Module 1: Potential Outcomes

Fall 2021

Matthew Blackwell

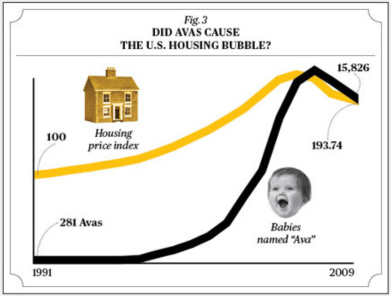
Gov 2003 (Harvard)

What is causal inference?

factual vs. counterfactual

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would it have gone up if the minimum wage increase not occurred?
- Does having girls affect a judge's rulings in court?
  - A judge with a daughter gave a pro-choice ruling.
  - Would they have done that if had a son instead?
- **Causal inference** is the study of these types of causal questions.

What isn't causal inference?



- Associations: parameters of the joint distribution of **observed data**.
  - Correlations, regression coefficients, odds ratios, etc.
  - Describes the world as it happened.
  - No meaningful "directionality", just a joint distribution.
- But causal questions are about **unobserved data**: counterfactuals!
  - Describes what would happen if we **changed** the world.
  - The backbone of most social science theorizing.

The observational-causal bridge

- Causal inference = missing data problem.
- **Assumptions** connect missing data to observed data.
    - Present Matt stays up until 3am prepping for class.
    - How would Present Matt have felt if he had gone to bed at 10pm?
    - Past Matt (w/ a 10pm bedtime) a good substitute? (Assumption!)
- How do we make assumptions crystal clear? Causal notation!
    - Special notation for counterfactuals and interventions.
    - Precisely state what data helps us learn about counterfactuals.

Motivation: Study of political canvassing

- Study of n voters
  - n_1 are canvassed
  - $n_0 = n - n_1$ are not canvassed
- For each voter $i \in \{1, 2, \ldots, n\}$, observe:
  - **Observed outcome** (turnout): Y_i
  - **Treatment variable**:

$$D_i = \begin{cases} 1 & \text{if treated (canvassed)} \\ 0 & \text{if control (not canvassed)} \end{cases}$$

  - **Pretreatment covariates**: $\mathbf{X}_i$
- Causal question of interest: does contact affect turnout?

Defining causal effects

- **Potential outcomes** formally encode counterfactuals (Neyman-Rubin)
  - $Y_i(1)$: outcome that unit i would have if treated.
  - $Y_i(0)$: outcome that unit i would have if untreated.
- Connect observed outcomes to potential outcomes (**consistency**)
  - $Y_i = Y_i(D_i)$: we observe the potential outcome of observed treatment.
- **Causal effect** for unit i: $\tau_i = Y_i(1) - Y_i(0)$.

Voters	Age	Gender	Contact	Turnout		Casual effect
i	X_{i1}	X_{i2}	D_i	$Y_i(1)$	$Y_i(0)$	$Y_i(1) - Y_i(0)$
1	25	M	1	0	???	
2	38	F	0	???	1	
3	67	F	0	???	1	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
n	43	M	1	1	???	

Fundamental problem of causal inference

- We only observe one potential outcome per unit.
  - $\rightsquigarrow Y_i(1) - Y_i(0)$ is never directly observed.
  - Can learn about the marginal distributions, not joint.
- Generalizes to non-binary treatments:
  - categorical: $Y_i(d)$ for $d = 0, 1, \ldots, K - 1$.
  - continuous (dose-response): $Y_i(d)$ for $d \in \mathbb{R}$
  - multivariate: $Y_i(d_1, \ldots, d_K)$ for $d_k \in \mathcal{D}_K$
- Causal inference is **missing data problem**
  - How do we infer the missing potential outcomes? (see rest of the course)

Key assumptions for defining effects

1. **Causal ordering**: $D_i \to Y_i$
    - No reverse causality or simultaneity.
2. **Consistency**: $Y_i = Y_i(d)$ if $D_i = d$
    - No hidden versions of treatment.
    - Or that treatment variance is **irrelevant** (Vanderweele, 2009)
3. **No interference between units**: $Y_i(D_1, D_2, \ldots, D_N) = Y_i(D_i)$
    - No causal effect of other units' treatment on other units' outcomes.

