Statistical Modeling and Causal Inference with R

Week 2: Potential Outcomes Framework

Manuel Bosancianu

Max Schaub

September 14, 2020

Hertie School of Governance

Today's focus

- ✓ Potential outcomes framework (POF)
- ✓ Biases in causal inference
- ✓ Assumptions underpinning POF
- ✓ Randomization (RCTs) in POF

Last week

In public policy, we are frequently interested in the causal effect* of an intervention (or phenomenon) on an important outcome.

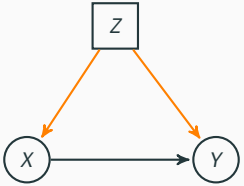
Assuming we have data to answer this, it's easy to compute the magnitude of association.

Example from France in 2014 (with ESS data): -0.14; the "contact hypothesis" (Allport, 1954) receives support.

Last week

Difficulties stem from the multiple dynamics that generate an association between the two. Naturally, *contact* $\Rightarrow$ *prejudice*.

What other reasons might there be for an association?



In this case, Z might be education.

Last week

Causal inference is primarily a process of *reasoning*:

- ✓ eliminating alternative (non-causal) explanations for the association
- ✓ making explicit the assumptions needed for a causal relationship

It requires an empirical step: marshalling data to show these arguments hold.

We rely on a *probabilistic* understanding of causality: causes change the probabilities of their effects.

Potential Outcomes Framework

Foundations

The POF is a very convenient way to illustrate concepts like *confounding* or *selection bias*.

Each individual *i* in a population can be exposed ($d_i = 1$) to a *cause*, or not exposed ($d_i = 0$) (e.g., contact with member of different ethnic group).

For each *i*, assume we can assess an outcome (e.g., prejudice level) *both* after being exposed and not exposed: $y_{d,i}$.

$y_{0,i}$ = non-exposure $\qquad$ $y_{1,i}$ = exposure

Notation

Angrist and Pischke (2015)	y_{0i} and y_{1i}
Cunningham (2021)	y_i^0 and y_i^1
Gerber and Green (2012)	$y_i(0)$ and $y_i(1)$

Notation differences

We follow Angrist and Pischke (2015) in notation for the rest of the class.

Individual treatment effect (ITE)

$$ITE = \delta_i = y_{1i} - y_{0i} \tag{1}$$

The *ITE* is a "comparison of two states of the world" (Cunningham, 2021): individual *i* exposed to contact, and not exposed to it.

Fundamental problem (Holland, 1986) of causal inference: in reality, we only see one of these two states.

Observable outcomes

Each observable outcome can be described by a *switching equation*:

$$y_i = d_i y_{1i} + (1 - d_i) y_{0i} \tag{2}$$

Once D is fixed, we only see *one of* the values: y_{0i} or y_{1i}.

The *ITE* is unattainable.

Average treatment effect (ATE)

For a population under study, we are typically interested in the ATE:

$$E[\delta_i] = E[y_{1i} - y_{0i}] = E[y_{1i}] - E[y_{0i}] \tag{3}$$

$E[X]$ is the expectation of a random variable, X, computed

$$E[X] = \sum x * p(X = x) \tag{4}$$

The *ATE* is unattainable as well.

Example (intermission)

You are studying a (very small) population of students, assigned to a dorm room with a co-ethnic ($contact = 0$) or a peer from a different ethnic group ($contact = 1$).

Student (i)	Prejudice	
	y_{0i}	y_{1i}
1	6	5
2	4	2
3	4	4
4	6	7
5	3	1
6	2	2
7	8	7
8	4	5

Potential outcomes for 8 students

Example (intermission)

Student (i)	Prejudice		
	y_{0i}	y_{1i}	δ_i
1	6	5	-1
2	4	2	-2
3	4	4	0
4	6	7	1
5	3	1	-2
6	2	2	0
7	8	7	-1
8	4	5	1

ITEs

$$ATE = E[\delta_i] = \frac{-1 + (-2) + 0 + 1 + (-2) + 0 + (-1) + 1}{8} = -0.5 \quad (5)$$

Example (intermission)

Imagine we now also know the treatment assignment of our subjects.

Student (i)	Prejudice			Contact
	y_{0i}	y_{1i}	δ_i	
1	6	5	-1	0
2	4	2	-2	1
3	4	4	0	0
4	6	7	1	0
5	3	1	-2	1
6	2	2	0	1
7	8	7	-1	0
8	4	5	1	0

With treatment assignment

Two more quantities: ATT and...

Average treatment effect among the treated (ATT):

$$\begin{aligned} ATT &= E[\delta_i | d_i = 1] \\ &= E[y_{1i} - y_{0i} | d_i = 1] \\ &= E[y_{1i} | d_i = 1] - E[y_{0i} | d_i = 1] \end{aligned} \tag{6}$$

What is the ATT in our example?

$$ATT = \frac{-2 + (-2) + 0}{3} = -1.333 \tag{7}$$

Two more quantities: ...ATU

Average treatment effect among the untreated (ATU):

$$\begin{aligned} ATU &= E[\delta_i | d_i = 0] \\ &= E[y_{1i} - y_{0i} | d_i = 0] \\ &= E[y_{1i} | d_i = 0] - E[y_{0i} | d_i = 0] \end{aligned} \tag{8}$$

What is the ATU in our example?

$$ATU = \frac{-1 + 0 + 1 + (-1) + 1}{5} = 0 \tag{9}$$

An attainable quantity: NATE

All of the quantities above depend on us having perfect information about alternative states of the world.

