

11. Matching

Ian Lundberg

Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

27 Sep 2022

Learning goals for today

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

1. Use matching methods for causal effects
 - ▶ Select a matching algorithm
 - ▶ Define a distance metric for multivariate matching
 - ▶ Evaluate matched sets

Matching: The big idea

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

$$\frac{1}{n_1} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

$$\frac{1}{n_1} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

Problem: We don't see Y_i^0

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

$$\frac{1}{n_1} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

Problem: We don't see Y_i^0

Solution: DAG + the g-formula

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

$$\frac{1}{n_1} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

Problem: We don't see Y_i^0

Solution: DAG + the g-formula

$$\frac{1}{n_1} \sum_{i:A_i=1} \left(Y_i^1 - E \left(Y \mid A = 0, \vec{L} = \vec{\ell}_i \right) \right)$$

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

$$\frac{1}{n_1} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

Problem: We don't see Y_i^0

Solution: DAG + the g-formula

$$\frac{1}{n_1} \sum_{i:A_i=1} \left(Y_i^1 - E \left(Y \mid A = 0, \vec{L} = \vec{\ell}_i \right) \right)$$

Matching: Estimate $E(Y \mid A = 0, \vec{L} = \vec{\ell}_i)$ from one or more untreated units with \vec{L} “near” $\vec{\ell}_i$

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

$$\frac{1}{n_1} \sum_{i:A_i=1} (Y_i^1 - Y_i^0)$$

Problem: We don't see Y_i^0

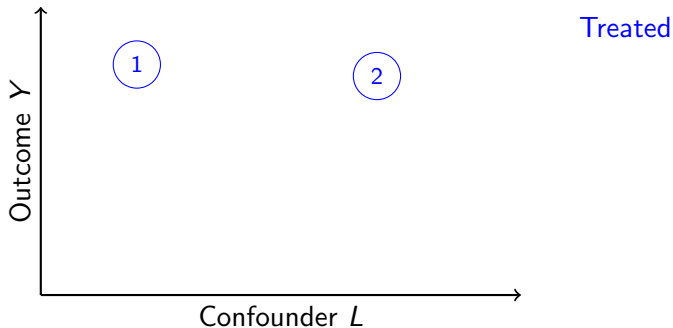
Solution: DAG + the g-formula

$$\frac{1}{n_1} \sum_{i:A_i=1} \left(Y_i^1 - E \left(Y \mid A = 0, \vec{L} = \vec{\ell}_i \right) \right)$$

Matching: Estimate $E(Y \mid A = 0, \vec{L} = \vec{\ell}_i)$ from one or more untreated units with \vec{L} “near” $\vec{\ell}_i$

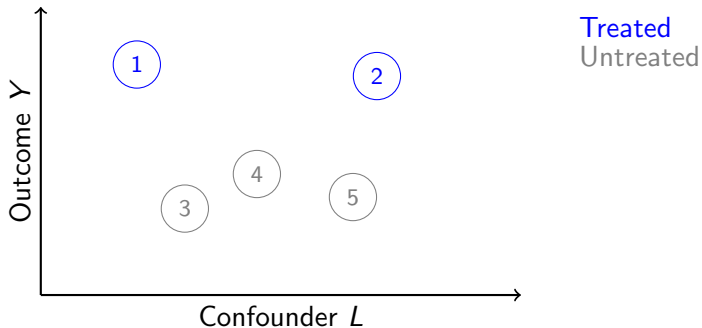
Debates: What does it mean to be “near”?

Matching: The big idea



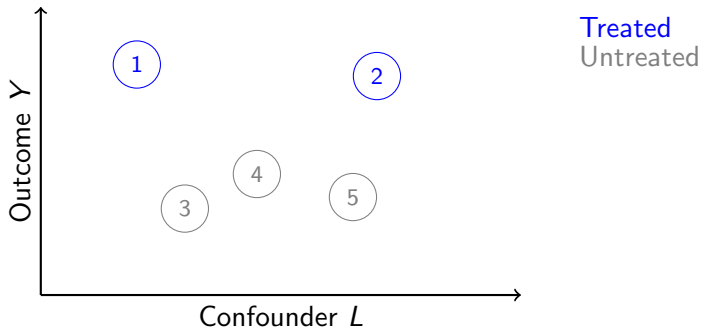
You have a some treated units.

Matching: The big idea



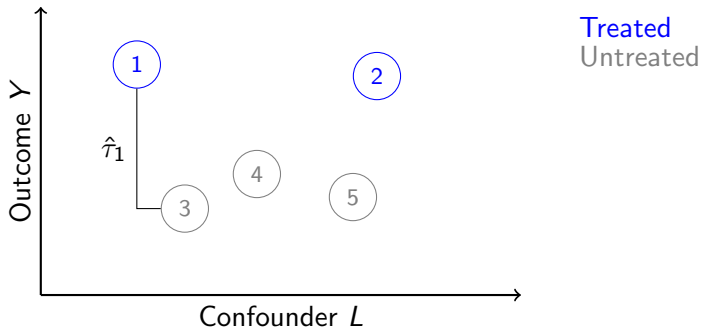
You go find some untreated units.

Matching: The big idea



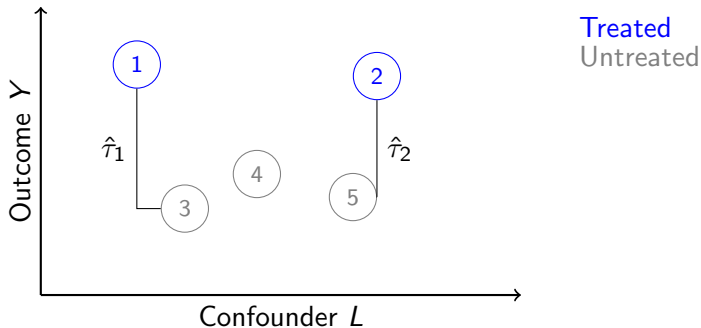
You find the closest matches along L .
You estimate each effect.

Matching: The big idea



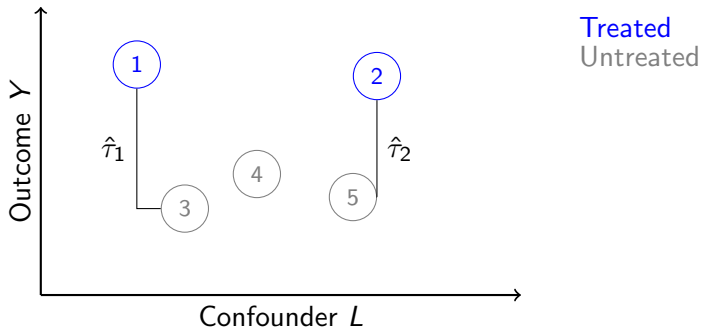
You find the closest matches along L .
You estimate each effect.

Matching: The big idea



You find the closest matches along L .
You estimate each effect.

