Business Analytics

Session 7a. Observational Studies

Renyu (Philip) Zhang

New York University Shanghai

Spring 2019

Sharing on Selection Bias

Essential Role of Identification in Causal Inference

- Causal inference requires good identification strategy (i.e., the treatment assignment mechanism with no selection bias).
- Treatment is randomized by the researcher ⋆⋆⋆
  - Lab experiments (Mendel's experiments on inheritance).
  - Field experiments (Oregon health insurance experiment).
- Treatment is haphazard (natural experiment) ⋆⋆
  - Weather, birthdays, child gender, arbitrary administrative rules, etc.
- Treatment is "as-if" random after statistical control ⋆
  - Regression, matching, weighting, etc.
- Treatment is self-selected and no plausible control is available :(

Observational Studies

- Randomization is the gold standard for causal inference, but it is sometimes infeasible.
- Cannot always randomize so we do observational studies, where we adjust for the observed covariates and hope that unobservables are balanced.
- Design observational study to approximate an experiment.

"*The planner of an observational study should always ask himself: How would the study be conducted if it were possible to do it by controlled experimentation.*" – Cochan (1965)

Good and Bad Observational Studies

- Randomized experiment: Well-defined treatment, clear distinction between covariates and outcomes, control of assignment mechanism.
- Good observational study: Well-defined treatment, clear distinction between covariates and outcomes, precise knowledge of assignment mechanism.
    - Can convincingly answer the following question: Why do two units who are identical on measured covariates receive different treatments?
- Bad observational study: Hard to say when treatment began or what the treatment really is. Distinction between covariates and outcomes is blurred, so problems that arise in experiments seem to be avoided but are in fact just ignored. No precise knowledge of assignment mechanism.

Assignment Mechanism of Good and Bad Observational Studies

- Randomized experiment: Completely random assignment.

- Good observational study: Assignment is not random, but circumstances for the study were chosen so that treatment seems haphazard, or at least not obviously related to potential outcomes (sometimes we refer to these as natural or quasi-experiments).

- Bad observational study: No attention given to assignment process, units self-select into treatment based on potential outcomes.

Examples of (Good and Bad) Observational Studies

- Impact of class size on academic performance.
  - Difficult to measure. Why?
  - Maimonides' rule: $n \leq 25$ kids, 1 teacher; $25 < n \leq 40$ kids, 1 teacher and 1 TA; $n > 40$, 2 teachers.
  - Regression discontinuity design.

- UC Berkeley's gender bias case.
  - It has been observed that girl's admission rate is lower than the boy's.
  - Omitted variable bias: Simpson's Paradox.

- Remove selection bias by conditioning.

Model for Observational Studies

- Subjects: $i = 1, 2, , \cdots, N$ (N is the sample size)
- Treatment: $W_i \in \{0, 1\}$, not necessarily randomly assigned
- Potential outcomes: $Y_i(1)$, $Y_i(0)$
- # of treated/untreated subjects: $N_1 = \sum_{i=1}^{N} W_i$ and $N_0 = N - N_1$
- Pre-treatment covariates for subject i: $X_i = (X_{i1}, X_{i2}, \cdots, X_{ik})$.
    - Predetermined and causally precedent with respect to W_i.
    - Typical covariates: Age, gender, etc.
    - X_i may be correlated with both W_i and Y_i, thus confounding the causal inference.
    - Excludes correlates that are potentially affected by W_i (post-treatment covariates).

Conditional Independence

- Conditioned on each covariate, treatment assignment is randomized (aka Conditional Independence/Ignorability, CI).
    - In natural experiments, treatment assignment is not random, but "almost random".

      $$\{Y_i(0), Y_i(1)\} \perp W_i | X_i = x \text{ for all } x$$

    - In a randomized experiment:

      $$\{Y_i(0), Y_i(1)\} \perp W_i$$

    - Common support assumption:

      $$0 < \mathbb{P}(W_i = 1 | X_i = x) < 1 \text{ for all } x$$

- Theorem. Under CI, there is no selection bias.

