MBA 753 : Causal Inference Methods for Business Analytics

Dr. Nivedita Bhaktha

19.08.2024

Agenda

- Housekeeping
- Average Treatment Effect
- Randomized Experiment
- Matching

Housekeeping

- Project Presentation

September

SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

Sl. No.	Topic
1	Introduction and refresher on multiple linear regression
2	Dummy variable predictors. Interaction effect and its interpretation.
3	From correlation to causality. True Experiments and Quasi-experiments.
4	Event studies
5	Difference-in-Differences method (DID).
6	Regression Discontinuity Design (RDD).
7	Instrumental Variable Regression (IV).
8	Matching Methods
9	Project Presentations

Average Treatment Effect

ATE Estimation

- Intuition: Make comparison across units using Y_i
  - Compare the average observed outcome under treatment to the average observed outcome under control
  - True ATE = 1.25
  - Difference in means =
  - In this example at least, difference in means ≠ ATE

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	5	1	5	?	?
2	2	1	2	?	?
3	0	0	?	0	?
4	1	0	?	1	?

Terminologies

- Average treatment effect on the treated (ATT)

$$E(\tau_i|D_i = 1) = E(Y_{1i} - Y_{0i}|D_i = 1)$$

- Average treatment effect on the untreated (ATU)

$$E(\tau_i|D_i = 0) = E(Y_{1i} - Y_{0i}|D_i = 0)$$

Estimation of ATE

- For a given sample, one obvious estimator of the ATE is the difference in group means (DIGM)

Let $D_i = 1 \; \forall \; i \; \in \{1, \ldots, m\}$ & $D_i = 0 \; \forall \; i \; \in \{m+1, \ldots, n\}$

$$DIGM = \frac{1}{m} \sum_{i=1}^{m} Y_i - \frac{1}{n-m} \sum_{i=m+1}^{n} Y_i$$

- Is DIGM an unbiased estimator of ATE?

Estimation of ATE

Estimation of ATE

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	5	1	5	2	3
2	2	1	2	1	1
3	0	0	1	0	1
4	1	0	1	1	0

$$\tau_{\text{ATE}} = \frac{3+1+1+0}{4} = 1.25$$

$$\tau_{\text{ATT}} = \frac{3+1}{2} = 2$$

$$\textbf{Selection bias} = \frac{2+1}{2} - \frac{0+1}{2} = 1$$

$$\tau_{\text{ATT}} + \textbf{Bias} = 2+1 = 3 = \textbf{DIGM}$$

Estimation of ATE

- DIGM is the unbiased estimator of ATE if
  - $\tau_{ATT} = \tau_{ATE}$
  - There is no selection bias i.e. $E(Y_{0i}|D_i = 1) = E(Y_{0i}|D_i = 0)$

- Selection Bias
  - Selection into treatment is often associated with potential outcomes
  - Selection bias can be positive or negative
  - (In general) Do not believe causal arguments based on simple differences between groups!

Solution for Selection Bias

- We need to know more about counterfactuals that we do not observe
  - Make assumptions about how certain units come to be “selected” for treatment

- Treatment Assignment: mechanism that determines which units are selected for treatment
  - **Random assignment**
  - **Selection on observable characteristics - matching**
  - Selection on unobservable characteristics – DID, RDD, IV

Randomized Experiments

The Randomized Experimental Approach

- The goal of experiments is to eliminate observable and unobservable confounders by design
- Assumption: The world is heterogeneous and ceteris paribus does not hold other than the treatment

→ Randomize to render confounders ineffectual

"Random assignment works not by eliminating individual differences but rather by ensuring that the mix of individuals being compared is the same"