Matching: The big idea



$$\widehat{\text{SATT}} = \frac{1}{2}(\hat{\tau}_1 + \hat{\tau}_2)$$

(Sample Average Treatment Effect on the Treated)

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. Completely transparent how we estimate \hat{Y}_i^0
(from the matched unit)

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Completely transparent how we estimate \hat{Y}_i^0
(from the matched unit)
3. Easy to explain

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Completely transparent how we estimate \hat{Y}_i^0
(from the matched unit)
3. Easy to explain
 - ▶ We had some treated units

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Completely transparent how we estimate \hat{Y}_i^0
(from the matched unit)
3. Easy to explain
 - ▶ We had some treated units
 - ▶ We found comparable control units

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Completely transparent how we estimate \hat{Y}_i^0
(from the matched unit)
3. Easy to explain
 - ▶ We had some treated units
 - ▶ We found comparable control units
 - ▶ We took a mean difference

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Completely transparent how we estimate \hat{Y}_i^0
(from the matched unit)
3. Easy to explain
 - ▶ We had some treated units
 - ▶ We found comparable control units
 - ▶ We took a mean difference
4. 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. Completely transparent how we estimate \hat{Y}_i^0
(from the matched unit)
3. Easy to explain
 - ▶ We had some treated units
 - ▶ We found comparable control units
 - ▶ We took a mean difference
4. Can assess quality of matches before we look at the outcome
5. Model-free*

Why matching is great

1. Completely transparent that Y_i^1 is observed
2. Completely transparent how we estimate \hat{Y}_i^0
(from the matched unit)
3. Easy to explain
 - ▶ We had some treated units
 - ▶ We found comparable control units
 - ▶ We took a mean difference
4. Can assess quality of matches before we look at the outcome
5. Model-free*
 - ▶ * but you have to define what makes a match “good”

Matching: A word of warning¹

¹Sekhon, J. S. (2009). [Opiates for the matches: Matching methods for causal inference](#). Annual Review of Political Science, 12(1), 487-508.

Matching: A word of warning¹

$$L \rightarrow A \rightarrow Y$$


Matching: A word of warning¹

Matching works!

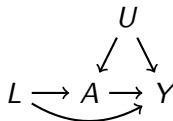
$$L \rightarrow A \rightarrow Y$$


Matching: A word of warning¹

Matching works!



No help!

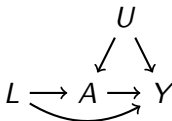


Matching: A word of warning¹

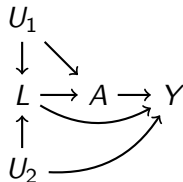
Matching works!



No help!



No help!



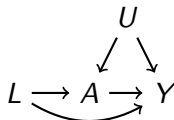
¹Sekhon, J. S. (2009). [Opiates for the matches: Matching methods for causal inference](#). Annual Review of Political Science, 12(1), 487-508.

Matching: A word of warning¹

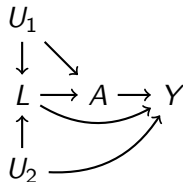
Matching works!



No help!



No help!



Matching is an estimation strategy.

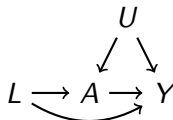
¹Sekhon, J. S. (2009). [Opiates for the matches: Matching methods for causal inference](#). Annual Review of Political Science, 12(1), 487-508.

Matching: A word of warning¹

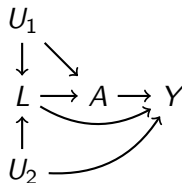
Matching works!



No help!



No help!



Matching is an estimation strategy.
It does not solve identification problems.

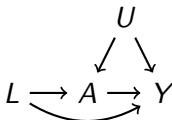
¹Sekhon, J. S. (2009). [Opiates for the matches: Matching methods for causal inference](#). Annual Review of Political Science, 12(1), 487-508.

Matching: A word of warning¹

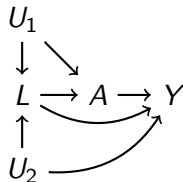
Matching works!



No help!



No help!



Matching is an estimation strategy.
It does not solve identification problems.
Matching is only as good as your DAG!

¹Sekhon, J. S. (2009). [Opiates for the matches: Matching methods for causal inference](#). Annual Review of Political Science, 12(1), 487-508.

Matching overview

Matching in univariate settings: Algorithms

Matching in multivariate settings: Distance metrics

After matching: Evaluate matched sets

Matching overview

Matching in univariate settings: Algorithms

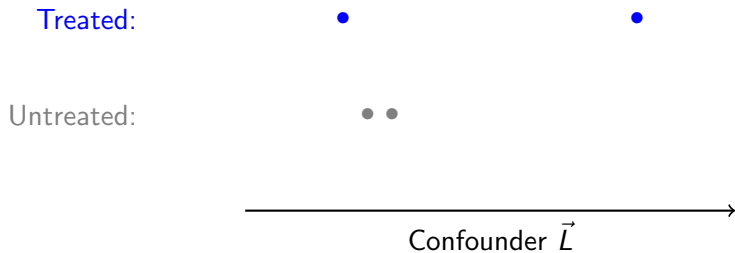
Matching in multivariate settings: Distance metrics

After matching: Evaluate matched sets

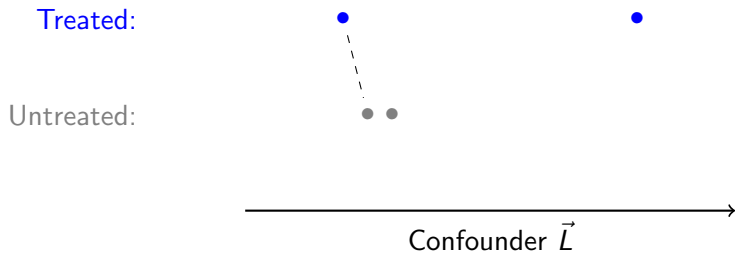
Matching in univariate settings: Algorithms

