Assumptions of Causal Inference

What assumption did we make in order to write down the potential outcomes in the table like this? **SUTVA**

Potential Outcomes Table

i	Name	T_i	Y_i	$Y_i(1)$	$Y_i(0)$
1	Benjamin Franklin	1	3	3	2
2	Thomas Jefferson	0	3	3	3
3	George Washington	1	5	5	2
4	Alexander Hamilton	0	2	1	2
5	James Madison	1	2	2	1
6	John Jay	0	3	4	3
7	Roger Sherman	1	3	3	4
8	Charles Pinckney	1	4	4	5
9	John Hancock	0	3	4	3
10	Robert Morris	1	1	1	3

Assumptions of Causal Inference

What assumption enabled us to write fill in Y_i the way we did, i.e. in terms of the corresponding potential outcomes?

Consistency

Observed Table

i	Name	T_i	Y_i	$Y_i(1)$	$Y_i(0)$
1	Benjamin Franklin	1	3	3	?
2	Thomas Jefferson	0	3	?	3
3	George Washington	1	5	5	?
4	Alexander Hamilton	0	2	?	2
5	James Madison	1	2	2	?
6	John Jay	0	3	?	3
7	Roger Sherman	1	3	3	?
8	Charles Pinckney	1	4	4	?
9	John Hancock	0	3	?	3
10	Robert Morris	1	1	1	?

Causal inference is a missing data problem!

What does exchangeability tell us about our *missingness mechanism*? **MCAR**

Defining the Estimand

Given our experiment and data, what are some potential causal quantities of interest we may be interested in?

- Average treatment effect: $E[Y_i(1) - Y_i(0)]$
- Stratum-specific average treatment effect: $E[Y_i(1) - Y_i(0)|X_i = x]$
- Generally, anything at all that can be written in terms of the potential outcomes in our table!

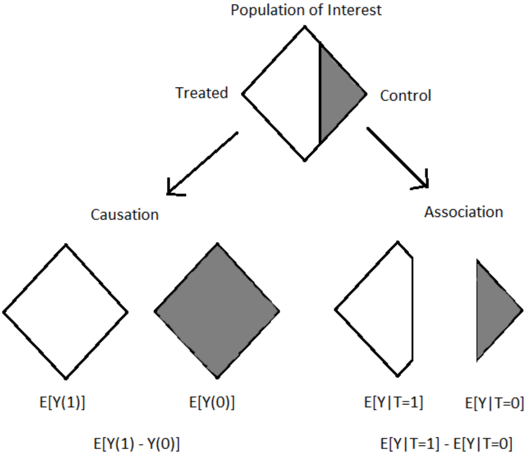
Estimating the ATE

Estimand $E[Y_i(1) - Y_i(0)]$

Estimator ?

Estimate?

Estimating the ATE



Under what assumption can we estimate $E[Y_i(1) - Y_i(0)]$ with $E[Y|T = 1] - E[Y|T = 0]$? **Exchangeability**

ESTIMATING THE ATE

Estimand $E[Y_i(1) - Y_i(0)]$

Estimator $E[Y|T = 1] - E[Y|T = 0]$

Estimate $3 - 2.75 = \mathbf{.25}$

ESTIMATING THE ATE

So what did we do about our missingness? **Mean Imputation**

i	Name	T_i	Y_i	$Y_i(1)$	$Y_i(0)$
1	Benjamin Franklin	1	3	3	2.75
2	Thomas Jefferson	0	3	3	3
3	George Washington	1	5	5	2.75
4	Alexander Hamilton	0	2	3	2
5	James Madison	1	2	2	2.75
6	John Jay	0	3	3	3
7	Roger Sherman	1	3	3	2.75
8	Charles Pinckney	1	4	4	2.75
9	John Hancock	0	3	3	3
10	Robert Morris	1	1	1	2.75

Are our imputed estimates biased? **No**

Link to Bivariate Regression

How does this estimator relate to our regression estimator, i.e. β_1 from $Y_i = \beta_0 + \beta_1 T_i$?

$$\begin{aligned} E[Y_i(1) - Y_i(0)] &= E[Y_i(1)] - E[Y_i(0)] \\ &= E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0] \\ &= E[Y_i|T_i = 1] - E[Y_i|T_i = 0] \\ &= \beta_1 \end{aligned}$$

They are the same!

Violation of Assumptions

Note that if exchangeability fails to hold, the simple difference-in-means estimate is no longer unbiased because:

$$E[Y_i(1)] - E[Y_i(0)] \neq E[Y_i(1)|T_i = 1] - E[Y_i(0)|T_i = 0]$$

Intuitively, it is wrong because we are comparing apples and oranges. We are imputing missing potential outcomes for apples using oranges.

Violation of Assumptions

If you estimate the ATE using regression, all of the assumptions needed for descriptive inference with OLS are needed:

- Random Sample of Observations
- Linear Conditional Expectation Function
- Constant Conditional Variance
- Normality of Conditional Y

Conditionally Randomized Experiment Example

Instead of executing the unemployment insurance experiment from before, you decide to condition treatment assignment on *employment status.* Treatment is randomized within levels of employment status.

$X = 1$ if employed, $X = 0$ if unemployed

Is X pre-treatment or post-treatment?

Quiz Break!