– Angrist and Pischke, 2015, p.16

Random Assignment

- If treatment is assigned at random, then we assume D_i and Y_{0i} & Y_{1i} to be independent

$$E(Y_{1i}|D_i = 1) = E(Y_{1i}) \,\&\, E(Y_{0i}|D_i = 0) = E(Y_{0i})$$

Also $E(Y_{0i}|D_i = 1) = E(Y_{0i}|D_i = 0) = E(Y_{0i}) \Longrightarrow$ Selection bias = 0

$$ATE = E(\tau_i) = E(Y_{1i}) - E(Y_{0i}) = E(Y_{1i}|D_i = 1) - \mathrm{E}(Y_{0i}|D_i = 0)$$
$$= DIGM$$

Randomized Experiment

- Randomization of the treatment makes the difference in group means an unbiased estimator of the true ATE
- On average, you can expect a randomized experiment to get the right answer
- This does not guarantee that the answer you get from any particular randomization will be exactly correct!
- The sample ATE is an unbiased estimator of population ATE only when the sample is drawn at random
  - In most situations, this is very rare!!
- Lots of uncertainty around estimation of ATE!

Threats to Internal and External Validity

- Internal Validity
  - Failure of randomization – pseudo randomization
  - Non-compliance with experimental protocol
  - Attrition - subjects dropping out of the study after randomization
  - Small samples
  - Hawthorne effect - Subjects behaving differently because they know they are being studied

- External Validity
  - Non-representative sample
  - Non-representative treatment

Matching

Motivation

- Randomization aids causal inference because in expectation it balances observed & unobserved confounders
- When we cannot randomize, we can design studies to capture the central strength of randomized experiments:
  - have treatment & control groups that are as comparable as possible
  - i.e. we can try to control for observed covariates

- Assume treatment is not randomized, but is independent of potential outcomes so long as other factors are held fixed
  - We are assuming that among units with the same values for some covariate X (i.e. conditional on X), the treatment is “as good as randomly” assigned

Example

- Does teacher training improve university applications?
  - Imagine that some school teachers take specialist training in how to prepare their students for university applications. Teachers select into the training program (i.e. they are not randomly assigned). You believe, however, that conditional on the type of school in which a teacher teaches, training is as good as random.

Y_i: Number of students applying for top universities
D_i: 1 if the teacher did the training, 0 otherwise
X_i: Whether the teacher is at a state, private, or public school

Example – is DIGM an unbiased estimator of ATE?

X_i, D_i joint distribution

	$D_i = 0$	$D_i = 1$
$X_i =$ State	0.30	0.05
$X_i =$ Private	0.15	0.15
$X_i =$ Public	0.05	0.30

Mean outcomes

	$D_i = 0$	$D_i = 1$
$X_i =$ State	0	2
$X_i =$ Private	3	4
$X_i =$ Public	5	5

$$\begin{aligned} \text{DIGM} &\equiv E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\ &= \frac{(0.05 \times 2 + 0.15 \times 4 + 0.3 \times 5)}{\frac{1}{2}} - \frac{(0.3 \times 0 + 0.15 \times 3 + 0.05 \times 5)}{\frac{1}{2}} \\ &= 3 \end{aligned}$$

Matching

- One way to deal with counterfactuals – treat it as missing data problem
- Matching: imputing missing outcomes
- For each unit i, find the "closest" unit j with opposite treatment status and impute j's outcome as the unobserved potential outcome for i

Types of Matching Methods

- **Nearest Neighbour Matching**
  - M:1 nearest neighbor matching
  - it matches control individuals to the treated group and discards controls who are not selected as matches
- Full Matching
- Subclassification

Example

- Do UN interventions Cause Peace?
  - Gilligan and Sergenti (2008) investigate whether UN peacekeeping operations have a causal effect on building sustainable peace after civil wars. They study 87 post-Cold-War conflicts, and evaluate whether peace lasts longer after conflict in 19 situations in which the UN had a peacekeeping mission compared to 68 situations where it did not.