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

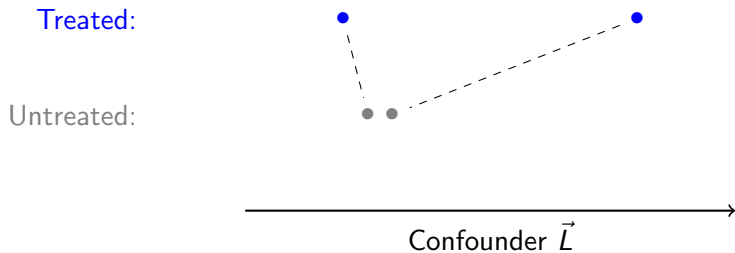
Caliper or no caliper matching



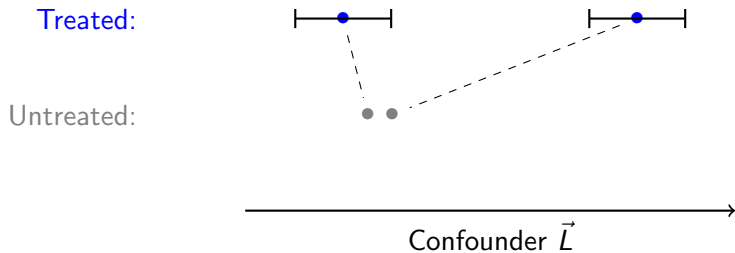
Caliper or no caliper matching



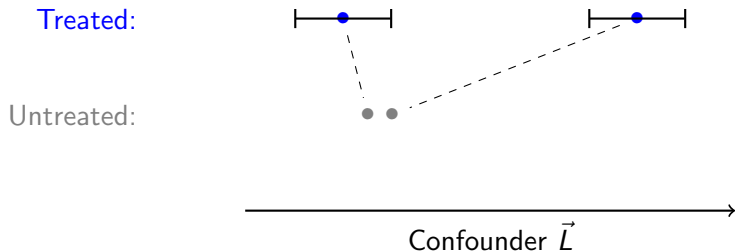
Caliper or no caliper matching



Caliper or no caliper matching

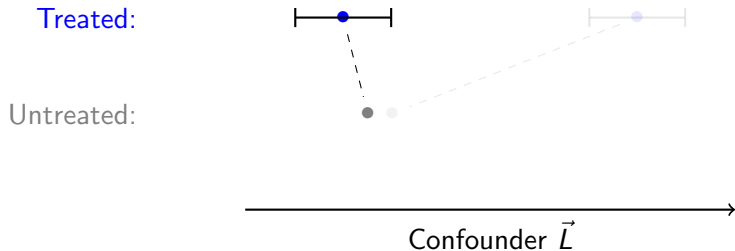


Caliper or no caliper matching



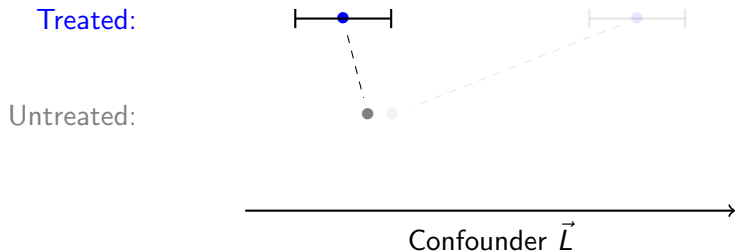
- Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius

Caliper or no caliper matching



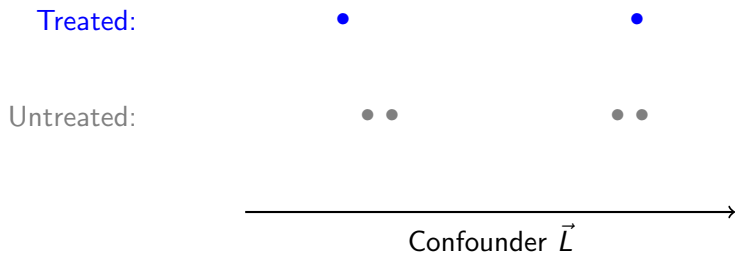
- Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius

Caliper or no caliper matching

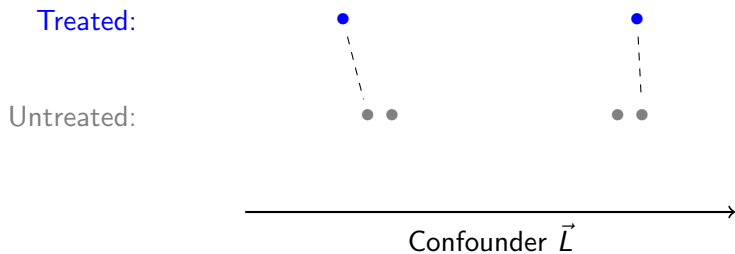


- ▶ Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius
- ▶ Feasible Sample Average Treatment Effect on the Treated (FSATT): Average among treated units for whom an acceptable match exists

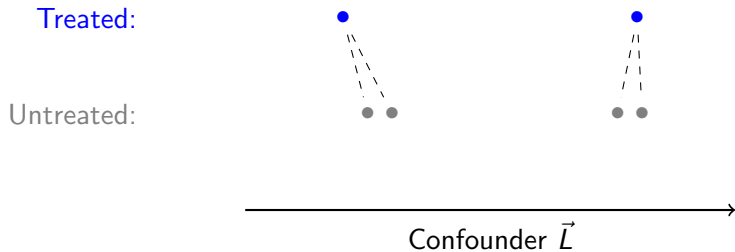
1:1 vs k :1 matching



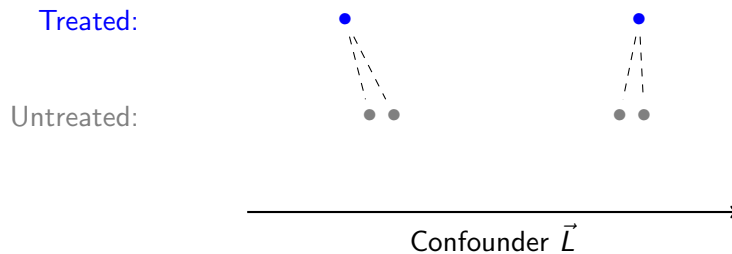
1:1 vs k :1 matching



1:1 vs k :1 matching

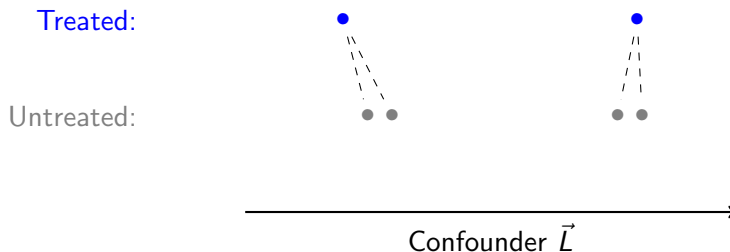


1:1 vs k :1 matching



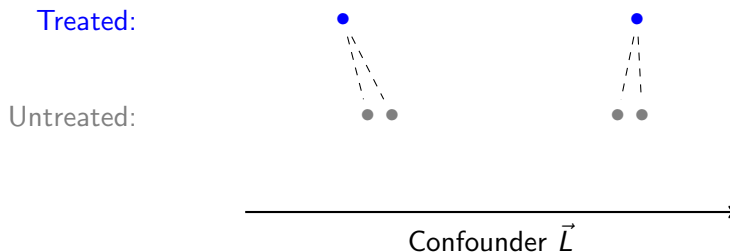
- Benefit of 2:1 matching
- Benefit of 1:1 matching

1:1 vs k :1 matching



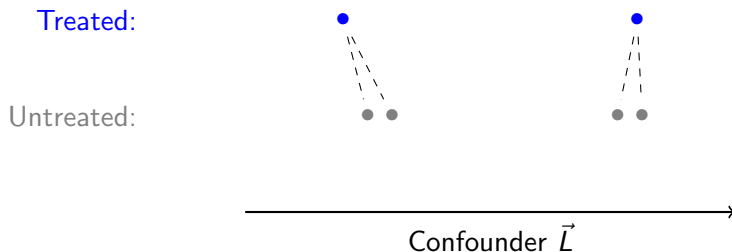
- ▶ Benefit of 2:1 matching
 - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of 1:1 matching

1:1 vs k :1 matching



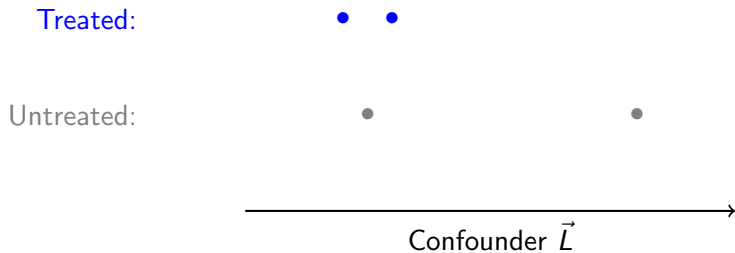
- ▶ Benefit of 2:1 matching
 - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of 1:1 matching
 - ▶ Lower bias. Only the best matches.

