Matching Intro

INFO/STSCI/ILRST 3900: Causal Inference

3 Oct 2023

Learning goals for today

At the end of class, you will be able to:

1. Explain how matching can be used to estimate causal effects
2. Explain bias variance trade-off in various matching procedures

Causal effect

What is the causal effect on income of a job training program?

Causal effect

What is the causal effect on income of a job training program?

- **Average Treatment Effect** (on everyone)

$$E(Y^{a=1}) - E(Y^{a=0})$$

Causal effect

What is the causal effect on income of a job training program?

- **Average Treatment Effect** (on everyone)

$$E(Y^{a=1}) - E(Y^{a=0})$$

- **Average Treatment Effect on the Treated** (ATT)

$$E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$$

Matching: The big idea

Matching: The big idea

Goal: $E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$ **ATT**

$$E(Y^{a=1} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=1} = \frac{1}{n_t} \sum_{i:A_i=1} Y_i$$

Matching: The big idea

Goal: $E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$ **ATT**

$$E(Y^{a=1} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=1} = \frac{1}{n_t} \sum_{i:A_i=1} Y_i$$

$$E(Y^{a=0} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=0} \not\approx \frac{1}{n_c} \sum_{i:A_i=0} Y_i$$

Matching: The big idea

Goal: $E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$ **ATT**

$$E(Y^{a=1} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=1} = \frac{1}{n_t} \sum_{i:A_i=1} Y_i$$

$$E(Y^{a=0} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=0} \not\approx \frac{1}{n_c} \sum_{i:A_i=0} Y_i$$

Problem: Control may be different than the treatment

Matching: The big idea

Goal: $E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$ **ATT**

$$E(Y^{a=1} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=1} = \frac{1}{n_t} \sum_{i:A_i=1} Y_i$$

$$E(Y^{a=0} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=0} \not\approx \frac{1}{n_c} \sum_{i:A_i=0} Y_i$$

**Problem:** Control may be different than the treatment

Potential Solution: Create a sample of **untreated** individuals, $\mathcal{M}$, which are similar to the treated group

Matching: The big idea

Goal: $E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$ **ATT**

$$E(Y^{a=1} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=1} = \frac{1}{n_t} \sum_{i:A_i=1} Y_i$$

$$E(Y^{a=0} \mid A = 1) \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=0} \not\approx \frac{1}{n_c} \sum_{i:A_i=0} Y_i$$

Problem: Control may be different than the treatment

Potential Solution: Create a sample of **untreated** individuals, $\mathcal{M}$, which are similar to the treated group

$$\frac{1}{n_m} \sum_{i \in \mathcal{M}} Y_i = \frac{1}{n_m} \sum_{i \in \mathcal{M}} Y_i^{a=0} \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=0}$$
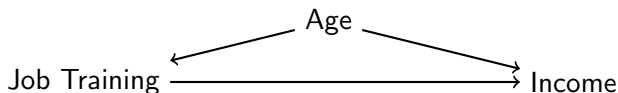
Example
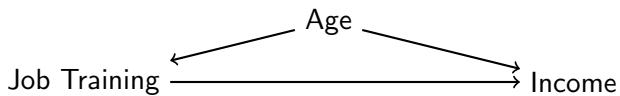
Example
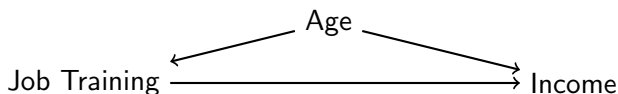
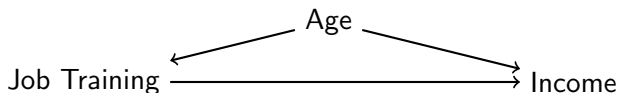- ▶ Conditional exchangeability holds when conditioning on Age!

$$E(Y^{a=0} \mid A = 1, \text{Age} = \ell) = E(Y^{a=0} \mid A = 0, \text{Age} = \ell)$$

- ▶ Estimate

$$E(Y^{a=0} \mid A = 1) = \underbrace{\sum_\ell Pr(\text{Age} = \ell \mid A = 1)E(Y^{a=0} \mid A = 1, \text{Age} = \ell)}_{\text{Weighted average of averages}}$$