Student (i)	Prejudice			Contact
	y_{0i}	y_{1i}	δ_i	
1	6			0
2		2		1
3	4			0
4	6			0
5		1		1
6		2		1
7	8			0
8	4			0

Information we *do* have

An attainable quantity: NATE

This allows us to compute a *naïve* estimate of the *ATE*:

$$\begin{aligned} NATE &= E[y_{1i}|d_i = 1] - E[y_{0i}|d_i = 0] \\ &= \frac{2+1+2}{3} - \frac{6+4+6+8+4}{5} \\ &= 1.666 - 5.6 \\ &= -3.933 \end{aligned} \tag{10}$$

Notice that in this case $NATE \neq ATE$.

Connections: ATE–ATT–ATU

Having perfect information allows us to understand the connection between *ATE*, *ATT*, and *ATU*.

$$ATE = p * ATT + (1 - p) * ATU, \text{ where } p = prob(D = 1) \tag{11}$$

In our example of inter-ethnic contact:

$$\begin{aligned} ATE &= \frac{3}{8}(-1.333) - \frac{5}{8}0 \\ &= -0.5 - 0 \\ &= -0.5 \end{aligned} \tag{12}$$

Connections: NATE–ATE–ATT–ATU

A bit of mathematical derivation (Cunningham, 2021, p. 90) lets us see how these are linked to the *NATE*.

$$NATE = ATE + E[Y_0|D = 1] - E[Y_0|D = 0] + (1 - p)(ATT - ATU) \tag{13}$$

The formula tells us how something we settle for (*NATE*) in actual analyses is related to something we strive for (*ATE*).

Defining biases

Two sources of bias

$$NATE = ATE + \underbrace{E[Y_0|D=1] - E[Y_0|D=0]}_{\textit{selection bias}} + \underbrace{(1-p)(ATT - ATU)}_{\textit{HTE bias}} \tag{14}$$

HTE = heterogeneous treatment effect

Selection bias: the difference in expected outcomes in the absence of treatment for the actual treatment and control group.

Selection bias

Student (i)	Prejudice			Contact
	y_{0i}	y_{1i}	δ_i	
1	6	5	-1	0
2	4	2	-2	1
3	4	4	0	0
4	6	7	1	0
5	3	1	-2	1
6	2	2	0	1
7	8	7	-1	0
8	4	5	1	0

$$
\begin{aligned}
\textit{Bias} &= \frac{4+3+2}{3} - \frac{6+4+6+8+4}{5} \\
&= 3 - 5.6 = -2.6
\end{aligned} \tag{15}
$$

HTE bias

HTE bias: the difference in *returns to treatment* (the treatment effect) between the treatment and control group, multiplied by the share of the population in control.

Student (i)	Prejudice			Contact
	y_{0i}	y_{1i}	δ_i	
1	6	5	-1	0
2	4	2	-2	1
3	4	4	0	0
4	6	7	1	0
5	3	1	-2	1
6	2	2	0	1
7	8	7	-1	0
8	4	5	1	0

HTE bias

$$\begin{aligned} \textit{Bias} &= \frac{5}{8}\left(\frac{-2+(-2)+0}{3} - \frac{-1+0+1+(-1)+1}{5}\right) \\ &= \frac{5}{8}(-1.333-0) = -0.833 \end{aligned} \tag{16}$$

Does it all add up?

$$\begin{aligned} \textit{NATE} &= \textit{ATE} + \textit{Selection bias} + \textit{HTE bias} \\ -3.933 &= -0.5 + (-2.6) + (-0.833) \\ -3.933 &= -3.933 \end{aligned} \tag{17}$$

Importance of biases

$$NATE = ATE + Selection\ bias + HTE\ bias \tag{18}$$

In our empirical work, we can't directly compute *ATE*, only *NATE*.

Using the POF, we have shown, though, that *NATE* is a good substitute for *ATE* only in the absence of selection bias and HTE bias.

Importance of biases

Student (i)	Prejudice			Contact
	y_{0i}	y_{1i}	δ_i	
1	6	5	-1	0
2	4	2	-2	1
3	4	4	0	0
4	6	7	1	0
5	3	1	-2	1
6	2	2	0	1
7	8	7	-1	0
8	4	5	1	0

Here, biases appeared because those who had inter-ethnic contact had lower levels of prejudice (as would be expected if individuals self-select into contact).

Tackling biases

If selection bias is due to differences in observables, it can easily be tackled.

If the bias is due to education, we can control for it statistically: contrast treatment and control groups with identical education level.

A bigger challenge is in how to control for all possible factors, some of which we might not be aware of.

