Matching Methods for Observational and Experimental Causal Inference

Gary King[1]

Institute for Quantitative Social Science
Harvard University

Facultad Latinoamericana de Ciencias Sociales, 5/19/2023

[1]GaryKing.org

Readings

Readings

- Matching, Causal Inference: bit.ly/causeI

Readings

- Matching, Causal Inference: bit.ly/causeI
- Coarsened Exact Matching (CEM) software, articles: GaryKing.org/cem

Readings

- Matching, Causal Inference: bit.ly/causeI
- Coarsened Exact Matching (CEM) software, articles: GaryKing.org/cem
- Matching in Experiments, including Seguro Popular: bit.ly/ExpMex

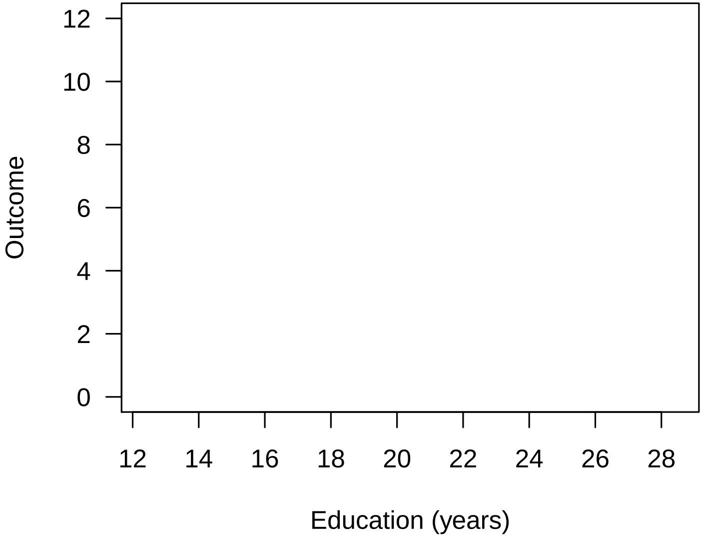
Matching to Reduce Model Dependence

Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

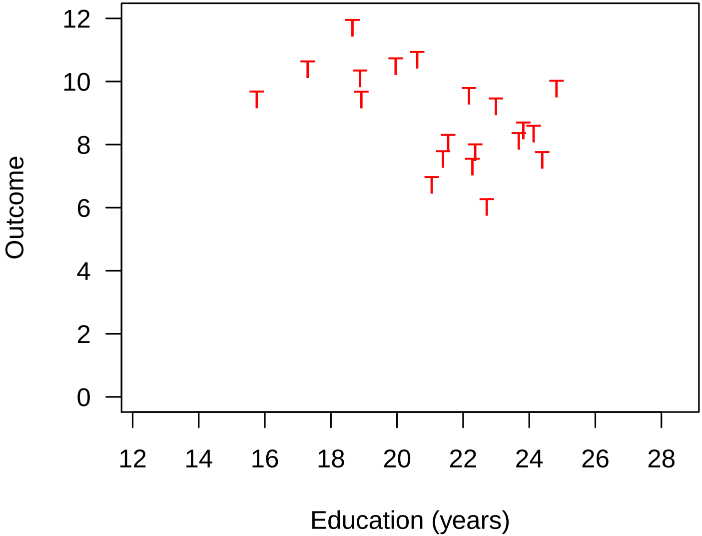
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



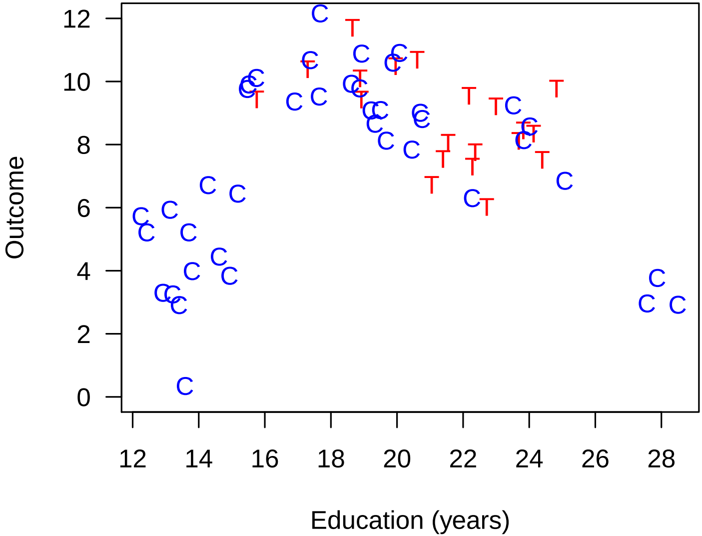
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



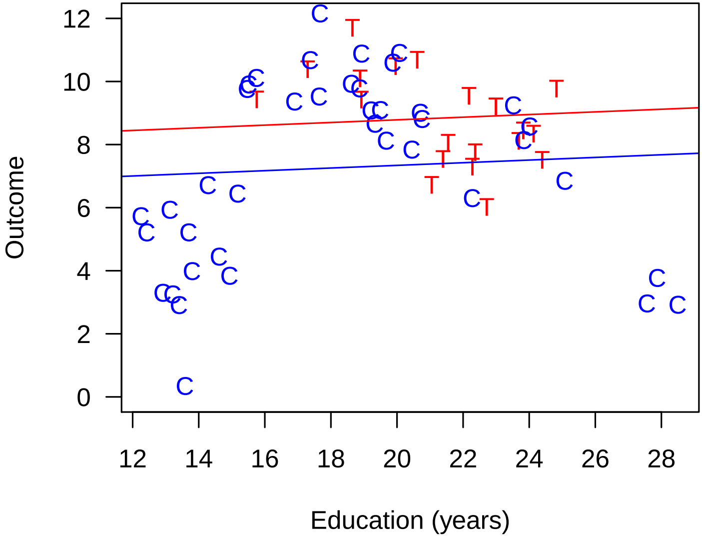
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



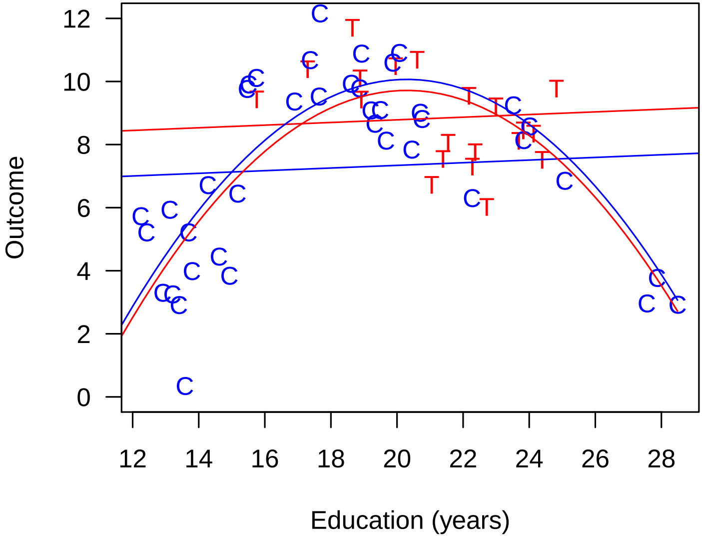
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



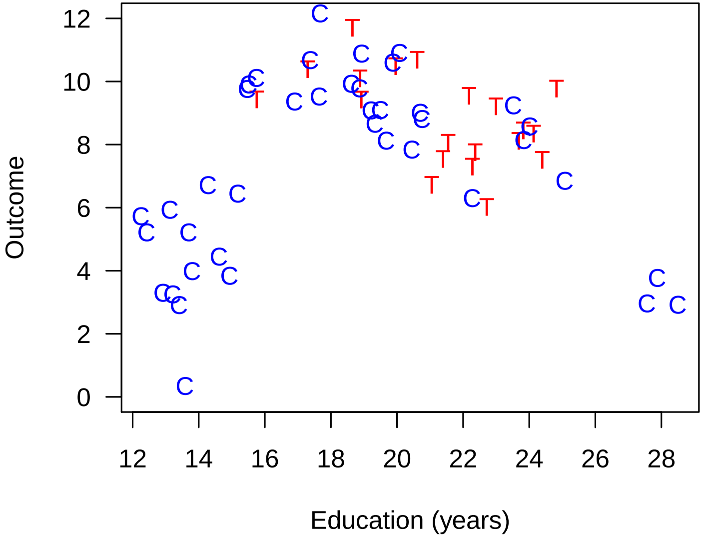
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



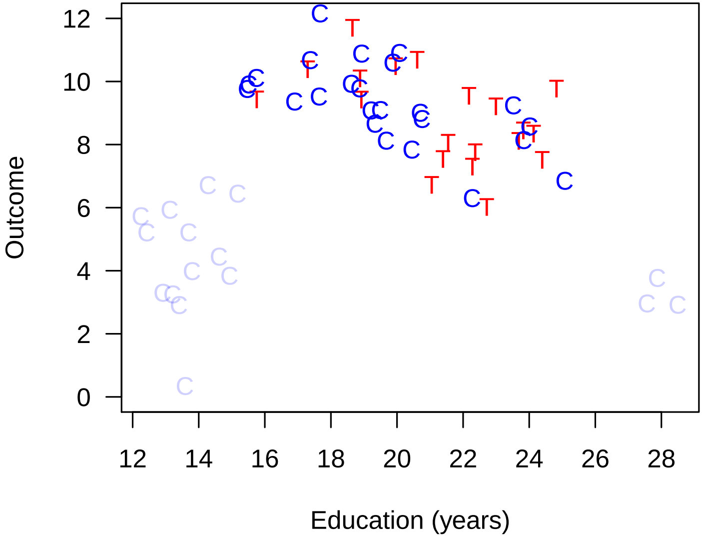
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



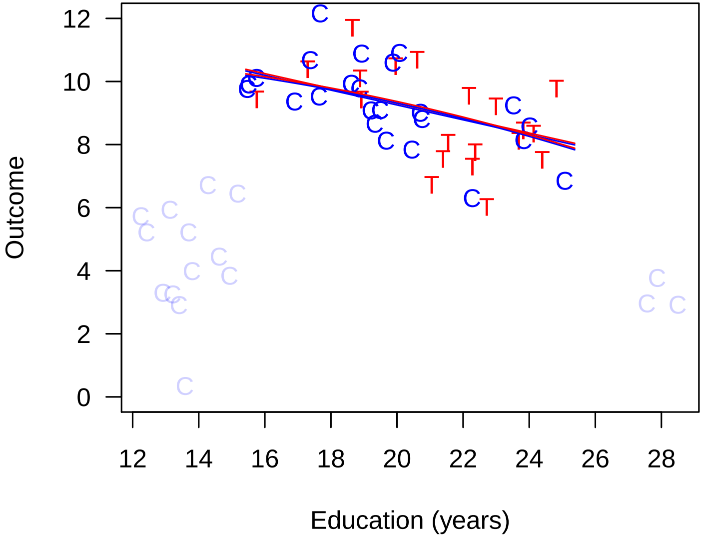
Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching to Reduce Model Dependence