Example

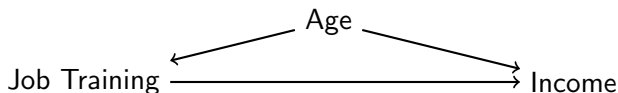▶ Conditional exchangeability holds when conditioning on Age!
$$E(Y^{a=0} \mid A = 1, \text{Age} = \ell) = E(Y^{a=0} \mid A = 0, \text{Age} = \ell)$$

▶ Estimate
$$E(Y^{a=0} \mid A = 1) = \underbrace{\sum_\ell Pr(\text{Age} = \ell \mid A = 1)E(Y^{a=0} \mid A = 1, \text{Age} = \ell)}_{\text{Weighted average of averages}}$$

$$E(Y^{a=0} \mid \mathcal{M}) = \sum_\ell Pr(\text{Age} = \ell \mid \mathcal{M})E(Y^{a=0} \mid A = 0, \text{Age} = \ell, \mathcal{M})$$

Example

$$\text{Job Training} \xleftarrow{\hspace{2cm}} \text{Age} \xrightarrow{\hspace{2cm}} \text{Income}$$

► Conditional exchangeability holds when conditioning on Age!

$$E(Y^{a=0} \mid A = 1, \text{Age} = \ell) = E(Y^{a=0} \mid A = 0, \text{Age} = \ell)$$

► Estimate

$$E(Y^{a=0} \mid A = 1) = \underbrace{\sum_{\ell} Pr(\text{Age} = \ell \mid A = 1) E(Y^{a=0} \mid A = 1, \text{Age} = \ell)}_{\text{Weighted average of averages}}$$

$$\begin{aligned} E(Y^{a=0} \mid \mathcal{M}) &= \sum_{\ell} Pr(\text{Age} = \ell \mid \mathcal{M}) E(Y^{a=0} \mid A = 0, \text{Age} = \ell, \mathcal{M}) \\ &= \sum_{\ell} Pr(\text{Age} = \ell \mid \mathcal{M}) E(Y^{a=0} \mid A = 0, \text{Age} = \ell) \end{aligned}$$

Example

- Conditional exchangeability holds when conditioning on Age!

$$E(Y^{a=0} \mid A = 1, \text{Age} = \ell) = E(Y^{a=0} \mid A = 0, \text{Age} = \ell)$$

- Estimate

$$E(Y^{a=0} \mid A = 1) = \underbrace{\sum_\ell Pr(\text{Age} = \ell \mid A = 1) E(Y^{a=0} \mid A = 1, \text{Age} = \ell)}_{\text{Weighted average of averages}}$$

$$\begin{aligned} E(Y^{a=0} \mid \mathcal{M}) &= \sum_\ell Pr(\text{Age} = \ell \mid \mathcal{M}) E(Y^{a=0} \mid A = 0, \text{Age} = \ell, \mathcal{M}) \\ &= \sum_\ell Pr(\text{Age} = \ell \mid \mathcal{M}) E(Y^{a=0} \mid A = 0, \text{Age} = \ell) \end{aligned}$$

- If we can make $Pr(\text{Age} = \ell \mid \mathcal{M}) \approx Pr(\text{Age} = \ell \mid A = 1)$, the two quantities should be the same

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

$$E(Y^{a=1} \mid A = 1) - E(Y^{a=0} \mid A = 1)$$

Potential Solution: Create a group of untreated individuals, $\mathcal{M}$, which have a **similar distribution of L** to the treated group

$$\frac{1}{n_m} \sum_{i \in \mathcal{M}} Y_i \approx \frac{1}{n_t} \sum_{i:A_i=1} Y_i^{a=0} \approx E(Y^{a=0} \mid A = 1)$$
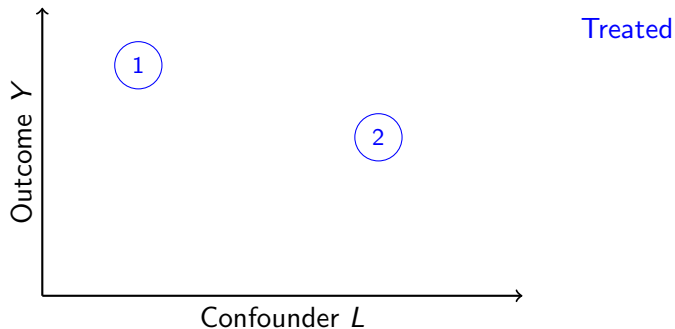
Detail: How?