1:1 vs k :1 matching

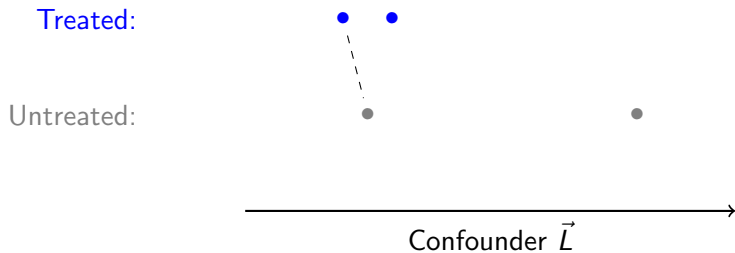


- ▶ Benefit of 2:1 matching
 - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of 1:1 matching
 - ▶ Lower bias. Only the best matches.
- ▶ Greater $k \rightarrow$ lower variance, higher bias

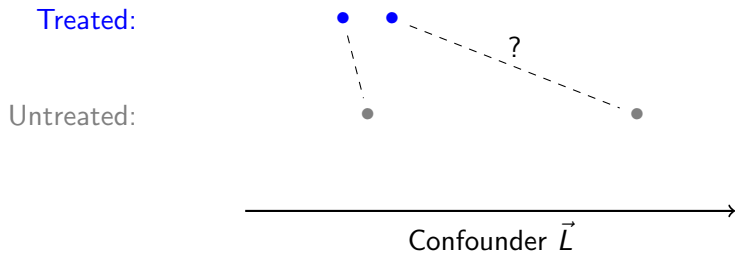
With replacement vs without replacement matching



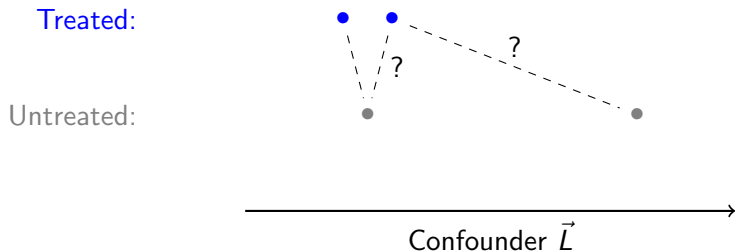
With replacement vs without replacement matching



With replacement vs without replacement matching

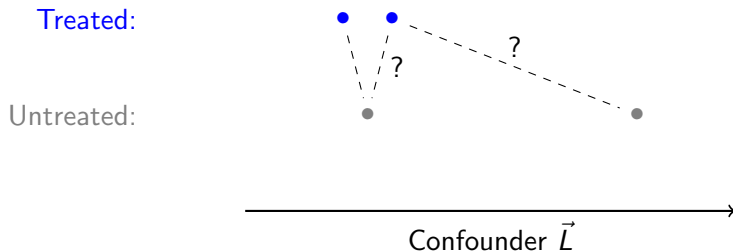


With replacement vs without replacement matching



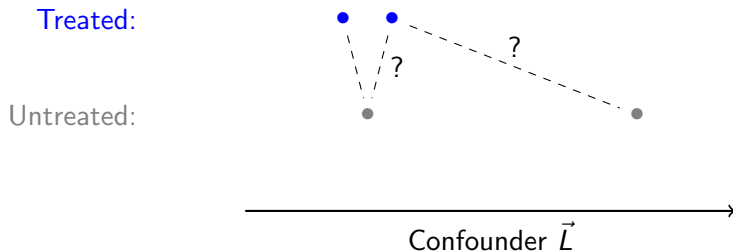
- Benefit of matching without replacement
- Benefit of matching with replacement

With replacement vs without replacement matching



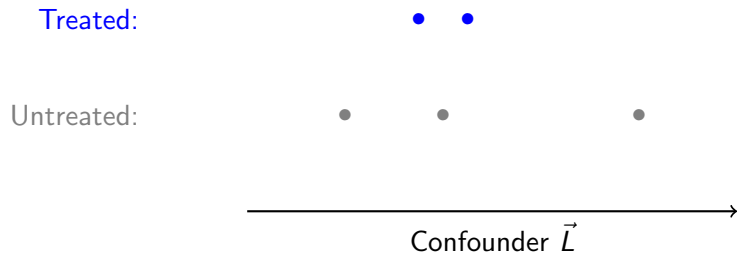
- ▶ Benefit of matching without replacement
 - ▶ Lower variance. Averaging over more cases.
- ▶ Benefit of matching with replacement

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.

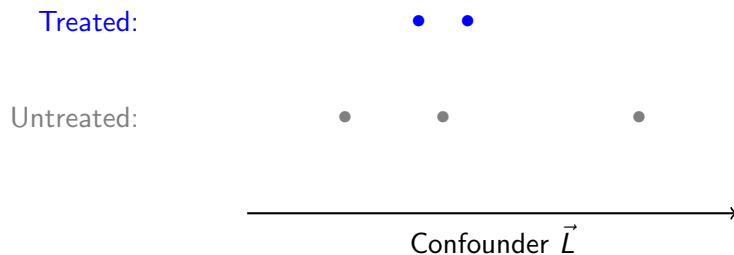
Greedy vs optimal matching²



²Gu, 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 matching²

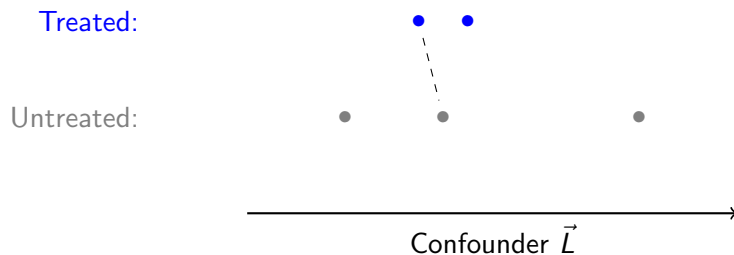
Greedy Matching:
Match sequentially



²Gu, 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 matching²

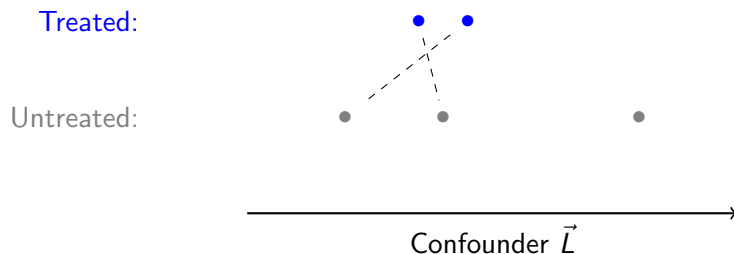
Greedy Matching:
Match sequentially



²Gu, 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 matching²

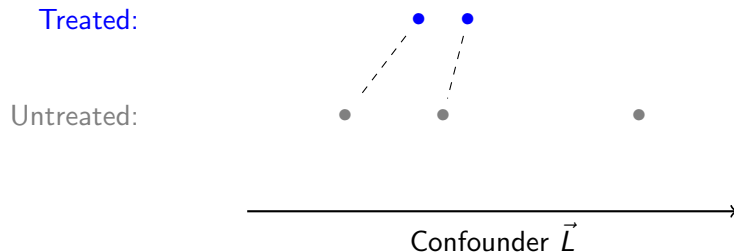
Greedy Matching:
Match sequentially



²Gu, 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 matching²

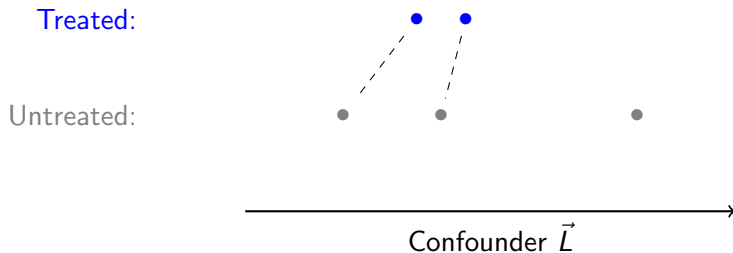
Optimal Matching:
Consider the whole set of matches



²Gu, 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 matching²

Optimal Matching:
Consider the whole set of matches



- Optimal is better. Just computationally harder.