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



The Problems Matching Solves

The Problems Matching Solves

Without Matching:

The Problems Matching Solves

Without Matching:

Imbalance

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates $=$ biased estimator

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates $=$ biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates $=$ biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on "plausibility" is probably worse[effrt]

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on "plausibility" is probably worse[effrt]
- conscientious effort doesn't avoid biases (Banaji 2013)[acc]

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates $=$ biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on "plausibility" is probably worse[effrt]
- conscientious effort doesn't avoid biases (Banaji 2013)[acc]
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)[exprt]

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on "plausibility" is probably worse[effrt]
- conscientious effort doesn't avoid biases (Banaji 2013)[acc]
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)[exprt]
- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)[tch]

The Problems Matching Solves

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

- Qualitative choice from unbiased estimates = biased estimator
  - e.g., Choosing from *results* of 50 randomized experiments
  - Choosing based on "plausibility" is probably worse[effrt]
- conscientious effort doesn't avoid biases (Banaji 2013)[acc]
- People do not have easy access to their own mental processes or feedback to avoid the problem (Wilson and Brekke 1994)[exprt]
- Experts overestimate their ability to control personal biases more than nonexperts, and more prominent experts are the most overconfident (Tetlock 2005)[tch]
- "Teaching psychology is mostly a waste of time" (Kahneman 2011)

The Problems Matching Solves

With~~out~~ Matching:

Imbalance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

The Problems Matching Solves

With~~out~~ Matching:

~~Im~~balance $\rightsquigarrow$ Model Dependence $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

The Problems Matching Solves

With~~out~~ Matching:

~~Im~~balance $\rightsquigarrow$ ~~Model Dependence~~ $\rightsquigarrow$ Researcher discretion $\rightsquigarrow$ Bias

The Problems Matching Solves

With~~out~~ Matching:

~~Im~~balance $\rightsquigarrow$ ~~Model Dependence~~ $\rightsquigarrow$ ~~Researcher discretion~~ $\rightsquigarrow$ Bias

The Problems Matching Solves

With~~out~~ Matching:

~~Im~~balance $\rightsquigarrow$ ~~Model Dependence~~ $\rightsquigarrow$ ~~Researcher discretion~~ $\rightsquigarrow$ ~~Bias~~

The Problems Matching Solves

With~~out~~ Matching:

~~Im~~balance $\rightsquigarrow$ ~~Model Dependence~~ $\rightsquigarrow$ ~~Researcher discretion~~ $\rightsquigarrow$ ~~Bias~~

A central project of statistics: Automating away human discretion

What's Matching?

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for treated observation i:

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for <u>treated</u> observation i:

$$\text{TE}_i = Y_i(1) - Y_i(0)$$

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for <u>treated</u> observation i:

$$\begin{aligned} \mathrm{TE}_i &= Y_i(1) - Y_i(0) \\ &= \text{observed} - \textit{unobserved} \end{aligned}$$

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for treated observation i:

$$\begin{aligned} \mathrm{TE}_i &= Y_i - Y_i(0) \\ &= \text{observed} - \textit{unobserved} \end{aligned}$$

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for <u>treated</u> observation i:

$$\begin{aligned} \text{TE}_i &= Y_i - Y_i(0) \\ &= \text{observed} - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j with a matched ($X_i \approx X_j$) control

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for <u>treated</u> observation i:

$$\begin{aligned} \mathrm{TE}_i &= Y_i - Y_i(0) \\ &= \text{observed} - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j with a matched ($X_i \approx X_j$) control
- Quantities of Interest:

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for <u>treated</u> observation i:

$$\begin{aligned} \text{TE}_i &= Y_i - Y_i(0) \\ &= \text{observed} - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j with a matched ($X_i \approx X_j$) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \underset{i:T_i=1}{\text{Mean}} (\text{TE}_i)$$

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for <u>treated</u> observation i:

$$\begin{aligned} \mathrm{TE}_i &= Y_i - Y_i(0) \\ &= \text{observed} - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j with a matched ($X_i \approx X_j$) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$\mathrm{SATT} = \underset{i:T_i=1}{\mathrm{Mean}} \left(\mathrm{TE}_i\right)$$

  2. FSATT: Feasible SATT (prune badly matched treateds too)

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for treated observation i:

$$\begin{aligned} \text{TE}_i &= Y_i - Y_i(0) \\ &= \text{observed} - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j with a matched $(X_i \approx X_j)$ control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \underset{i:T_i=1}{\text{Mean}}(\text{TE}_i)$$

  2. FSATT: Feasible SATT (prune badly matched treateds too)
- Big convenience: Follow preprocessing with whatever statistical method you'd have used without matching

What's Matching?

- Y_i dep var, T_i (1=treated, 0=control), X_i confounders
- Treatment Effect for treated observation i:

$$\begin{aligned} \text{TE}_i &= Y_i - Y_i(0) \\ &= \text{observed} - \textit{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j with a matched ($X_i \approx X_j$) control
- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \underset{i:T_i=1}{\text{Mean}}(\text{TE}_i)$$

  2. FSATT: Feasible SATT (prune badly matched treateds too)
- Big convenience: Follow preprocessing with whatever statistical method you'd have used without matching
- Pruning nonmatches makes control vars matter less: reduces imbalance, model dependence, researcher discretion, & bias

Matching: Finding Hidden Randomized Experiments

Matching: Finding Hidden Randomized Experiments

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Complete Randomization

Matching: Finding Hidden Randomized Experiments

Types of Experiments

	Complete Randomization	*Fully Blocked*

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed		
Unobserved		

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	
Unobserved		

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	
Unobserved	On average	

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization*

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for:

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for:
imbalance,

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence,

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for:
imbalance, model dependence, power,

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for:
imbalance, model dependence, power, efficiency,

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for:
imbalance, model dependence, power, efficiency, bias,

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs,

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness.

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

⇝ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*
- Other methods: *fully blocked*

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\leadsto$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*
- Other methods: *fully blocked*
- Other matching methods dominate PSM

Matching: Finding Hidden Randomized Experiments

Types of Experiments

Balance Covariates:	*Complete Randomization*	*Fully Blocked*
Observed	On average	Exact
Unobserved	On average	On average

$\rightsquigarrow$ *Fully blocked* dominates *complete randomization* for: imbalance, model dependence, power, efficiency, bias, research costs, robustness. E.g., Imai, King, Nall 2009: SEs 600% smaller!

Goal of Each Matching Method (in Observational Data)

- PSM: *complete randomization*
- Other methods: *fully blocked*
- Other matching methods dominate PSM (wait, it gets worse)

Method 1: Mahalanobis Distance Matching

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)

2. Estimation Difference in means or a model

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$

2. Estimation Difference in means or a model

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)

2. Estimation Difference in means or a model

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit

2. Estimation Difference in means or a model

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
2. Estimation Difference in means or a model

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*

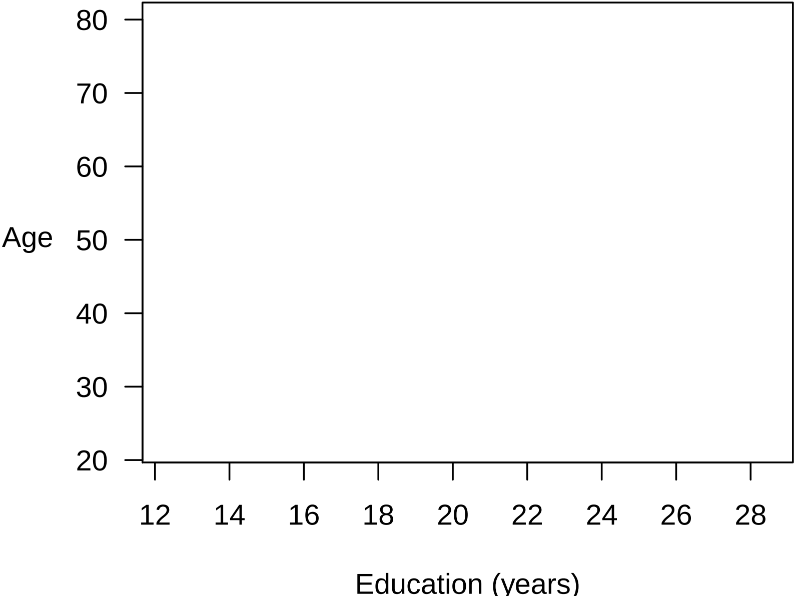
2. Estimation Difference in means or a model

Method 1: Mahalanobis Distance Matching

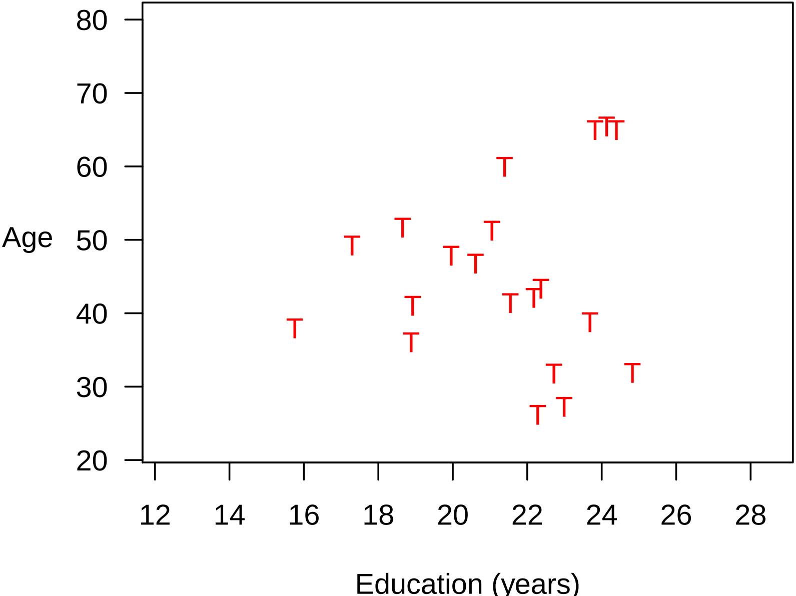
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*
   - (Many adjustments available to this basic method)
2. Estimation Difference in means or a model

