Advanced Quantitative Research Methodology, Lecture Notes: Matching Methods for Causal Inference[1]

Gary King[2]

Institute for Quantitative Social Science
Harvard University

[2] GaryKing.org

Matching Overview

Matching Overview

- Current practice:

Matching Overview

- Current practice:
  - ⇝ "Matching As Nonparametric Preprocessing For Reducing Model Dependence In Parametric Causal Inference" (Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart)

Matching Overview

- Current practice:
  - ⇝ "Matching As Nonparametric Preprocessing For Reducing Model Dependence In Parametric Causal Inference" (Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart)
- Current practice violates current statistical theory.

Matching Overview

- Current practice:
  - ⇝ "Matching As Nonparametric Preprocessing For Reducing Model Dependence In Parametric Causal Inference" (Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart)
- Current practice violates current statistical theory. So let's change the theory:

Matching Overview

- Current practice:
  - ⇝ "Matching As Nonparametric Preprocessing For Reducing Model Dependence In Parametric Causal Inference" (Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart)
- Current practice violates current statistical theory. So let's change the theory:
  - ⇝ "A Theory of Statistical Inference for Matching Methods in Applied Causal Research"
    (Stefano Iacus, Gary King, Giuseppe Porro)

Matching Overview

- Current practice:
  - ⇝ "Matching As Nonparametric Preprocessing For Reducing Model Dependence In Parametric Causal Inference" (Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart)
- Current practice violates current statistical theory. So let's change the theory:
  - ⇝ "A Theory of Statistical Inference for Matching Methods in Applied Causal Research"
    (Stefano Iacus, Gary King, Giuseppe Porro)
- The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:

Matching Overview

- Current practice:
  - ⇝ "Matching As Nonparametric Preprocessing For Reducing Model Dependence In Parametric Causal Inference" (Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart)
- Current practice violates current statistical theory. So let's change the theory:
  - ⇝ "A Theory of Statistical Inference for Matching Methods in Applied Causal Research" (Stefano Iacus, Gary King, Giuseppe Porro)
- The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:
  - ⇝ "Why Propensity Scores Should Not Be Used for Matching" (Gary King, Richard Nielsen)

Matching Overview

- Current practice:
  - $\rightsquigarrow$ "Matching As Nonparametric Preprocessing For Reducing Model Dependence In Parametric Causal Inference" (Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart)
- Current practice violates current statistical theory. So let's change the theory:
  - $\rightsquigarrow$ "A Theory of Statistical Inference for Matching Methods in Applied Causal Research" (Stefano Iacus, Gary King, Giuseppe Porro)
- The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:
  - $\rightsquigarrow$ "Why Propensity Scores Should Not Be Used for Matching" (Gary King, Richard Nielsen)
- Matching methods optimize either imbalance ($\approx$ bias) or # units pruned ($\approx$ variance); users need both simultaneously':

Matching Overview

- Current practice:
  - ⇝ "Matching As Nonparametric Preprocessing For Reducing Model Dependence In Parametric Causal Inference" (Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart)
- Current practice violates current statistical theory. So let's change the theory:
  - ⇝ "A Theory of Statistical Inference for Matching Methods in Applied Causal Research" (Stefano Iacus, Gary King, Giuseppe Porro)
- The most popular method (propensity score matching, used in 53,200 articles!) sounds magical:
  - ⇝ "Why Propensity Scores Should Not Be Used for Matching" (Gary King, Richard Nielsen)
- Matching methods optimize either imbalance ($\approx$ bias) or # units pruned ($\approx$ variance); users need both simultaneously':
  - ⇝ "The Balance-Sample Size Frontier in Matching Methods for Causal Inference" (Gary King, Christopher Lucas and Richard Nielsen)

Overview of Matching for Causal Inference

Overview of Matching for Causal Inference

- Goal: reduce model dependence

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:
    - If each treated unit exactly matches a control unit w.r.t. X, then: (1) treated and control groups are identical, (2) X is no longer a confounder, (3) no need to worry about the functional form ($\bar{Y}_T - \bar{Y}_C$ is good enough).

Overview of Matching for Causal Inference

- Goal: reduce model dependence
- A nonparametric, non-model-based approach
- Makes parametric models work better rather than substitute for them (i.e,. matching is not an estimator; its a preprocessing method)
- Should have been called pruning (no bias is introduced if pruning is a function of T and X, but not Y)
- Apply model to preprocessed (pruned) rather than raw data
- Violates the "more data is better" principle, but that only applies when you know the DGP
- Overall idea:
    - If each treated unit exactly matches a control unit w.r.t. X, then: (1) treated and control groups are identical, (2) X is no longer a confounder, (3) no need to worry about the functional form ($\bar{Y}_T - \bar{Y}_C$ is good enough).
    - If treated and control groups are better balanced than when you started, due to pruning, model dependence is reduced

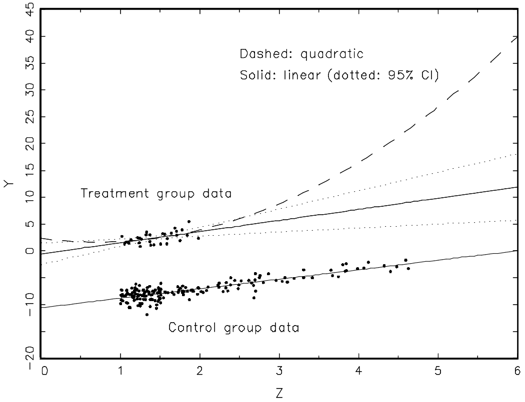
Model Dependence: A Simpler Example

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

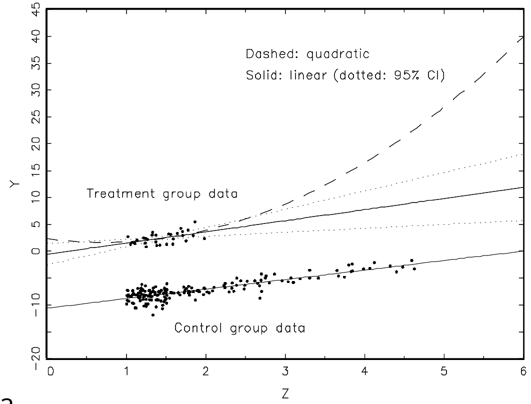
Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)



Model Dependence: A Simpler Example

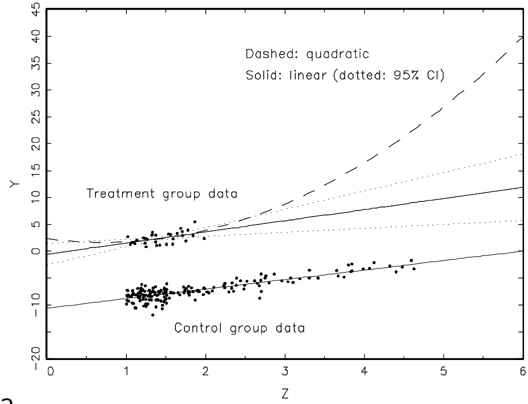
(King and Zeng, 2006: fig.4 *Political Analysis*)



What to do?

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

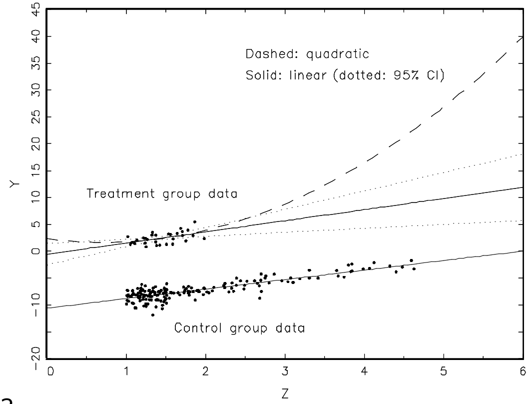


What to do?

- Preprocess I: Eliminate extrapolation region

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

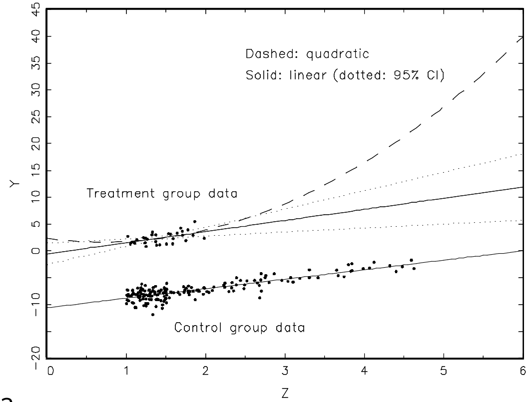


What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune) within interpolation region

Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

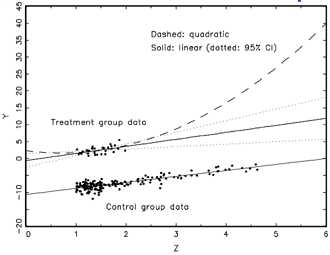


What to do?

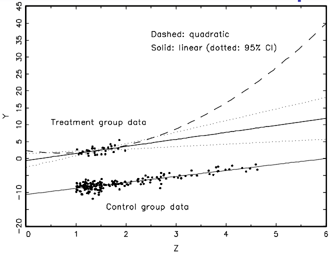
- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune) within interpolation region
- Model remaining imbalance (as you would w/o matching)

Remove Extrapolation Region, then Match

Remove Extrapolation Region, then Match

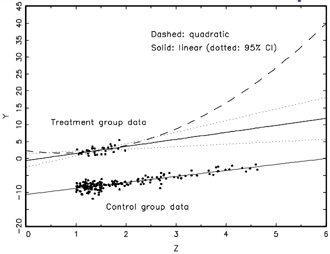


Remove Extrapolation Region, then Match



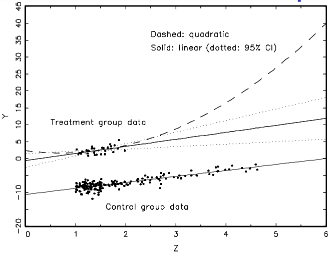
- Must remove data (selecting on X) to avoid extrapolation.

Remove Extrapolation Region, then Match



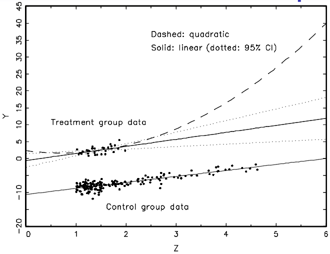
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$

Remove Extrapolation Region, then Match



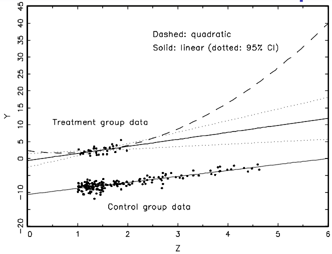
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
    1. Exact match, so support is defined only at data points

Remove Extrapolation Region, then Match



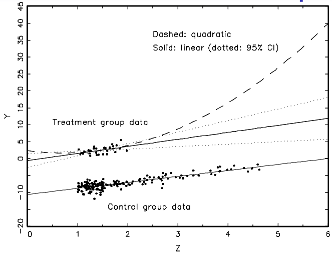
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach

Remove Extrapolation Region, then Match



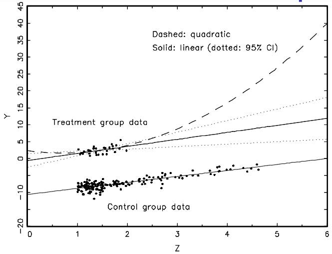
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
    1. Exact match, so support is defined only at data points
    2. Less but still conservative: convex hull approach
        - let T^* and X^* denote subsets of T and X s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$

Remove Extrapolation Region, then Match



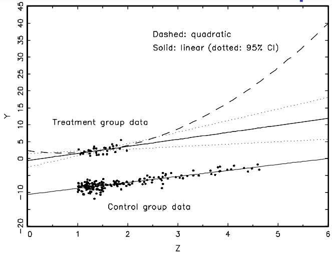
- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let T^* and X^* denote subsets of T and X s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use X^* as estimate of common support (deleting remaining observations)

Remove Extrapolation Region, then Match



- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let T^* and X^* denote subsets of T and X s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use X^* as estimate of common support (deleting remaining observations)
  3. Other approaches, based on distance metrics, pscores, etc.