Assumptions of causal identification

SUTVA: stable unit-treatment value assumption

Also called *non-interference* by Gerber and Green (2012).

As its name says, it requires that the "treatment value" is the same across all units in the population:

1. the treatment is of uniform intensity across units*
2. no externalities: one subject's potential outcome is not affected by the treatment assignment of other subjects (both in terms of mechanism and treatment effect)

In our student dorm room contact experiment, how would SUTVA be violated?

Excludability

The only reason for the change in potential outcomes is the treatment.

In our contact experiment, this would be violated if other agents (parents, campus organizations) focus efforts more on the control group.

It also breaks down when we introduce measurement asymmetries: e.g. more skilled enumerators to measure prejudice in treatment group.

Value of random assignment

Random assignment

There are a few types (simple, complete, cluster, block), but we will discuss simple/complete sampling.

Defining feature: $p(d_i = 1)$ is identical for all subjects. If population is N and m are assigned to treatment, $p = \frac{m}{N}$.

Under such conditions, treatment status is independent of potential outcomes, and all background attributes, $\boldsymbol{X}$ (*ignorability*).

$$Y_{0i}, Y_{1i}, \boldsymbol{X} \perp\!\!\!\perp D_i \tag{19}$$

Mechanics of random assignment (RA)

Randomly select m individuals out of N (for our student contact study) in treatment.

Therefore, potential outcomes for this group of m are the same (in expectation) as that for the population of N.

$$E[Y_{1i}|D_i = 1] = E[Y_{1i}] \tag{20}$$

The same is the case for the $N - m$ allocated to control group:

$$E[Y_{1i}|D_i = 0] = E[Y_{1i}] \tag{21}$$

Mechanics of random assignment (RA)

Putting Eq. 20 and 21 together reveals that under RA, treatment and control groups have the same expected potential outcome.

$$E[Y_{1i}|D_i = 1] = E[Y_{1i}|D_i = 0] \tag{22}$$

We also have

$$E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0] \tag{23}$$

Mechanics of random assignment (RA)

We defined *ATE* as

$$ATE = E[Y_{1i}] - E[Y_{0i}] \tag{24}$$

With RA (under Eq. 20 and a similar one for $E[Y_{0i}]$), we can re-write the *ATE* as

$$ATE = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \tag{25}$$

The two expectations in Eq. 25 are quantities we can observe in real life ($NATE = ATE$)!

RA as bias killer

$$NATE = ATE + \underbrace{E[Y_0|D=1] - E[Y_0|D=0]}_{\textit{selection bias}} + \underbrace{(1-p)(ATT - ATU)}_{\textit{HTE bias}} \tag{26}$$

With RA, we know that $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$ (Eq. 23): *selection bias* disappears.

$$\begin{cases} ATT = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \\ ATU = E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0] \end{cases} \tag{27}$$

RA as bias killer

Based on Eq. 27

$$\begin{aligned} ATT - ATU &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] - (E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0]) \\ &= \underbrace{E[Y_{1i}|D_i = 1] - E[Y_{1i}|D_i = 0]}_{=0 (Eq. 22)} + \underbrace{E[Y_{0i}|D_i = 0] - E[Y_{0i}|D_i = 1]}_{=0 (Eq. 23)} \end{aligned} \tag{28}$$

HTE bias is also addressed by mechanics of RA.

$$NATE = ATE + \underbrace{E[Y_0|D = 1] - E[Y_0|D = 0]}_{=0} + \underbrace{(1 - p)(ATT - ATU)}_{=0} \tag{29}$$

NATE *unbiased* under RA

In this sense, under RA, NATE is an unbiased estimator of the ATE.

If θ is the parameter we want to estimate, such as the ATE, and $\hat{\theta}$ is an estimator for this, such as the NATE...

$$\hat{\theta} \textit{ is unbiased if } E[\hat{\theta}] = \theta \tag{30}$$

$E[\hat{\theta}]$ is the average estimate we would get if we apply the estimator to all possible realizations of the experiment or observational study.

Recap

Conclusions

The POF gives us:

- ✓ the notation to clearly express important concepts like *causal effect* or *bias*
- ✓ a world of perfect information, to illustrate the connections between these concepts

The real world falls short of this; we want *ATE*, but have to settle for *NATE*.

$$NATE = ATE + Selection\ bias + HTE\ bias \tag{31}$$

Conclusions

Random assignment is a situation where $NATE = ATE$, by eliminating bias.

If properly implemented, treatment and control groups should be similar in everything except the *treatment* itself.

Regression tries to achieve the same outcome, by controlling for all possible relevant differences between treatment and control groups...

Thank you for the kind attention!

References

Allport, G. W. (1954). *The Nature of Prejudice*. Reading, MA: Addison-Wesley.

Angrist, J. D., & Pischke, J.-S. (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton, NJ: Princeton University Press.

Cunningham, S. (2021). *Causal Inference: The Mixtape*. New Haven, CT: Yale University Press.

Gerber, A. S., & Green, D. P. (2012). *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton & Co.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960.