- Last two combined: **SUTVA** (stable unit-treatment variation assumption)

Manipulation

- $Y_i(d)$ is the value that Y would take under D_i set to d.
  - To be well-defined, D_i should be manipulable at least in principle.
- $\rightsquigarrow$ common motto: **“No causation without manipulation”** Holland (1986)
- Tricky causal problems: immutable characteristics such as race, sex, etc.
  - What is the effect of being a man on my political views?
  - What’s the hypothetical manipulation? Very tricky!
- Common alternative: focus on places where we can manipulate these characteristics:
  - Effect of perceived race/gender on legislator replies to constituent mail.
  - Effect of elective female versus male legislators on policy outcomes.
  - Differential effects of treatment by race or gender.

Estimands

- Ideal world: estimate unit causal effects $Y_i(1) - Y_i(0)$
    - But... **FPOCI**! Almost always unidentified without strong assumptions
- **Sample average treatment effect** (SATE):

$$\text{SATE} = \frac{1}{n}\sum_{i=1}^{n}[Y_i(1) - Y_i(0)]$$

    - Average outcomes if everyone is treated vs. no one.
    - We'll spend a lot time trying to identify this.
- **Sample average treatment effect for the treated** (SATT):

$$\text{SATT} = \frac{1}{n_1}\sum_{i=1}^{n} D_i(Y_i(1) - Y_i(0)) = \frac{1}{n_1}\sum_{i=1}^{n} D_i(Y_i - Y_i(0))$$

    - Useful for potentially harmful treatments we may want to remove.

Samples versus Populations

- SATE and SATT are specific to a particular study $i = 1, \ldots, n$.
  - Well-defined even without "repeated sampling" magical thinking
  - Called **finite-sample** or **finite population** inference.
- What if there is a larger population we would like to target?
  - Assume units are a random sample from a large/infinite population.
  - Called the **superpopulation** or sometimes just **population** inference.
- **Population average treatment effects**:

$$\text{PATE} = \mathbb{E}[Y_i(1) - Y_i(0)]$$

$$\text{PATT} = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$$

Other estimands

- Conditional average treatment effect (CATE):

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid \mathbf{X}_i = \mathbf{x}]$$

  - Useful detecting heterogeneous effects for theory testing or targeting.

- Multiple treatments:
  - Controlled direct effect: $\mathbb{E}[Y_i(1, d_2) - Y_i(0, d_2)]$
  - Subtle but important differences from CATE!

- Non-additive effects:
  - **Quantile treatment effects**:
    - Example: $\text{median}(Y_i(1)) - \text{median}(Y_i(0))$
    - How does treated shift a particular quantile of the outcome distribution?
  - **Odds-ratio**:

$$\frac{\mathbb{P}[Y_i(1) = 1]/\mathbb{P}[Y_i(1) = 0]}{\mathbb{P}[Y_i(0) = 1]/\mathbb{P}[Y_i(0) = 0]}$$

More complicated setup: truncation/attrition

- Setting: effect of a **job training program** D_i on **wages** Y_i
- Truncation by "death" problem:
  - Wages only defined for **employed** respondents ($S_i = 1$)
  - But employed are not comparable to unemployed
  - **Post-treatment bias**: program might affect employment.
  - If program increases employment, it might seem like the program decreases wages.
- Don't adjust for post-treatment variables! (collider/selection bias)

Principal Stratification

- We only observe Y_i when $S_i = 1$.
- Potential variables:
  - Potential employment: $S_i(1), S_i(0)$
  - Potential wages: $Y_i(d, s) \to Y_i(1, 0)$, $Y_i(0, 0)$ do not exist.
- Four **principal strata** defined by $(S_i(0), S_i(1))$:
  1. $(1, 1)$: always employed (regardless of program).
  2. $(0, 0)$: never employed (regardless of program).
  3. $(0, 1)$: helped (employed only when treated).
  4. $(1, 0)$: hurt (unemployed only when treated).
- Can't tell which units in which strata.
- Effect of interest is the effect among always employed:

$$\mathbb{E}[Y_i(1, 1) - Y_i(0, 1) \mid S_i(1) = S_i(0) = 1]$$

To sum up

- Causal inference is about comparing **counterfactuals**.
- Potential outcomes represent these counterfactuals mathematically.
- Many, many possible **causal** quantities of interest (any contrast of POs).
- Up next: randomized experiments and tests for causal effects.