1. Potential Outcomes

ISS5096 || ECI
Jaewon ("Jay-one") Yoo
National Tsing Hua University

Why learn causal inference?

- Social science theories are almost always **causal** in nature.
- We should understand when our methods can have a causal interpretation.
- Charles Manski (Northwestern Econ.): "data + assumptions = conclusions"
    - Causal inference is about making assumptions and conclusions more **transparent!**
- The old way was "kitchen sink" regression + causal weasel words:
    - "associated with", "leads to", "the [causal?] effect of", "[in—decreases]", "more likely", "encourages", "is linked to", "predicts"
- **Causal (credibility) revolution**: pick 1) a causal estimand and 2) a research design to identify it.

Outline of Topics to be Covered

- Applied Econometrics
  - Regression with Panel Data
  - Regression with a Binary Dependent Variable
  - Instrumental Variables Regression
  - Experiments and Quasi-Experiments
- Assessing Treatment Effects
  - Linear regression
  - Matching
  - Instrumental variables
  - Difference-in-differences
  - Regression discontinuity
- We may also discuss recent advancements
  - Double ML / Meta-Learners
  - Synthetic DiD
  - Heterogeneous Effects

Teaching Staff

- Instructor: Jaewon Yoo
- Teaching Assistant: Fifi Ding

Learning Resources

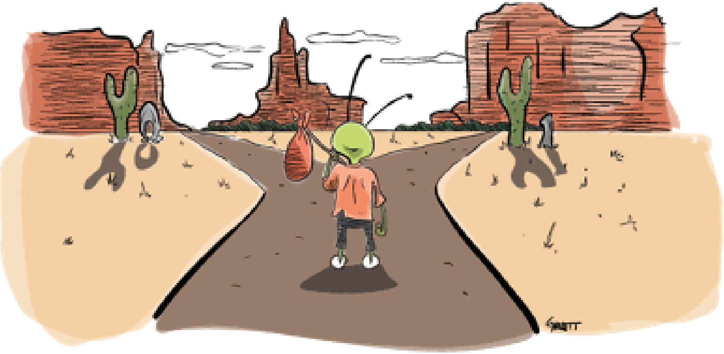
- Lecture: general theoretical and practical issues.
- Round-table discussions and paper presentations.
- Canvas site: place for all the course materials.
    - Lecture notes, guidelines, etc.
    - Submitting assignments.
- MS Teams: logical and social discussion, DMs for help/study groups.
- Office Hours: ask even more questions.

Textbooks

- Responsibility = material covered in lectures.
- Good books that I'll draw upon:
  - Imbens & Rubin: fairly technical, but covers basics well.
  - Hernan & Robins: slightly less technical, more biostat influence.
  - Angrist & Pischke: universal classic, opinionated, most readable.
  - Morgan & Winship: good combo of potential outcomes and graphs.
- Also check out:
  - The Effect by Nick Huntington-Klein
  - Causal Inference: The Mixtape by Scott Cunningham
  - Mastering Metrics by Angrist & Pischke
  - The Book of Why by Judea Pearl
  - Causal ML by Chernozhukov et al.

Work

- Homework assignments (40%)
  - Taken from one of our textbooks (The Effect), we'll point out the relevant chapters for you to review before completing each assignment.
- Final research project (40%)
  - Short research paper ($<$ 20 pg) either applying or extending a method from the class.
  - Milestones throughout semester: submit half-page proposal by Week 5, a midterm report by Week 10.
- One page summaries & paper presentations (10%)
  - Total 14 paper presentations throughout the semester (peer-reviewed).
  - Summaries are check-based submissions to ensure everyone's on the same page. They will not be graded for content or quality.
- Participation (5%)
  - Answering questions and being part of the discussion.
  - Not really intended to hurt your final grade.
- Attendance (5%)



Source: *Chapter 1* of Mastering Metrics by J. Angrist & J. Pischke

What is causal inference?

Factual vs. **Counterfactual**

What is causal inference?

Factual vs. **Counterfactual**

- Q: Does having girls affect a judge's rulings in court?
    - A judge with a daughter gave a pro-choice ruling.
      (p.s., pro-choice: belief that everyone has a right to choose when & whether or NOT to have children $\rightsquigarrow$ pro-abortion)
    - Would they have done that if had a son instead?

What is causal inference?

