Matching

Ian Lundberg
Soc 114

Winter 2025

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
2. Reason about choosing regression vs matching

Matching: The big idea

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

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

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

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

Problem: We don't see Y_i^0

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

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

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

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

Problem: We don't see Y_i^0

Solution: DAG + the g-formula

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

Matching: The big idea

Goal: Sample Average Treatment Effect on the Treated

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

Problem: We don't see Y_i^0

Solution: DAG + the g-formula

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

Matching: Estimate $\mathrm{E}(Y \mid A = 0, \vec{L} = \vec{\ell}_i)$ 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

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

Problem: We don't see Y_i^0

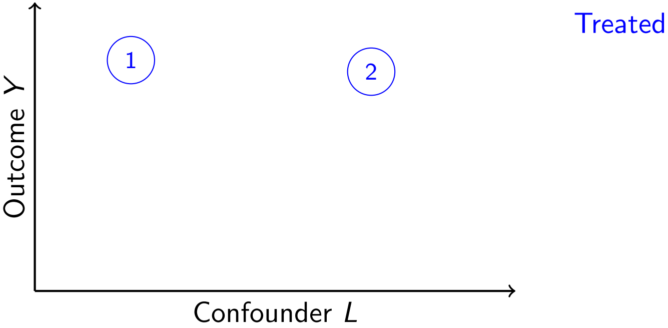
Solution: DAG + the g-formula

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

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

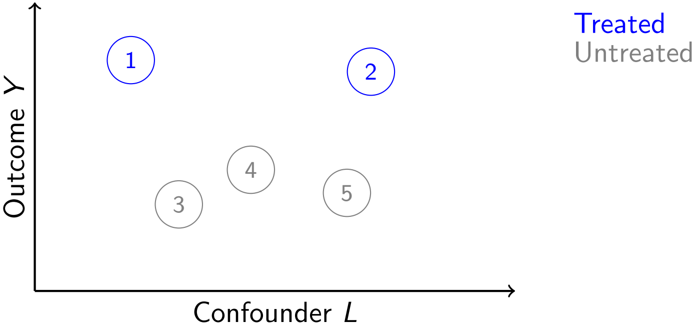
Debates: What does it mean to be "near"?

Matching: The big idea



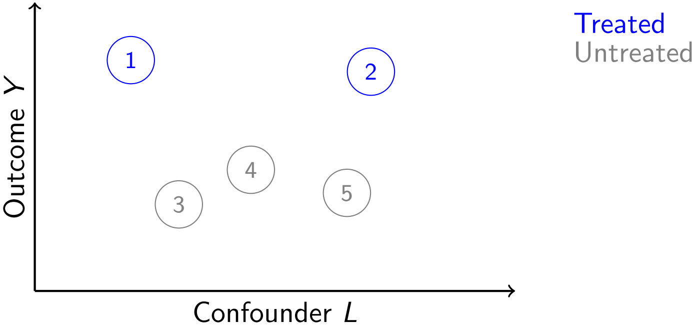
You have a some treated units.

Matching: The big idea



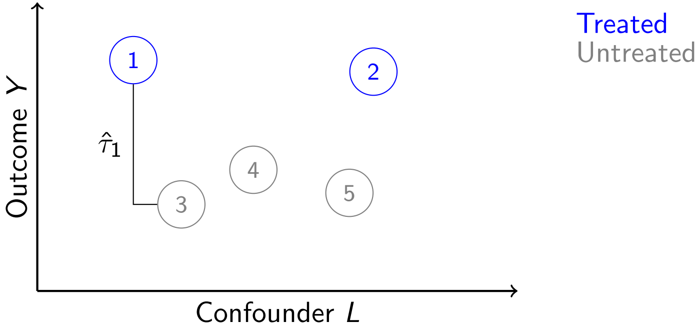
You go find some untreated units.

Matching: The big idea



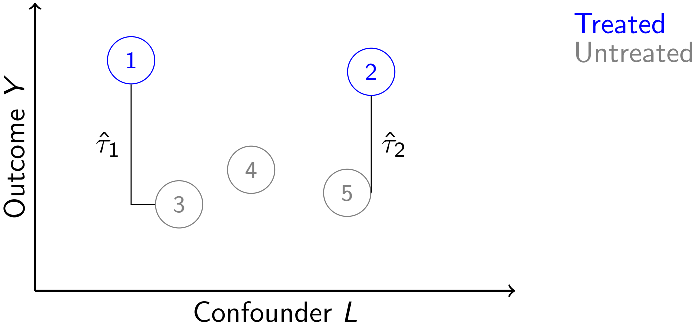
You find the closest matches along L.
You estimate each effect.

