GOV 51 Section

Week 3: Matching, and Unit 1 (Causal Inf.) Review

Pranav Moudgalya

Harvard College

Topics

- Review of matching
- Causal inference (DiD, IV, Matching, & assumptions) review
- Project discussion

Motivating Matching

- What is matching? and why do we do it?

Motivating Matching

- What is matching? and why do we do it?
- Matching in causal inference is a method used to reduce confounding by pairing treated and untreated units with similar observed characteristics, ensuring that comparisons more closely approximate a randomized experiment. It is useful when estimating causal effects in observational studies where random assignment is not possible, helping to control for pre-treatment differences and improve the validity of causal claims.

Motivating Matching

- What is matching? and why do we do it?
- Matching in causal inference is a method used to reduce confounding by pairing treated and untreated units with similar observed characteristics, ensuring that comparisons more closely approximate a randomized experiment. It is useful when estimating causal effects in observational studies where random assignment is not possible, helping to control for pre-treatment differences and improve the validity of causal claims.
- Essentially: how close can we get to a clone to get a true counter-factual

Motivating Matching

- What is matching? and why do we do it?
- Matching in causal inference is a method used to reduce confounding by pairing treated and untreated units with similar observed characteristics, ensuring that comparisons more closely approximate a randomized experiment. It is useful when estimating causal effects in observational studies where random assignment is not possible, helping to control for pre-treatment differences and improve the validity of causal claims.
- Essentially: how close can we get to a clone to get a true counter-factual
- Useful when DiD & IV aren't appropriate *and* we have quality covariate data.

Matching Assumptions

- In addition to probabilistic treatment, ignorability, and SUTVA...

Matching Assumptions

- In addition to probabilistic treatment, ignorability, and SUTVA...
- Valid matched set!

$$\underbrace{\mathbb{E}\left(Y_i(0) \mid T_i = 1\right)}_{\text{unobserved}} = \mathbb{E}\left(Y_i(0) \mid T_i = 0, i \in \mathcal{M}\right)$$

Matching Assumptions

- In addition to probabilistic treatment, ignorability, and SUTVA...
- Valid matched set!

$$\underbrace{\mathbb{E}\left(Y_i(0) \mid T_i = 1\right)}_{\text{unobserved}} = \mathbb{E}\left(Y_i(0) \mid T_i = 0, i \in \mathcal{M}\right)$$

- What does that mean?

Matching Assumptions

- In addition to probabilistic treatment, ignorability, and SUTVA...
- Valid matched set!

$$\underbrace{\mathbb{E}\left(Y_i(0) \mid T_i = 1\right)}_{\text{unobserved}} = \mathbb{E}\left(Y_i(0) \mid T_i = 0, i \in \mathcal{M}\right)$$

- What does that mean?
- Expected outcome for treated individuals had they not been treated (which is unobserved) can be estimated using the expected outcome of similar untreated individuals (from the matched set M), making the untreated group a valid counterfactual.

Matching Example

Matching gives us the Average Treatment Effect on the Treated (ATT), why?

Matching Example

Matching gives us the Average Treatment Effect on the Treated (ATT), why?

Unit	**Pre-treatment Covariates**		**Treatment**	**Observed**	**Outcomes**
	Treated	**Matched**	**Indicator**	**Treated**	**Matched**
1	X_1,	$\mathbf{X_1^M}$	1, 0	$Y_1(1)$	$Y_1^M(0)$
2	X_2,	$\mathbf{X_2^M}$	1, 0	$Y_2(1)$	$Y_2^M(0)$
3	X_3,	$\mathbf{X_3^M}$	1, 0	$Y_3(1)$	$Y_3^M(0)$
4	X_4,	$\mathbf{X_4^M}$	1, 0	$Y_4(1)$	$Y_4^M(0)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	X_N,	$\mathbf{X_N^M}$	1, 0	$Y_N(1)$	$Y_N^M(0)$

How to Match?

- Various ways to match

How to Match?

- Various ways to match
- Absolute distance matching (can be weighted)

How to Match?

- Various ways to match
- Absolute distance matching (can be weighted)
- Mahalonobis weighting

How to Match?

- Various ways to match
- Absolute distance matching (can be weighted)
- Mahalonobis weighting

How To Do Mahalonobis Matching?

Given a set of covariates X, the Mahalanobis distance between observations i and j is:

$$d_M(i,j) = \sqrt{(X_i - X_j)^T S^{-1} (X_i - X_j)} \tag{1}$$

where:

- X_i, X_j are covariate vectors.
- S is the empirical covariance matrix.
- S^{-1} accounts for correlations, ensuring scale-invariance.

How To Do Mahalonobis Matching?

1. Compute $d_M(i, j)$ for all treated-control pairs.
2. Perform matching:
   - **Greedy Matching:** Iteratively select closest pairs.
   - **Optimal Matching:** Minimize total distance.
3. Check balance using standardized mean differences.
4. Estimate treatment effects using matched pairs.

Just joking!

- You will never worry about implementation: use `MatchIt` in R to do this automatically.

Just joking!

- You will never worry about implementation: use `MatchIt` in R to do this automatically.

```
library(MatchIt)

m.out <- matchit(mailer ~ age + voted96 + newvote +
                   dem,
                 data = gotv,
                 distance = "mahalanobis",
                 replace = TRUE,
                 exact = ~white)

matchsum <- summary(m.out)
```

Review of Causal Inference

- So far, we have discussed at least three causal inference mechanisms:

Review of Causal Inference

- So far, we have discussed at least three causal inference mechanisms:
  - Difference-in-Difference

Review of Causal Inference

- So far, we have discussed at least three causal inference mechanisms:
  - Difference-in-Difference
  - Instrumental Variables

Review of Causal Inference

- So far, we have discussed at least three causal inference mechanisms:
  - Difference-in-Difference
  - Instrumental Variables
  - Matching

Review of Causal Inference

- So far, we have discussed at least three causal inference mechanisms:
  - Difference-in-Difference
  - Instrumental Variables
  - Matching
- Let's put your knowledge to the test!
- Jeopardy game: https://jeopardylabs.com/play/gov-51-causal-inference-jeopard-review

Final Project

Time to chat with others: what are you interested in studying? what kind of data would you use / create? what kind of methods would you use?

Final Project

Time to chat with others: what are you interested in studying? what kind of data would you use / create? what kind of methods would you use?

Next week(s), focusing on regression.

Final Project

Time to chat with others: what are you interested in studying? what kind of data would you use / create? what kind of methods would you use?

Next week(s), focusing on regression.

Problem Set I released, due by **February 20th, 2020 by 11:59pm via Gradescope**. We will only grade your compiled PDF. Come to OH if you need help!