Remove Extrapolation Region, then Match



- Must remove data (selecting on X) to avoid extrapolation.
- Options to find "common support" of $p(X|T=1)$ and $P(X|T=0)$
  1. Exact match, so support is defined only at data points
  2. Less but still conservative: convex hull approach
     - let T^* and X^* denote subsets of T and X s.t. $\{1-T^*, X^*\}$ falls within the convex hull of $\{T, X\}$
     - use X^* as estimate of common support (deleting remaining observations)
  3. Other approaches, based on distance metrics, pscores, etc.
  4. Easiest: Coarsened Exact Matching, no separate step needed

Matching within the Interpolation Region

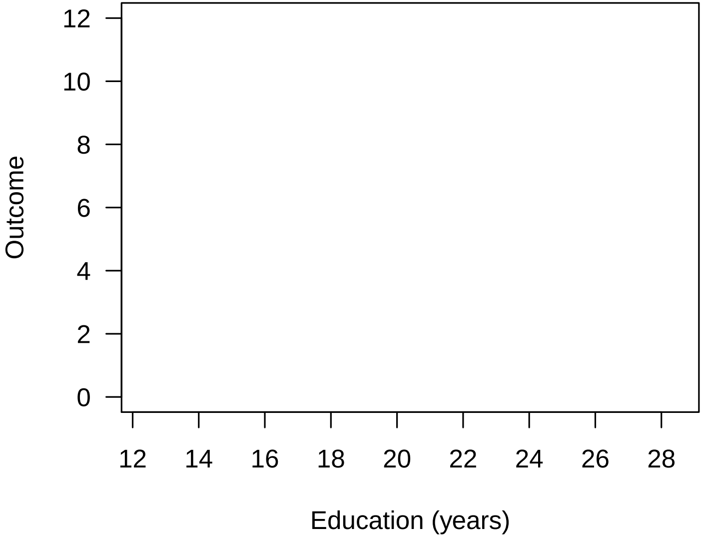
(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)

Matching within the Interpolation Region

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

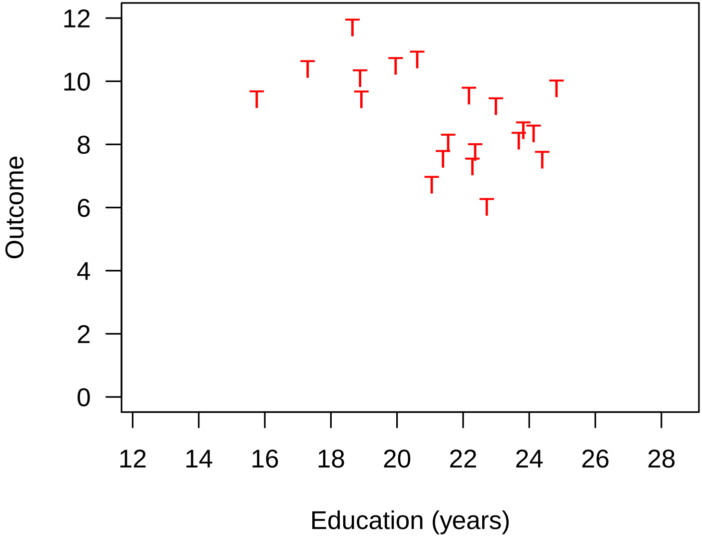
Matching within the Interpolation Region

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



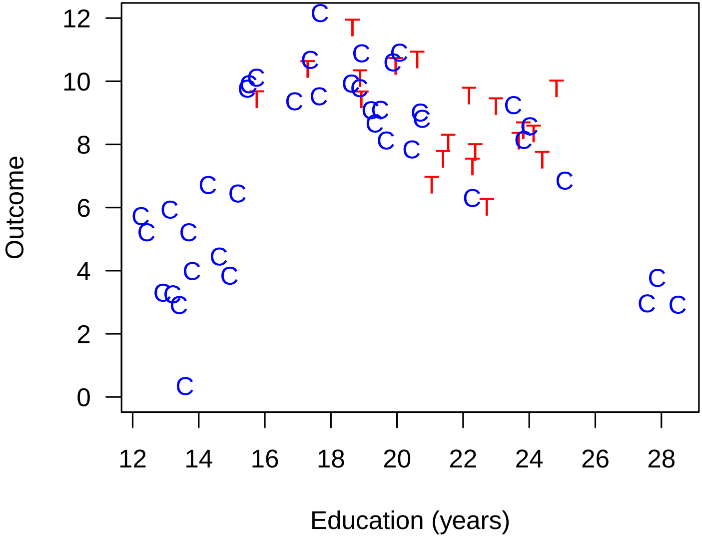
Matching within the Interpolation Region

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



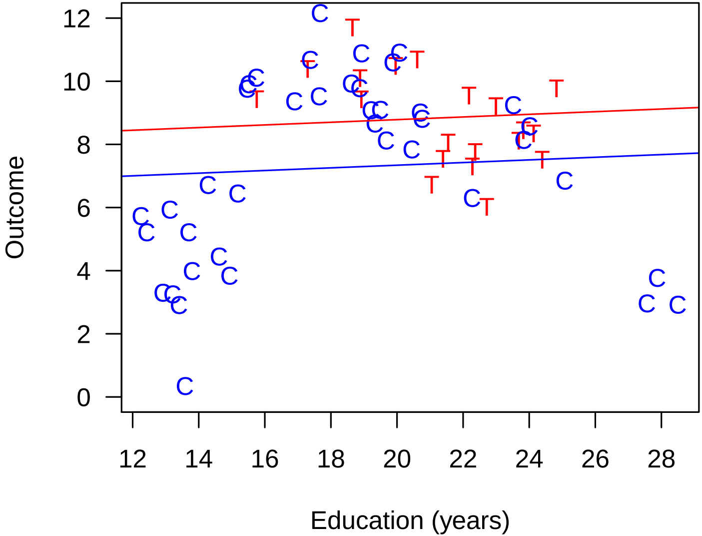
Matching within the Interpolation Region

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



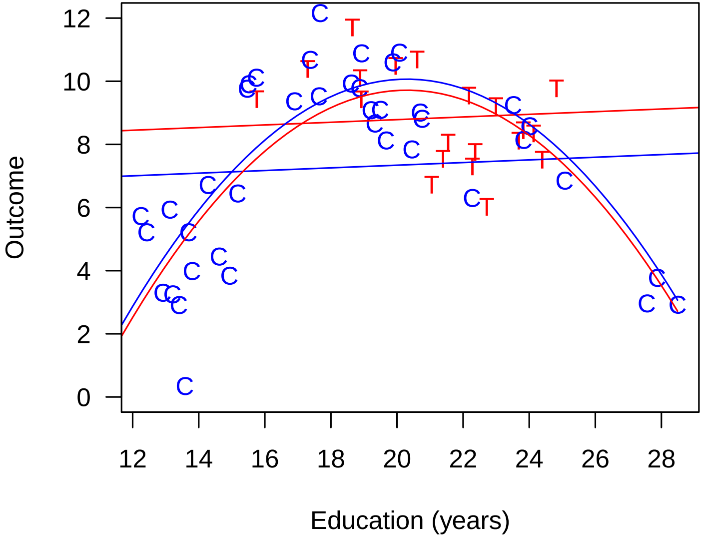
Matching within the Interpolation Region

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



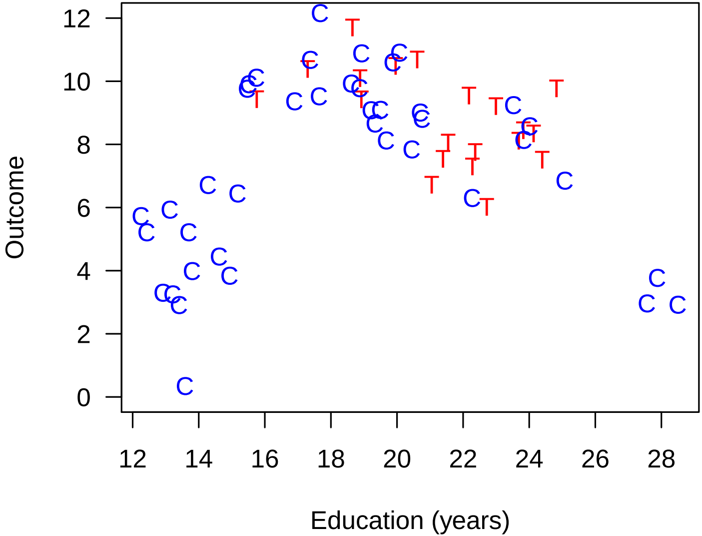
Matching within the Interpolation Region

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



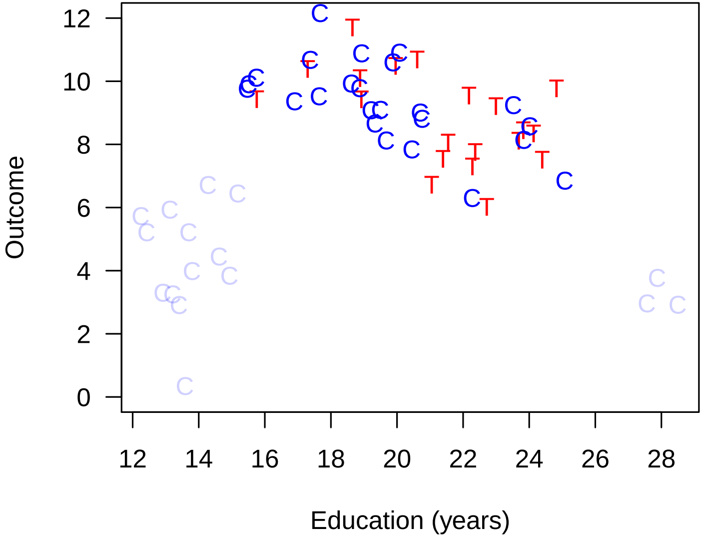
Matching within the Interpolation Region

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



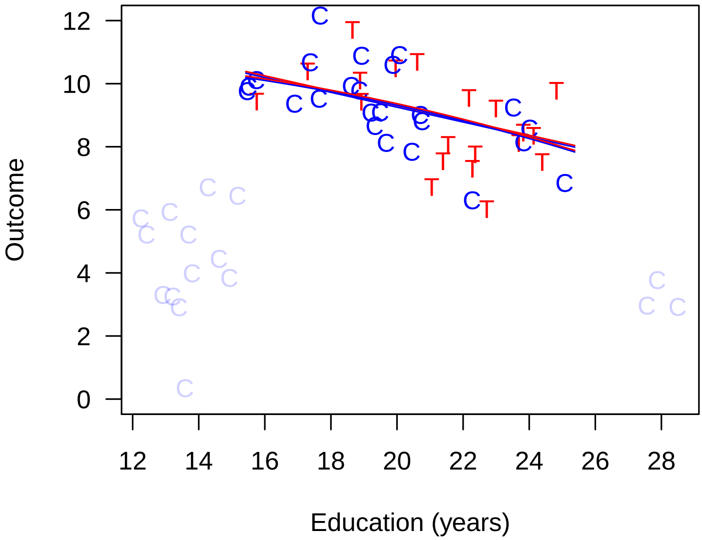
Matching within the Interpolation Region

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



Matching within the Interpolation Region

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



Matching within the Interpolation Region

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

Matching reduces model dependence, bias, and variance

Empirical Illustration: Carpenter, AJPS, 2002

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.
- seven oversight variables (median adjusted ADA scores for House and Senate Committees as well as for House and Senate floors, Democratic Majority in House and Senate, and Democratic Presidency).