Matching: The big idea



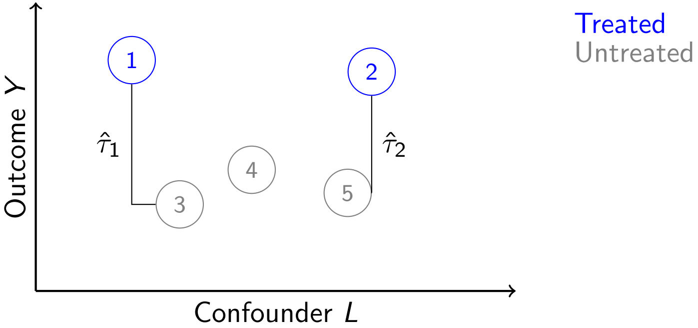
You find the closest matches along L.
You estimate each effect.

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 warning[1]

[1] Sekhon, 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 warning[1]

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

Matching: A word of warning[1]

Matching works!

$$
\begin{tikzcd}
L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y
\end{tikzcd}
$$

Matching: A word of warning[1]

Matching works!

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} & & & U \arrow[dl] \arrow[dr] & \\ L \arrow[rr] \arrow[rrrr, bend right] & & A \arrow[rr] & & Y \end{tikzcd}$$

Matching: A word of warning[1]

Matching works!

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} & & U \arrow[dl] \arrow[dr] & \\ L \arrow[r] \arrow[rrr, bend right] & A \arrow[rr] & & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} U_1 \arrow[d] \arrow[dr] & & \\ L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \\ U_2 \arrow[u] \arrow[urr, bend right] & & \end{tikzcd}$$

[1] Sekhon, 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 warning[1]

Matching works!

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} & & U \arrow[dl] \arrow[dr] & \\ L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} U_1 \arrow[d] \arrow[dr] & & \\ L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \\ U_2 \arrow[u] \arrow[urr, bend right] & & \end{tikzcd}$$

Matching is an estimation strategy.

[1] Sekhon, 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 warning[1]

Matching works!

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} & & U \arrow[dl] \arrow[d] \\ L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} U_1 \arrow[d] \arrow[dr] & & \\ L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \\ U_2 \arrow[u] \arrow[urr, bend right] & & \end{tikzcd}$$

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

[1] Sekhon, 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 warning[1]

Matching works!

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} & & U \arrow[dl] \arrow[dr] & \\ L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & & Y \end{tikzcd}$$

No help!

$$\begin{tikzcd} U_1 \arrow[d] \arrow[dr] & & \\ L \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y \\ U_2 \arrow[u] \arrow[urr, bend right] & & \end{tikzcd}$$

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

[1] Sekhon, 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

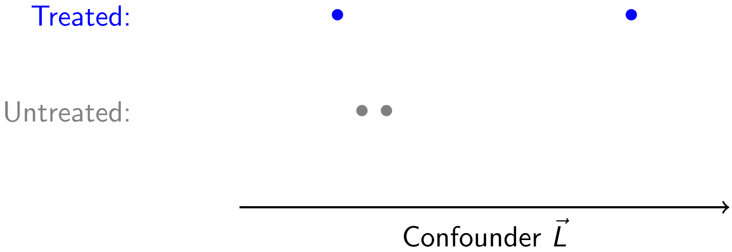
Matching overview

- Matching in univariate settings: Algorithms
- Matching in multivariate settings: Distance metrics
- After matching: Evaluate matched sets

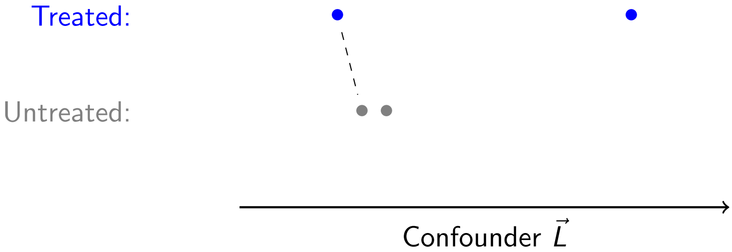
Matching in univariate settings: Algorithms