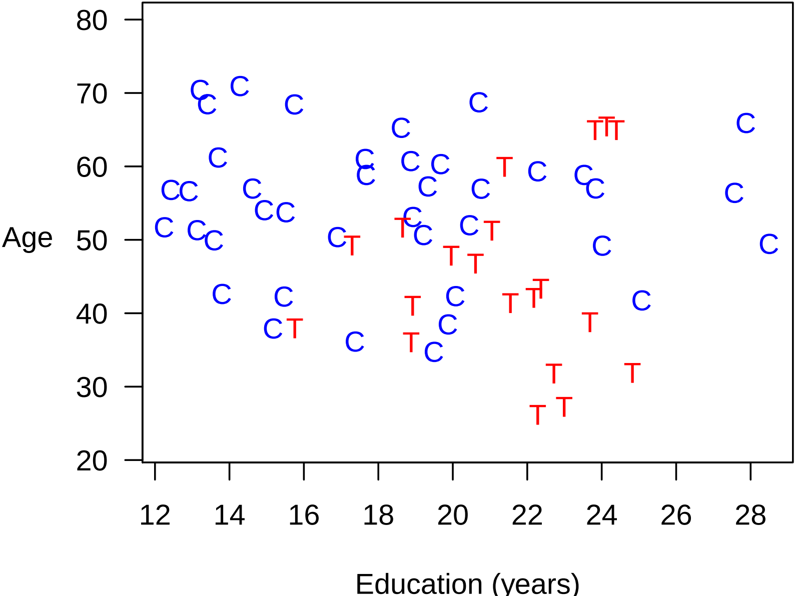
Mahalanobis Distance Matching



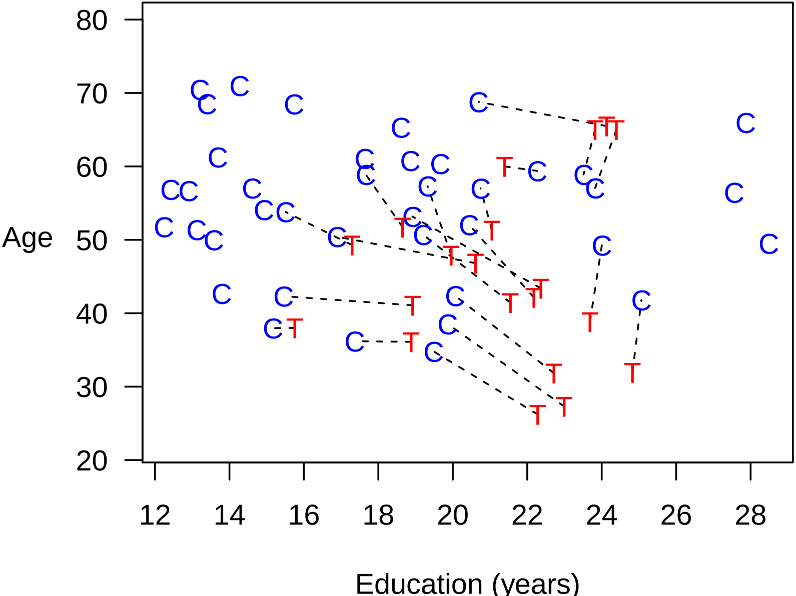
Mahalanobis Distance Matching



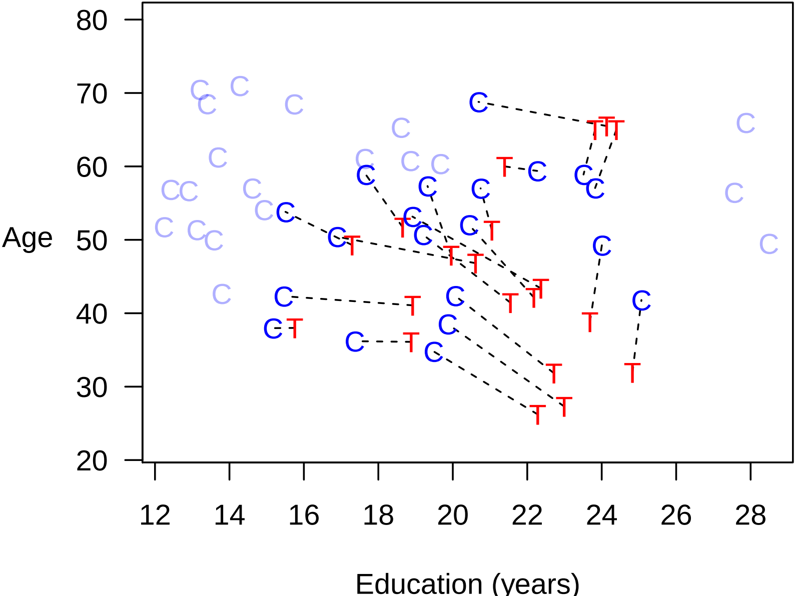
Mahalanobis Distance Matching



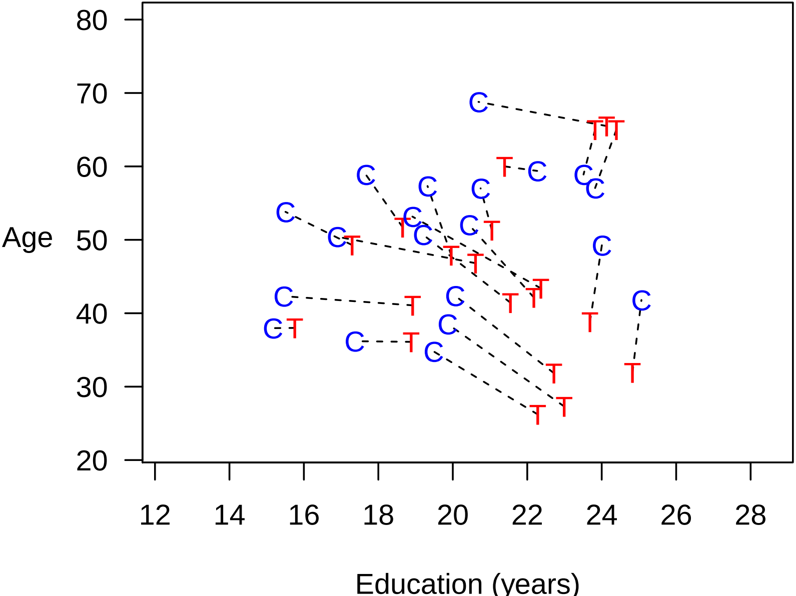
Mahalanobis Distance Matching



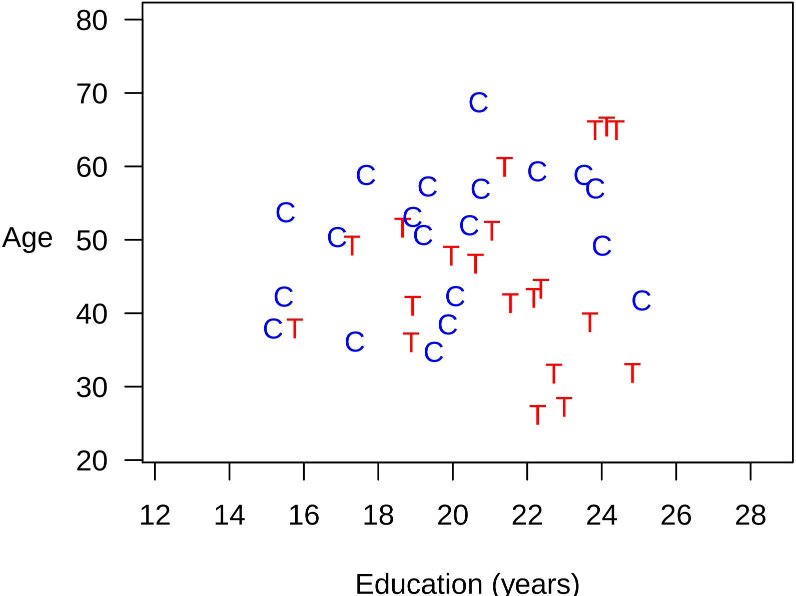
Mahalanobis Distance Matching



Mahalanobis Distance Matching

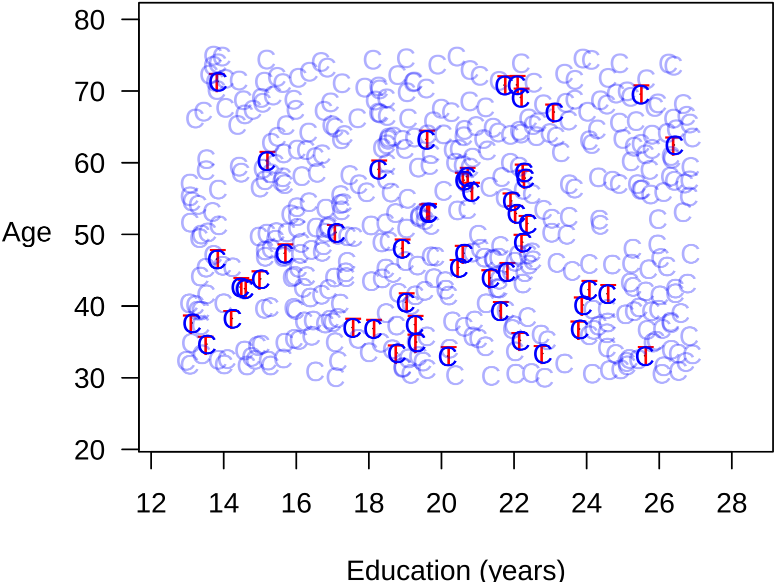


Mahalanobis Distance Matching

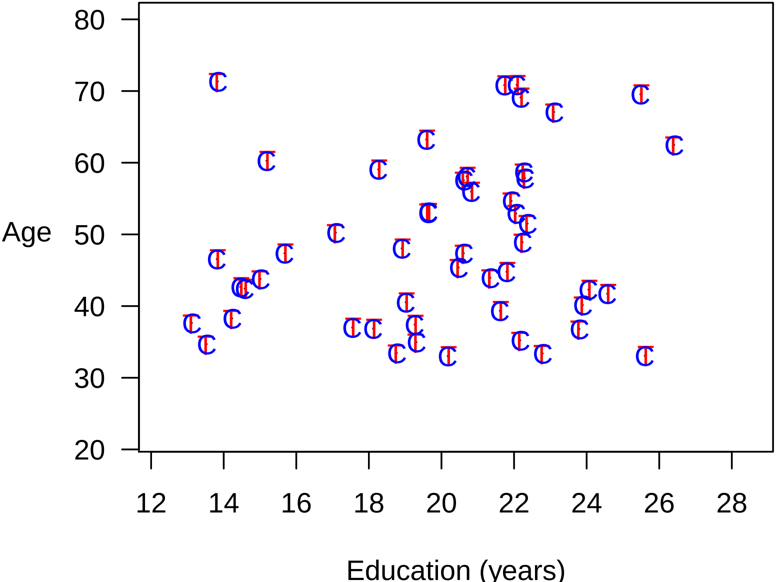


Best Case: Mahalanobis Distance Matching

Best Case: Mahalanobis Distance Matching



Best Case: Mahalanobis Distance Matching



Method 2: Coarsened Exact Matching
(Most powerful easy-to-use approach)

Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)

2. Estimation Difference in means or a model

Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
2. Estimation Difference in means or a model

Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)

2. Estimation Difference in means or a model

Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened X, $C(X)$
2. Estimation Difference in means or a model

Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
2. Estimation Difference in means or a model

Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units

2. Estimation Difference in means or a model

Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model

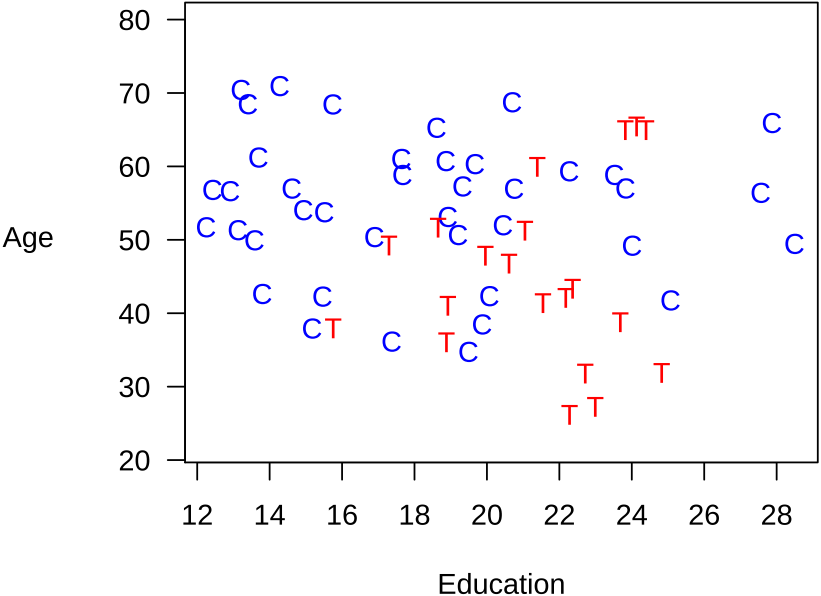
Method 2: Coarsened Exact Matching (Most powerful easy-to-use approach)

(Approximates Fully Blocked Experiment)

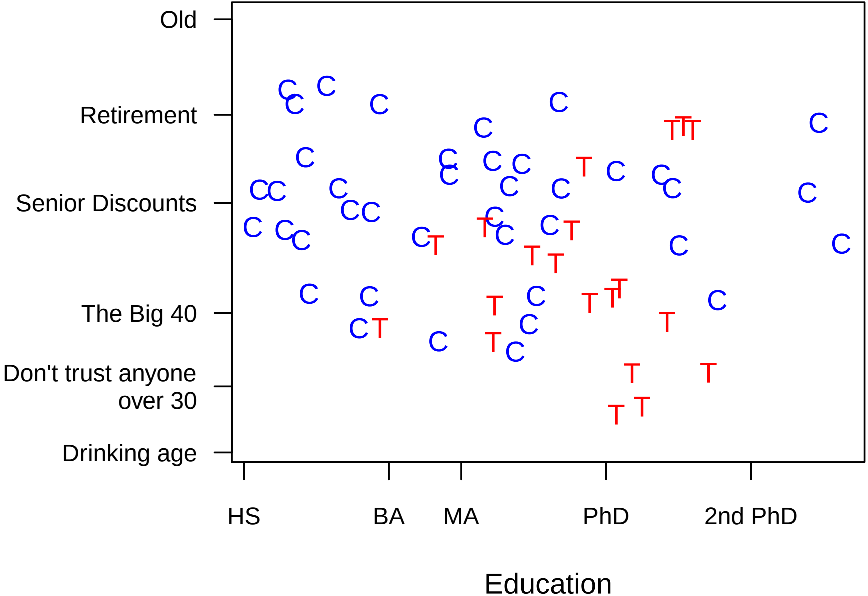
1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with 0 treated or 0 control units
   - Pass on original (uncoarsened) units except those pruned
2. Estimation Difference in means or a model
   - Weight controls in each stratum to equal treateds