Example

Job training

Ind	Age	Y^{Train}	Y^{NoTrain}
1	20	19	?
2	25	63	?
3	38	65	?
4	38	43	?

Example

Job training

Ind	Age	Y^{Train}	$Y^{NoTrain}$
1	20	19	?
2	25	63	?
3	38	65	?
4	38	43	?

No job training

Ind	Age	$Y^{NoTrain}$
1	19	82
2	18	39
3	20	49
4	20	56
5	24	33
6	26	82
7	26	35
8	38	35
9	28	83
10	30	79
11	25	63
12	32	52
13	34	58
14	34	70
15	35	47
16	37	42
17	37	83
18	38	33
19	39	37
20	39	60

Matching: The big idea

You have a some treated units.

Matching: The big idea
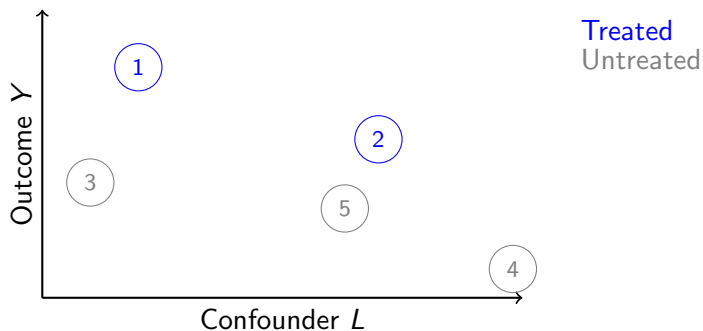
You go find some untreated units.

Matching: The big idea
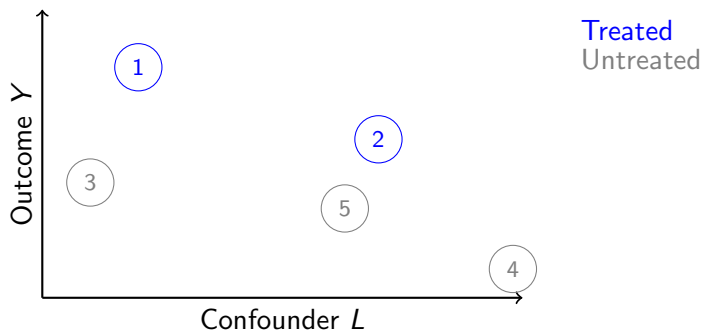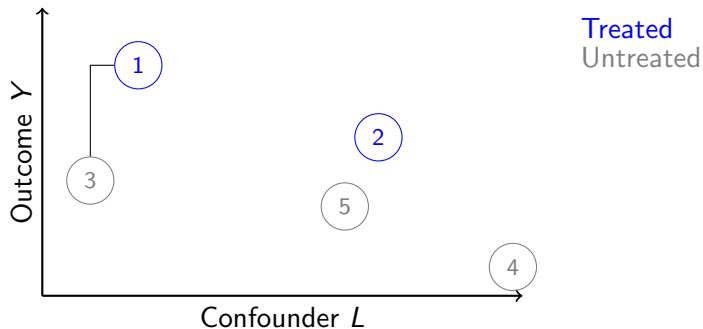
You find the closest matches along L
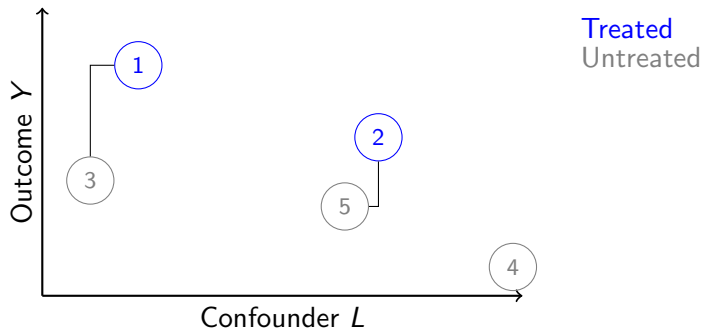
Matching: The big idea

You find the closest matches along *L*

Matching: The big idea

You find the closest matches along *L*
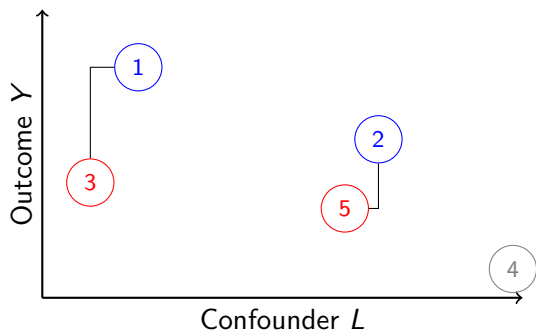
Matching: The big idea