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

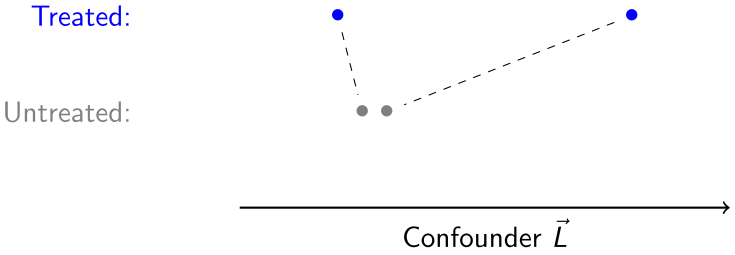
Caliper or no caliper matching



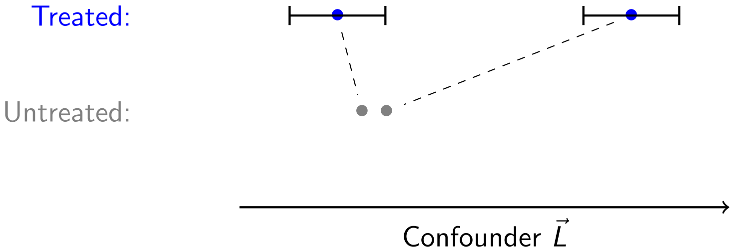
Caliper or no caliper matching



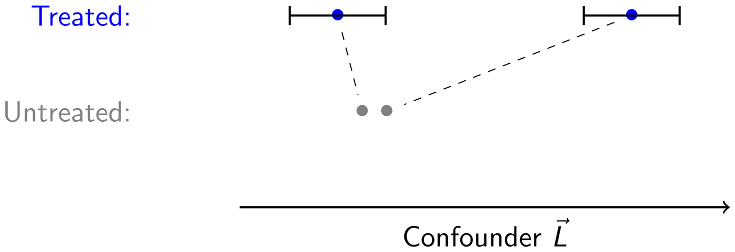
Caliper or no caliper matching



Caliper or no caliper matching

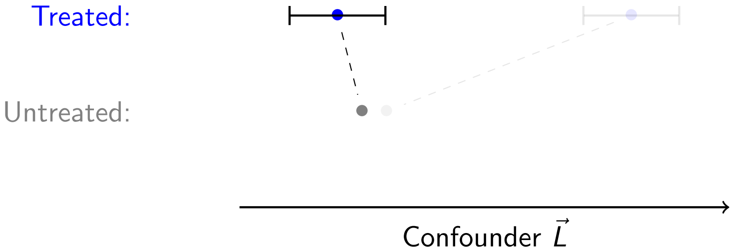


Caliper or no caliper matching



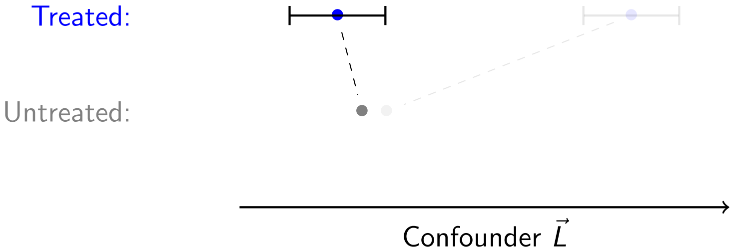
- Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius

Caliper or no caliper matching



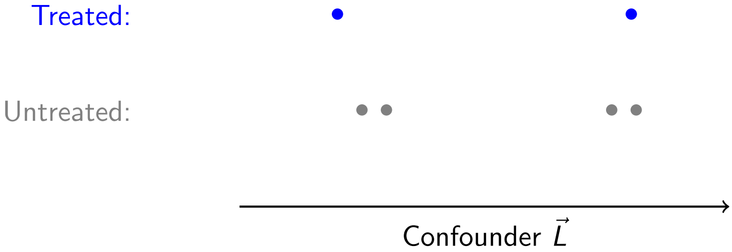
- Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius

Caliper or no caliper matching

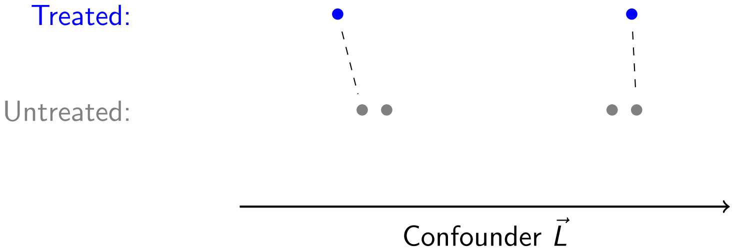


- Caliper: A radius around a treated unit such that we would rather drop the unit than make a match beyond that radius
- Feasible Sample Average Treatment Effect on the Treated (FSATT): Average among treated units for whom an acceptable match exists