²Gu, 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 (recap)

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

Matching overview

Matching in univariate settings: Algorithms

Matching in multivariate settings: Distance metrics

After matching: Evaluate matched sets

Matching overview

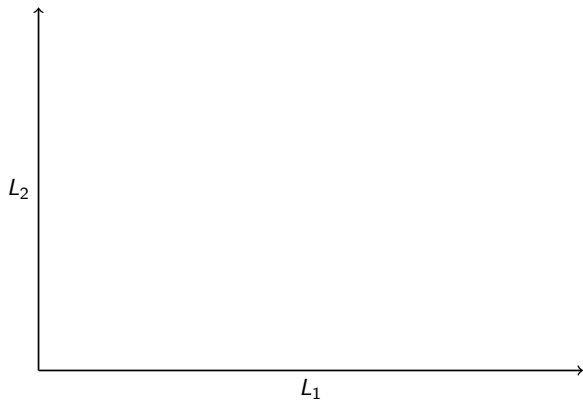
Matching in univariate settings: Algorithms

Matching in multivariate settings: Distance metrics

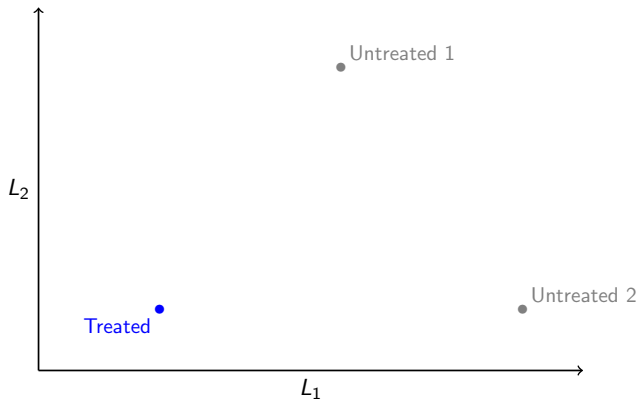
After matching: Evaluate matched sets

What if \vec{L} is multivariate?

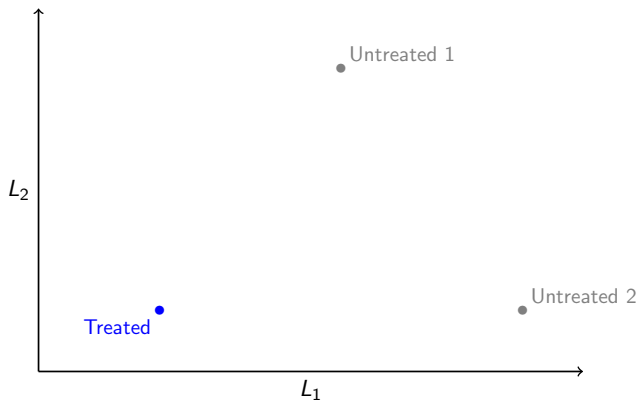
What if \vec{L} is multivariate?



What if \vec{L} is multivariate?

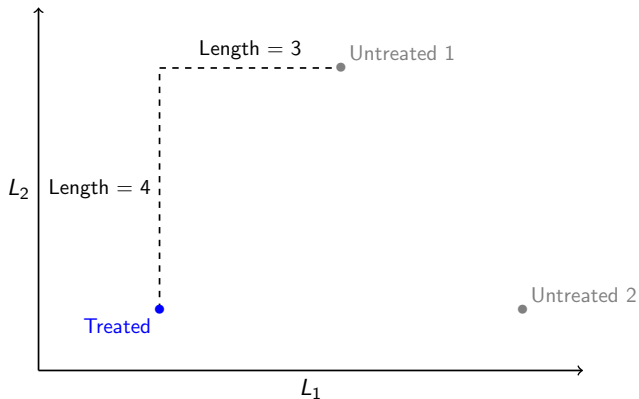


What if \vec{L} is multivariate?



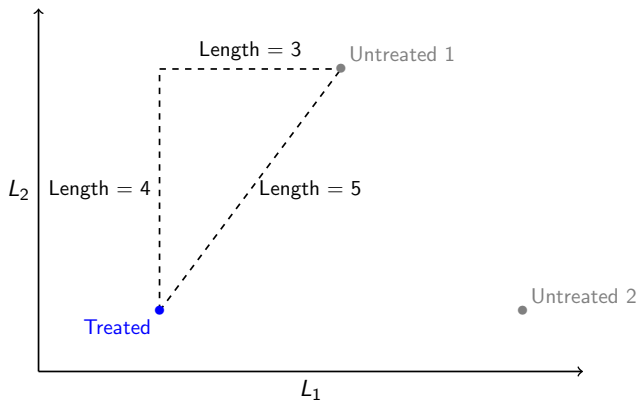
Which untreated unit should be the match?

What if \vec{L} is multivariate?



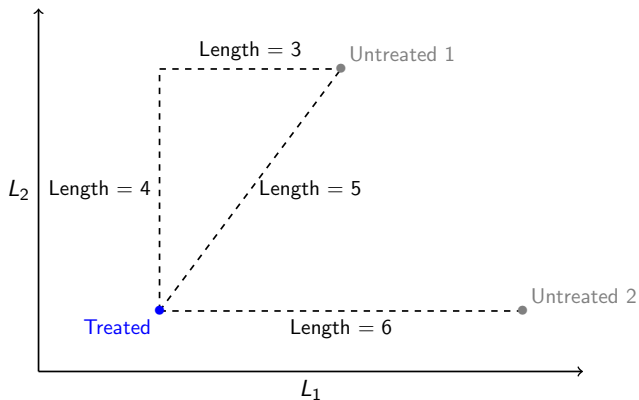
Which untreated unit should be the match?

What if \vec{L} is multivariate?



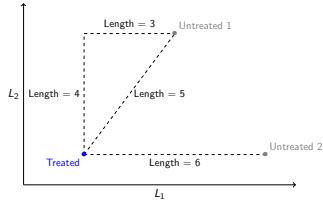
Which untreated unit should be the match?

What if \vec{L} is multivariate?

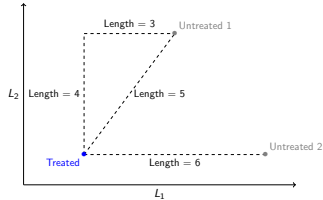


Which untreated unit should be the match?