Empirical Illustration: Carpenter, AJPS, 2002

- Hypothesis: Democratic senate majorities slow FDA drug approval time
- $n = 408$ new drugs (262 approved, 146 pending).
- lognormal survival model.
- seven oversight variables (median adjusted ADA scores for House and Senate Committees as well as for House and Senate floors, Democratic Majority in House and Senate, and Democratic Presidency).
- 18 control variables (clinical factors, firm characteristics, media variables, etc.)

Evaluating Reduction in Model Dependence

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- Match: prune 49 units (2 treated, 17 control units).

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- Match: prune 49 units (2 treated, 17 control units).
- run 262,143 possible specifications and calculates ATE for each.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- Match: prune 49 units (2 treated, 17 control units).
- run 262,143 possible specifications and calculates ATE for each.
- Look at *variability* in ATE estimate across specifications.

Evaluating Reduction in Model Dependence

- Focus on the causal effect of a Democratic majority in the Senate (identified by Carpenter as not robust).
- Match: prune 49 units (2 treated, 17 control units).
- run 262,143 possible specifications and calculates ATE for each.
- Look at *variability* in ATE estimate across specifications.
- (Normal applications would only use one or a few specifications.)

Reducing Model Dependence

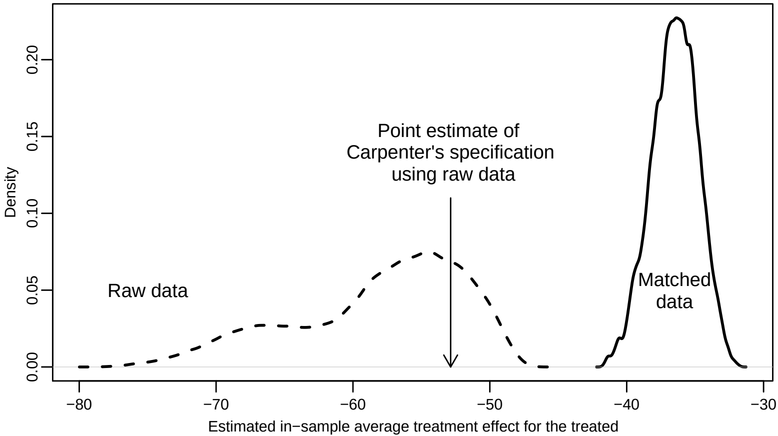


Figure: SATT Histogram: Effect of Democratic Senate majority on FDA drug approval time, across 262, 143 specifications.

Another Example: Jeffrey Koch, AJPS, 2002

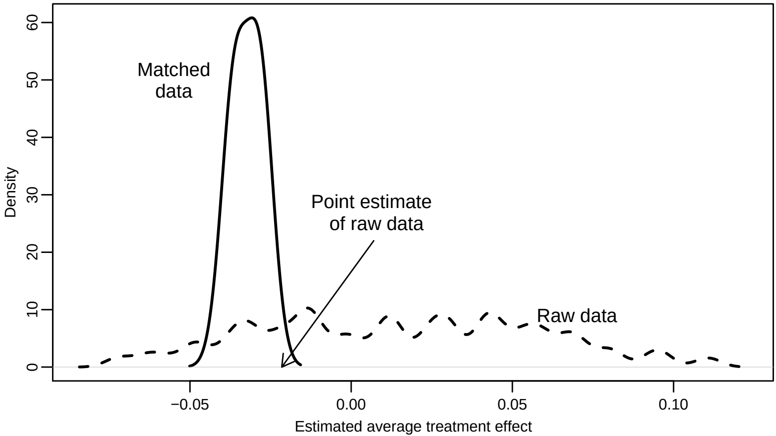


Figure: SATT Histogram: Effect of being a highly visible female Republican candidate across 63 possible specifications with the Koch data

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[eff]

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[eff]
- 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[eff]
- 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[eff]
- 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[eff]
- 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 <u>treated</u> 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:

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

What's Matching?

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

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

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:

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

  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 <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:

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

  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:

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

  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	

Matching: Finding Hidden Randomized Experiments

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

$\leadsto$ *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

$\rightsquigarrow$ *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*

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*
- 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

$\rightsquigarrow$ *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

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)

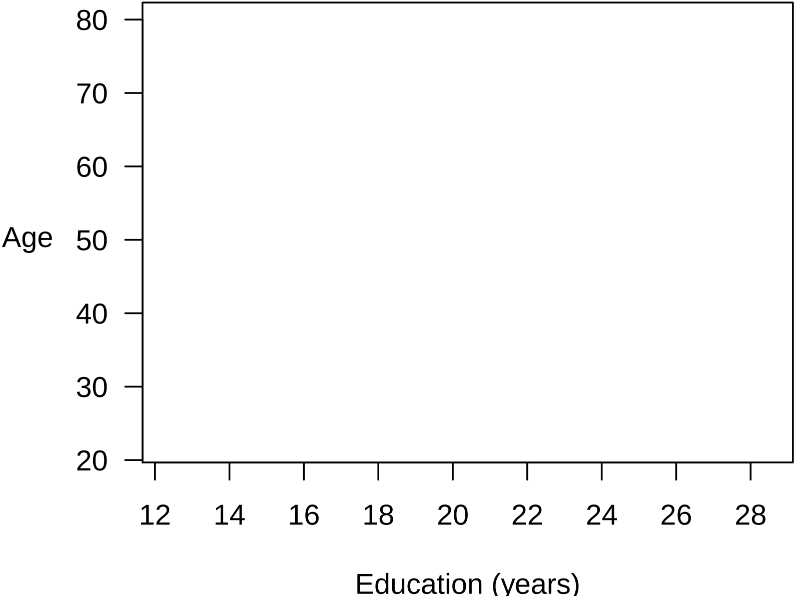
1. Preprocess (Matching)
   - Distance$(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
   - (Mahalanobis is for methodologists; in applications, use Euclidean!)
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance>*caliper*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

Method 1: Mahalanobis Distance Matching

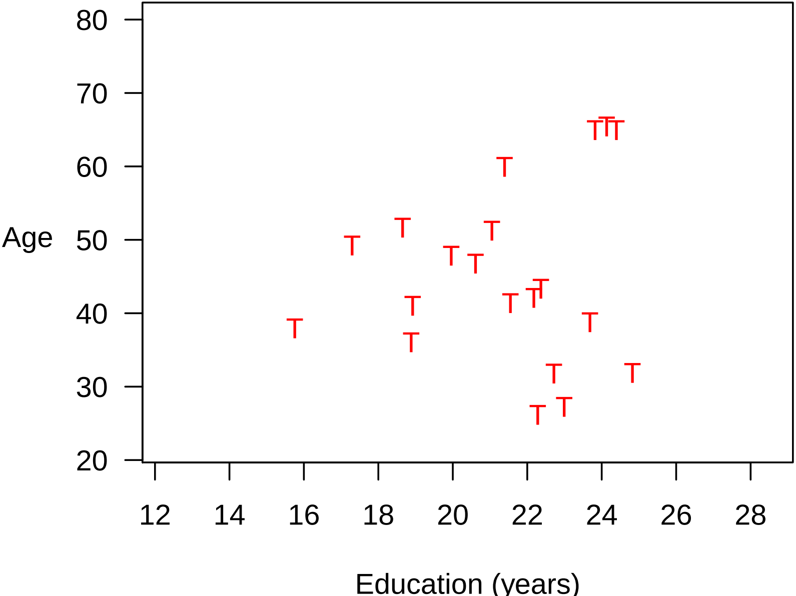
(Approximates Fully Blocked Experiment)