Factual vs. **Counterfactual**

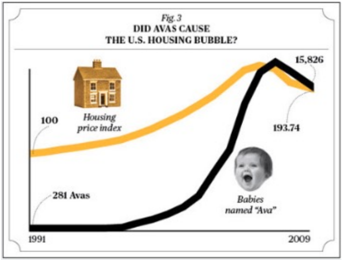
- Q: Does having girls affect a judge's rulings in court?
  - A judge with a daughter gave a pro-choice ruling.
    (p.s., pro-choice: belief that everyone has a right to choose when & whether or NOT to have children $\rightsquigarrow$ pro-abortion)
  - Would they have done that if had a son instead?
- Q: Does having a CEO with a hobby of flying airplanes lead to better innovation outcomes? (Sunder et al. 2017. JFE)
  - Innovation outcomes are higher for companies Led by Pilot CEOs.
  - Would the innovation outcomes have been the same w/o a pilot CEO?

What is causal inference?

Factual vs. **Counterfactual**

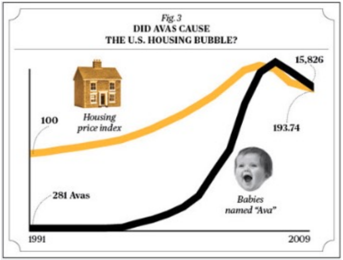
- Q: Does having girls affect a judge's rulings in court?
  - A judge with a daughter gave a pro-choice ruling.
    (p.s., pro-choice: belief that everyone has a right to choose when & whether or NOT to have children $\rightsquigarrow$ pro-abortion)
  - Would they have done that if had a son instead?
- Q: Does having a CEO with a hobby of flying airplanes lead to better innovation outcomes? (Sunder et al. 2017. JFE)
  - Innovation outcomes are higher for companies Led by Pilot CEOs.
  - Would the innovation outcomes have been the same w/o a pilot CEO?
- **Causal inference** is the study of these types of causal questions.

What isn't causal inference?



Source: https://www.bloomberg.com/news/articles/2011-12-01/correlation-or-causation

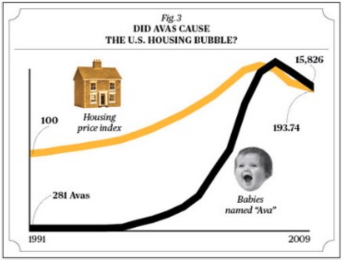
What isn't causal inference?



Source: https://www.bloomberg.com/news/articles/2011-12-01/correlation-or-causation

- 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.

What isn't causal inference?



Source: https://www.bloomberg.com/news/articles/2011-12-01/correlation-or-causation

- 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.

The observational-causal bridge

- Causal inference = missing data problem.
- **Assumptions** connect missing data to observed data.
    - `Present Jaewon` stays up until 3 am prepping for class.
    - How would `Present Jaewon` have felt if he had gone to bed at 10 pm?
    - `Past Jaewon` (w/ a 10pm bedtime) a good substitute? (Assumption!)

The observational-causal bridge

- Causal inference = missing data problem.
- **Assumptions** connect missing data to observed data.
    - `Present Jaewon` stays up until 3 am prepping for class.
    - How would `Present Jaewon` have felt if he had gone to bed at 10 pm?
    - `Past Jaewon` (w/ a 10pm bedtime) a good substitute? (Assumption!)
- How do we make assumptions crystal clear? $\rightsquigarrow$ `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.
  - *Canvassing?*: A systematic initiation of direct contact with individuals, commonly used during political campaigns (think of it as political advertising!)
  - n_1 are canvassed.
  - $n_0 = n - n_1$ are not canvassed.

Motivation: Study of Political Canvassing

- Study of n voters.
  - *Canvassing?*: A systematic initiation of direct contact with individuals, commonly used during political campaigns (think of it as political advertising!)
  - n_1 are canvassed.
  - $n_0 = n - n_1$ are not canvassed.
- For each voter $i \in \{1, 2, \ldots, n\}$ we observe:
  - **Observed outcome** (turnout): Y_i
    - *Turnout?:* The percentage of eligible voters who participated in an election.
  - **Treatment variable**:

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

  - **Pretreatment covariates**: X_i

Motivation: Study of Political Canvassing

- Study of n voters.
  - *Canvassing?*: A systematic initiation of direct contact with individuals, commonly used during political campaigns (think of it as political advertising!)
  - n_1 are canvassed.
  - $n_0 = n - n_1$ are not canvassed.
- For each voter $i \in \{1, 2, \ldots, n\}$ we observe:
  - **Observed outcome** (turnout): Y_i
    - *Turnout?:* The percentage of eligible voters who participated in an election.
  - **Treatment variable**:

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

  - **Pretreatment covariates**: X_i
- Causal question of interest: does contact/canvassing 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.