Matching: The big idea
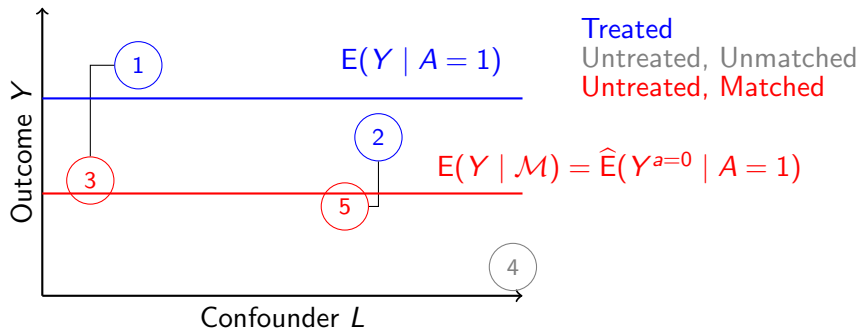
Why matching is great

Why matching is great

1. Completely transparent that Y_i^1 is observed

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Easy to explain

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Easy to explain
   - We had some treated units

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Easy to explain
   - We had some treated units
   - We found a set of control units which are comparable

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Easy to explain
   - ▶ We had some treated units
   - ▶ We found a set of control units which are comparable
   - ▶ We compared the means

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Easy to explain
   - ▶ We had some treated units
   - ▶ We found a set of control units which are comparable
   - ▶ We compared the means
3. Can assess quality of matches before we look at the outcome

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Easy to explain
   - ▶ We had some treated units
   - ▶ We found a set of control units which are comparable
   - ▶ We compared the means
3. Can assess quality of matches before we look at the outcome
4. Model-free*

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Easy to explain
   - ▶ We had some treated units
   - ▶ We found a set of control units which are comparable
   - ▶ We compared the means
3. Can assess quality of matches before we look at the outcome
4. Model-free*
   - ▶ * but you have to define what makes a match "good"

Bias vs variance

The idea of matching is straightforward, but the details matter!

Bias vs variance

The idea of matching is straightforward, but the details matter!

---

[1]Figure from:
http://scott.fortmann-roe.com/docs/BiasVariance.html
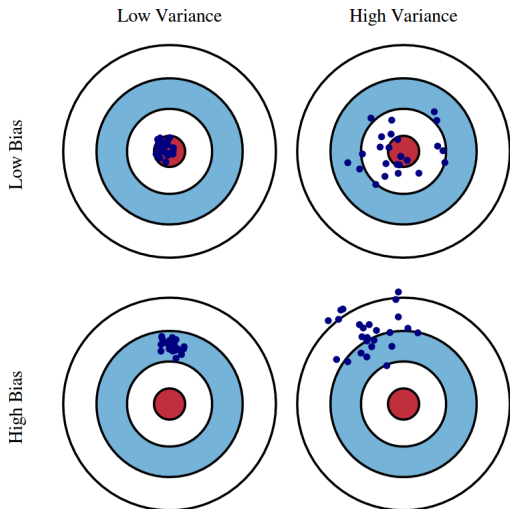
Matching in univariate settings: Algorithms

- Caliper or no caliper
- 1:1 vs k:1
- With replacement vs without replacement
- Greedy vs optimal

Caliper or no caliper matching
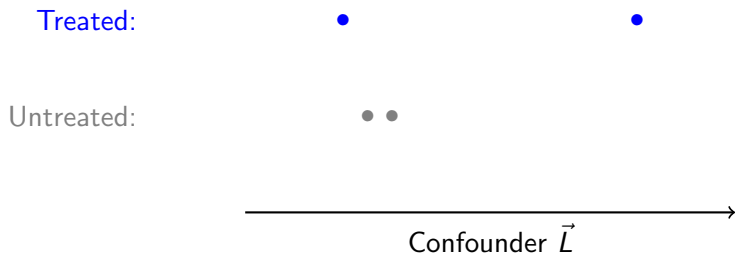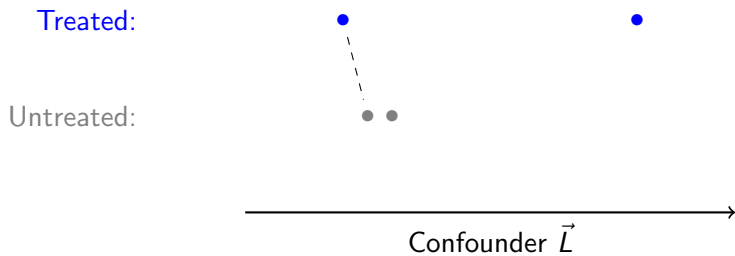
Caliper or no caliper matching