Coarsened Exact Matching

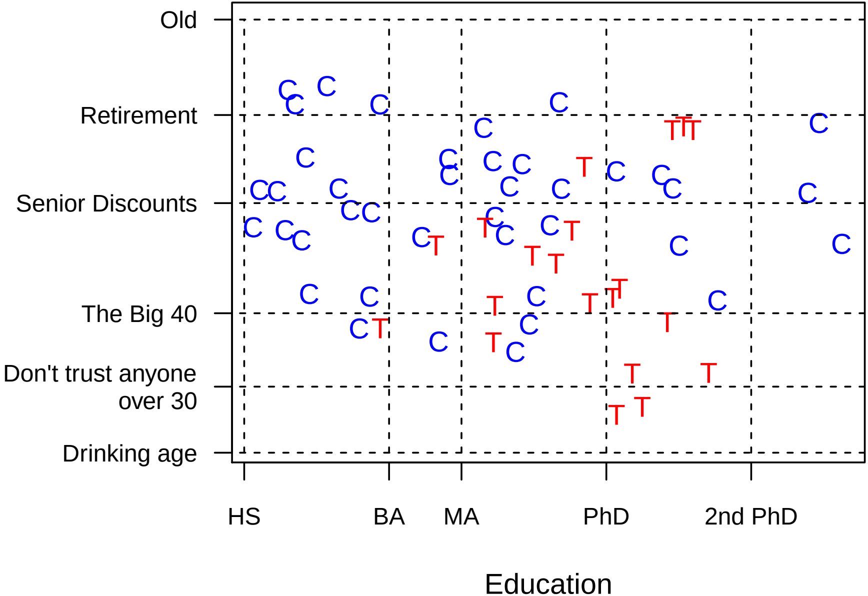
Coarsened Exact Matching



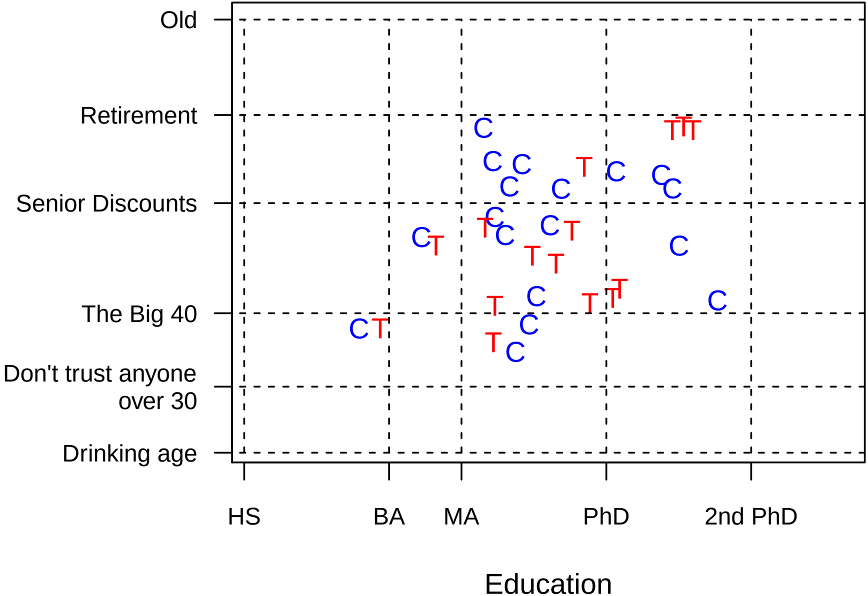
Coarsened Exact Matching



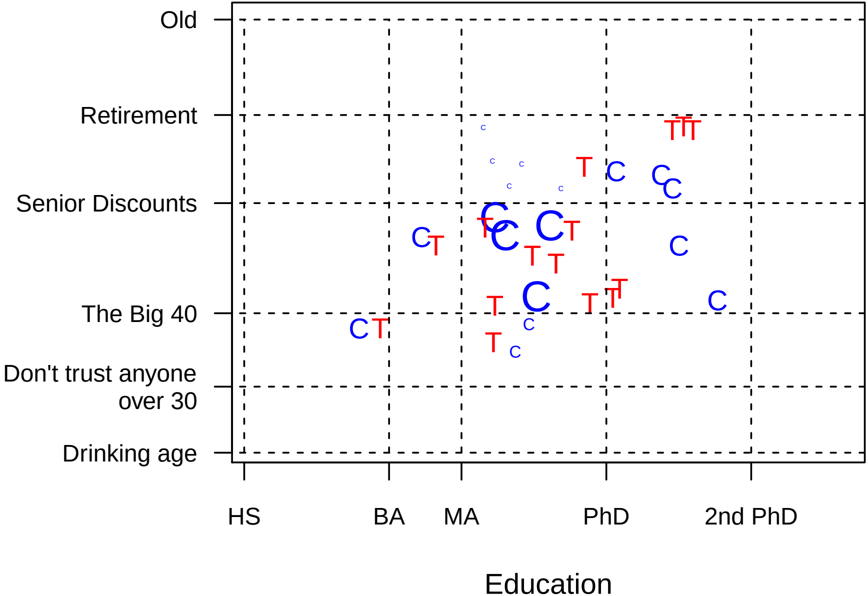
Coarsened Exact Matching



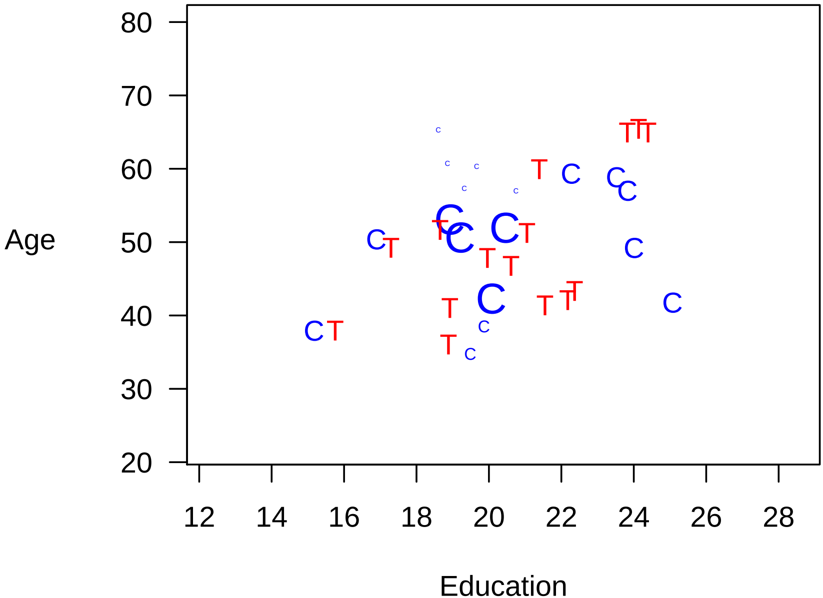
Coarsened Exact Matching



Coarsened Exact Matching

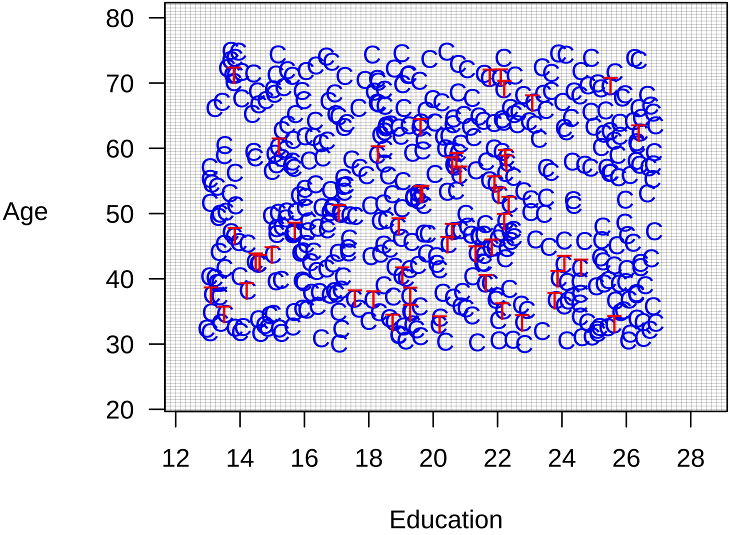


Coarsened Exact Matching

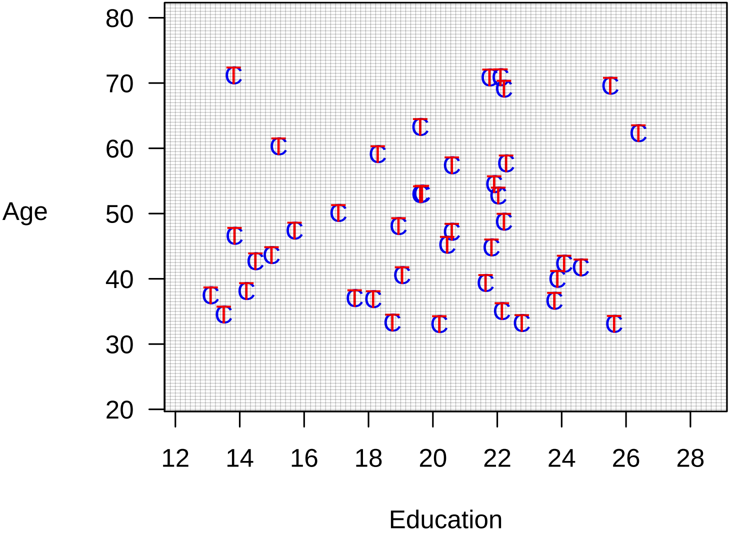


Best Case: Coarsened Exact Matching

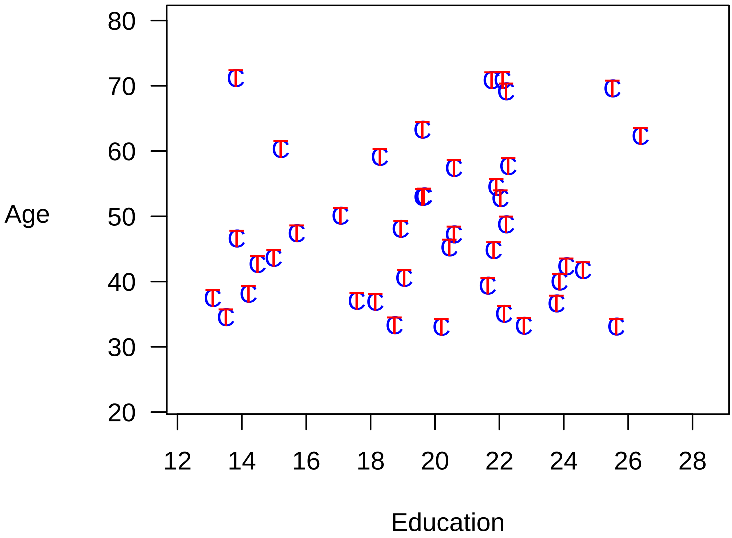
Best Case: Coarsened Exact Matching



Best Case: Coarsened Exact Matching



Best Case: Coarsened Exact Matching



Method 3: Propensity Score Matching

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)

2. Estimation Difference in means or a model

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
2. Estimation Difference in means or a model

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - $\text{Distance}(X_c, X_t) = |\pi_c - \pi_t|$

2. Estimation Difference in means or a model

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
    - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
    - $\text{Distance}(X_c, X_t) = |\pi_c - \pi_t|$
    - Match each treated unit to the nearest control unit
2. Estimation Difference in means or a model

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - $\text{Distance}(X_c, X_t) = |\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused

2. Estimation Difference in means or a model

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

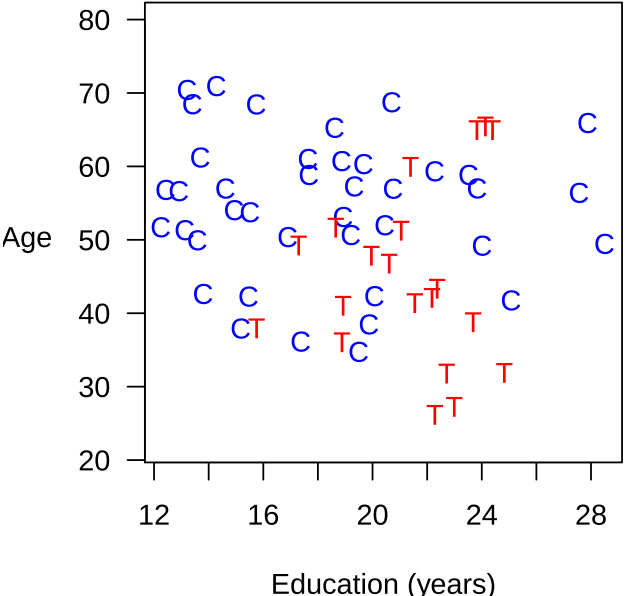
1. Preprocess (Matching)
   - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - $\text{Distance}(X_c, X_t) = |\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*
2. Estimation Difference in means or a model

Method 3: Propensity Score Matching

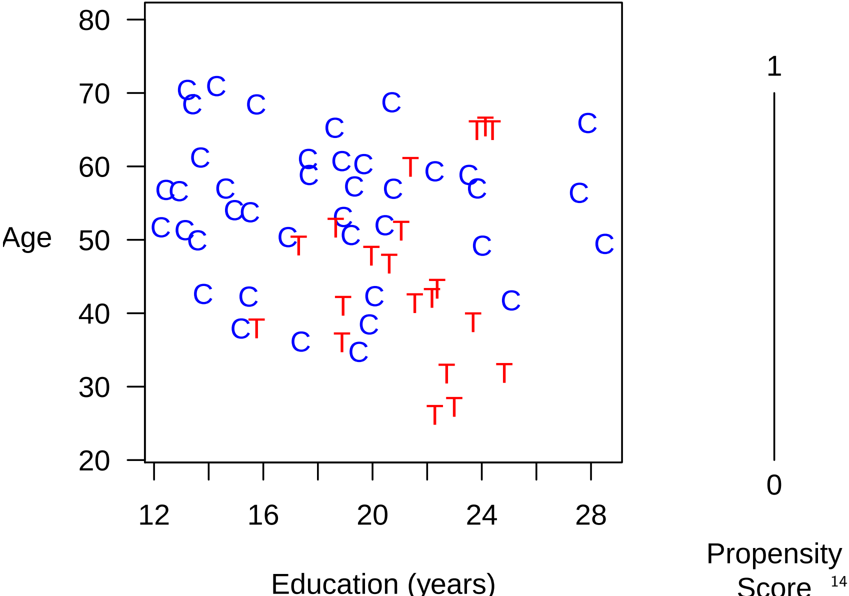
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
    - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
    - $\text{Distance}(X_c, X_t) = |\pi_c - \pi_t|$
    - Match each treated unit to the nearest control unit
    - Control units: not reused; pruned if unused
    - Prune matches if Distance$>$*caliper*
    - (Many adjustments available to this basic method)
2. Estimation Difference in means or a model