Y_i: Peace duration (measured in months)
D_i: 1 if the UN intervened post-conflict, 0 otherwise
$X_{1,i}$: Region of conflict (categorical)
$X_{2,i}$: Democracy in the past (binary, based on polity)
$X_{3,i}$: Ethnic Fractionalization (continuous)

Example source: Lecy & Fusi, ds4ps.org

Example – Matching M = 1 with replacement

NN 1:1 Matching

Country	D	EthFrac	Region	Y_{0i}	Y_{1i}
Liberia	1	83	SS Africa	?3	51
Sierra Leone	1	77	SS Africa	?11	35
Zaire	1	90	SS Africa	?3	23
Chad	0	83	SS Africa	3	
Senegal	0	72	SS Africa	11	
Niger	0	73	SS Africa	11	

What is the $\hat{\tau}_{ATT}$?

$$
\begin{aligned}
\hat{\tau}_{\text{ATT}} &= \frac{1}{N_1} \sum_{D_i=1} (Y_i - Y_{j(i)}) \\
&= (51-3) \times 1/3 + (35-11) \times 1/3 + (23-3) \times 1/3 \\
&= 30.7
\end{aligned}
$$

Example – Matching M = 2 with replacement

NN 2:1 Matching

Country	D	EthFrac	Region	Y_{0i}	Y_{1i}
Liberia	1	83	SS Africa	?7	51
Sierra Leone	1	77	SS Africa	?11	35
Zaire	1	90	SS Africa	?7	23
Chad	0	83	SS Africa	3	
Senegal	0	72	SS Africa	11	
Niger	0	73	SS Africa	11	

What is the $\hat{\tau}_{ATT}$?

$$
\begin{aligned}
\hat{\tau}_{\text{ATT}} &= \frac{1}{N_1} \sum_{D_i=1} \left(Y_i - \frac{1}{M} \sum_{m=1}^{M} Y_{j_m(i)}\right) \\
&= (51 - 7) \times 1/3 + (35 - 11) \times 1/3 + (23 - 7) \times 1/3 \\
&= 28
\end{aligned}
$$

Matching – Many questions

- If we are selecting matches, how many?
  - One best match
  - K best matches
  - All acceptable matches
- Matching with or without replacement?
  - Without replacement - running out of controls
  - With replacement - use the same control multiple times, giving it a weight equal to the number of times it has matched
- How do we use weights?
  - Constructing weights for observations

Matching – Bias-Variance Trade-off

- How many – few or more
  - Few matches – better matches – less bias
  - More matches – less sampling variation – lower standard errors
- With/out replacement –
  - With – reduces bias – each control has more influence – sampling variation is larger
- Weighting –
  - Weight each observation separately
    - Kernel matching
    - Inverse probability weighting

Matching on more variables

- Commonly we will want to match on many X variables, not just one or two
- Adding more covariates creates a problem - We have to define how we measure whether two units are "close" to one another

Treated case:

- Haiti, with polity = -6, region = Latin America, and **ethfrac = 1**

Control cases:

- Panama, with polity = 8, region = Latin America, **ethfrac = 3**
- Egypt, with polity = -7, region = N Africa, **ethfrac = 4**
- El Salvador, with polity = -6, region = Latin America, **ethfrac = 26**

Multiple Matching Variables

- Distance Matching
  - Mahalanobis Distance: distance from the centroid
  - Entropy Balancing: *enforce restrictions* on the distance between treatment and control
  - Curse of dimensionality
    - Limit the number of matching variables
    - Extremely large sample sizes
    - Using caliper or bandwidth for match quality