"As-if" Randomization

- Causal inference in observational studies rests on CI assumption.
- Intuition: Find the strata (i.e., groups) of X in which the assignment of W_i is randomized.
- The goal is to approximate a randomized experiment within each stratum.
- Plausibility of CI: Can we argue that variation in treatment assignment within each stratum of X is random?
- Commonly used methods:
  - Regression
  - Matching
  - Biggest issue with either method: Cannot balance unobservables.

Controlling for Covariates using Regression

- The life depends on adjusted health condition of each individual as well.
$$Y_i \approx \hat{\beta}_0 + \hat{\beta}_1 W_i + \hat{\beta}_2 X_i$$
- $\hat{\beta}_1$ is still interpreted as the population-level ATE.
  - Assumption: The treatment effect is the same across different individuals.
- Adding the covariate adjusted health condition $(0, 1, 2, 3, 4, 5)$ gives us a better estimate of the baseline $Y(0)$:
  - $Y(0) \approx \hat{\beta}_0 + \hat{\beta}_2 X$.
  - $Y(1) \approx \hat{\beta}_0 + \hat{\beta}_1 + \hat{\beta}_2 X$.
- If randomization is not perfect, controlling for observed covariates can reduce estimation bias.
  - $\text{Cov}(W, X) \neq 0$, e.g., sicker patients are more likely to be given the drug.
  - Variation in Y is explained by variations in X, but not the variation in W.
  - This is called an omitted variable bias.
  - Simpson's Paradox.
- Controlling for X helps remove the omitted variable bias.

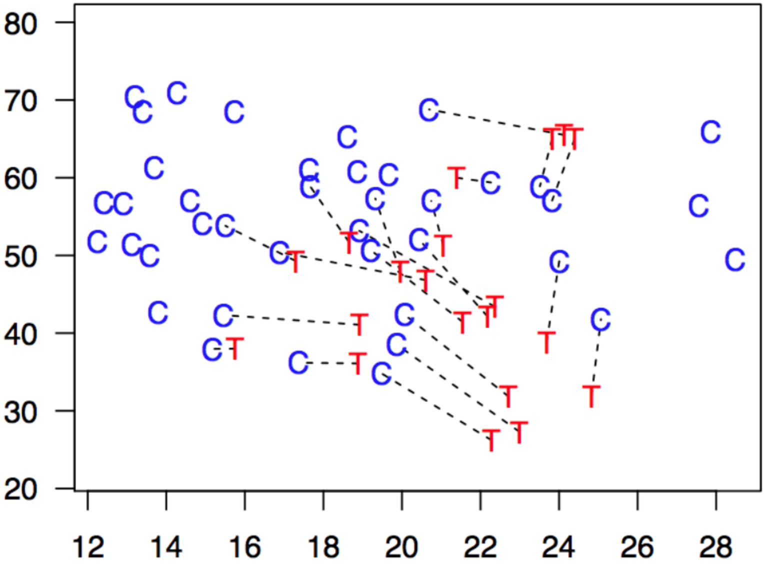
Interactions

- What if the treatment effect depends on the observed covariates?
  - The effect of the drug depends on a patient's health condition.
- Interactions between the covariates and treatment assignment.

$$Y_i \approx \hat{\beta}_0 + \hat{\beta}_W W_i + \hat{\beta}_X X_i + \hat{\beta}_{WX} W_i X_i$$

- Interpretation:
  - $Y(0) \approx \hat{\beta}_0 + \hat{\beta}_X X$
  - $Y(1) \approx \hat{\beta}_0 + \hat{\beta}_X X + (\hat{\beta}_W + \hat{\beta}_{WX} X) W$
  - The estimated causal effect$\approx \hat{\beta}_W + \hat{\beta}_{WX} X$

Matching



Matching

- Idea: Impute missing potential outcomes using observed outcomes of "closest" units (i.e., nearest neighbors).