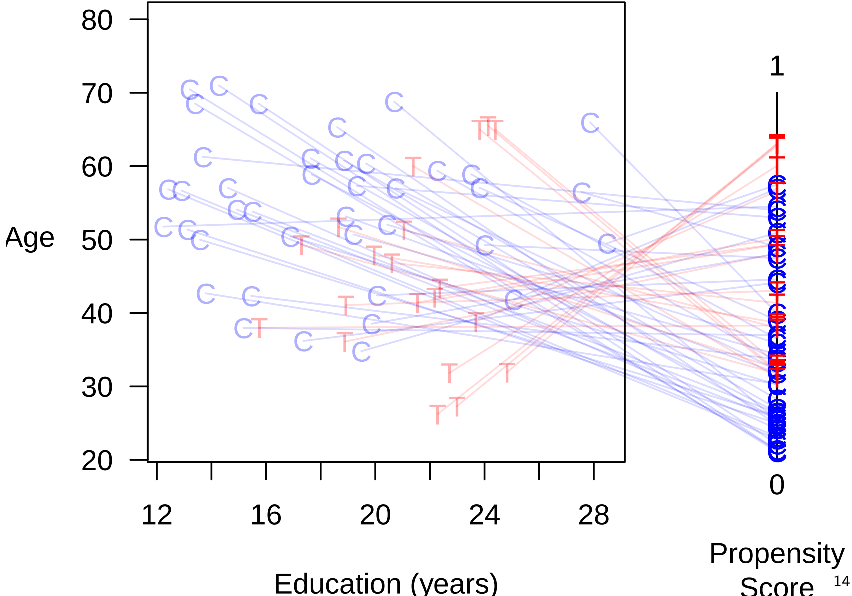
Propensity Score Matching



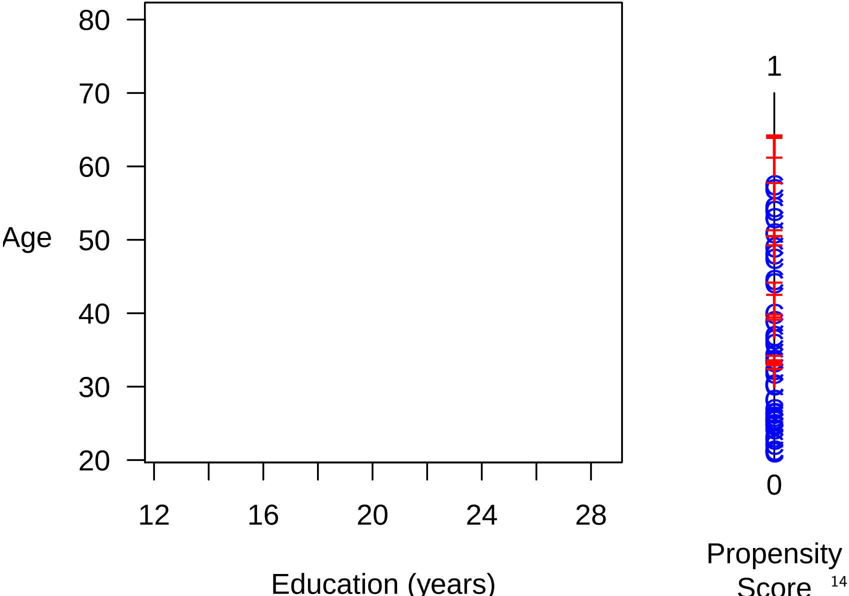
Propensity Score Matching



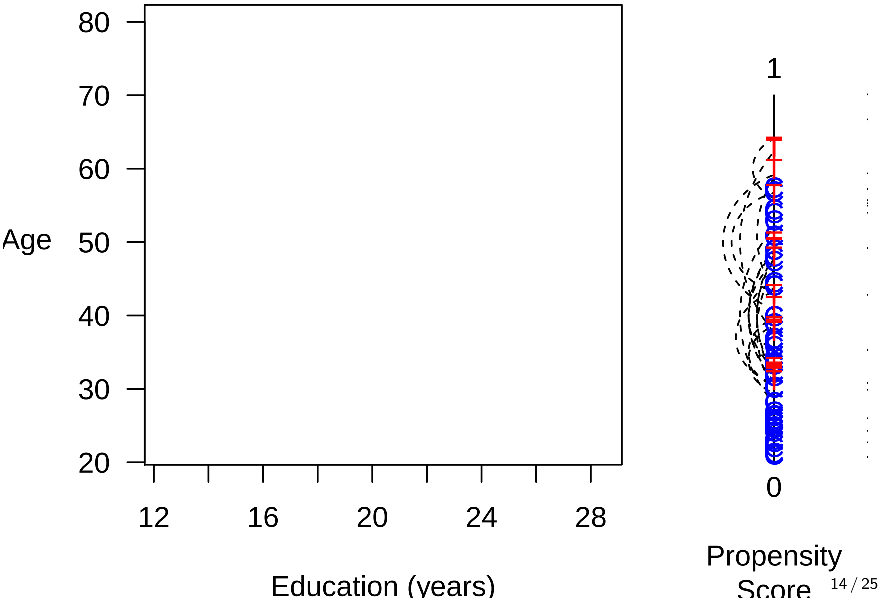
Propensity Score Matching



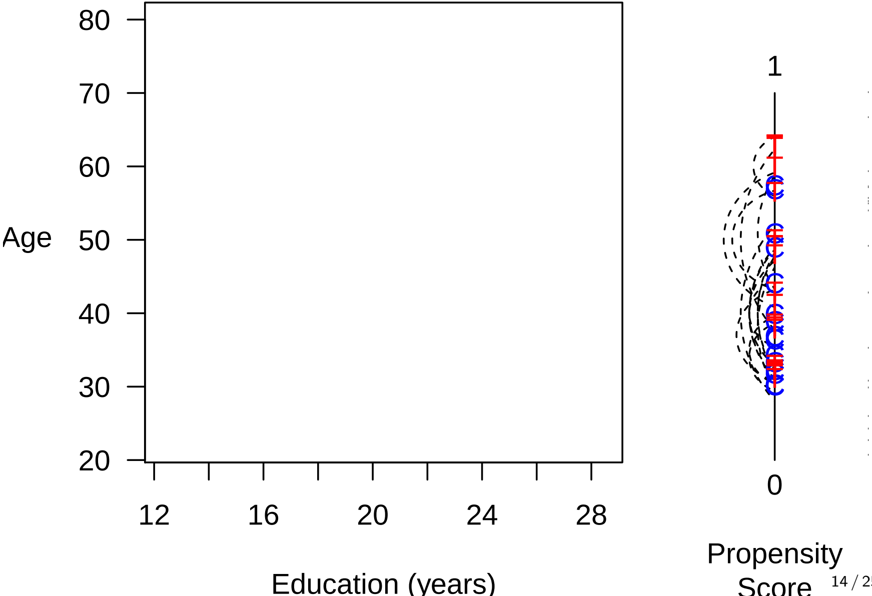
Propensity Score Matching



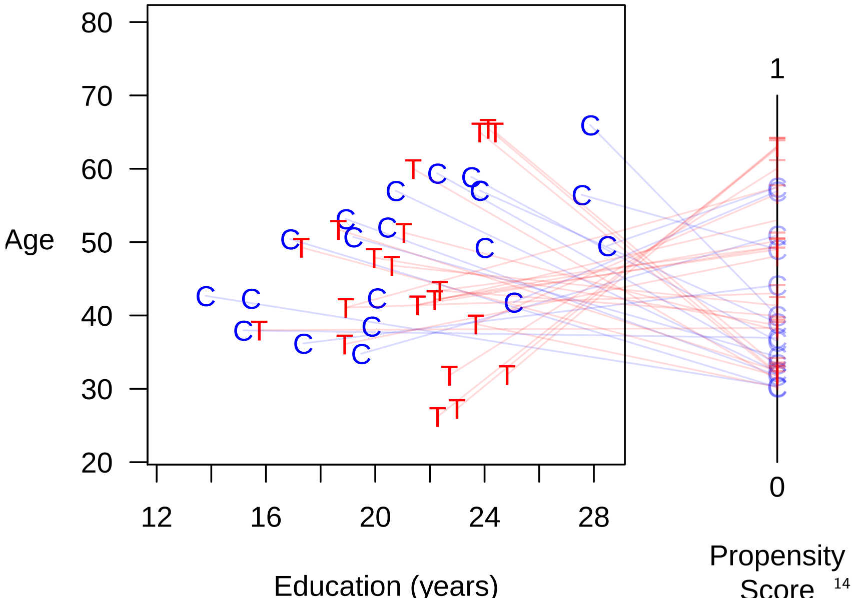
Propensity Score Matching



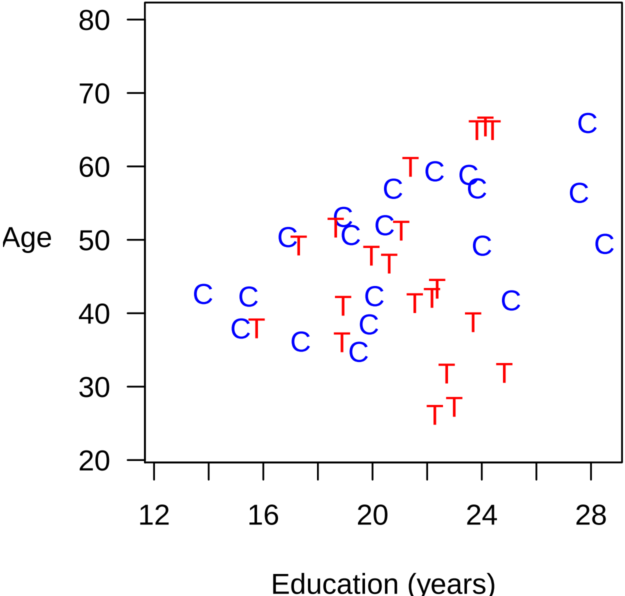
Propensity Score Matching



Propensity Score Matching

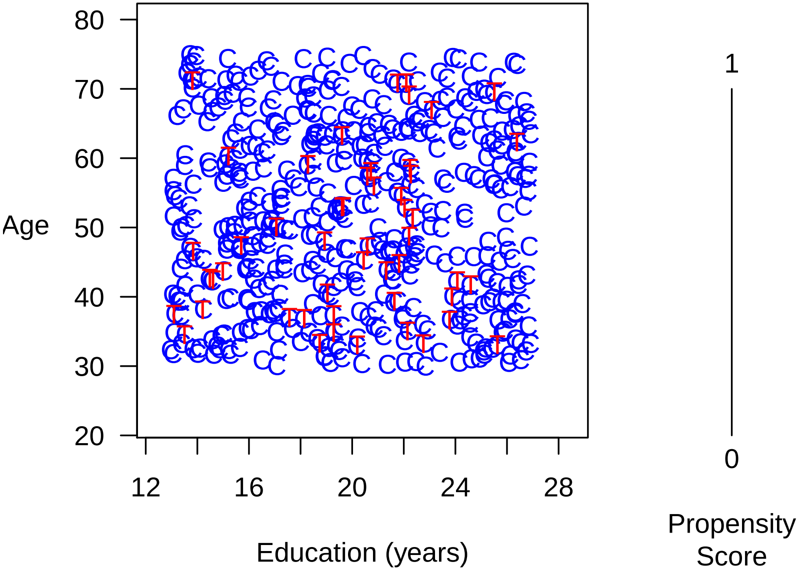


Propensity Score Matching

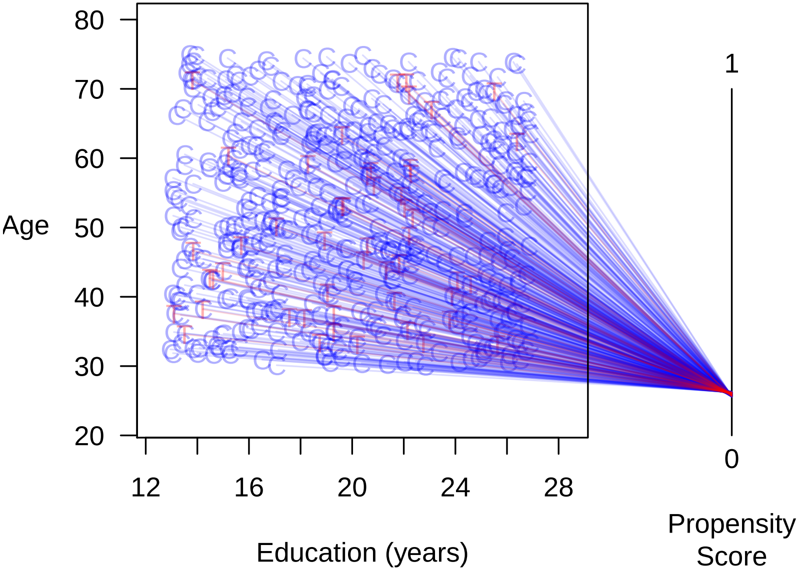


Best Case: Propensity Score Matching

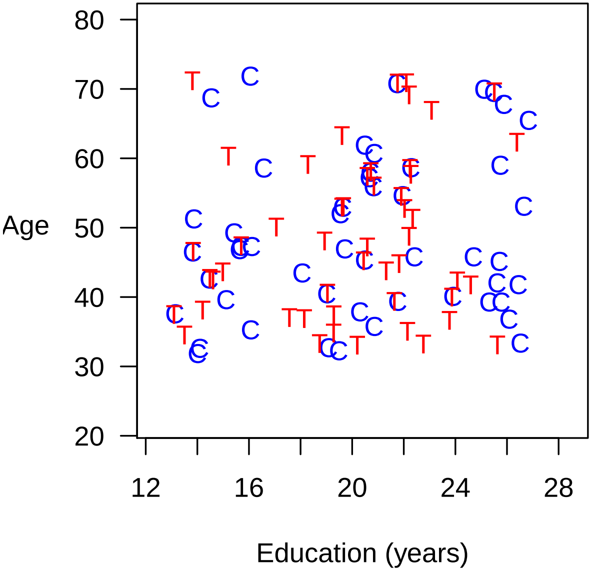
Best Case: Propensity Score Matching



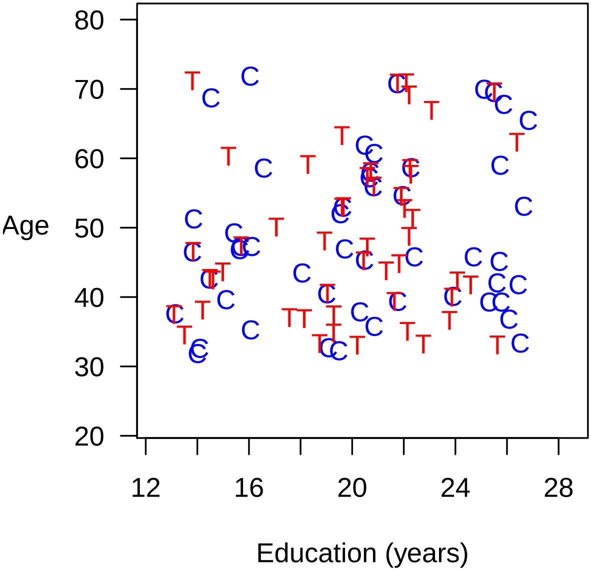
Best Case: Propensity Score Matching



Best Case: Propensity Score Matching



Best Case: Propensity Score Matching is Suboptimal



PSM's Statistical Properties

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking

PSM's Statistical Properties

1. <u>Low Standards:</u> Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
    - *Efficient* relative to complete randomization, but
    - *Inefficient* relative to (the more powerful) full blocking
    - Other methods usually dominate:
      $X_c = X_t \implies \pi_c = \pi_t$ but
      $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning)

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata)

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random $\rightsquigarrow$ Imbalance

PSM's Statistical Properties

1. <u>Low Standards:</u> Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. <u>The PSM Paradox:</u> When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random $\rightsquigarrow$ Imbalance $\rightsquigarrow$ Inefficency

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random $\rightsquigarrow$ Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random $\rightsquigarrow$ Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random $\rightsquigarrow$ Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias
   - If the data have no good matches, the paradox won't be a problem but you're cooked anyway.

PSM's Statistical Properties

1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random $\rightsquigarrow$ Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias
   - If the data have no good matches, the paradox won't be a problem but you're cooked anyway.
   - Doesn't PSM solve the curse of dimensionality problem?

PSM's Statistical Properties

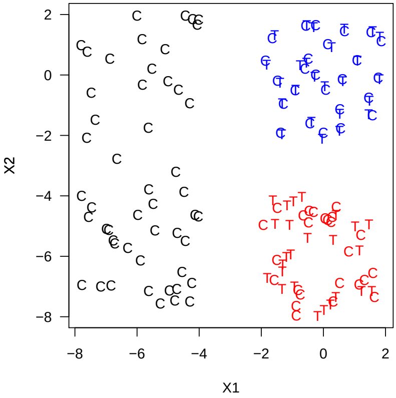
1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\rightsquigarrow$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ pruning at random $\rightsquigarrow$ Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias
   - If the data have no good matches, the paradox won't be a problem but you're cooked anyway.
   - Doesn't PSM solve the curse of dimensionality problem? Nope.

PSM's Statistical Properties

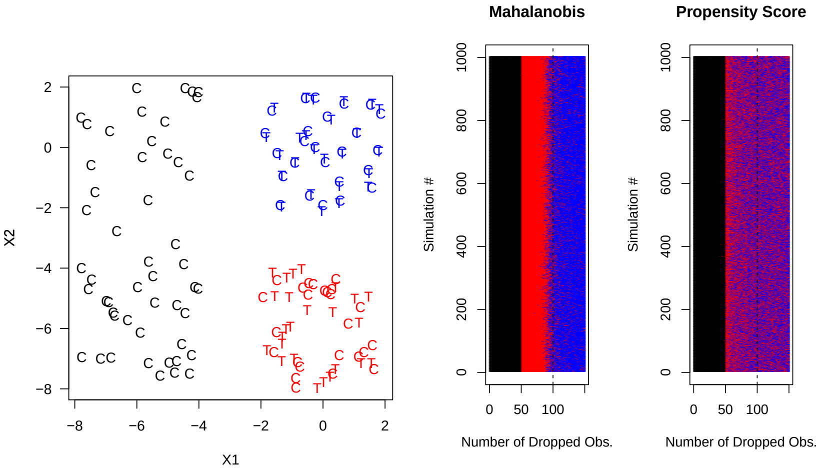
1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods usually dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The PSM Paradox: When you do "better," you do worse
   - Background: Random matching increases imbalance
   - When PSM approximates complete randomization (to begin with or, after some pruning) $\leadsto$ all $\hat{\pi} \approx 0.5$ (or constant within strata) $\leadsto$ pruning at random $\leadsto$ Imbalance $\leadsto$ Inefficency $\leadsto$ Model dependence $\leadsto$ Bias
   - If the data have no good matches, the paradox won't be a problem but you're cooked anyway.
   - Doesn't PSM solve the curse of dimensionality problem? Nope. The PSM Paradox gets worse with more covariates

PSM is Blind Where Other Methods Can See

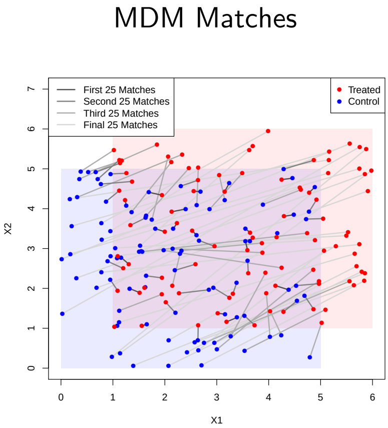
PSM is Blind Where Other Methods Can See

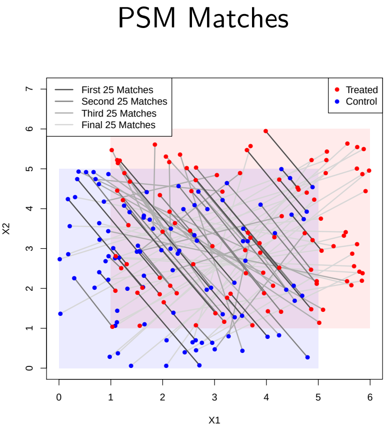


PSM is Blind Where Other Methods Can See



What Does PSM Match?