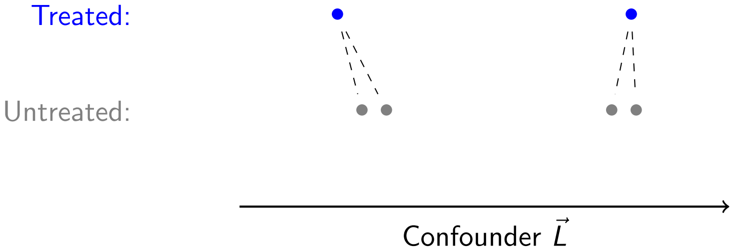
1:1 vs k:1 matching



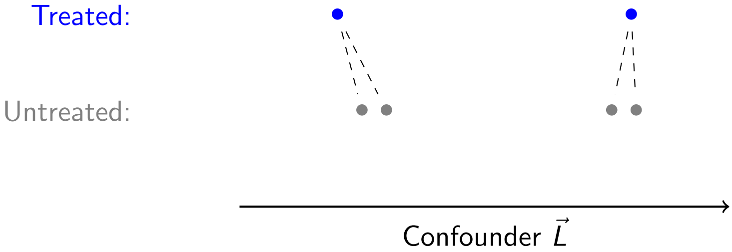
1:1 vs k:1 matching



1:1 vs k:1 matching

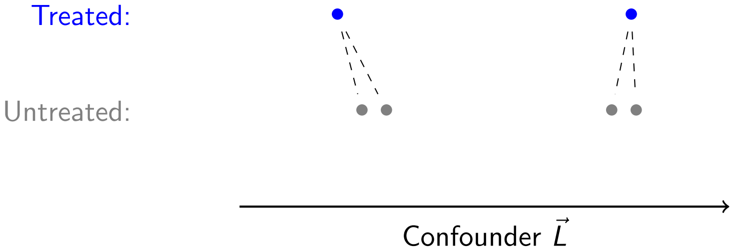


1:1 vs k:1 matching



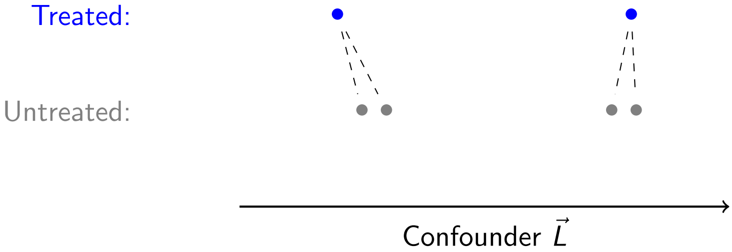
- Benefit of 2:1 matching
- Benefit of 1:1 matching

1:1 vs k:1 matching



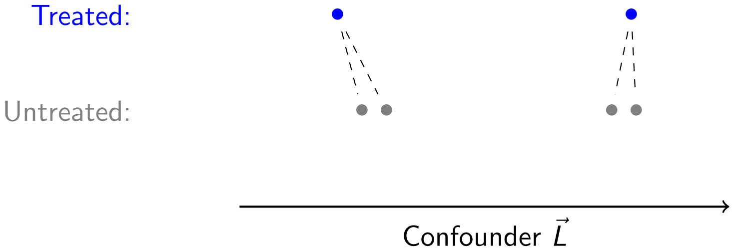
- Benefit of 2:1 matching
  - Lower variance. Averaging over more cases.
- Benefit of 1:1 matching

1:1 vs k:1 matching



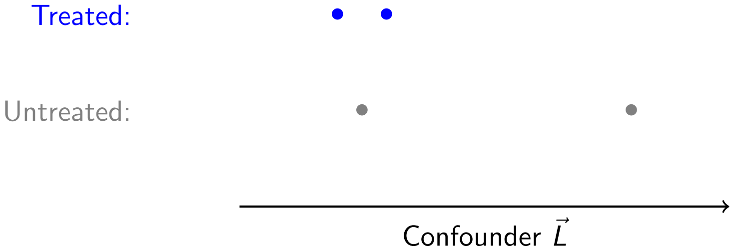
- Benefit of 2:1 matching
  - Lower variance. Averaging over more cases.
- Benefit of 1:1 matching
  - Lower bias. Only the best matches.

1:1 vs k:1 matching

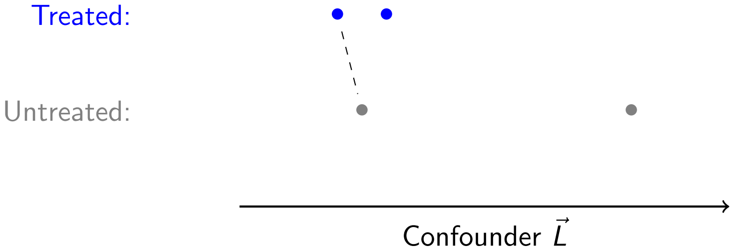


- Benefit of 2:1 matching
  - Lower variance. Averaging over more cases.
- Benefit of 1:1 matching
  - Lower bias. Only the best matches.
- Greater $k \rightarrow$ lower variance, higher bias

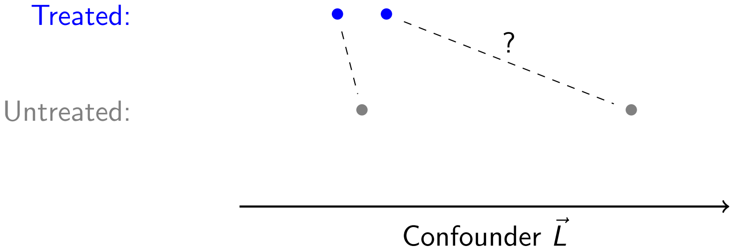
With replacement vs without replacement matching