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

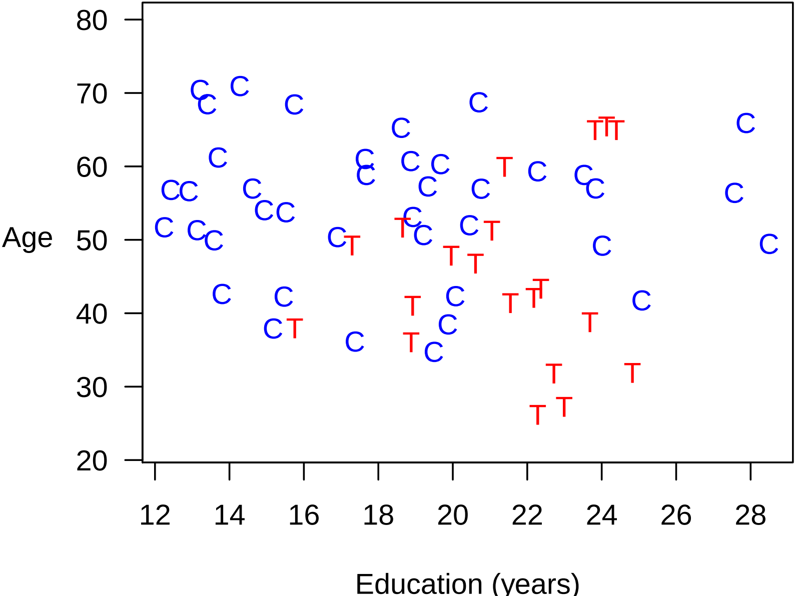
Mahalanobis Distance Matching



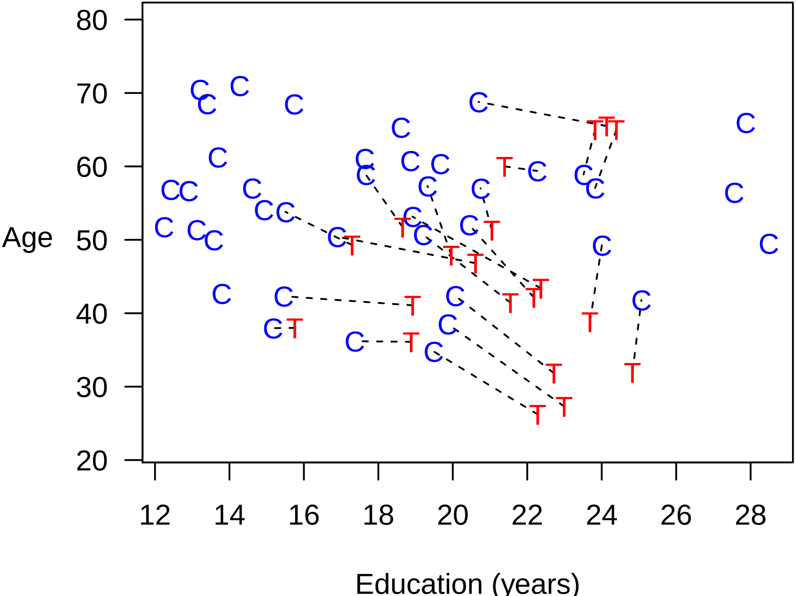
Mahalanobis Distance Matching



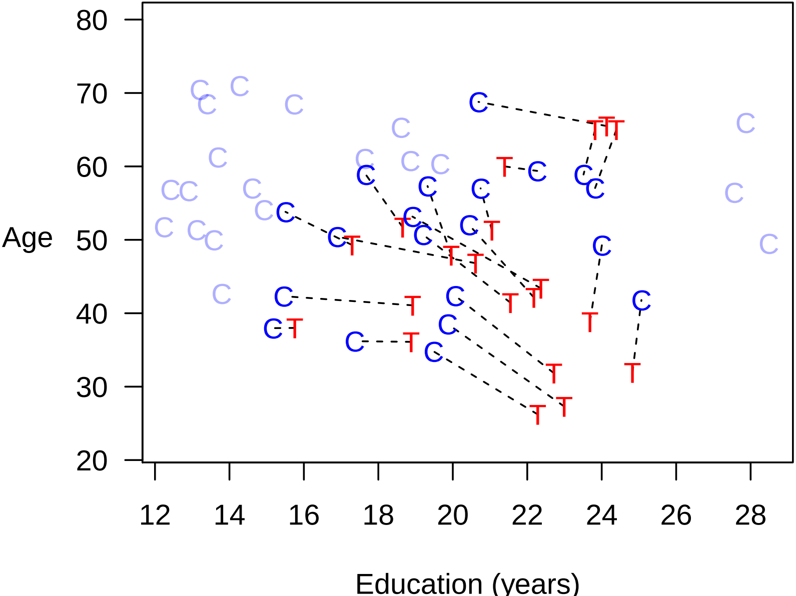
Mahalanobis Distance Matching



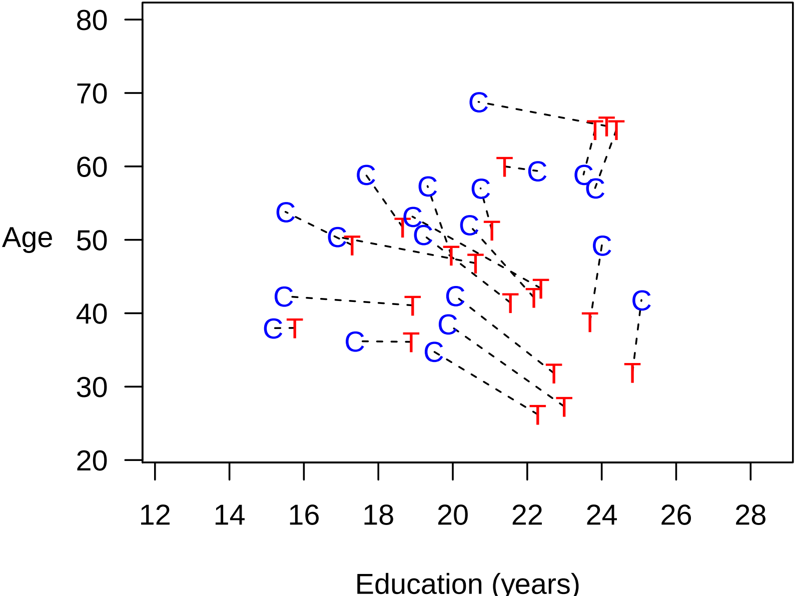
Mahalanobis Distance Matching



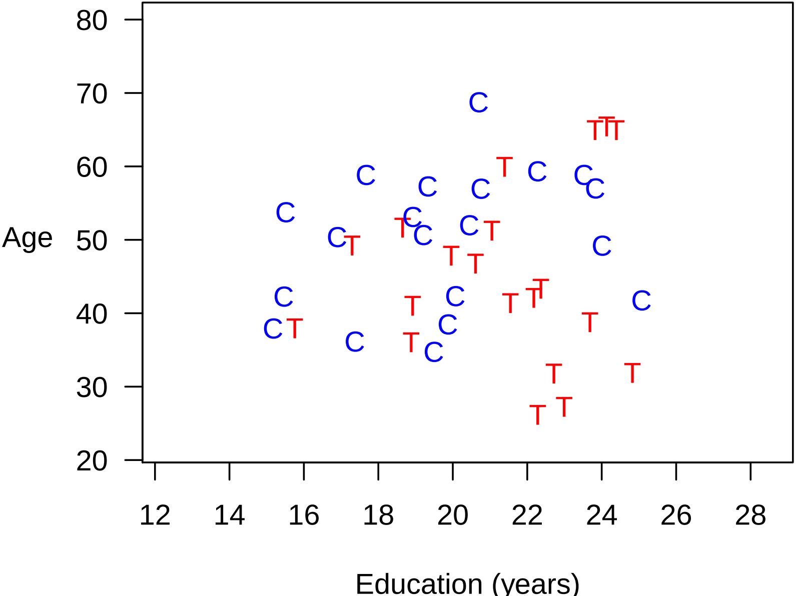
Mahalanobis Distance Matching



Mahalanobis Distance Matching

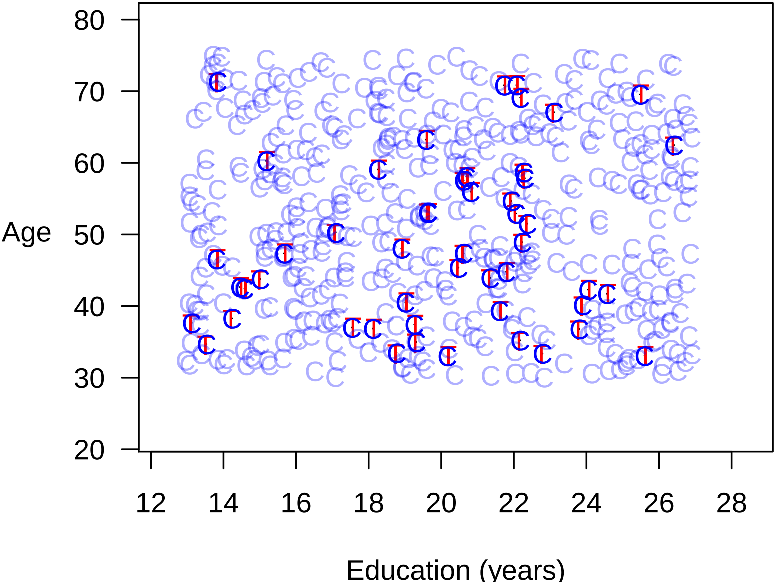


Mahalanobis Distance Matching

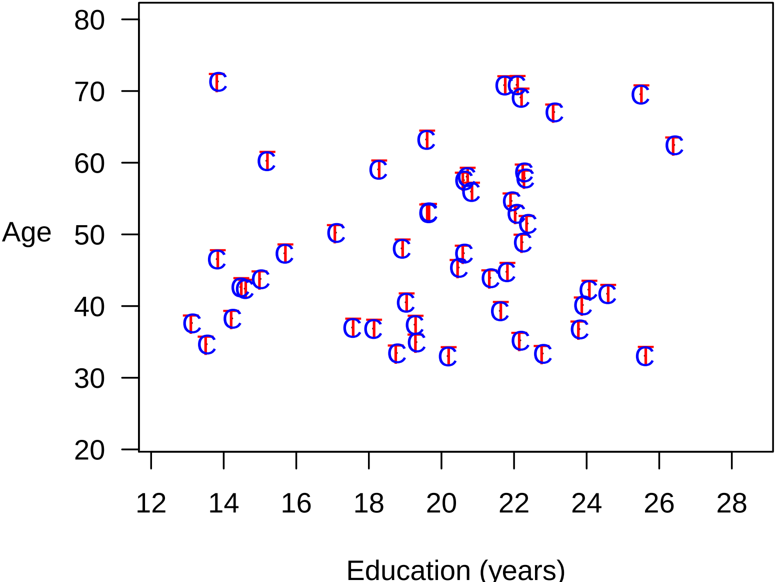


Best Case: Mahalanobis Distance Matching

Best Case: Mahalanobis Distance Matching



Best Case: Mahalanobis Distance Matching



Method 2: Coarsened Exact Matching

Method 2: Coarsened Exact Matching

(Approximates Fully Blocked Experiment)

Method 2: Coarsened Exact Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)

2. Estimation Difference in means or a model

Method 2: Coarsened Exact Matching

(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

(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

(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

(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

(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

(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

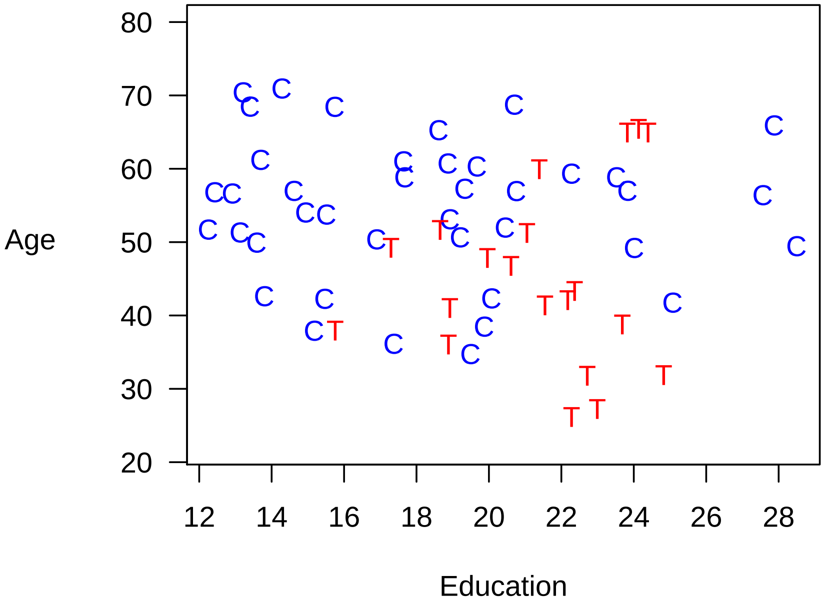
Method 2: Coarsened Exact Matching

(Approximates Fully Blocked Experiment)

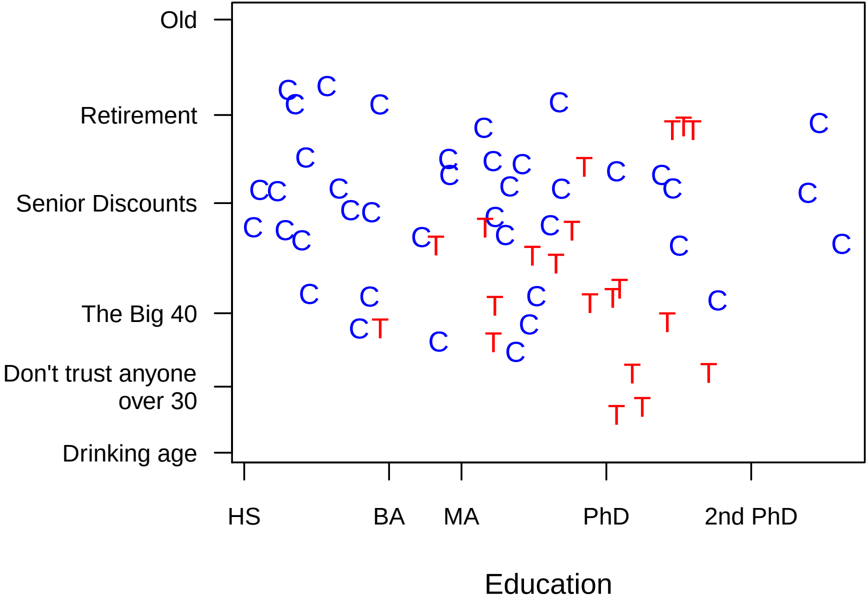
1. Preprocess (Matching)
    - Temporarily coarsen X as much as you're willing
        - e.g., Education (grade school, high school, college, graduate)
    - Apply exact matching to the coarsened X, $C(X)$
        - Sort observations into strata, each with unique values of $C(X)$
        - Prune any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