What if \vec{L} is multivariate? We need a **distance metric**



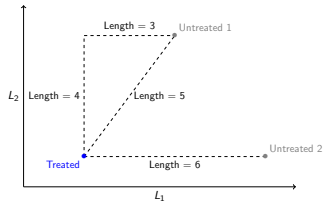
What if \vec{L} is multivariate? We need a **distance metric**



► Manhattan distance:

► Euclidean distance:

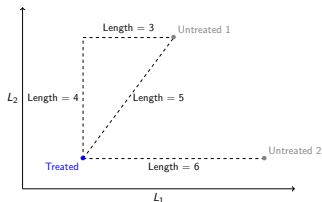
What if \vec{L} is multivariate? We need a **distance metric**



► Manhattan distance: $d(i, j) = \sum_p |L_{pi} - L_{pj}|$

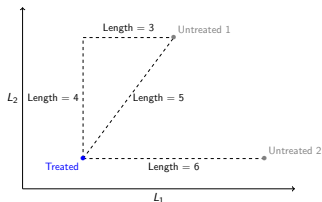
► Euclidean distance:

What if \vec{L} is multivariate? We need a **distance metric**



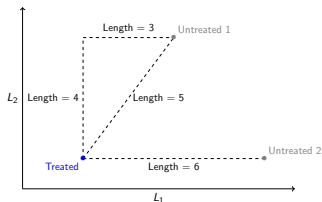
- ▶ Manhattan distance: $d(i, j) = \sum_p |L_{pi} - L_{pj}|$
 - ▶ $d(\text{Treated}, \text{Untreated 1}) = 3 + 4 = 7$
 - ▶ $d(\text{Treated}, \text{Untreated 2}) = 6 + 0 = 6 \checkmark$
- ▶ Euclidean distance:

What if \vec{L} is multivariate? We need a **distance metric**



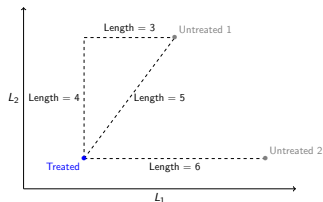
- ▶ Manhattan distance: $d(i,j) = \sum_p |L_{pi} - L_{pj}|$
 - ▶ $d(\text{Treated}, \text{Untreated 1}) = 3 + 4 = 7$
 - ▶ $d(\text{Treated}, \text{Untreated 2}) = 6 + 0 = 6 \checkmark$
- ▶ Euclidean distance: $d(i,j) = \sqrt{\sum_p (L_{pi} - L_{pj})^2}$

What if \vec{L} is multivariate? We need a **distance metric**



- ▶ Manhattan distance: $d(i,j) = \sum_p |L_{pi} - L_{pj}|$
 - ▶ $d(\text{Treated}, \text{Untreated 1}) = 3 + 4 = 7$
 - ▶ $d(\text{Treated}, \text{Untreated 2}) = 6 + 0 = 6 \checkmark$
- ▶ Euclidean distance: $d(i,j) = \sqrt{\sum_p (L_{pi} - L_{pj})^2}$
 - ▶ $d(\text{Treated}, \text{Untreated 1}) = \sqrt{3^2 + 4^2} = 5 \checkmark$
 - ▶ $d(\text{Treated}, \text{Untreated 2}) = \sqrt{6^2 + 0^2} = 6$

What if \vec{L} is multivariate? We need a **distance metric**



- ▶ Manhattan distance: $d(i, j) = \sum_p |L_{pi} - L_{pj}|$
 - ▶ $d(\text{Treated}, \text{Untreated 1}) = 3 + 4 = 7$
 - ▶ $d(\text{Treated}, \text{Untreated 2}) = 6 + 0 = 6 \checkmark$
- ▶ Euclidean distance: $d(i, j) = \sqrt{\sum_p (L_{pi} - L_{pj})^2}$
 - ▶ $d(\text{Treated}, \text{Untreated 1}) = \sqrt{3^2 + 4^2} = 5 \checkmark$
 - ▶ $d(\text{Treated}, \text{Untreated 2}) = \sqrt{6^2 + 0^2} = 6$
- ▶ It depends on the distance metric!

A common distance metric: Mahalanobis distance

A common distance metric: Mahalanobis distance

Motivated by two principles

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100
 - ▶ Suppose L_2 ranges uniformly from 0 to 1

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100
 - ▶ Suppose L_2 ranges uniformly from 0 to 1
 - ▶ We might correct for this so L_1 doesn't dominate the distance

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100
 - ▶ Suppose L_2 ranges uniformly from 0 to 1
 - ▶ We might correct for this so L_1 doesn't dominate the distance
- ▶ Principle 2: Address correlations

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100
 - ▶ Suppose L_2 ranges uniformly from 0 to 1
 - ▶ We might correct for this so L_1 doesn't dominate the distance
- ▶ Principle 2: Address correlations
 - ▶ Suppose L_1 and L_2 are very correlated

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100
 - ▶ Suppose L_2 ranges uniformly from 0 to 1
 - ▶ We might correct for this so L_1 doesn't dominate the distance
- ▶ Principle 2: Address correlations
 - ▶ Suppose L_1 and L_2 are very correlated
 - ▶ Suppose L_3 is independent

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100
 - ▶ Suppose L_2 ranges uniformly from 0 to 1
 - ▶ We might correct for this so L_1 doesn't dominate the distance
- ▶ Principle 2: Address correlations
 - ▶ Suppose L_1 and L_2 are very correlated
 - ▶ Suppose L_3 is independent
 - ▶ We should care about a correlation-corrected distance

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100
 - ▶ Suppose L_2 ranges uniformly from 0 to 1
 - ▶ We might correct for this so L_1 doesn't dominate the distance
- ▶ Principle 2: Address correlations
 - ▶ Suppose L_1 and L_2 are very correlated
 - ▶ Suppose L_3 is independent
 - ▶ We should care about a correlation-corrected distance

A common distance metric: Mahalanobis distance

Motivated by two principles

- ▶ Principle 1: Address unequal variances
 - ▶ Suppose L_1 ranges uniformly from 0 to 100
 - ▶ Suppose L_2 ranges uniformly from 0 to 1
 - ▶ We might correct for this so L_1 doesn't dominate the distance
- ▶ Principle 2: Address correlations
 - ▶ Suppose L_1 and L_2 are very correlated
 - ▶ Suppose L_3 is independent
 - ▶ We should care about a correlation-corrected distance

$$d(i,j) = \sqrt{(\vec{L}_i - \vec{L}_j)^T \Sigma^{-1} (\vec{L}_i - \vec{L}_j)}$$

where $\Sigma = V(\vec{L})$, the variance-covariance matrix

A common distance metric: Exact matching

A common distance metric: Exact matching

- ▶ Equivalent to nonparametric stratification

A common distance metric: Exact matching

- ▶ Equivalent to nonparametric stratification
- ▶ Infinite distance if any confounder is different!

$$d(i, j) = \begin{cases} 0 & \text{if } \vec{L}_i = \vec{L}_j \\ \infty & \text{if } \vec{L}_i \neq \vec{L}_j \end{cases}$$

Often leads to **no matches at all**

A common distance metric: Coarsened exact matching³

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- Define \tilde{L} to be a coarsened version of \vec{L}

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). *Political Analysis*, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). *Political Analysis*, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories
- ▶ Match exactly on $\tilde{\vec{L}}$

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). *Political Analysis*, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories
- ▶ Match exactly on $\tilde{\vec{L}}$

$$d(i, j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). *Political Analysis*, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- ▶ Define \tilde{L} to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories
- ▶ Match exactly on \tilde{L}

$$d(i, j) = \begin{cases} 0 & \text{if } \tilde{L}_i = \tilde{L}_j \\ \infty & \text{if } \tilde{L}_i \neq \tilde{L}_j \end{cases}$$