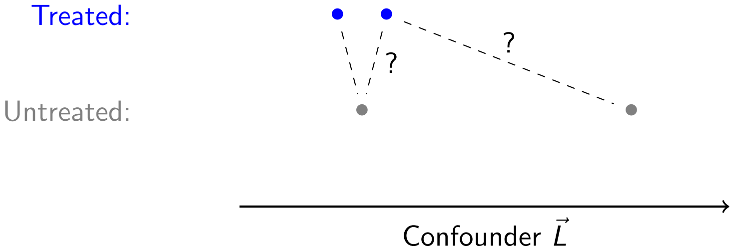
With replacement vs without replacement matching



With replacement vs without replacement matching

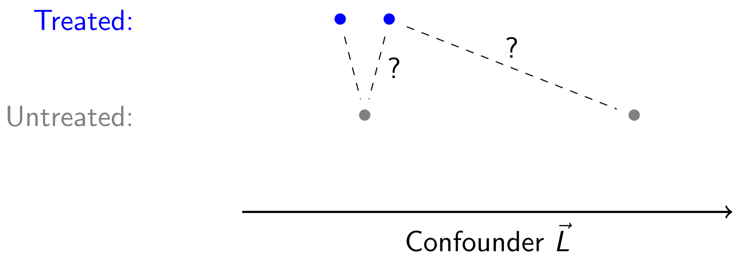


With replacement vs without replacement matching



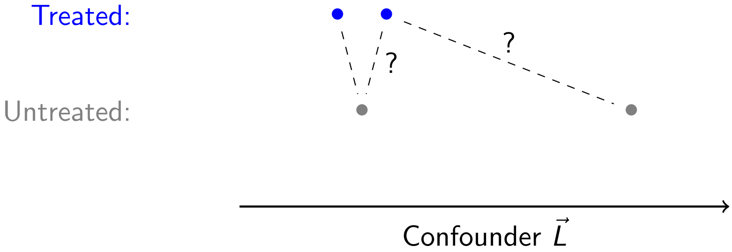
- Benefit of matching without replacement
- Benefit of matching with replacement

With replacement vs without replacement matching



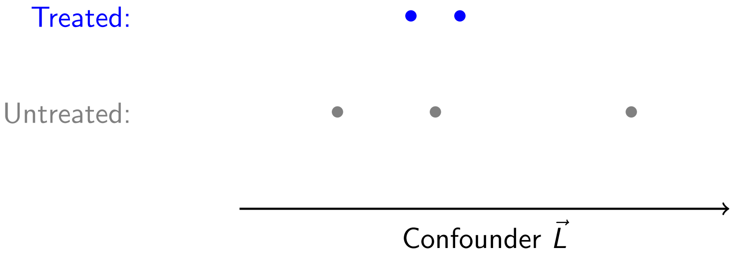
- Benefit of matching without replacement
  - Lower variance. Averaging over more cases.
- Benefit of matching with replacement

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.

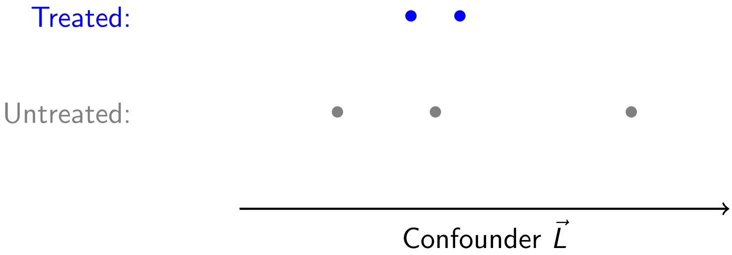
Greedy vs optimal matching[2]



[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.

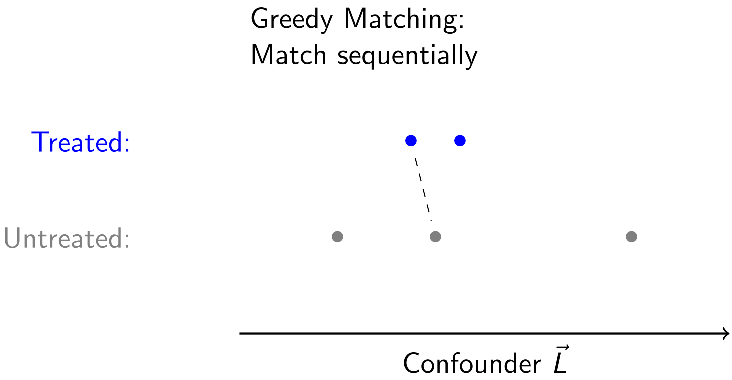
Greedy vs optimal matching[2]

Greedy Matching:
Match sequentially



[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.

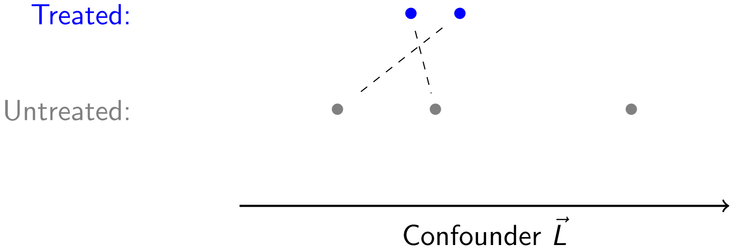
Greedy vs optimal matching[2]



[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.