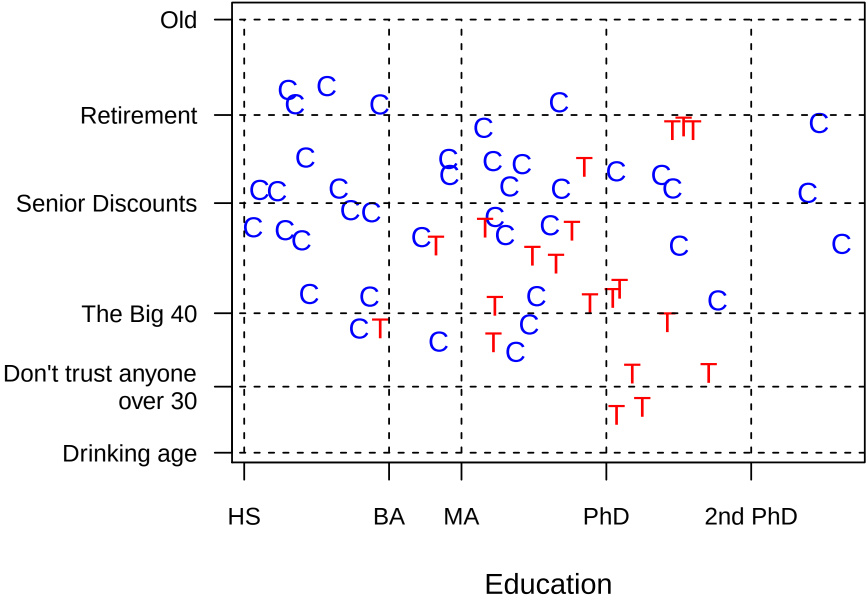
Coarsened Exact Matching



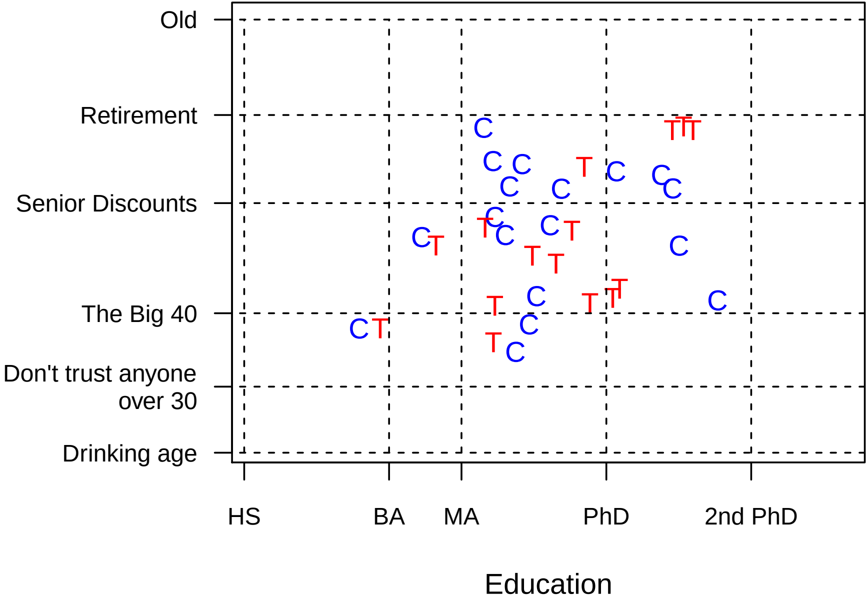
Coarsened Exact Matching



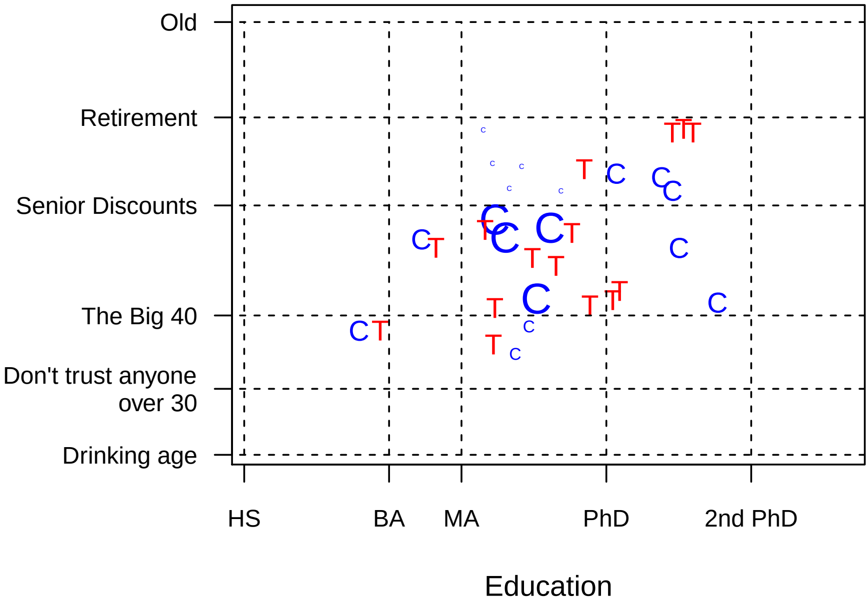
Coarsened Exact Matching



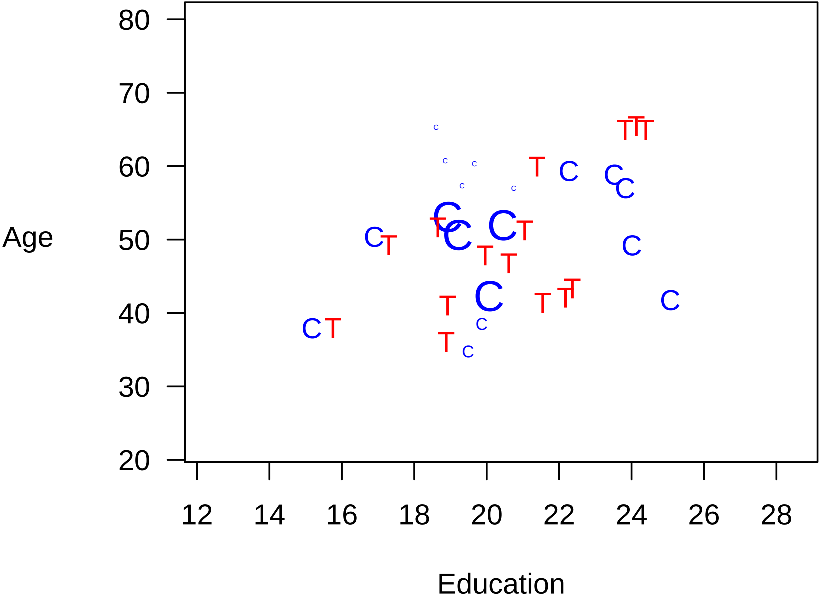
Coarsened Exact Matching



Coarsened Exact Matching

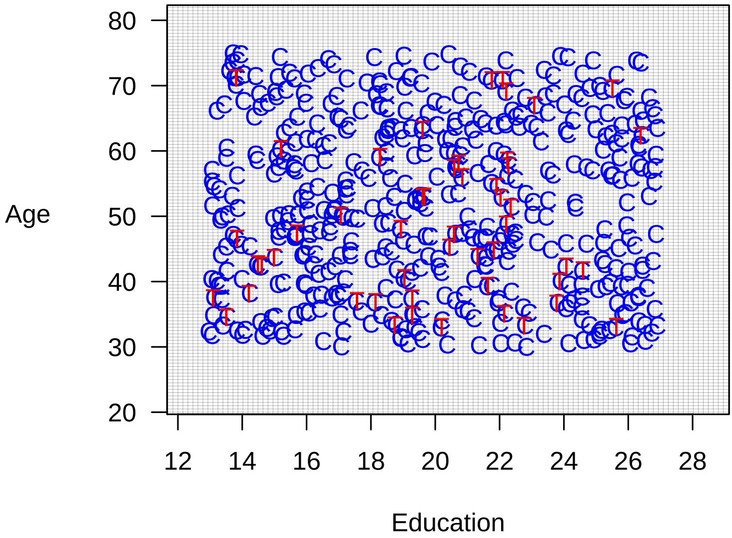


Coarsened Exact Matching

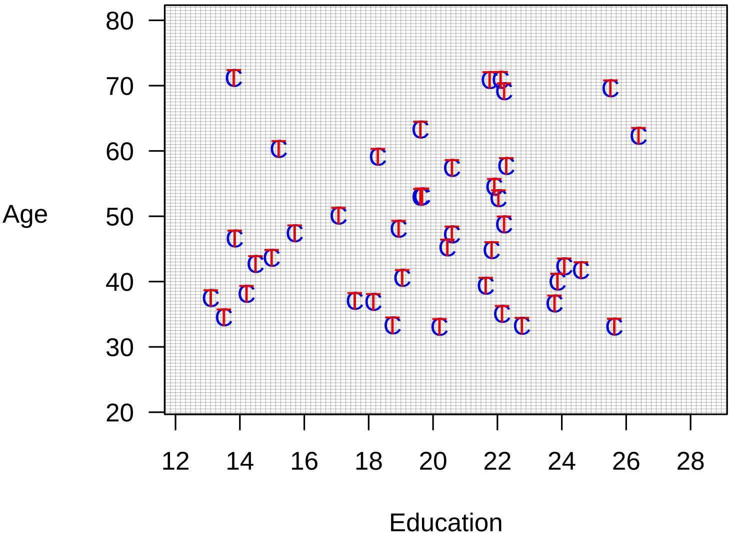


Best Case: Coarsened Exact Matching

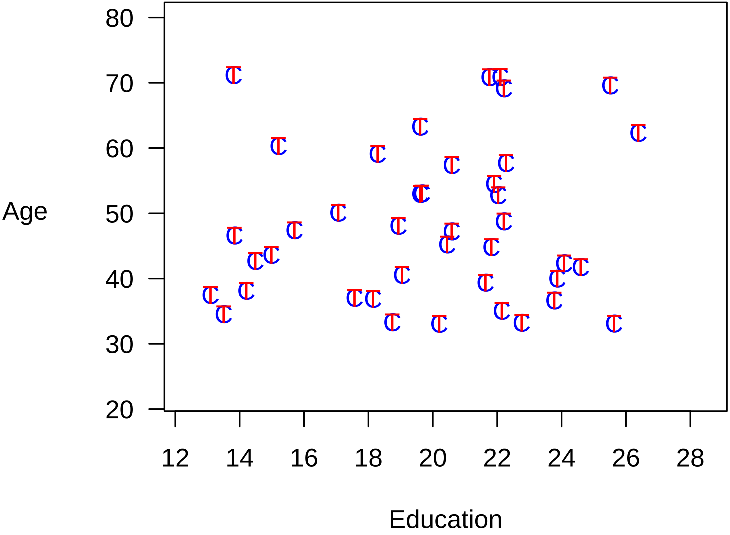
Best Case: Coarsened Exact Matching



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 $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
   - 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 $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
    - 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*

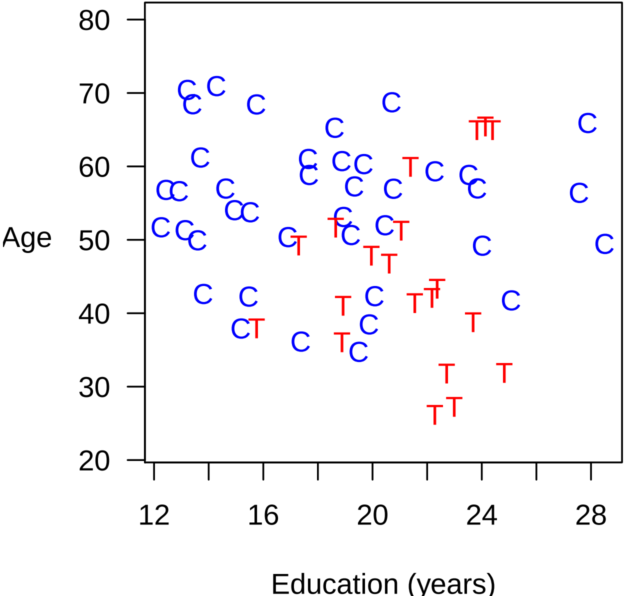
2. Estimation Difference in means or a model

