The Potential Outcome Framework

PS200D Quant Methods in Politics IV, Spring 2015

Chad Hazlett

UCLA

chazlett@ucla.edu

Purpose, Scope, and Examples

Goal in causal inference is to assess the causal effect some potential cause (e.g. an institution, intervention, policy, or event) on some outcome.

Examples of such research questions include: What is the effect of

- political institutions on corruption?
- voting technology on voting fraud?
- incumbency status on vote shares?
- peace-keeping missions on peace?
- mass media on voter preferences?
- church attendance on turnout?

The Neyman-Rubin Potential Outcome Model

Much of the progress on causal inference in recent years made possible by the Neyman-Rubin causal model, aka the Potential Outcomes Model (POM).

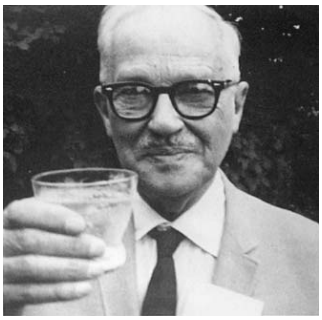


Figure : Neyman

Figure : Rubin

The Potential Outcomes Approach in a Nutshell

How would you know something had a causal effect? What does that mean?

The POM answer:

- Suppose unit i takes a treatment and you measure the outcome. Let's call this Y_{1i}, or the treatment potential outcome
- You'd like to know how Y_{1i} compares to the value of Y unit i *would have had, if it had not taken the treatment*. Call this Y_{0i}, or the non-treatment potential outcome
- Then unit i's *treatment effect* is $\tau_i = Y_{1i} - Y_{0i}$.
- The trouble is you do not get to see both Y_{1i} and Y_{0i} for unit i. This is the fundamental problem of causal inference.
- Everything we do will be about filling in the "missing" potential outcome for each unit through various assumptions and statistical tricks. That is it!

Fundamental Quantities (so far)

- Y_{1i} and Y_{0i}, the potential outcomes
- D_i, the treatment for unit i
- Observed outcome, $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$
- $\tau_i = Y_{1i} - Y_{0i}$
- We will soon be interested in $\mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}[\tau] = ATE$

Which of these are observable?

Warning: regression goggles

- tempting to think of the (observed) Y_i as "responding" to observed D_i.
- but here we see the observed Y_i as just a revelation of either Y_{1i} or Y_{0i}.
- Put differently, the potential outcomes $\{Y_{1i}, Y_{0i}\}$
  - are not *changed* by D_i.
  - they may even be independent of D_i (which is great)
  - D_i just switches which one you can see.

Quick Intuitions

Before turning to formalisms, let's lock in some intuitions

Recall D_i is a "switch" that determines whether you see Y_{1i} or Y_{0i}

Test: Is $\mathbb{E}[Y_{i1}|D_i = 1] = \mathbb{E}[Y_i|D_i = 1]$?

Scenario 1: Someone is throwing the D_i switch randomly.

- how would the Y_{1i} you see (when $D_i = 1$) compare to the Y_{1i} you don't see? And for Y_{0i}?
- what can we say about $\mathbb{E}[Y_{1i}|D_i = 1]$ compared to $\mathbb{E}[Y_{1i}|D_i = 0]$ and $\mathbb{E}[Y_{1i}]$?
- So can we estimate $\mathbb{E}[Y_{1i}]$? And $\mathbb{E}[Y_{0i}]$?
- What does a sample estimator of $\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]$ tell us?

Quick Intuitions II

Scenario 2: Suppose $Y_{1i} = Y_{0i}$ for all i. But now, someone sees Y_{1i}, and chooses $D_i = 1$ more often when Y_{1i} is higher.

- How would the Y_{1i} you see compare to Y_{1i} you don't see?
- So does $\mathbb{E}[Y_{1i}|D_i = 1] = \mathbb{E}[Y_{1i}|D_i = 0] = \mathbb{E}[Y_{1i}]$?
- How would the Y_{0i} you see compare to the ones you don't?
- If you estimate $\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]$, what will we get? Does this tell us anything about $\mathbb{E}[Y_{1i} - Y_{0i}]$?