Caliper or no caliper matching

Caliper or no caliper matching

Caliper or no caliper matching

Treated:

Untreated:

Confounder $\vec{L}$

▶ Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius

Caliper or no caliper matching

Treated:

Untreated:

Confounder $\vec{L}$

▶ Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius
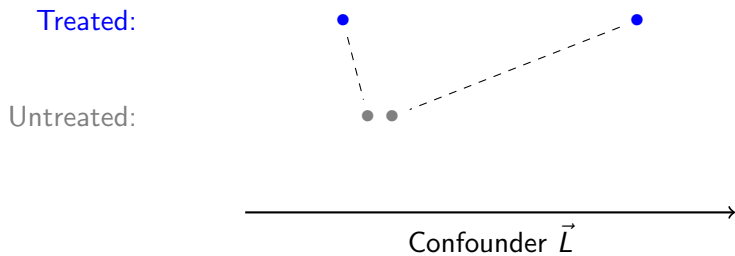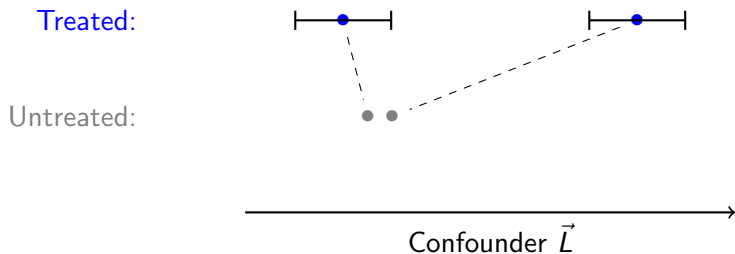
Caliper or no caliper matching
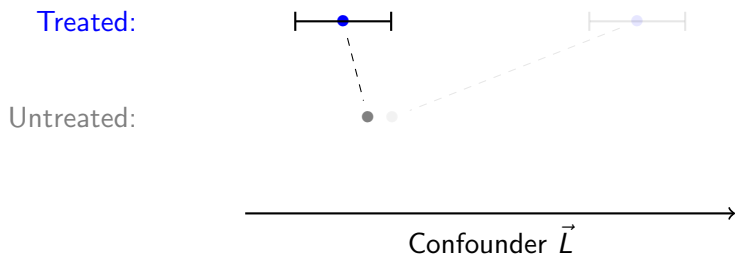
- ▶ Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius
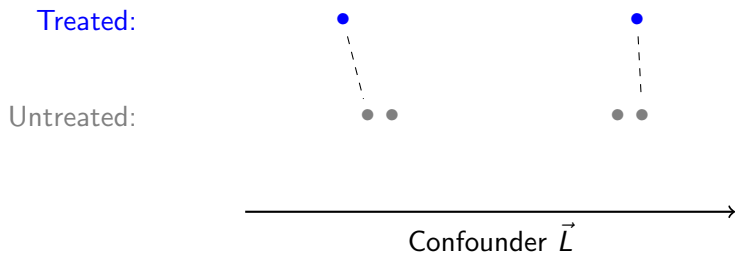- ▶ Feasible Sample Average Treatment Effect on the Treated (FSATT): Average among treated units for whom an acceptable match exists

1:1 vs k:1 matching

Treated:

Untreated:

Confounder $\vec{L}$
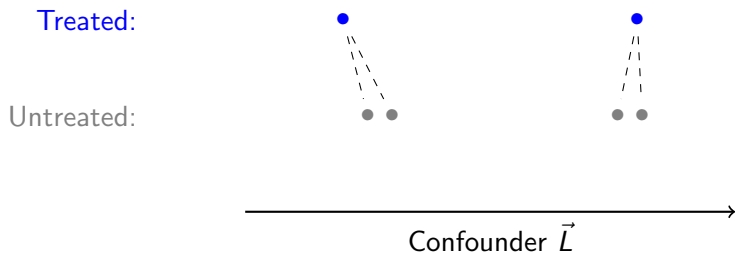
1:1 vs k:1 matching