Method 3: Propensity Score Matching

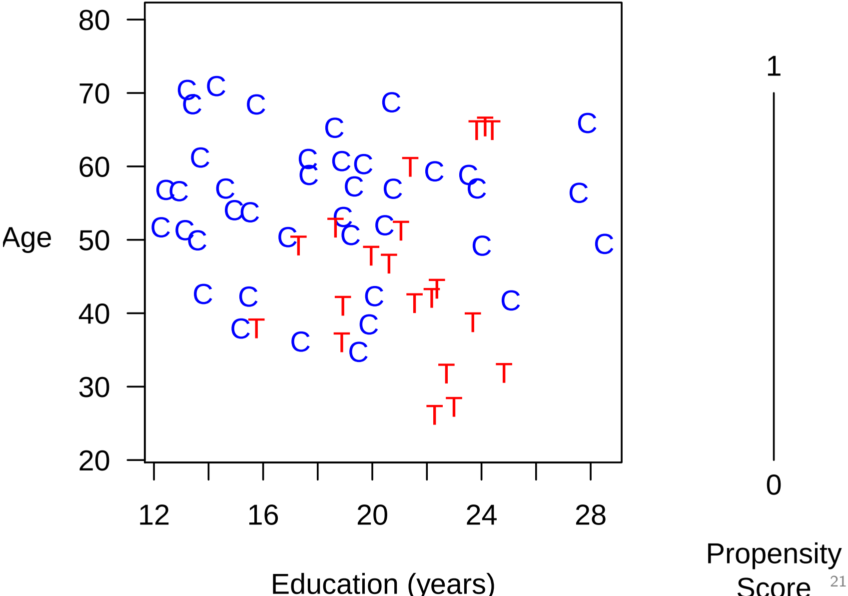
(Approximates Completely Randomized Experiment)