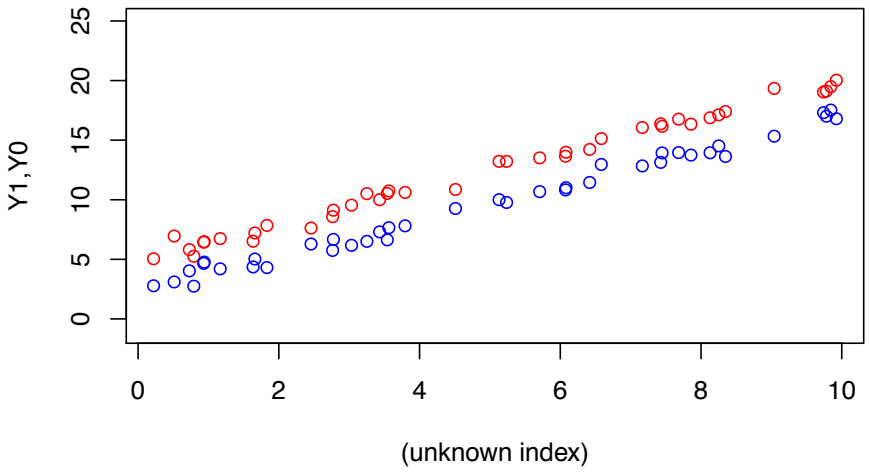
Scenario 3: Suppose you don't know anything about how D_i is assigned w.r.t. Y_{1i} and Y_{0i}.

- Can we say anything about how $\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0]$ compares to $\mathbb{E}[Y_{1i} - Y_{0i}]$?

Which of these scenarios are we usually in with observational data?

Visual Practice

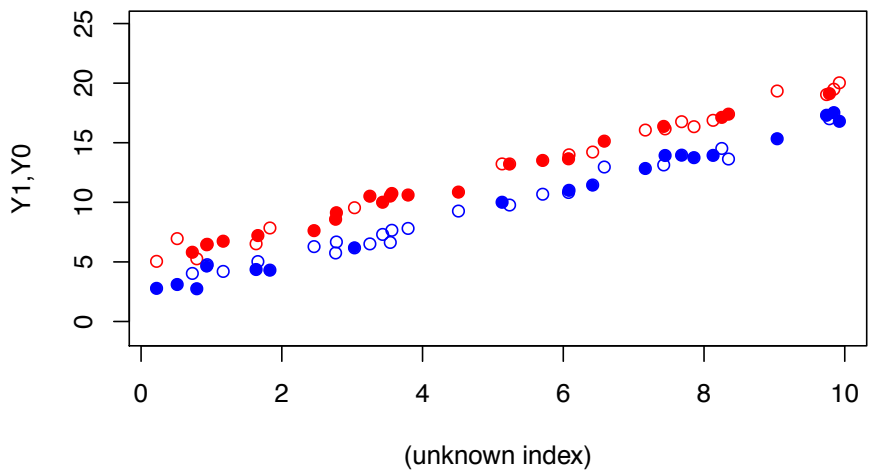
Let's looks at Y_{1i} and Y_{0i} alone first.



True $\mathbb{E}[Y_{1i} - Y_{0i}] = ATE = 3$

Random Assignment of Treatment

Now suppose $D_i = 1$ is randomly assigned. We see only the filled dots:

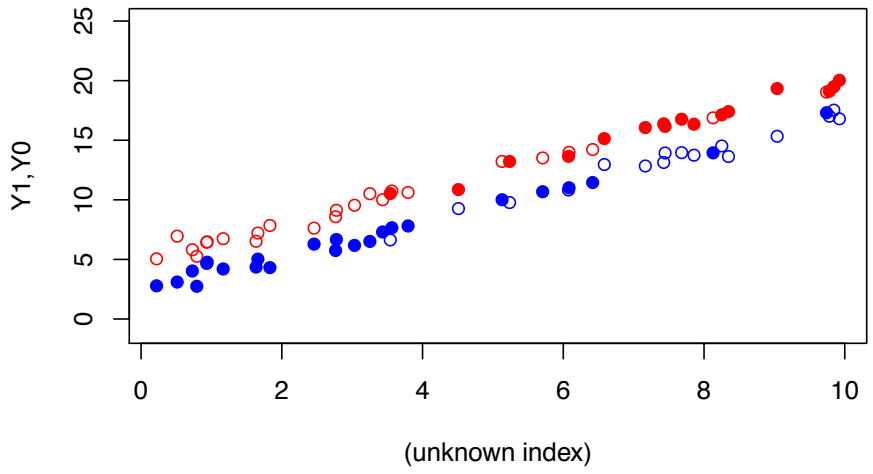


We get

$$\hat{\mathbb{E}}[Y_i|D_i = 1] - \hat{\mathbb{E}}[Y_i|D_i = 0] = \widehat{ATE} = 3.01$$

Non-Random Assignment of Treatment 1

Suppose $Pr(D_i = 1)$ increases to the right. Now we observe (filled dots):

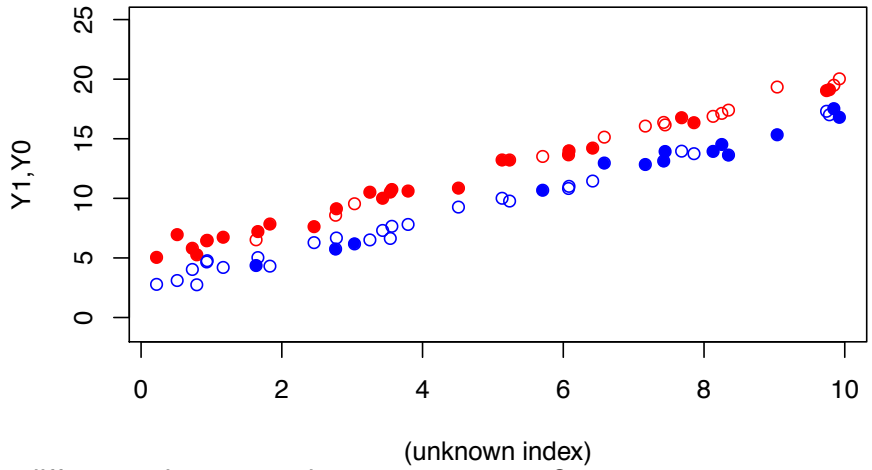


What difference in means do you expect?

$$\hat{\mathbb{E}}[Y_i|D_i = 1] - \hat{\mathbb{E}}[Y_i|D_i = 0] = \widehat{ATE} = 9.97$$

Non-Random Assignment of Treatment 2

Now suppose $Pr(D_i = 1)$ decreases to the right. We observe (filled dots):



What difference in means do you expect now?

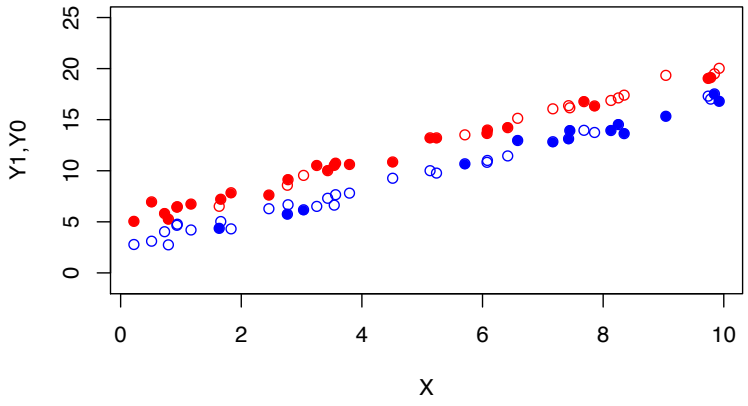
$$\hat{\mathbb{E}}[Y_i|D_i = 1] - \hat{\mathbb{E}}[Y_i|D_i = 0] = \widehat{ATE} = -1.59$$