1:1 vs k:1 matching

1:1 vs k:1 matching

Treated:

Untreated:
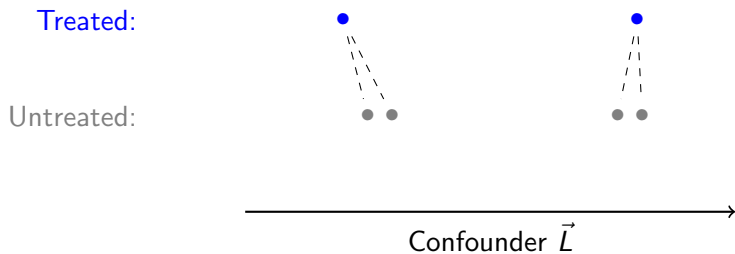
Confounder $\vec{L}$
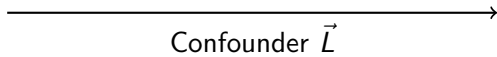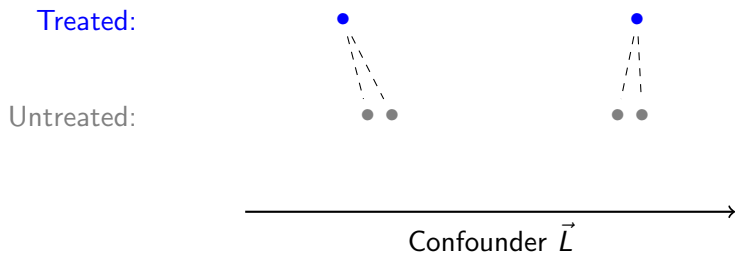
- ▶ Benefit of 2:1 matching

- ▶ Benefit of 1:1 matching
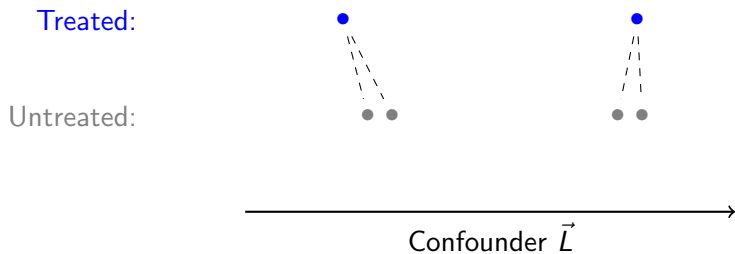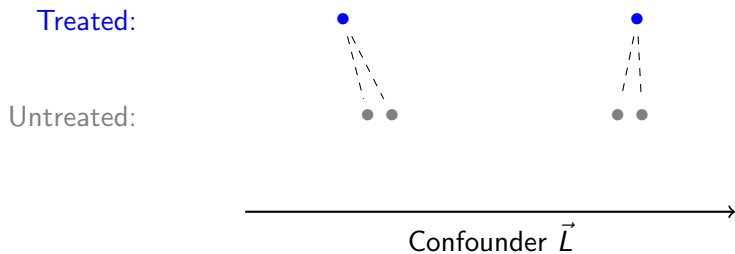
1:1 vs k:1 matching

Treated:

Untreated:

Confounder $\vec{L}$

- ▶ Benefit of 2:1 matching
  - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of 1:1 matching

1:1 vs k:1 matching


Treated:

Untreated:

Confounder $\vec{L}$

- ▶ Benefit of 2:1 matching
  - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of 1:1 matching
  - ▶ Lower bias. Only the best matches.