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	???	

The 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 $d = 0, 1, \dots, K-1$
  - Continuous (dose-response): $Y_i(d)$ for $d \in \mathbb{R}$
  - Multivariate: $Y_i(d_1, \dots, d_K)$ for $d_k \in D_K$

The 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 $d_k \in D_K$
- Again, causal inference is **missing data problem!**
  - How do we infer the missing potential outcomes?
    (stick around for the rest of the course)

Key assumptions for defining effects

1. **Causal ordering:** $D_i \rightarrow Y_i$
    - No reverse causality or simultaneity.
2. **Consistency:** $Y_i = Y_i(d)$ if $D_i = d$
    - For each unit, there are no different "versions" of each treatment level.
    - 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)

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!

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.

Causal estimands

- Ideal world: estimate unit causal effects $Y_i(1) - Y_i(0)$
    - But... **almost always unidentified** without strong assumptions

Causal estimands

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

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

    - Average outcomes if everyone is treated vs. no one.
    - We'll spend a lot time trying to identify this.

Causal estimands

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

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

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

$$\text{SATT} = \frac{1}{n_1}\sum_{i=1}^{n} D_i(Y_i(1) - Y_i(0)) = \frac{1}{n_1}\sum_{i=1}^{n} D_i(Y_i - Y_i(0)) \tag{2}$$

  - We will be looking at this when we have noncompliance issues.

Samples versus Populations

- SATE and SATT are specific to a particular study $i = 1, \ldots, n$.
    - 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.

Samples versus Populations

- SATE and SATT are specific to a particular study $i = 1, \ldots, n$.
  - 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:**

$$\text{PATE} = \mathbb{E}[Y_i(1) - Y_i(0)] \tag{3}$$

$$\text{PATT} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i = 1] \tag{4}$$

Other causal estimands

- Conditional average treatment effect (CATE):

$$\mathbb{E}[Y_i(1) - Y_i(0) | \mathbf{X_i} = x] \tag{5}$$

  - 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: $median(Y_i(1)) - median(Y_i(0))$
    - How does treated shift a particular quantile of the outcome distribution?
  - **Odds-ratio:**

$$\frac{\mathbb{P}[Y_i(1) = 1]/\mathbb{P}[Y_i(1) = 0]}{\mathbb{P}[Y_i(0) = 1]/\mathbb{P}[Y_i(0) = 0]} \tag{6}$$

More complicated setup: Truncation by death

- Set up:
  - Units: patients
  - Treatment: new medicine
  - Outcome: cholesterol level
  - Truncation: patient death

- Truncation by "death" problem (Zhang and Rubin 2003, J. Educ. Behav. Stat.):
  - Cholesterol level is undefined for the dead.
  - Survivors in the treatment group are likely not comparable to those in the control group.
  - Post-treatment bias: treatment may also affect survival!
    - $\rightsquigarrow$ If the treatment saves the lives of the people with high cholesterol, it may appear that the treatment increases cholesterol!

Another Truncation Problem

- RQ: effect of a **job training program** D_i on **wages** Y_i
- Truncation by "death" problem:
  - Wages can be observed only for those that are **employed**.
  - But employed individuals are likely not comparable to those that are unemployed.
- **Issue?:** program (D_i) might also affect employment status (S_i).
  - If program increases employment, it might seem like the program decreases wages.

Principal stratification (Frangakis and Rubin, 2002. Biometrics)

- Q: How can we think about the causal effect of D_i on Y_i under the truncation by death problem?
  - We only observe Y_i when $S_i = 1$ (i.e., employed).
- Potential variables:
  - Potential employment: $S_i(1)$, $S_i(0)$
  - Potential wages: $Y_i(d, s) \longrightarrow Y_i(1, 0)$; $Y_i(0, 0)$ 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).
- Causal effect is defined only for always employed:

$$\mathbb{E}[Y_i(1, 1) - Y_i(0, 1) | S_i(1) = S_i(0) = 1] \tag{7}$$

- Can't tell which principal stratum each unit belongs to. Why?

Takeaways

1. Causal inference is about comparing **counterfactuals**.

2. Potential outcome (PO) represents these counterfactuals mathematically.
    - $\rightsquigarrow$ Allows us to identify (then estimate) causal estimands of interests!

3. Many, many possible **causal** quantities of interest (any contrast of POs).

Have a Great Weekend! :)

Contact Information:
jaewon.yoo@iss.nthu.edu.tw
https://j1yoo.github.io/