$$\widehat{ATT} = \frac{1}{N_1} \sum_{D_i=1} (Y_i(1) - Y_{j(i)}(0)),$$

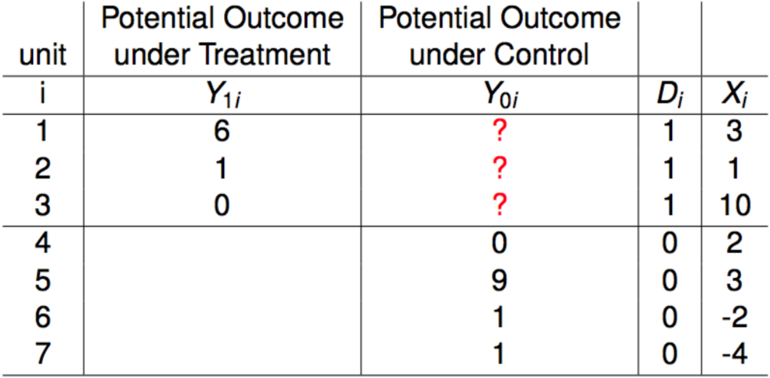
where $j(i)$ is the individual with covariate closest to i.

- Alternatively, we can use the average of M closest matches:

$$\widehat{ATT} = \frac{1}{N_1} \sum_{D_i=1} \left\{ Y_i(1) - \left(\frac{1}{M} \sum_{m=1}^{M} Y_{j_m(i)}(0) \right) \right\}$$

where $j_m(i)$ is the individual with covariate m^{th} closest to i.

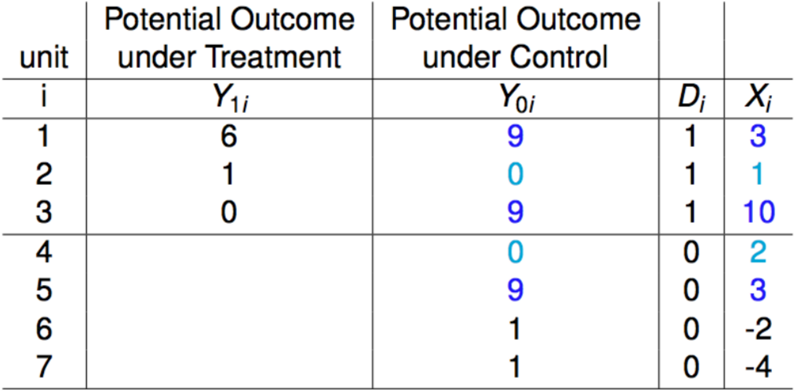
Matching Example



unit	Potential Outcome under Treatment	Potential Outcome under Control		
i	Y_{1i}	Y_{0i}	D_i	X_i
1	6	?	1	3
2	1	?	1	1
3	0	?	1	10
4		0	0	2
5		9	0	3
6		1	0	-2
7		1	0	-4

What is the estimated value of ATT?

Matching Example



unit	Potential Outcome under Treatment	Potential Outcome under Control		
i	Y_{1i}	Y_{0i}	D_i	X_i
1	6	9	1	3
2	1	0	1	1
3	0	9	1	10
4		0	0	2
5		9	0	3
6		1	0	-2
7		1	0	-4

$$ATE \approx \widehat{ATE} = \frac{1}{3}((6-9)+(1-0)+(0-9)) = -3.7$$

Propensity Score Matching

- Issue with the neighborhood matching: Curse of dimensionality.
- Solution: Propensity score matching.
- Probability of receiving treatment given X_i:

$$\pi(X_i) = \mathbb{P}(D_i = 1 | X_i)$$

- Theorem. If CI holds and $0 < \mathbb{P}(D_i = 1 | X_i = x) < 1$ for all x, we have

$$\{Y_i(0), Y_i(1)\} \perp W_i | \pi(X_i).$$

  - Among the subjects with the same propensity score, X_i is identically distributed between the treated and untreated.
  - Sufficient to just condition on $\pi(X_i)$, instead of the whole X_i.

Propensity Score Matching

Procedure of Propensity Score Matching

1. Estimate $\pi(X_i)$: Logistic regression (or other classification models).

2. Match treated and untreated using propensity scores (instead of directly using covariates).

3. Balance check: Check the covariate and propensity score distribution difference between the treated and untreated in the matched sample.

4. Estimate ATT: $t-$test between treated and untreated in the matched sample.