The potential outcomes model

PSCI 2301: Quantitative Political Science II

Prof. Brenton Kenkel

brenton.kenkel@gmail.com

Vanderbilt University

January 22, 2025

Recap

Last week we discussed:

1. Causal statements as counterfactual statements
2. Ingredients of a causal analysis
3. Building blocks of correlational analysis
    - Mean of a variable
    - Variance of a variable
    - Covariance between two variables

Today's agenda

- Potential outcomes model of causality
  - → Mathematical formalization of causality as counterfactuals
- Key requirement for causal inference: the **independence condition**
  - → Potential outcomes don't predict whether you actually get treatment
  - → When this holds, difference of means = average treatment effect
- Ways to make the independence condition plausible
  1. Randomize treatment assignment in an experiment
  2. Control for **confounding variables** that affect treatment assignment and potential outcomes
  3. Identify subpopulations with "as-if random" treatment assignment

The potential outcomes model

Ingredients of the model

We have a population of $\mathcal{N}$ units

→ e.g., the roughly 260 million adults in the USA

We only observe the sample of $i = 1, \ldots, N$ units

→ e.g., the roughly 8000 adults surveyed in the 2020 ANES

Each unit is in the treatment group, in which case we say $D_i = 1$, or else the comparison group, $D_i = 0$

→ e.g., treatment ($D_i = 1$): watches Tucker Carlson regularly
comparison ($D_i = 0$): does not

For each unit we observe the outcome we want to explain, Y_i

→ e.g., the person's opinion of Donald Trump on a 0–100 scale

Potential outcomes

Key assumption: For every unit, there are two “potential outcomes”

- Y_{1i}: outcome i would experience if in treatment group
- Y_{0i}: outcome i would experience if in comparison group

The **treatment effect** for unit i is the difference in potential outcomes:

$$\tau_i = Y_{1i} - Y_{0i}$$

→ How much more/less does i like Trump if they watch Tucker, vs if they don’t?

Potential outcomes notation

Some writers, including Holland, would call these $Y_i(1)$ and $Y_i(0)$ instead. I personally prefer that, but to reduce confusion I will stay close to the *Mastering ’Metrics* notation.

Potential outcomes: Hidden assumptions

1. What matters is *that* you get the treatment, not *how*
   - e.g., "watch Tucker voluntarily" results in same outcome as "watch Tucker because the RIPS worker made me"
   - When this isn't true, must consider as separate types of treatment
2. One unit's outcome doesn't depend on another's treatment assignment
   - e.g., my opinion of Trump if I don't watch Tucker (Y_{0i}) doesn't change depending on whether my wife watches Tucker
   - When this isn't true, need to aggregate data and/or consider as separate types of treatment

Jointly known as stable unit treatment value assumption (SUTVA)

The fundamental problem of causal inference

For any given unit, we only observe one of the two potential outcomes

Example — the underlying reality

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	90	85	5
2	1	100	100	0
3	0	5	30	-25
4	0	0	0	0

The fundamental problem of causal inference

For any given unit, we only observe one of the two potential outcomes

Example — what's observable

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	90	**?**	**?**
2	1	100	**?**	**?**
3	0	**?**	30	**?**
4	0	**?**	0	**?**

$\rightsquigarrow$ we cannot calculate or observe unit-level treatment effects

Average treatment effects

Typical statistical goal is to estimate the average treatment effect

$$\mathbb{E}[\tau_i] = \mathbb{E}[Y_{1i} - Y_{0i}] = \mathbb{E}[Y_{1i}] - \mathbb{E}[Y_{0i}].$$

Example — what's the average treatment effect?

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	90	85	5
2	1	100	100	0
3	0	5	30	-25
4	0	0	0	0

$$\mathbb{E}[\tau_i] = \frac{5 + 0 - 25 + 0}{4} = \frac{-20}{4} = -5$$

Estimating treatment effects

The fundamental problem, again

We want to figure out the average difference between:

- Y_{1i}: outcome for i if in treatment group
- Y_{0i}: outcome for i if in comparison group

Fundamental problem of causal inference $\rightsquigarrow$ only see one of these per unit

So we can only directly calculate:

- Average of Y_{1i} among those who *actually end up* in treatment group
- Average of Y_{0i} among those who *actually end up* in comparison group

But these units might be unlike each other in many ways!