- ▶ Benefit: Very transparent

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories
- ▶ Match exactly on $\tilde{\vec{L}}$

$$d(i, j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

- ▶ Benefit: Very transparent
- ▶ Drawback: Ignores \vec{L} within $\tilde{\vec{L}}$

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories
- ▶ Match exactly on $\tilde{\vec{L}}$

$$d(i, j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

- ▶ Benefit: Very transparent
- ▶ Drawback: Ignores \vec{L} within $\tilde{\vec{L}}$
- ▶ Best used as preprocessing for regression

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories
- ▶ Match exactly on $\tilde{\vec{L}}$

$$d(i, j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

- ▶ Benefit: Very transparent
- ▶ Drawback: Ignores \vec{L} within $\tilde{\vec{L}}$
- ▶ Best used as preprocessing for regression
 - ▶ Conduct coarsened exact matching

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories
- ▶ Match exactly on $\tilde{\vec{L}}$

$$d(i, j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

- ▶ Benefit: Very transparent
- ▶ Drawback: Ignores \vec{L} within $\tilde{\vec{L}}$
- ▶ Best used as preprocessing for regression
 - ▶ Conduct coarsened exact matching
 - ▶ Conduct regression on the matched dataset

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). Political Analysis, 20(1), 1-24.

A common distance metric: Coarsened exact matching³

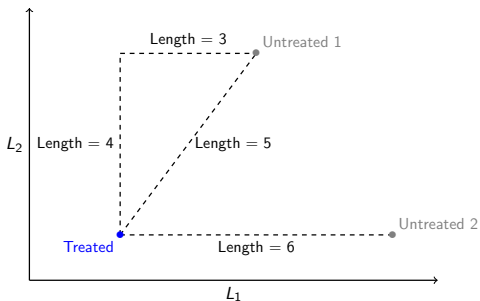
- ▶ Define $\tilde{\vec{L}}$ to be a coarsened version of \vec{L}
 - ▶ Example: Convert all continuous values to decile categories
- ▶ Match exactly on $\tilde{\vec{L}}$

$$d(i, j) = \begin{cases} 0 & \text{if } \tilde{\vec{L}}_i = \tilde{\vec{L}}_j \\ \infty & \text{if } \tilde{\vec{L}}_i \neq \tilde{\vec{L}}_j \end{cases}$$

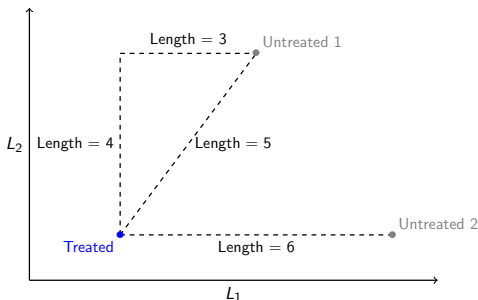
- ▶ Benefit: Very transparent
- ▶ Drawback: Ignores \vec{L} within $\tilde{\vec{L}}$
- ▶ Best used as preprocessing for regression
 - ▶ Conduct coarsened exact matching
 - ▶ Conduct regression on the matched dataset
 - ▶ See paper

³Iacus, S. M., King, G., & Porro, G. (2012). [Causal inference without balance checking: Coarsened exact matching](#). Political Analysis, 20(1), 1-24.

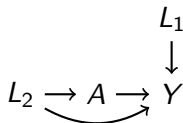
A common distance metric: Propensity scores



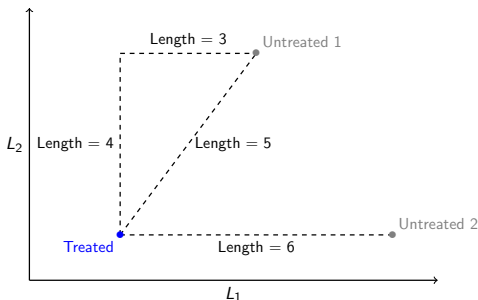
A common distance metric: Propensity scores



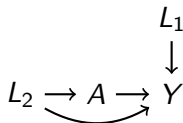
Now suppose only L_2 is related to treatment. L_1 doesn't matter.



A common distance metric: Propensity scores

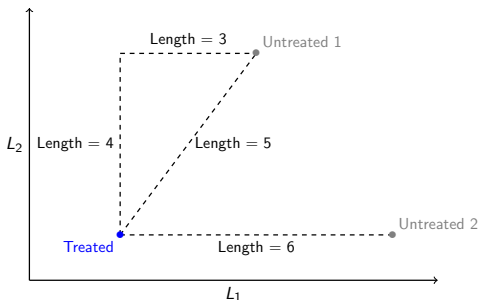


Now suppose only L_2 is related to treatment. L_1 doesn't matter.

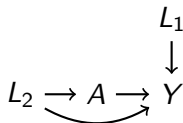


Which match do you pick?

A common distance metric: Propensity scores



Now suppose only L_2 is related to treatment. L_1 doesn't matter.



Which match do you pick? Untreated 2! Perfect match.

A common distance metric: Propensity scores

Propensity score: $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$

⁴Rosenbaum, P. R., & Rubin, D. B. (1983). [The central role of the propensity score in observational studies for causal effects](#). *Biometrika*, 70(1), 41-55.

A common distance metric: Propensity scores

Propensity score: $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$

- Univariate summary of all confounders

⁴Rosenbaum, P. R., & Rubin, D. B. (1983). [The central role of the propensity score in observational studies for causal effects](#). *Biometrika*, 70(1), 41-55.

A common distance metric: Propensity scores

Propensity score: $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$

- ▶ Univariate summary of all confounders
- ▶ In expectation, a sample balanced on π is balanced on \vec{L}
 - ▶ Rosenbaum & Rubin theorem⁴

⁴Rosenbaum, P. R., & Rubin, D. B. (1983). [The central role of the propensity score in observational studies for causal effects](#). *Biometrika*, 70(1), 41-55.

A common distance metric: Propensity scores

Propensity scores can be nonparametric or parametric.

A common distance metric: Propensity scores

Propensity scores can be nonparametric or parametric.

- Nonparametric: $\hat{\pi}_i$ is the proportion treated in the sample stratum $\vec{L} = \vec{\ell}_i$.

A common distance metric: Propensity scores

Propensity scores can be nonparametric or parametric.

- ▶ Nonparametric: $\hat{\pi}_i$ is the proportion treated in the sample stratum $\vec{L} = \vec{\ell}_i$.
- ▶ Parametric: Often estimated as
 - ▶ Fit logistic regression

$$\text{logit} \left(P(A = 1 \mid \vec{L}) \right) = \alpha + \vec{L}\vec{\beta}$$

A common distance metric: Propensity scores

Propensity scores can be nonparametric or parametric.

- ▶ Nonparametric: $\hat{\pi}_i$ is the proportion treated in the sample stratum $\vec{L} = \vec{\ell}_i$.
- ▶ Parametric: Often estimated as
 - ▶ Fit logistic regression

$$\text{logit} \left(P(A = 1 \mid \vec{L}) \right) = \alpha + \vec{L}\vec{\beta}$$

- ▶ Predict the probability of treatment

$$\hat{\pi}_i = \text{logit}^{-1} \left(\hat{\alpha} + \vec{\ell}_i \hat{\vec{\beta}} \right)$$

A common distance metric: Propensity scores

Propensity scores can be nonparametric or parametric.