1:1 vs k:1 matching


Confounder $\vec{L}$

- ▶ Benefit of 2:1 matching
  - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of 1:1 matching
  - ▶ Lower bias. Only the best matches.
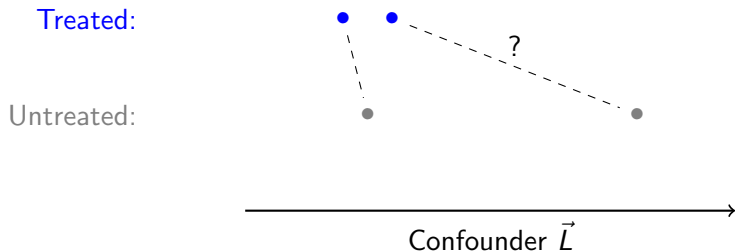- ▶ Greater $k \rightarrow$ lower variance, higher bias

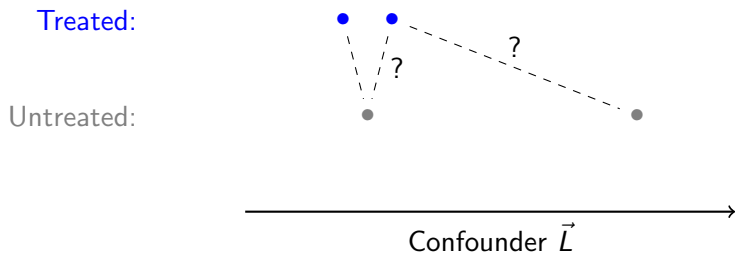# With replacement vs without replacement matching
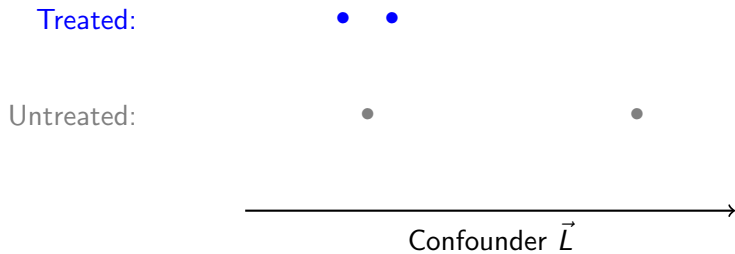
With replacement vs without replacement matching

Treated:

Untreated:

Confounder $\vec{L}$

With replacement vs without replacement matching

With replacement vs without replacement matching

Treated:

Untreated:

Confounder $\vec{L}$
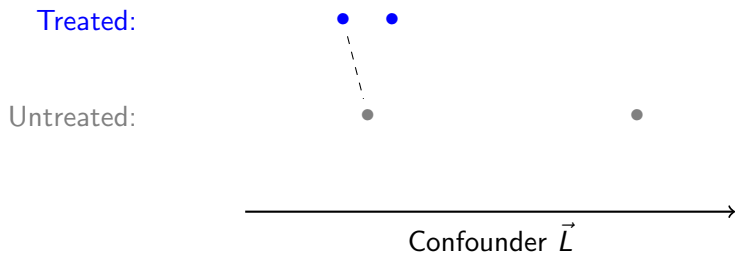
▶ Benefit of matching without replacement

▶ Benefit of matching with replacement
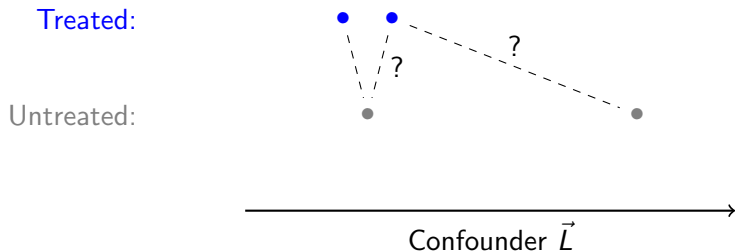
With replacement vs without replacement matching

- ▶ Benefit of matching without replacement
    - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of matching with replacement
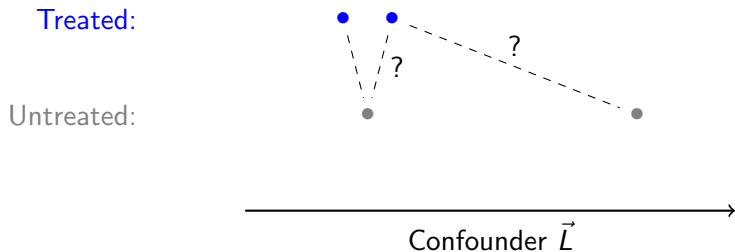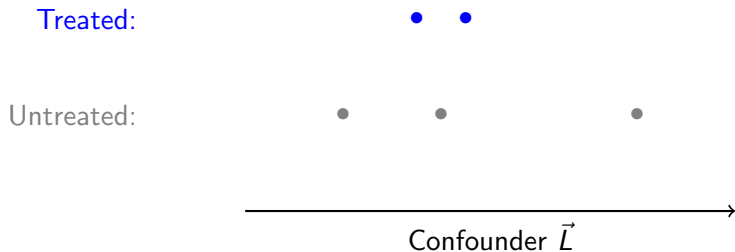
With replacement vs without replacement matching

- ▶ Benefit of matching without replacement
  - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of matching with replacement
  - ▶ Lower bias. Better matches.