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

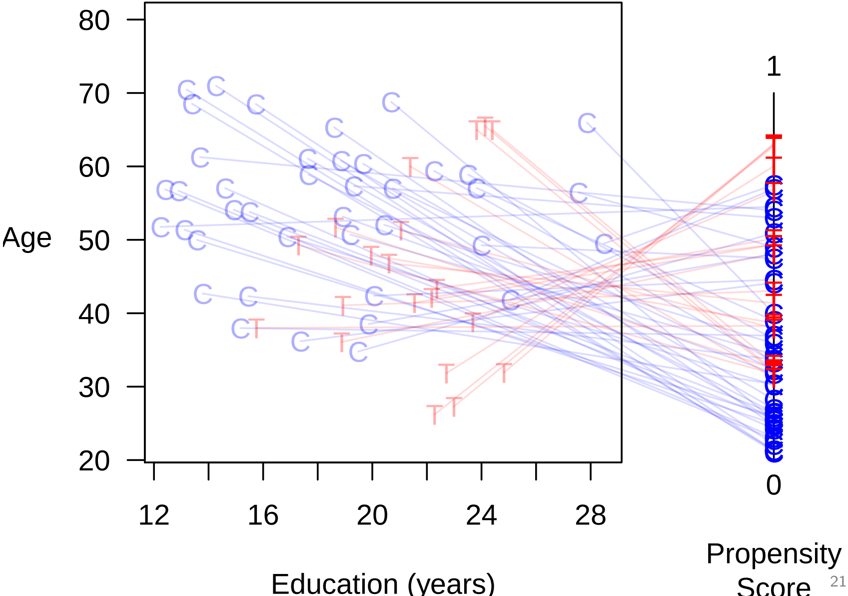
Propensity Score Matching



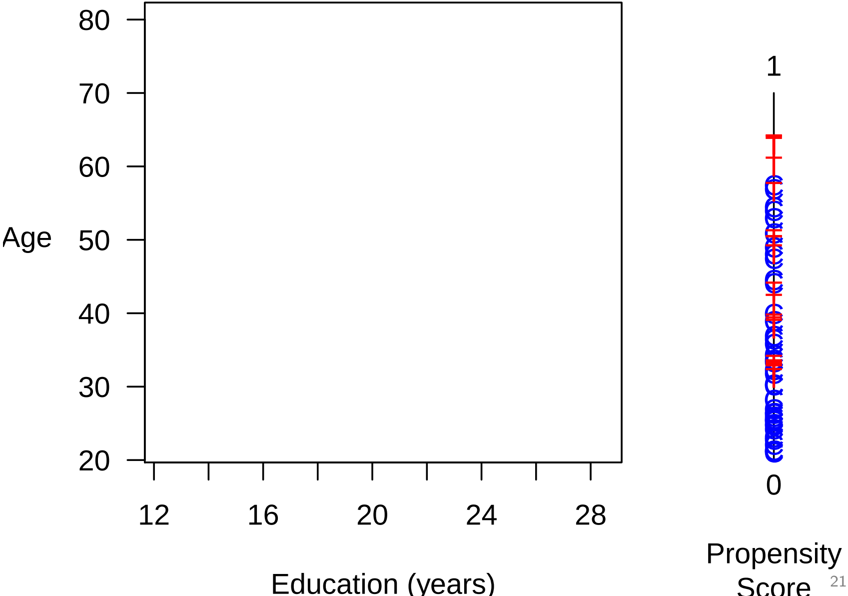
Propensity Score Matching



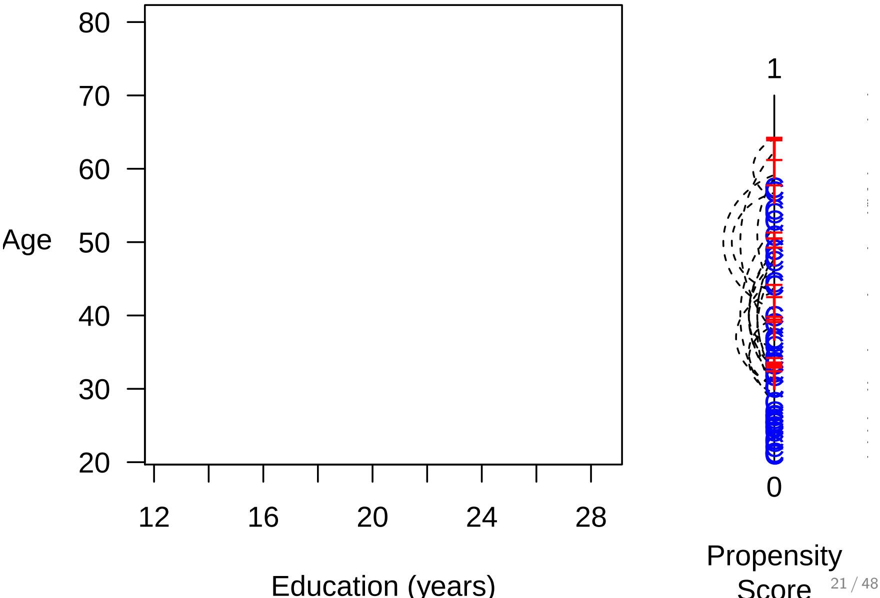
Propensity Score Matching



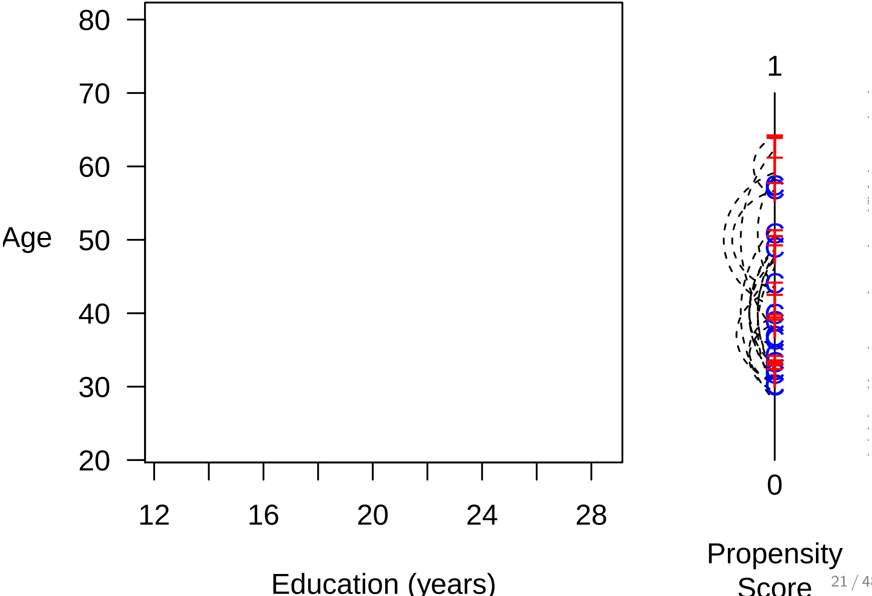
Propensity Score Matching



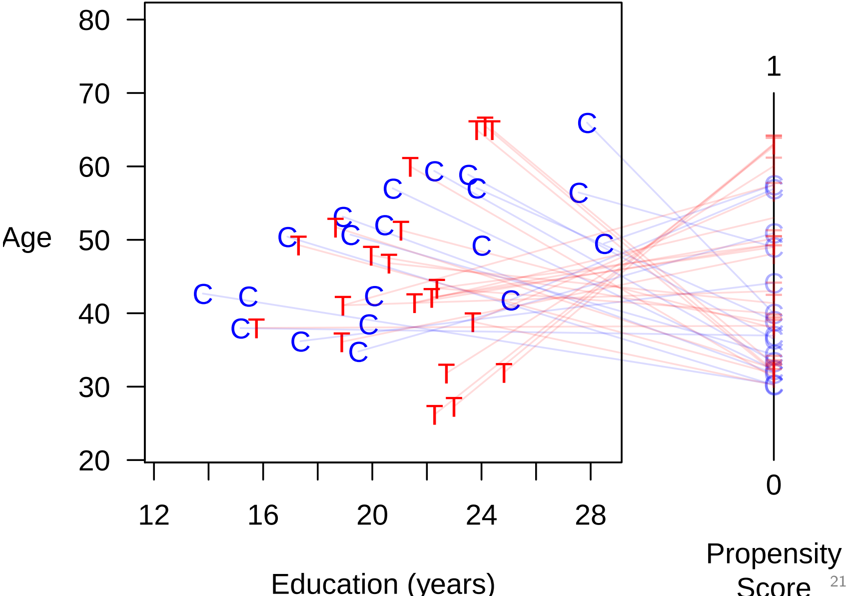
Propensity Score Matching



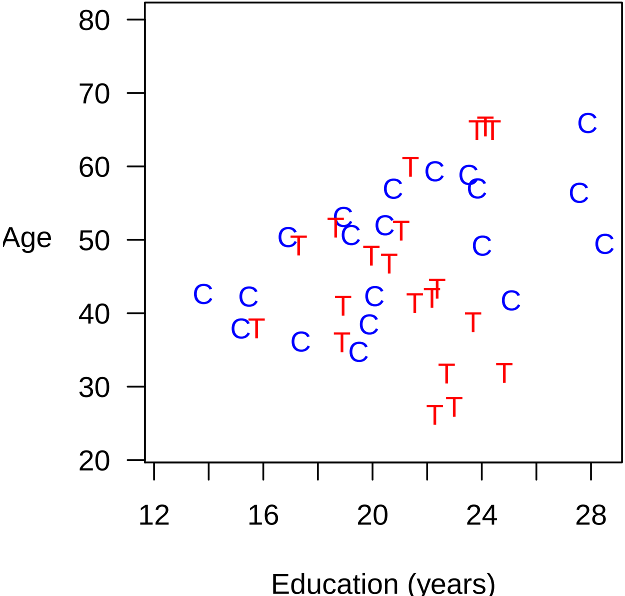
Propensity Score Matching



Propensity Score Matching

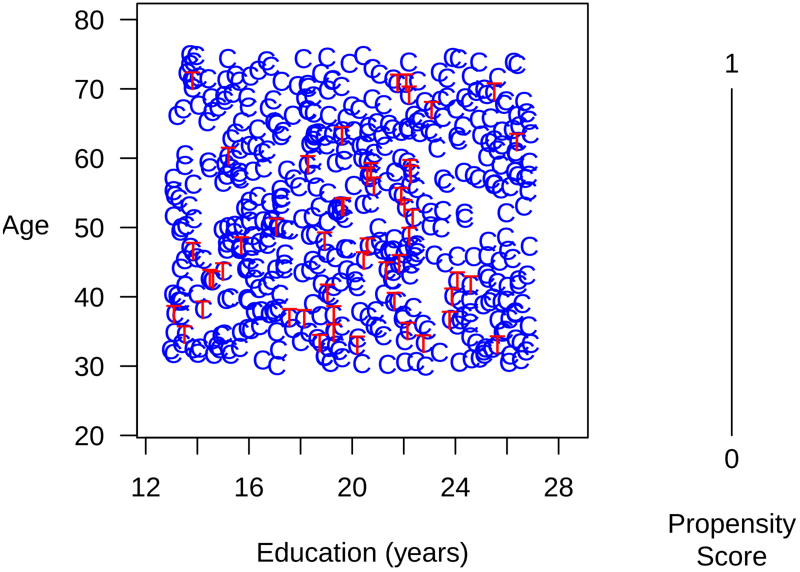


Propensity Score Matching

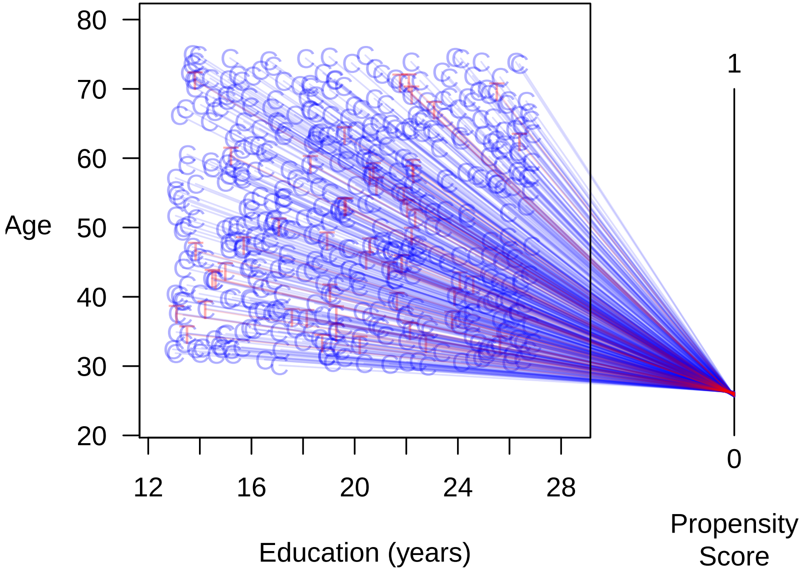


Best Case: Propensity Score Matching

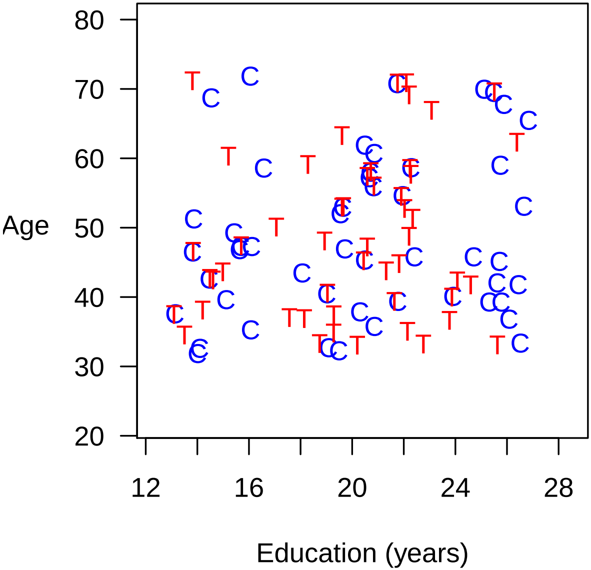
Best Case: Propensity Score Matching



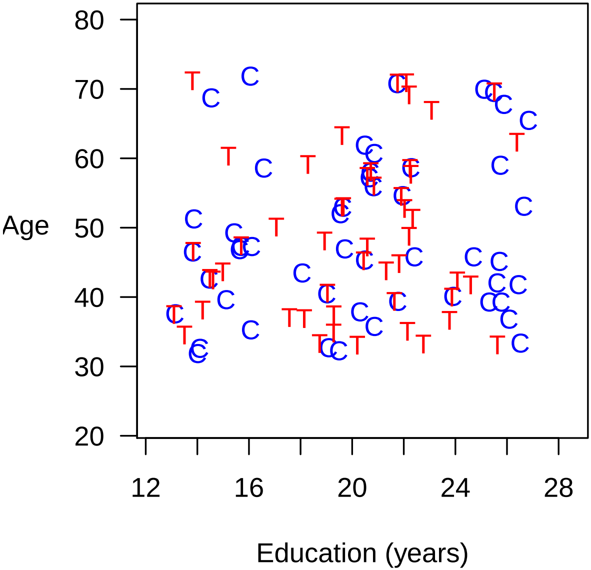
Best Case: Propensity Score Matching



Best Case: Propensity Score Matching



Best Case: Propensity Score Matching is Suboptimal



Background: Random Pruning Increases Imbalance

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
    - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
    - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
      2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
      2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
    - $\implies$ random pruning increases imbalance

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- Continuous example

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\Longrightarrow$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; X any fixed variable; with n units

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\Longrightarrow$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; X any fixed variable; with n units
  - Measure of imbalance: squared difference in means d^2, where $d = \bar{X}_t - \bar{X}_c$

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\Longrightarrow$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; X any fixed variable; with n units
  - Measure of imbalance: squared difference in means d^2, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\implies$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; X any fixed variable; with n units
  - Measure of imbalance: squared difference in means d^2, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\rightsquigarrow$ n declines $\rightsquigarrow$ $E(d^2)$ increases

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\Longrightarrow$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; X any fixed variable; with n units
  - Measure of imbalance: squared difference in means d^2, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\rightsquigarrow$ n declines $\rightsquigarrow$ $E(d^2)$ increases
  - $\Longrightarrow$ random pruning increases imbalance

Background: Random Pruning Increases Imbalance

Deleting data only helps if you're careful!

- "Random pruning": pruning process is independent of X
- Discrete example
  - Sex-balanced dataset: treateds M_t, F_t, controls M_c, F_c
  - Randomly prune 1 treated & 1 control $\rightsquigarrow$ 4 possible datasets:
    2 balanced $\{M_t, M_c\}$, $\{F_t, F_c\}$
    2 imbalanced $\{M_t, F_c\}$, $\{F_t, M_c\}$
  - $\Longrightarrow$ random pruning increases imbalance
- Continuous example
  - Dataset: $T \in \{0, 1\}$ randomly assigned; X any fixed variable; with n units
  - Measure of imbalance: squared difference in means d^2, where $d = \bar{X}_t - \bar{X}_c$
  - $E(d^2) = V(d) \propto 1/n$ (note: $E(d) = 0$)
  - Random pruning $\rightsquigarrow$ n declines $\rightsquigarrow$ $E(d^2)$ increases
  - $\Longrightarrow$ random pruning increases imbalance
- Result is completely general (see math in the paper)

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

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 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
   - 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 $\rightsquigarrow$ Ineffecincy $\rightsquigarrow$ 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

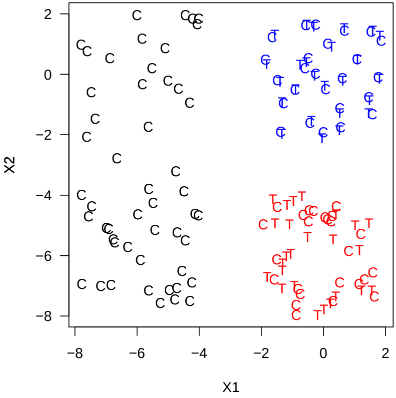
1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The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

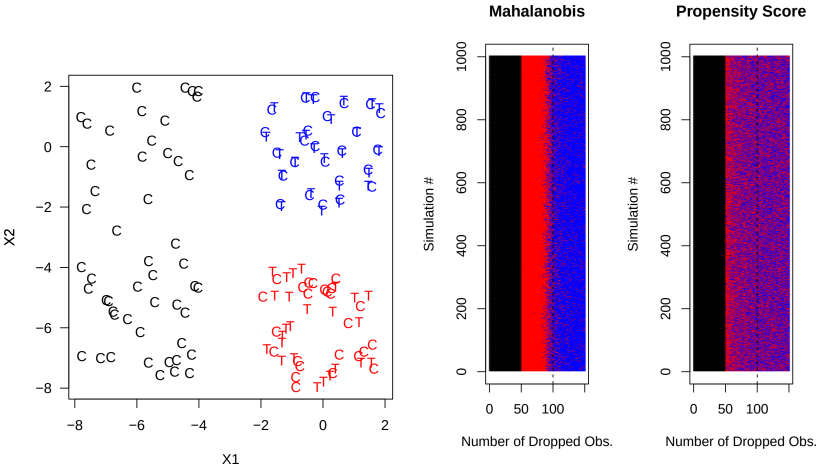
1. Low Standards: Sometimes helps, never optimizes
   - *Efficient* relative to complete randomization, but
   - *Inefficient* relative to (the more powerful) full blocking
   - Other methods dominate:
     $X_c = X_t \implies \pi_c = \pi_t$ but
     $\pi_c = \pi_t \not\Longrightarrow X_c = X_t$
2. The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

PSM is Blind Where Other Methods Can See

PSM is Blind Where Other Methods Can See

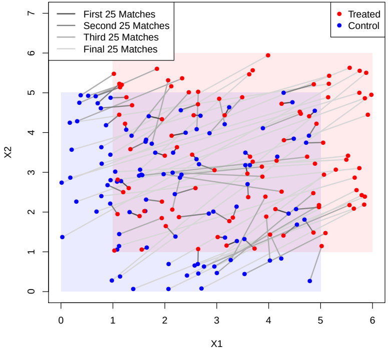


PSM is Blind Where Other Methods Can See

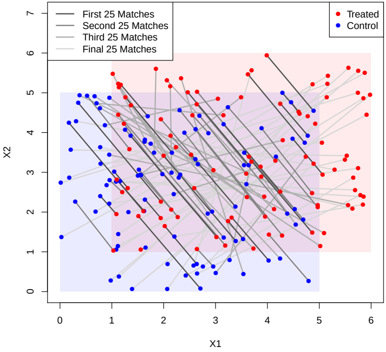


What Does PSM Match?