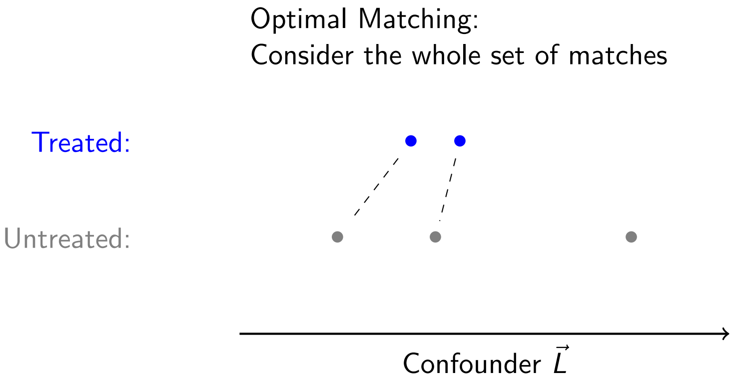
Greedy vs optimal matching[2]

Greedy Matching:
Match sequentially



[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.

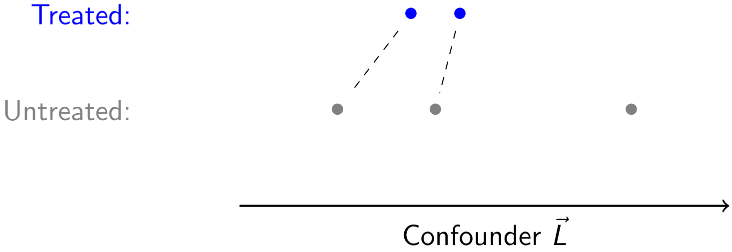
Greedy vs optimal matching[2]



[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



- 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 (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

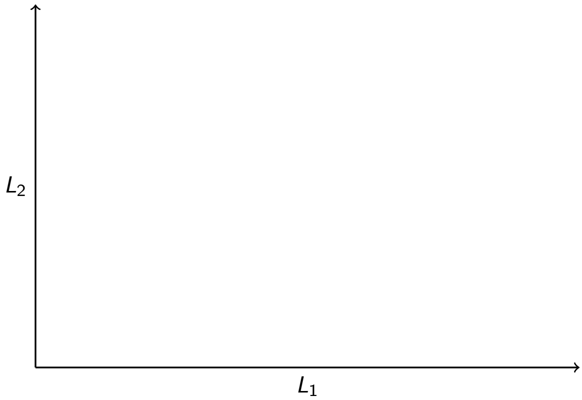
Matching in univariate settings: Algorithms

Matching in multivariate settings: Distance metrics

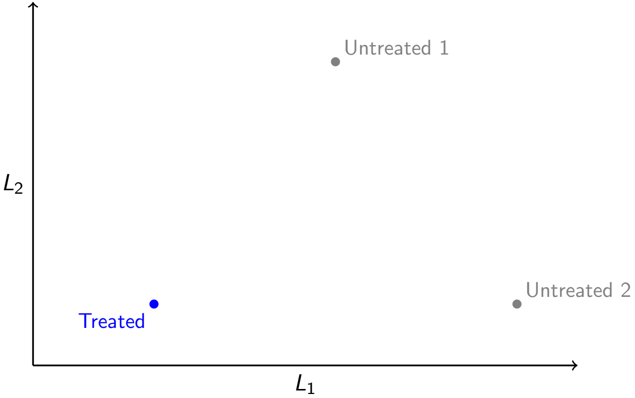
After matching: Evaluate matched sets

What if $\vec{L}$ is multivariate?

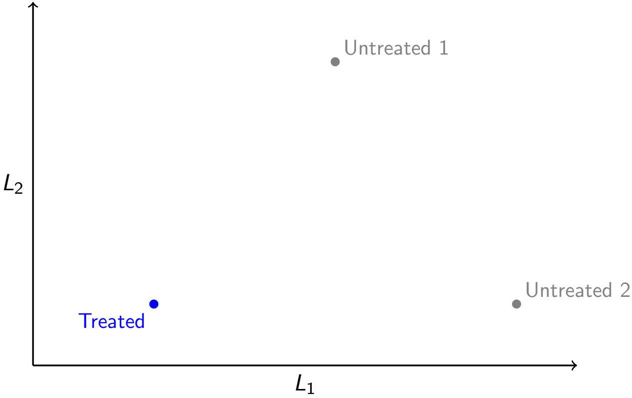
What if $\vec{L}$ is multivariate?



What if $\vec{L}$ is multivariate?