- Propensity Score Matching

Recap

Summary

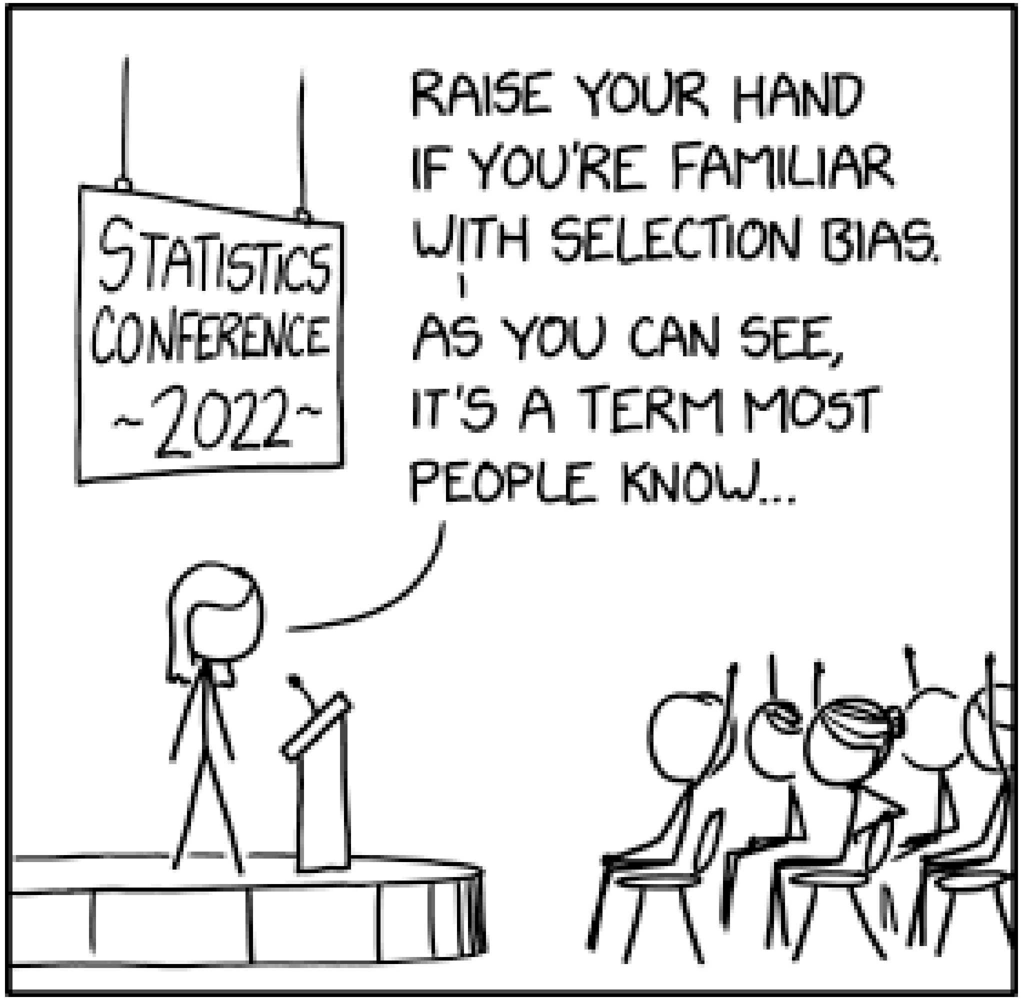
- Causality is defined by potential outcomes, not by observed outcomes
- The difference in means is only an unbiased estimator for the ATE when there is no selection bias
- Randomized experiments are often thought of as the gold standard for causal inference
- By assuming treatments are "as good as random" conditional on X, we can make causal claims from non-experimental data

- Objectives achieved:
  - Can understand potential outcomes framework and define causal effect
  - Can realize the importance of random assignment
  - Can match on observed confounders

References

- Scott Cunningham, Causal Inference: The Mix Tape, Yale University Press.
- Cook, T. D., & Campbell, D. T. (2007). *Experimental and quasi-experimental designs for generalized causal inference*.
- Angrist, J. D., & Pischke, J. S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.

Thank You ☺


STATISTICS CONFERENCE ~2022~
RAISE YOUR HAND IF YOU'RE FAMILIAR WITH SELECTION BIAS.
AS YOU CAN SEE, IT'S A TERM MOST PEOPLE KNOW...