Under our new experiment, can we still estimate the ATE using the procedure from before? **No**

What assumption have we violated? **Exchangeability**

ASSUMPTIONS

Conditional Exchangeability

- The observed treatment is independent of the *potential outcomes*, conditional on levels of X.
- $Y_i(0), Y_i(1) \perp\!\!\!\perp T_i | X_i$ (Hernan and Robins)
- $P(T_i | Y_i(0), Y_i(1), X_i) = P(T_i | X_i)$ (Rubin)

In our example, does this hold?

Yes - because treatment was conditionally randomly assigned!

When would this not hold?

Assumptions

Conditional exchangeability implies that *within* the levels of X (Strata), we have exchangeability.

$$
\begin{aligned}
\mathrm{E}[Y_i(1) - Y_i(0)|X_i = x] &= \mathrm{E}[Y_i(1)|X_i] - \mathrm{E}[Y_i(0)|X_i = x] \\
&= \mathrm{E}[Y_i(1)|T_i = 1, X_i = x] - \mathrm{E}[Y_i(0)|T_i = 0, X_i = x] \\
&= \mathrm{E}[Y_i|T_i = 1, X_i = x] - \mathrm{E}[Y_i|T_i = 0, X_i = x]
\end{aligned}
$$

Observed Table

i	Name	X_i	T_i	Y_i	$Y_i(1)$	$Y_i(0)$
1	Benjamin Franklin	1	1	3	3	?
2	Thomas Jefferson	1	0	3	?	3
3	George Washington	1	1	5	5	?
4	Alexander Hamilton	1	0	2	?	2
5	James Madison	1	1	2	2	?
6	John Jay	1	0	3	?	3
7	Roger Sherman	0	1	3	3	?
8	Charles Pinckney	0	1	4	4	?
9	John Hancock	0	0	3	?	3
10	Robert Morris	0	1	1	1	?

What does conditional exchangeability tell us about our *missingness mechanism*? **MAR**

Estimating the Stratum-specific ATE

Estimand $E[Y_i(1) - Y_i(0) | X_i = x]$

Estimator $E[Y_i | T_i = 1, X_i = x] - E[Y_i | T_i = 0, X_i = x]$

Estimate $\begin{cases} \text{for } x = 1 & \frac{10}{3} - \frac{8}{3} = \frac{2}{3} \\ \text{for } x = 0 & \frac{8}{3} - 3 = -\frac{1}{3} \end{cases}$

Estimating the Stratum-specific ATE

So what did we do about our missingness? **Conditional Mean Imputation**

i	Name	X_i	T_i	Y_i	$Y_i(1)$	$Y_i(0)$
1	Benjamin Franklin	1	1	3	3	$\frac{8}{3}$
2	Thomas Jefferson	1	0	3	$\frac{10}{3}$	3
3	George Washington	1	1	5	5	$\frac{8}{3}$
4	Alexander Hamilton	1	0	2	$\frac{10}{3}$	2
5	James Madison	1	1	2	2	$\frac{8}{3}$
6	John Jay	1	0	3	$\frac{10}{3}$	3
7	Roger Sherman	0	1	3	3	3
8	Charles Pinckney	0	1	4	4	3
9	John Hancock	0	0	3	$\frac{8}{3}$	3
10	Robert Morris	0	1	1	1	3

Are our imputed estimates biased? **No**

ESTIMATING THE ATE

But wait! What if we wanted to estimate the ATE?

Estimand $E[Y_i(1) - Y_i(0)]$

Estimator ?

Estimate ?

Estimating the ATE

There are several ways to calculate an overall ATE in a conditionally randomized experiment that capture the idea of combining strata-specific ATEs:

- Weight strata-specific ATEs by size of strata:
  - Standardization
  - IP weighting
  - Interactive regression
- Weight strata-specific ATEs by size of strata and within-strata variance of treatment indicator:
  - Additive regression

Estimating the ATE with Standardization

Estimand $E[Y_i(1) - Y_i(0)]$

Estimator $\sum_x [E[Y_i|T_i = 1, X_i = x] - E[Y_i|T_i = 0, X_i = x]] \cdot P(X_i = x)$

Estimate $\frac{2}{3} \cdot \frac{3}{5} + \frac{-1}{3} \cdot \frac{2}{5} = \frac{4}{15}$

Link to Interactive Regression

It turns out that an interactive regression model will give you the same results as standardization:

$$\mathrm{E}[Y|T=t, X=x] = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot x + \beta_3 \cdot t \cdot x$$

We can interpret $\beta_1 + \beta_3 \cdot x$ as a conditional difference-in-means between treated and control groups:

$$\mathrm{E}[Y|T=1, X=x] \text{ - } \mathrm{E}[Y|T=0, X=x]$$

Note that for $X = x$, we get out stratum-specific ATEs:

- $X = 0 \Rightarrow = \beta_1$
- $X = 1 \Rightarrow = \beta_1 + \beta_3$

Link Interactive Regression

To get the overall ATE, we can use the law of total expectation:

$$E[Y_i(1) - Y_i(0)] = \beta_1 + \beta_3 \cdot E_X[X]$$

Therefore:

$$\widehat{ATE} = \hat{\beta}_1 + \hat{\beta}_3 \cdot \bar{X}$$

In practice:

- Recenter X by subtracting off $\bar{X}$ (de-meaning)
- Re-run interactive regression above using demeaned X
- $\widehat{ATE}$ will now be $\hat{\beta}_1$

Conclusion

Questions?