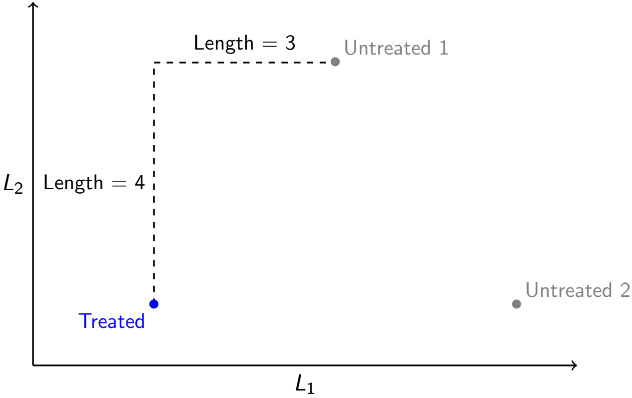


What if $\vec{L}$ is multivariate?



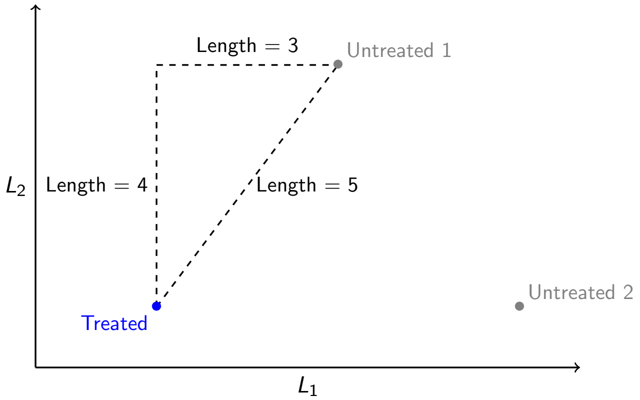
Which untreated unit should be the match?

What if $\vec{L}$ is multivariate?



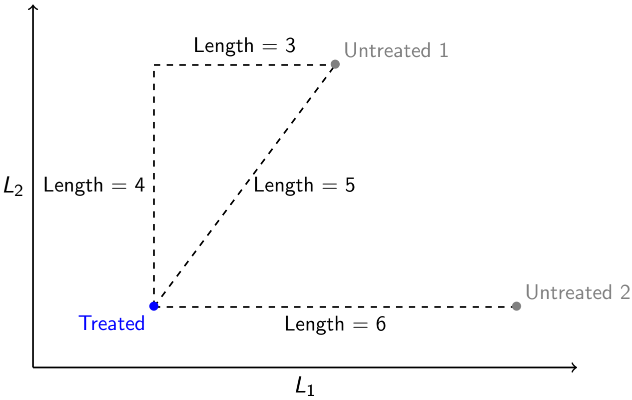
Which untreated unit should be the match?

What if $\vec{L}$ is multivariate?



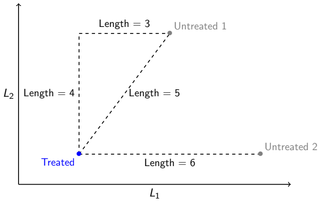
Which untreated unit should be the match?

What if $\vec{L}$ is multivariate?

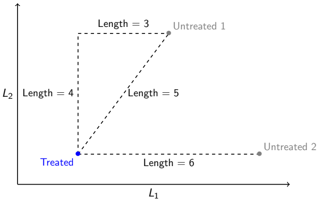


Which untreated unit should be the match?

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

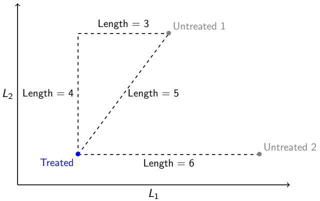


What if $\vec{L}$ is multivariate? We need a **distance metric**



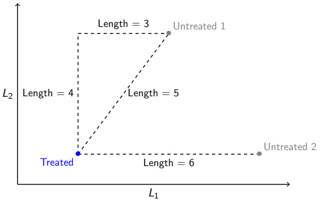
- Manhattan distance:
- Euclidean distance:

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



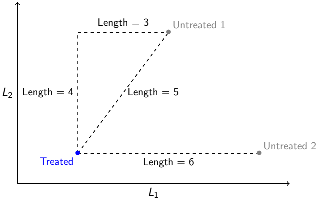
- Manhattan distance: $d(i,j) = \sum_p |L_{pi} - L_{pj}|$
- Euclidean distance:

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



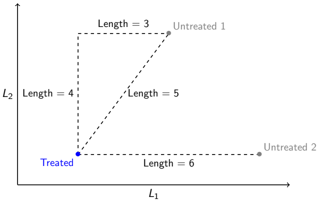
- Manhattan distance: $d(i,j) = \sum_p |L_{pi} - L_{pj}|$
  - $d(\text{Treated, Untreated 1}) = 3 + 4 = 7$
  - $d(\text{Treated, Untreated 2}) = 6 + 0 = 6$ ✓
- Euclidean distance:

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



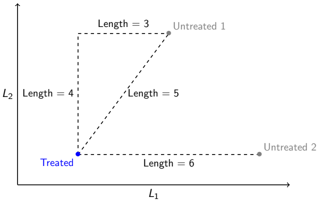
- Manhattan distance: $d(i,j) = \sum_p |L_{pi} - L_{pj}|$
  - $d(\text{Treated, Untreated 1}) = 3 + 4 = 7$
  - $d(\text{Treated, Untreated 2}) = 6 + 0 = 6$ ✓
- Euclidean distance: $d(i,j) = \sqrt{\sum_p (L_{pi} - L_{pj})^2}$

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



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

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



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

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

where $\Sigma = \text{V}(\vec{L})$, 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 matching[3]

[3] Iacus, 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 matching[3]

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

[3] Iacus, 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 matching[3]

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

[3] Iacus, 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 matching[3]

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

[3] Iacus, 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 matching[3]

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

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

[3] Iacus, 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 matching[3]

- 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

[3] Iacus, 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 matching[3]

- 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 $\tilde{\vec{L}}$

[3] Iacus, 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 matching[3]

- 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

[3] Iacus, 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 matching[3]

- 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

[3] Iacus, 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 matching[3]

- 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

[3] Iacus, 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 matching[3]

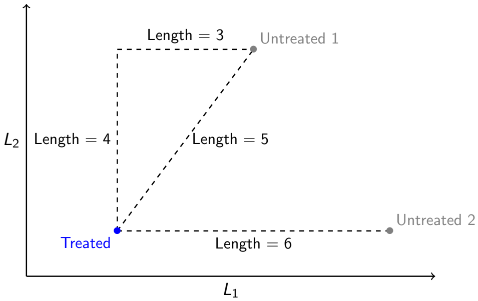
- Define $\tilde{\vec{L}}$ to be a coarsened version of $\vec{L}$
  - Example: Convert all continuous values to decile categories
- Match exactly on $\tilde{\vec{L}}$

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

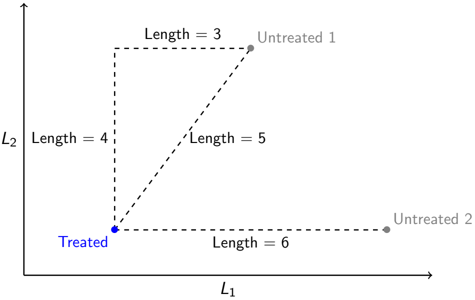
- Benefit: Very transparent
- Drawback: Ignores $\vec{L}$ within $\tilde{\vec{L}}$
- Best used as preprocessing for regression
  - Conduct coarsened exact matching
  - Conduct regression on the matched dataset
  - See paper

[3] Iacus, 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



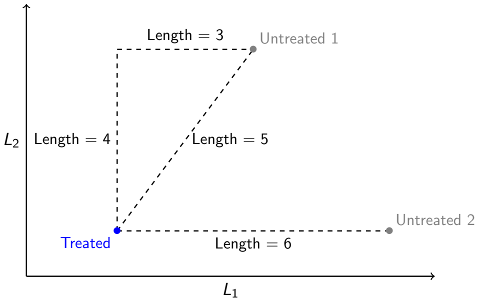
A common distance metric: Propensity scores



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

$$\begin{tikzcd}
 & & L_1 \arrow[d] \\
L_2 \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y
\end{tikzcd}$$

A common distance metric: Propensity scores

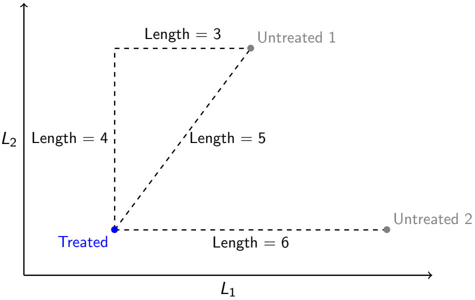


Now suppose only L_2 is related to treatment. L_1 doesn't matter.

$$
\begin{tikzcd}
 & & L_1 \arrow[d] \\
L_2 \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y
\end{tikzcd}
$$

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.

$$\begin{tikzcd}
 & & L_1 \arrow[d] \\
L_2 \arrow[r] \arrow[rr, bend right] & A \arrow[r] & Y
\end{tikzcd}$$

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

A common distance metric: Propensity scores

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

[4] Rosenbaum, 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: $\pi_i = \text{P}(A = 1 \mid \vec{L} = \vec{\ell}_i)$

- Univariate summary of all confounders

[4] Rosenbaum, 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: $\pi_i = \mathrm{P}(A = 1 \mid \vec{L} = \vec{\ell}_i)$

- Univariate summary of all confounders
- In expectation, a sample balanced on π is balanced on $\vec{L}$
  - Rosenbaum & Rubin theorem[4]

[4] Rosenbaum, 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

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

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

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

  - 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

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

  - 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

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 distanace
- 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
2. Reason about choosing regression vs matching