What can we do about this? Nothing, yet

Preview of Things to Come

Again, say $Pr(D_i = 1)$ decreases to the right

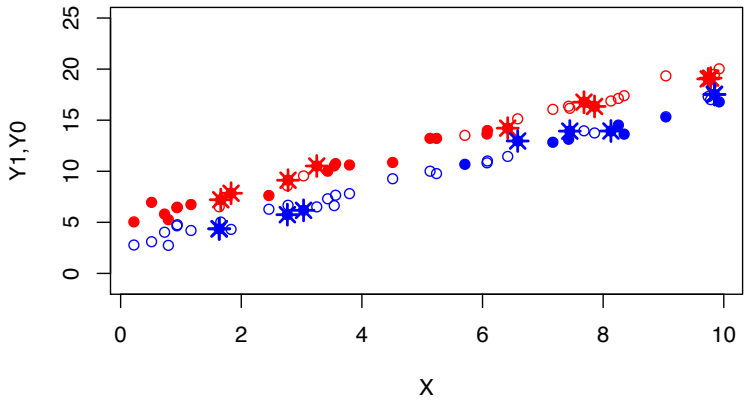
But now suppose X is an observable, and D is random conditional on X



Preview of Things to Come

Again, say $Pr(D_i = 1)$ decreases to the right

But now suppose X is an observable, and D is random conditional on X



Having observed X what if we compare treated to control only at starred points?

$$[\overline{Y}_i|D = 1] - [\overline{Y}_i|D = 0] = \widehat{ATE} = 2.63$$

Notation and Definitions

Unit-level effects (τ_i) are fundamentally unidentifiable. On occasion we will be able to identify various average effects. Some estimands include:

- Treatment effect, $\tau_i = Y_{1i} - Y_{0i}$
- Average Treatment Effect (ATE):

$$ATE = \mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}[\tau_i] = \int \tau p(\tau) d\tau$$

- Average treatment effect on the treated (ATT):

$$ATT = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] = \int \tau p(\tau|D = 1) d\tau$$

- Average treatment effect on the controls (ATC):

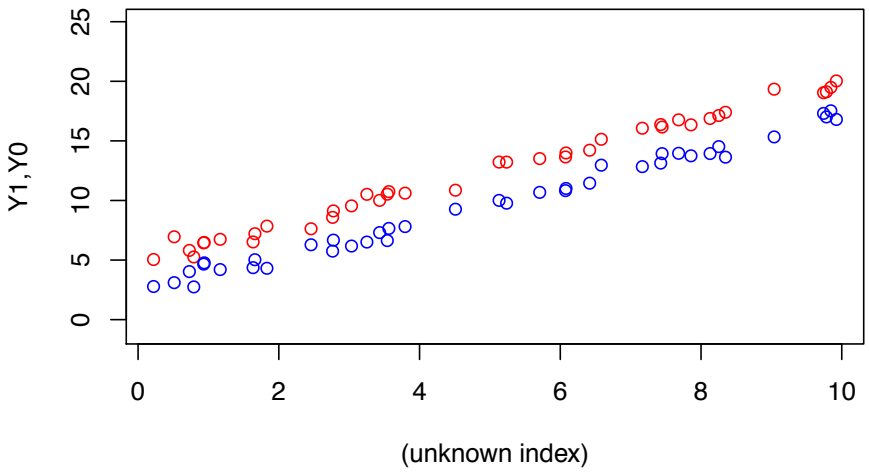
$$ATC = \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 0] = \int \tau p(\tau|D = 0) d\tau$$

- Average treatment effect for sub-groups ($ATE(X)$):

$$ATE(X) = \mathbb{E}[Y_{1i} - Y_{0i}|X = x] = \int \tau p(\tau|X = x) d\tau$$

Practice

Looking at this plot, describe the meaning of the *ATE*, *ATT*, *ATC*, & *ATE*(X) graphically



A Common Assumption: SUTVA

The use of notation $\{Y_{1i}, Y_{0i}\}$ implicitly means we only care about treatment status of i and not of other units, j

- We also defined observed to depend only on unit i's treatment: $Y_i = Y_{1i}D_i + Y_{0i}(1 - D_i)$
- And we defined the causal effect τ_i as $Y_{1i} - Y_{0i}$

Known as Stable Unit Treatment Value Assumption (SUTVA), "no interference", or "individualized treatment response".

It is easy to break: disease, vaccination, conflict, education, anti-corruption monitoring, ...