Difference of means ≠ Average treatment effect

Back to our Tucker-and-Trump example, remember that $\mathbb{E}[\tau_i] = -5$

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	90	85	5
2	1	100	100	0
3	0	5	30	-25
4	0	0	0	0

Difference of means ≠ Average treatment effect

Much different if we just look at average opinions of viewers and non-viewers

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	90	**?**	**?**
2	1	100	**?**	**?**
3	0	**?**	30	**?**
4	0	**?**	0	**?**

Average Trump opinion among viewers: $\mathbb{E}[Y_{1i} \mid D_i = 1] = \frac{90+100}{2} = 95$

Average Trump opinion among non-viewers: $\mathbb{E}[Y_{0i} \mid D_i = 0] = \frac{30+0}{2} = 15$

Difference of means is $+80$, whereas average treatment effect is -5

This is the precise sense in which correlation is not causation

Why the difference of means may be misleading

1. Self selection

Predisposition to like Trump either way $\rightsquigarrow$ More likely to watch Tucker

- He has appeared at Trump rallies
- His commentary favors Trump and his agenda

This is an example of self selection into treatment

→ Subset of units who get the treatment aren't representative of full population

People make the choice they think will be best for them

→ Difference of means may overstate *or* understate the true causal effect

Another self selection example

What's the effect of attending a code boot camp on a worker's salary?

Imagine the effect for *everyone* is +$10k

… but only people dissatisfied with their current salary sign up

i	D_i	Y_{1i}	Y_{0i}
1	1	\$50k	\$40k
2	1	\$55k	\$45k
3	0	\$120k	\$110k
4	0	\$130k	\$120k

Another self selection example

What's the effect of attending a code boot camp on a worker's salary?

Imagine the effect for *everyone* is +$10k

… but only people dissatisfied with their current salary sign up

i	D_i	Y_{1i}	Y_{0i}
1	1	$50k	**?**
2	1	$55k	**?**
3	0	**?**	$110k
4	0	**?**	$120k

From diff in means, would incorrectly look like boot camp <u>lowers</u> salary

Why the difference of means may be misleading

2. Confounding variables

Older people watch more TV news and like Trump more

$\rightsquigarrow$ Tucker watchers may like Trump more than average, simply because old

This is an example of a **confounding variable**

Variable X_i is a confounder if it meets both conditions:

1. Affects whether units are in treatment or comparison group
2. Affects at least one potential outcome (Y_{1i} and/or Y_{0i})

When <u>isn't</u> the difference in means misleading?

What we want to estimate, but can't directly — average treatment effect

$$\mathbb{E}[Y_{1i}] - \mathbb{E}[Y_{0i}]$$

What we can easily estimate — difference of means

$$\mathbb{E}[Y_{1i} \mid D_i = 1] - \mathbb{E}[Y_{0i} \mid D_i = 0]$$

Q: When will these coincide with each other?

A: When the potential outcomes in each subgroup — treatment and comparison — are representative of the population as a whole

Representativeness of the subgroups

Avg treatment effect = diff of means when the subgroups are representative

Comes down to covariances: $\mathbb{C}[Y_{1i}, D_i] = 0$ and $\mathbb{C}[Y_{0i}, D_i] = 0$

→ called the **independence condition**

In words: Having an above- or below-average *potential outcome* doesn't at all predict whether you're in the treatment or comparison group

- Important: This is a condition on *potential*, not observed, outcomes
- Fundamental prob of causal inference $\rightsquigarrow$ condition is not testable

Why the independence condition works (1/2)

Remember from a week ago:

$$\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = \frac{\mathbb{C}[Y_i, D_i]}{\mathbb{V}[D_i]}$$

Consequences of the independence condition:

$$\begin{aligned}
\mathbb{C}[Y_{1i}, D_i] = 0 \quad &\Rightarrow \quad \mathbb{E}[Y_{1i} \mid D_i = 1] - \mathbb{E}[Y_{1i} \mid D_i = 0] = \frac{0}{\mathbb{V}[D_i]} = 0 \\
&\Rightarrow \quad \mathbb{E}[Y_{1i} \mid D_i = 1] = \mathbb{E}[Y_{1i} \mid D_i = 0]
\end{aligned}$$

Same line of reasoning for the other potential outcome:

$$\mathbb{C}[Y_{0i}, D_i] = 0 \quad \Rightarrow \quad \mathbb{E}[Y_{0i} \mid D_i = 1] = \mathbb{E}[Y_{0i} \mid D_i = 0]$$

Why the independence condition works (2/2)

We've seen so far that independence implies

$$\mathbb{E}[Y_{1i} \mid D_i = 1] = \mathbb{E}[Y_{1i} \mid D_i = 0]$$
$$\mathbb{E}[Y_{0i} \mid D_i = 1] = \mathbb{E}[Y_{0i} \mid D_i = 0]$$