- ▶ Nonparametric: $\hat{\pi}_i$ is the proportion treated in the sample stratum $\vec{L} = \vec{\ell}_i$.
- ▶ Parametric: Often estimated as
 - ▶ Fit logistic regression

$$\text{logit} \left(P(A = 1 \mid \vec{L}) \right) = \alpha + \vec{L}\vec{\beta}$$

- ▶ Predict the probability of treatment

$$\hat{\pi}_i = \text{logit}^{-1} \left(\hat{\alpha} + \vec{\ell}_i \hat{\vec{\beta}} \right)$$

Propensity score distance for matching:

$$d(i, j) = |\hat{\pi}_i - \hat{\pi}_j|$$

Why propensity scores are nice

Why propensity scores are nice

- Multivariate \vec{L}_i to univariate π_i

Why propensity scores are nice

- ▶ Multivariate \vec{L}_i to univariate π_i
 - ▶ Easy to reason about

Why propensity scores are nice

- ▶ Multivariate \vec{L}_i to univariate π_i
 - ▶ Easy to reason about
 - ▶ Can directly visualize the univariate matches

Why propensity scores are nice

- ▶ Multivariate \vec{L}_i to univariate π_i
 - ▶ Easy to reason about
 - ▶ Can directly visualize the univariate matches
- ▶ Mathematical guarantees in expectation

Why propensity scores are nice

- ▶ Multivariate \vec{L}_i to univariate π_i
 - ▶ Easy to reason about
 - ▶ Can directly visualize the univariate matches
- ▶ Mathematical guarantees in expectation
- ▶ Intuitive: Prioritizes covariates that predict treatment

Propensity score problems: King & Nielsen⁵

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip
 - ▶ Imbalance on \vec{L} is zero in every sample

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip
 - ▶ Imbalance on \vec{L} is zero in every sample
- ▶ Experimental ideal 2: Completely randomized experiment

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip
 - ▶ Imbalance on \vec{L} is zero in every sample
- ▶ Experimental ideal 2: Completely randomized experiment
 - ▶ Assign each unit a known $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip
 - ▶ Imbalance on \vec{L} is zero in every sample
- ▶ Experimental ideal 2: Completely randomized experiment
 - ▶ Assign each unit a known $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$
 - ▶ Randomize treatment weighted by π_i

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip
 - ▶ Imbalance on \vec{L} is zero in every sample
- ▶ Experimental ideal 2: Completely randomized experiment
 - ▶ Assign each unit a known $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$
 - ▶ Randomize treatment weighted by π_i
 - ▶ Imbalance on \vec{L} is zero in expectation, but not in any particular sample

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip
 - ▶ Imbalance on \vec{L} is zero in every sample
- ▶ Experimental ideal 2: Completely randomized experiment
 - ▶ Assign each unit a known $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$
 - ▶ Randomize treatment weighted by π_i
 - ▶ Imbalance on \vec{L} is zero in expectation, but not in any particular sample

In a randomized experiment, we would always choose (1).

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip
 - ▶ Imbalance on \vec{L} is zero in every sample
- ▶ Experimental ideal 2: Completely randomized experiment
 - ▶ Assign each unit a known $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$
 - ▶ Randomize treatment weighted by π_i
 - ▶ Imbalance on \vec{L} is zero in expectation, but not in any particular sample

In a randomized experiment, we would always choose (1).

But propensity score matching approximates (2).

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity score problems: King & Nielsen⁵

Motivated by two designs for randomized experiments

- ▶ Experimental ideal 1: Fully blocked
 - ▶ First, stratify units on \vec{L}
 - ▶ Then, randomize A within blocks by a coin flip
 - ▶ Imbalance on \vec{L} is zero in every sample
- ▶ Experimental ideal 2: Completely randomized experiment
 - ▶ Assign each unit a known $\pi_i = P(A = 1 \mid \vec{L} = \vec{\ell}_i)$
 - ▶ Randomize treatment weighted by π_i
 - ▶ Imbalance on \vec{L} is zero in expectation, but not in any particular sample

In a randomized experiment, we would always choose (1).

But propensity score matching approximates (2).

King & Nielsen claim: Better to match on \vec{L} directly

⁵King, G., & Nielsen, R. (2019). [Why propensity scores should not be used for matching](#). Political Analysis, 27(4), 435-454.

Propensity scores: My view

⁶Brand, J. E., & Xie, Y. (2010). [Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education.](#) *American Sociological Review*, 75(2), 273-302.

Propensity scores: My view

There is no uniformly superior method. Always application-specific.

⁶Brand, J. E., & Xie, Y. (2010). [Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education.](#) American Sociological Review, 75(2), 273-302.

Propensity scores: My view

There is no uniformly superior method. Always application-specific.

- ▶ Propensity scores are most popular

⁶Brand, J. E., & Xie, Y. (2010). [Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education.](#) American Sociological Review, 75(2), 273-302.

Propensity scores: My view

There is no uniformly superior method. Always application-specific.

- ▶ Propensity scores are most popular
- ▶ Other matching distances might be better in finite samples

⁶Brand, J. E., & Xie, Y. (2010). [Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education.](#) American Sociological Review, 75(2), 273-302.

Propensity scores: My view

There is no uniformly superior method. Always application-specific.

- ▶ Propensity scores are most popular
- ▶ Other matching distances might be better in finite samples

Why might you use propensity scores?

⁶Brand, J. E., & Xie, Y. (2010). [Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education.](#) American Sociological Review, 75(2), 273-302.

Propensity scores: My view

There is no uniformly superior method. Always application-specific.

- ▶ Propensity scores are most popular
- ▶ Other matching distances might be better in finite samples

Why might you use propensity scores?

- ▶ Sometimes they are substantively meaningful⁶

⁶Brand, J. E., & Xie, Y. (2010). [Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education.](#) American Sociological Review, 75(2), 273-302.

Propensity scores: My view

There is no uniformly superior method. Always application-specific.

- ▶ Propensity scores are most popular
- ▶ Other matching distances might be better in finite samples

Why might you use propensity scores?

- ▶ Sometimes they are substantively meaningful⁶
- ▶ Useful in other ways (e.g., weighting)

⁶Brand, J. E., & Xie, Y. (2010). [Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education.](#) American Sociological Review, 75(2), 273-302.

Multivariate distances: Recap

When matching on multivariate \vec{L} , you have to define the distance between each pair of confounder values $\vec{\ell}$ and $\vec{\ell}'$

- ▶ Manhattan distance
- ▶ Euclidean distance
- ▶ Mahalanobis distance
- ▶ Exact distance
- ▶ Coarsened exact distance
- ▶ Propensity score distance

There is no right answer! Depends on the setting.

Matching overview

Matching in univariate settings: Algorithms

Matching in multivariate settings: Distance metrics

After matching: Evaluate matched sets

Matching overview

Matching in univariate settings: Algorithms

Matching in multivariate settings: Distance metrics

After matching: Evaluate matched sets

Evaluate the matched sets

Why we match: So \vec{L} follows a similar distribution

- ▶ in the treated sample
- ▶ in the untreated sample

Whatever method, you should check that.

- ▶ Compare means of \vec{L} across groups
- ▶ Possibly compare interactive cells
- ▶ Ideally, before looking at Y !

Learning goals for today

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

1. Use matching methods for causal effects
 - ▶ Select a matching algorithm
 - ▶ Define a distance metric for multivariate matching
 - ▶ Evaluate matched sets

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!