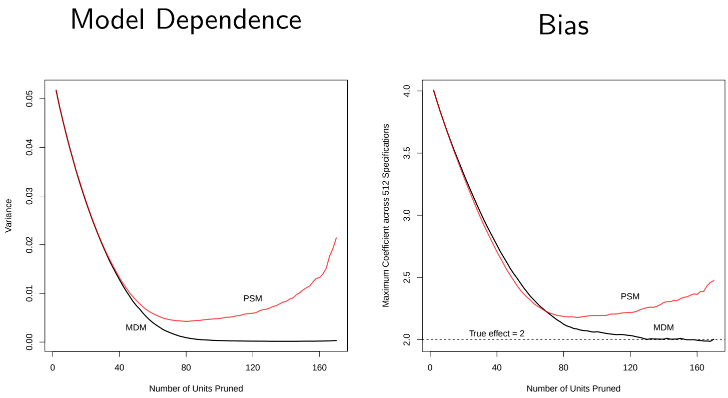




Controls: $X_1, X_2 \sim \text{Uniform}(0{,}5)$

Treateds: $X_1, X_2 \sim \text{Uniform}(1{,}6)$

PSM Increases Model Dependence & Bias

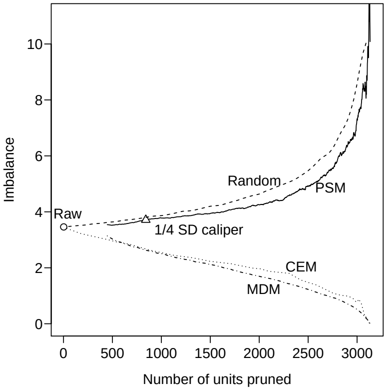


$$Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i$$
$$\epsilon_i \sim N(0,1)$$

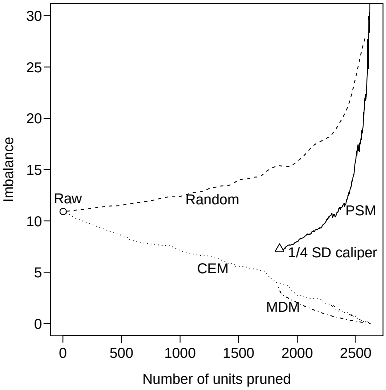
The Propensity Score Paradox in Real Data

The Propensity Score Paradox in Real Data

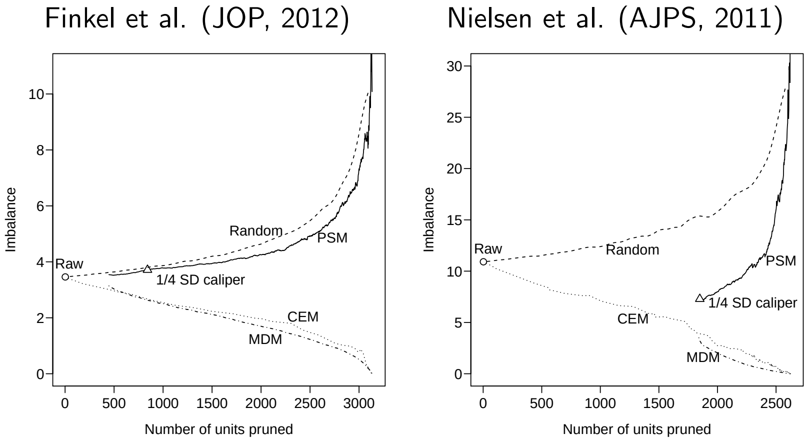
Finkel et al. (JOP, 2012)



Nielsen et al. (AJPS, 2011)



The Propensity Score Paradox in Real Data



Similar pattern for $>$ 20 other real data sets we checked

Seguro Popular Evaluation

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government "tie the hands" of their successors?

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government "tie the hands" of their successors?
    - Commission an independent evaluation

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government "tie the hands" of their successors?
  - Commission an independent evaluation
  - Like in science: make themselves vulnerable to being proven wrong

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government "tie the hands" of their successors?
  - Commission an independent evaluation
  - Like in science: make themselves vulnerable to being proven wrong
  - If we show SP is a success: elimination would be difficult

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government “tie the hands” of their successors?
  - Commission an independent evaluation
  - Like in science: make themselves vulnerable to being proven wrong
  - If we show SP is a success: elimination would be difficult
  - If SP is a failure: who cares about extending it

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government "tie the hands" of their successors?
  - Commission an independent evaluation
  - Like in science: make themselves vulnerable to being proven wrong
  - If we show SP is a success: elimination would be difficult
  - If SP is a failure: who cares about extending it
  - (They are true believers in SP)

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government “tie the hands” of their successors?
  - Commission an independent evaluation
  - Like in science: make themselves vulnerable to being proven wrong
  - If we show SP is a success: elimination would be difficult
  - If SP is a failure: who cares about extending it
  - (They are true believers in SP)
- The largest randomized health policy experiment in history

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government "tie the hands" of their successors?
  - Commission an independent evaluation
  - Like in science: make themselves vulnerable to being proven wrong
  - If we show SP is a success: elimination would be difficult
  - If SP is a failure: who cares about extending it
  - (They are true believers in SP)
- The largest randomized health policy experiment in history
- One of the largest policy experiments to date

Seguro Popular Evaluation

- Frenk and Fox asked: How can one democratically elected government "tie the hands" of their successors?
  - Commission an independent evaluation
  - Like in science: make themselves vulnerable to being proven wrong
  - If we show SP is a success: elimination would be difficult
  - If SP is a failure: who cares about extending it
  - (They are true believers in SP)
- The largest randomized health policy experiment in history
- One of the largest policy experiments to date
- 148 geographic areas, 1,380 localities, $\approx 118,569$ households, and $\approx 534,457$ people

Design Summary

Design Summary

1. Define 12,284 "health clusters" that tile Mexico's 31 states; each includes a health clinic and catchment area

Design Summary

1. Define 12,284 "health clusters" that tile Mexico's 31 states; each includes a health clinic and catchment area
2. Persuaded 13 of 31 states to participate (7,078 clusters)

Design Summary

1. Define 12,284 "health clusters" that tile Mexico's 31 states; each includes a health clinic and catchment area
2. Persuaded 13 of 31 states to participate (7,078 clusters)
3. Match clusters in pairs on background characteristics

Design Summary

1. Define 12,284 “health clusters” that tile Mexico’s 31 states; each includes a health clinic and catchment area
2. Persuaded 13 of 31 states to participate (7,078 clusters)
3. Match clusters in pairs on background characteristics
4. Select 74 pairs (based on necessary political criteria, closeness of the match, likelihood of compliance)

Design Summary

1. Define 12,284 "health clusters" that tile Mexico's 31 states; each includes a health clinic and catchment area
2. Persuaded 13 of 31 states to participate (7,078 clusters)
3. Match clusters in pairs on background characteristics
4. Select 74 pairs (based on necessary political criteria, closeness of the match, likelihood of compliance)
5. Randomly assign one in each pair to receive encouragement to affiliate, better health facilities, drugs, and doctors

Design Summary

1. Define 12,284 “health clusters” that tile Mexico’s 31 states; each includes a health clinic and catchment area
2. Persuaded 13 of 31 states to participate (7,078 clusters)
3. Match clusters in pairs on background characteristics
4. Select 74 pairs (based on necessary political criteria, closeness of the match, likelihood of compliance)
5. Randomly assign one in each pair to receive encouragement to affiliate, better health facilities, drugs, and doctors
6. Conduct baseline survey of each cluster’s health facility

Design Summary

1. Define 12,284 "health clusters" that tile Mexico's 31 states; each includes a health clinic and catchment area
2. Persuaded 13 of 31 states to participate (7,078 clusters)
3. Match clusters in pairs on background characteristics
4. Select 74 pairs (based on necessary political criteria, closeness of the match, likelihood of compliance)
5. Randomly assign one in each pair to receive encouragement to affiliate, better health facilities, drugs, and doctors
6. Conduct baseline survey of each cluster's health facility
7. Survey ≈32,000 random households in 50 of the 74 treated and control unit pairs (chosen based on likelihood of compliance with encouragement and similarity of the clusters within pair)

Design Summary

1. Define 12,284 "health clusters" that tile Mexico's 31 states; each includes a health clinic and catchment area
2. Persuaded 13 of 31 states to participate (7,078 clusters)
3. Match clusters in pairs on background characteristics
4. Select 74 pairs (based on necessary political criteria, closeness of the match, likelihood of compliance)
5. Randomly assign one in each pair to receive encouragement to affiliate, better health facilities, drugs, and doctors
6. Conduct baseline survey of each cluster's health facility
7. Survey $\approx$32,000 random households in 50 of the 74 treated and control unit pairs (chosen based on likelihood of compliance with encouragement and similarity of the clusters within pair)
8. Repeat surveys in 10 months to measure outcome

Matching in Seguro Popular Experiment: Triply Robust

Matching in Seguro Popular Experiment: Triply Robust

Design has three parts

Matching in Seguro Popular Experiment: Triply Robust

Design has three parts

1. Matching pairs (of health clusters) on observed covariates

Matching in Seguro Popular Experiment: Triply Robust

Design has three parts

1. Matching pairs (of health clusters) on observed covariates
2. Randomization of treatment within pairs

Matching in Seguro Popular Experiment: Triply Robust

Design has three parts

1. Matching pairs (of health clusters) on observed covariates
2. Randomization of treatment within pairs
3. If necessary statistically adjust for differences

Matching in Seguro Popular Experiment: Triply Robust

Design has three parts

1. Matching pairs (of health clusters) on observed covariates
2. Randomization of treatment within pairs
3. If necessary statistically adjust for differences

Triple Robustness

If matching or randomization or statistical analysis is right, but the other two are wrong, results are still unbiased

Matching in Seguro Popular Experiment: Triply Robust

Design has three parts

1. Matching pairs (of health clusters) on observed covariates
2. Randomization of treatment within pairs
3. If necessary statistically adjust for differences

Triple Robustness

If matching or randomization or statistical analysis is right, but the other two are wrong, results are still unbiased

Two Additional Checks if Triple Robustness Fails

Matching in Seguro Popular Experiment: Triply Robust

Design has three parts

1. Matching pairs (of health clusters) on observed covariates
2. Randomization of treatment within pairs
3. If necessary statistically adjust for differences

Triple Robustness

If matching or randomization or statistical analysis is right, but the other two are wrong, results are still unbiased

Two Additional Checks if Triple Robustness Fails

1. If one of the three works, then "effect of SP" on time 0 outcomes (measured in baseline survey) must be zero

Matching in Seguro Popular Experiment: Triply Robust

Design has three parts

1. Matching pairs (of health clusters) on observed covariates
2. Randomization of treatment within pairs
3. If necessary statistically adjust for differences

Triple Robustness

If matching or randomization or statistical analysis is right, but the other two are wrong, results are still unbiased

Two Additional Checks if Triple Robustness Fails

1. If one of the three works, then “effect of SP” on time 0 outcomes (measured in baseline survey) must be zero
2. If we lose pairs, we check for selection bias by rerunning this check

Concluding Remarks

Concluding Remarks

- Don't use propensity score matching

Concluding Remarks

- Don't use propensity score matching
  - Approximates complete, not fully blocked, experiments

Concluding Remarks

- Don't use propensity score matching
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence

Concluding Remarks

- Don't use propensity score matching
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
- Use this Simple and Powerful Method: CEM

Concluding Remarks

- Don't use propensity score matching
    - Approximates complete, not fully blocked, experiments
    - Ignores information; exacerbates model dependence
- Use this Simple and Powerful Method: CEM
- Randomized Experiments: Better with matched pairs

For more information, articles, & software

GaryKing.org