Greedy vs optimal matching[2]

Treated:

Untreated:

Confounder $\vec{L}$

---

[2]Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405-420.

Greedy vs optimal matching[2]

Greedy Matching:
Match sequentially

Treated:

Untreated:

Confounder $\vec{L}$

[2]Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405-420.
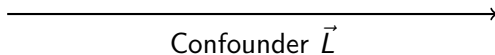
Greedy vs optimal matching[2]

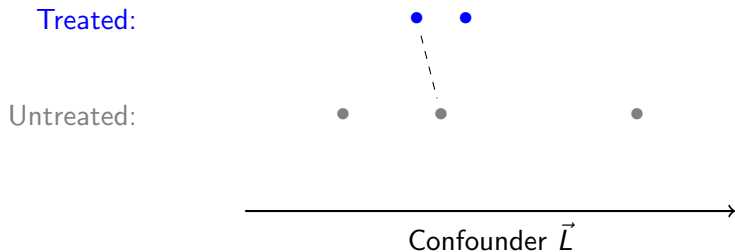Greedy Matching:
Match sequentially

Treated:

Untreated:

Confounder $\vec{L}$

---

[2]Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405-420.

Greedy vs optimal matching[2]

Greedy Matching:
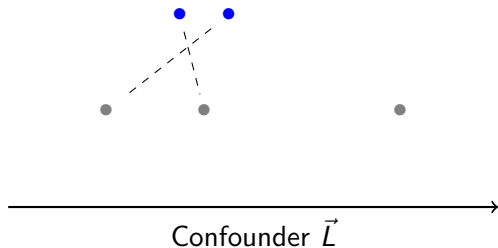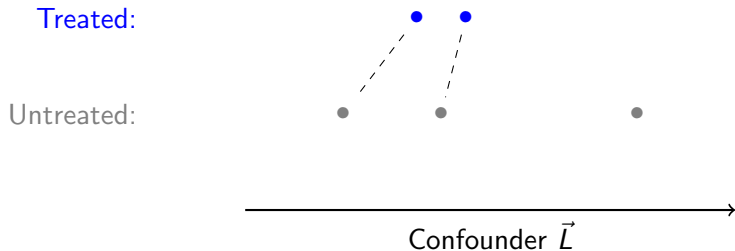Match sequentially

Treated:

Untreated:

Confounder $\vec{L}$

[2]Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405-420.

Greedy vs optimal matching[2]

Optimal Matching:
Consider the whole set of matches

Treated:

Untreated:

Confounder $\vec{L}$

---

[2]Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405-420.

Greedy vs optimal matching[2]
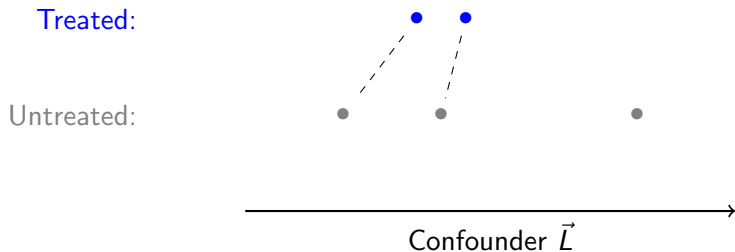

Optimal Matching:
Consider the whole set of matches

Treated:

Untreated:

Confounder $\vec{L}$

▶ Optimal is better. Just computationally harder.

---

[2]Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. Journal of Computational and Graphical Statistics, 2(4), 405-420.

Matching in univariate settings: Algorithms

- ► Caliper or no caliper
- ► 1:1 vs k:1
- ► With replacement vs without replacement
- ► Greedy vs optimal

Matching in univariate settings: Algorithms

- Caliper or no caliper
- 1:1 vs k:1
- With replacement vs without replacement
- Greedy vs optimal

Many reasonable choices, good choices depend on the data you have

Learning goals for today

At the end of class, you will be able to:

1. Explain how matching can be used to estimate causal effects
2. Explain bias variance trade-off in various matching procedures