Without this you get a proliferation of potential outcomes. E.g. with four units you could define:

$$\{Y_i(0,0,0,0), Y_i(0,0,0,1), ..., Y_i(1,1,1,1)\}$$

- how many potential outcomes for each unit, i?
- And how many causal effects for i?

Potential Outcomes with Interference

Definition (SUTVA)

Unit i's potential outcomes depend on D_i and not on $D_{j \neq i}$

- e.g. $Y_3(0, 0, D_3, 0)$ is same as $Y_3(0, 1, D_3, 1)$, $Y_3(1, 1, D_3, 0)$,...
- So just write $Y_3(D_3)$
- Also limits the causal effects for each unit: simply $\tau_i = Y_{1i} - Y_{0i}$

This is an example of an **exclusion restriction**: we use outside information to rule out possibility of certain causal effects.

- assumes that you taking the treatment has no effect on my Y_1 or Y_0
- traditional models (e.g. regression) do same: Y_i depends on X_i not X_j.

Causal inference in the presence of interference is an area of active research. e.g. sometimes "spillover" or "saturation" is deliberately varied

Some practice

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

1. What is the ATE? $\mathbb{E}[Y_{1i} - Y_{0i}] = 1/4 \times (6 - 1 + 0 + 3) = 2$

(Note: average effect is positive, but τ_i are negative for some units)

2. What are the ATT and ATC?

$$\mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1] = .5(6 - 1) = 2.5$$

$$\mathbb{E}[Y_{1i} - Y_{0i} | D_i = 0] = .5(0 + 3) = 1.5$$

Naive Comparison: Difference in Means

You saw earlier how simple comparison of observed outcomes can be misleading. Let's use POM to see this more rigorously.

The Difference in Means estimator is unbiased for

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

What does the POM tell us about this quantity? One terrific decomposition:

$$\begin{aligned}\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] &= \mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0] \\ &= \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] + (\mathbb{E}[Y_{i0}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0])\end{aligned}$$

What are the terms in blue & green?

Selection Bias

$$\begin{aligned}\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] &= \mathbb{E}[Y_{1i}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0] \\ &= \mathbb{E}[Y_{1i} - Y_{0i}|D_i = 1] + (\mathbb{E}[Y_{i0}|D_i = 1] - \mathbb{E}[Y_{0i}|D_i = 0])\end{aligned}$$

Example: $D =$ Church Attendance, $Y =$ Political Participation

- Church goers likely to differ from non-Church goers on a range of background characteristics (e.g. civic duty)
- Turnout for churchgoers could be higher than for non-churchgoers even if churchgoers never attended church, or church had zero mobilizing effect ($\mathbb{E}[Y_0|D = 1] - \mathbb{E}[Y_0|D = 0] > 0$)

Example: Human rights treaties

- Countries willing to sign a human rights treaty will often be those with better human rights already.
- Human rights better in signatory countries even if they had not signed ($\mathbb{E}[Y_0|D = 1] - \mathbb{E}[Y_0|D = 0] > 0$)

The Assignment Mechanism

To fill in the missing potential outcomes using observed ones, assumptions about the *assignment mechanism* comes into play.

With the math above we can get the *ATT* if we know "D_i uncorrelated with Y_{0i}"

In our graphical example, we saw that identification relied on being able to make assumptions about D_i's assigment:

- When we know it was random, we knew unobserved POs would look just like the observed ones.
- When we don't know anything, we're in trouble.
- We previewed how knowledge that D_i is random conditional on X can be used.

Those represent important assignment mechanisms: random assignment, selection on observables, and selection on unobservables.

Summing Up: The Neyman-Rubin / Potential Outcome Model

- Defines causality through counterfactual comparisons
- In so doing, forces us to think about potential outcomes rather than just the observed outcome.
- There is no progress without estimating missing potential outcomes.
  - how do the potential outcomes we see differ from those we don't?
  - assignment mechanism can tell us, describing $p(D, Y_1, Y_0, X)$
- Does not assume homogenous effects (though we will often average it away, or into sub-groups at least)
- SUTVA is implicit in the notation
- Multiple definitions of "causal effect", need to be precise about estimand.

From here: random assignment, then random-conditional-on-observables.