This in turn means the subgroup averages equal the *overall* averages:

$$\mathbb{E}[Y_{1i} \mid D_i = 1] = \mathbb{E}[Y_{1i} \mid D_i = 0] = \mathbb{E}[Y_{1i}]$$
$$\mathbb{E}[Y_{0i} \mid D_i = 1] = \mathbb{E}[Y_{0i} \mid D_i = 0] = \mathbb{E}[Y_{0i}]$$

(due to the law of iterated expectations)

Therefore, avg treatment effect = diff of means:

$$\mathbb{E}[Y_{1i} \mid D_i = 1] - \mathbb{E}[Y_{0i} \mid D_i = 0] = \mathbb{E}[Y_{1i}] - \mathbb{E}[Y_{0i}]$$

Ways to make the independence condition plausible

1. Experimentally manipulate treatment assignment

Ideal: Getting treatment is unrelated to potential outcomes

- No self selection
- No possible confounding influence from external variables

Cleanest way to ensure this: randomize assignment to treatment

Ways to make the independence condition plausible

1. Experimentally manipulate treatment assignment

Relevant methods: Lab experiment, field experiment

Pros

- Independence condition highly plausible
- Easy for audience to understand methodology

Cons

- Some treatments are too impractical, expensive, or unethical to manipulate experimentally
- **External validity:** How much do results in an artificial setting carry over to real politics?

Ways to make the independence condition plausible

2. Condition on confounding variables

With confounding variables, independence may not hold unconditionally

…but it may hold within <u>subgroups</u> defined by the confounding variables

The conditional independence condition

For all possible confounder values x,

$$\mathbb{C}[Y_{1i}, D_i \mid X_i = x] = 0 \quad \text{and} \quad \mathbb{C}[Y_{0i}, D_i \mid X_i = x] = 0$$

→ e.g., age confounds the Tucker watching–Trump opinion relationship

→ instead of comparing all watchers to all non-watchers, only make comparisons within groups of similar-aged people

Ways to make the independence condition plausible

2. Condition on confounding variables

Relevant methods: Matching, regression, instrumental variables (kinda)

Pros

- Only uses observational data, no expensive or difficult manipulations needed
- Studying real-world outcomes $\rightsquigarrow$ low external validity concern

Cons

- Can you really observe and measure all the confounders?
- Harder to convey findings to non-scientist audience
- Many different ways to analyze same data, hard to know which is most accurate

Ways to make the independence condition plausible

3. Look for "natural experiments"

What if you can't run an experiment but also can't control for all confounders?

Look for a subpopulation where treatment assignment is **"as-if random"**

Example: Effect of military service on lifetime earnings

People who join the military are quite different than those who don't
→ Independence unlikely to hold in full population

But those selected in the draft aren't very different than eligible people not selected
→ Compare earnings of eligible+selected to those of eligible+unselected

See the 1990 study by Josh Angrist, co-author of *Mastering 'Metrics*

Ways to make the independence condition plausible

3. Look for "natural experiments"

Relevant methods: Instrumental variables (kinda), differences in differences, regression discontinuity, synthetic control

Pros

- Only uses observational data, no expensive or difficult manipulations needed
- Cleaner causal inference than simply conditioning on confounders

Cons

- Hard to find good natural experiments
- Still can quibble with independence condition
- External validity: how much does the effect in the subpopulation generalize?

Wrapping up

What we did today

1. Modeled causal effects in terms of potential outcomes
    - Potential outcome if in treatment group: Y_{1i}
    - Potential outcome if in comparison group: Y_{0i}
    - Individual treatment effect: $\tau_i = Y_{1i} - Y_{0i}$
    - Average treatment effect: $\mathbb{E}[\tau_i] = \mathbb{E}[Y_{1i}] - \mathbb{E}[Y_{0i}]$
2. Discussed how to estimate average treatment effects
    - Independence condition: zero covariance b/w potential outcomes and treatment assignment
    - Ways to make this condition plausible
        a. Randomized experiment
        b. Conditioning on confounders
        c. Restriction to subpopulation with "as-if random" assignment

To do for next week

1. Problem Set 1 due 11:59pm on Friday, 1/24
2. Read the polisci paper “Social Pressure and Voter Turnout”
3. Read pages 12–33 of *Mastering ’Metrics*
4. Think about research questions that <u>would</u> and <u>would not</u> be feasible to answer experimentally