MDM Matches

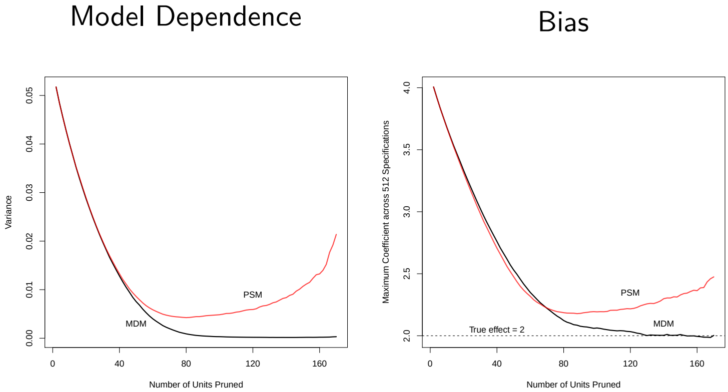


PSM Matches



$$\text{Controls: } X_1, X_2 \sim \text{Uniform}(0{,}5)$$
$$\text{Treateds: } X_1, X_2 \sim \text{Uniform}(1{,}6)$$

PSM Increases Model Dependence & Bias

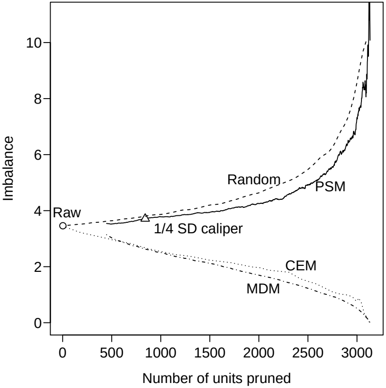


$$Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i$$
$$\epsilon_i \sim N(0,1)$$

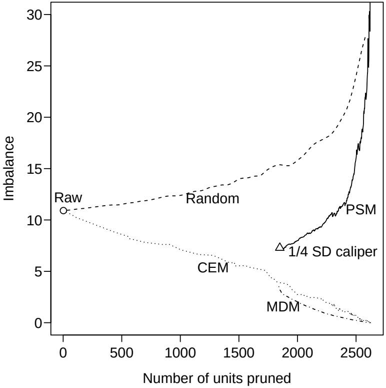
The Propensity Score Paradox in Real Data

The Propensity Score Paradox in Real Data

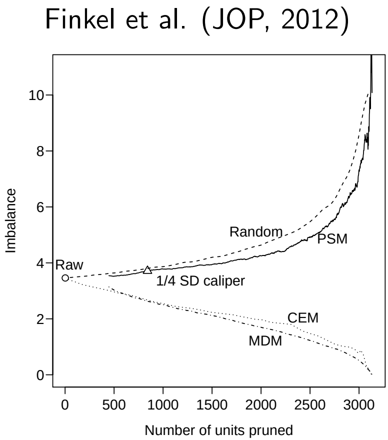
Finkel et al. (JOP, 2012)

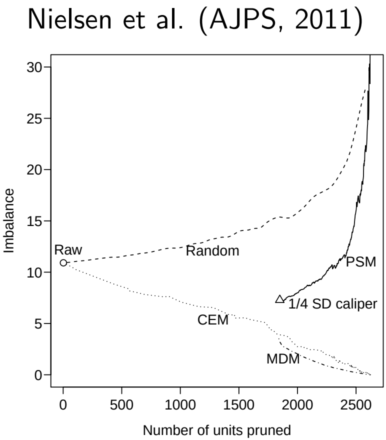


Nielsen et al. (AJPS, 2011)



The Propensity Score Paradox in Real Data





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

The Matching Frontier

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier $=$ matched dataset with lowest imbalance for each n

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier $=$ matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier $=$ matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - All solutions are optimal

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - All solutions are optimal
  - No iteration or diagnostics required

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- To use, make 3 choices:
  1. Imbalance metric, e.g.:
     - Average Mahalanobis Distance (average distance from each unit to the closest in the other treatment regime)
     - Difference of multivariate histograms (L_1):
  2. Quantity of interest: SATT (prune Cs only) or FSATT
  3. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - All solutions are optimal
  - No iteration or diagnostics required
  - No cherry picking possible

How hard is the frontier to calculate?

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal
  - do not require evaluating every subset

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets

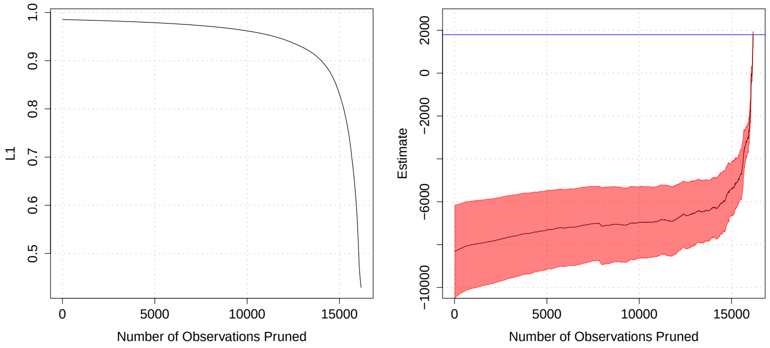
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for <u>all</u> $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for <u>each</u> sample size $n = N, N-1, \ldots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)

How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for all $\binom{N}{n}$ subsets of rows of X_0
  - Choose subset with lowest imbalance
- Evaluations needed to compute the entire frontier:
  - $\binom{N}{n}$ evaluations for each sample size $n = N, N-1, \dots, 1$
  - The combination is the (gargantuan) "power set"
  - e.g., $N > 300$ requires more imbalance evaluations than elementary particles in the universe
  - $\rightsquigarrow$ It's **hard** to calculate!
- We develop algorithms for the (optimal) frontier which:
  - runs very fast
  - operate as "greedy" but we prove are optimal
  - do not require evaluating every subset
  - work with very large data sets
  - is the exact frontier (no approximation or estimation)
  - $\rightsquigarrow$ It's **easy** to calculate!

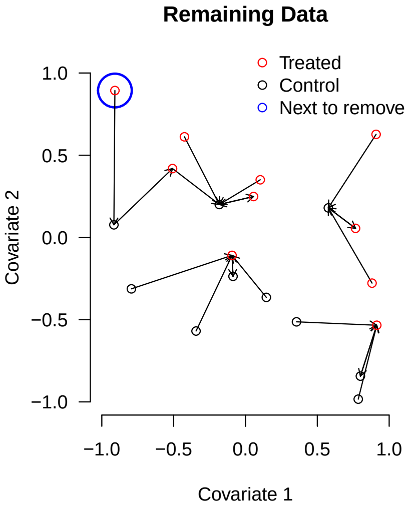
Job Training Data: Frontier and Causal Estimates

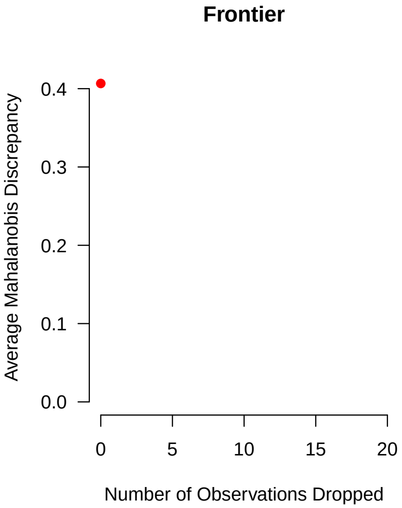


- 185 Ts; pruning most 16,252 Cs won't increase variance much
- Huge bias-variance trade-off after pruning most Cs
- Estimates converge to experiment after removing bias
- No mysteries: basis of inference clearly revealed

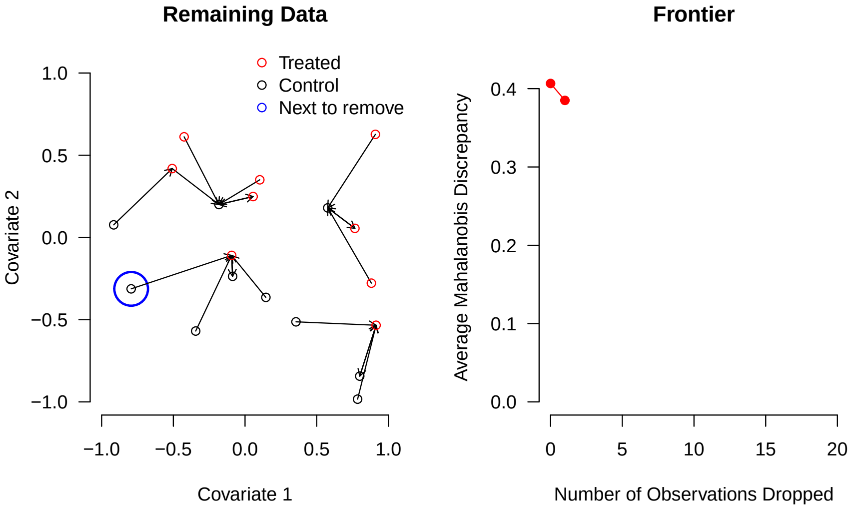
Constructing the FSATT Mahalanobis Frontier

Constructing the FSATT Mahalanobis Frontier

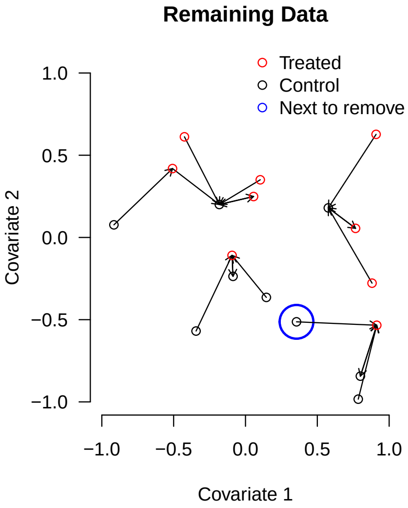


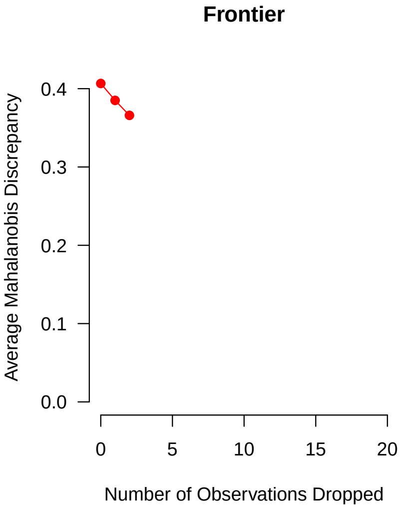


Constructing the FSATT Mahalanobis Frontier

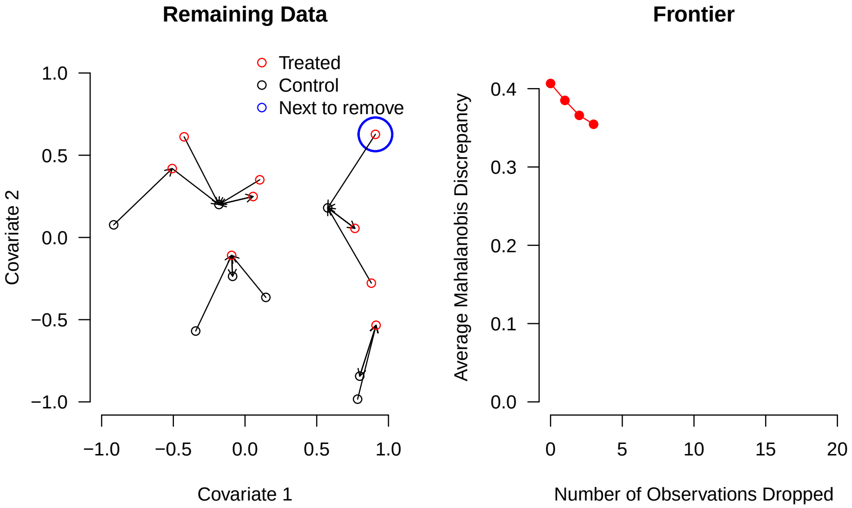


Constructing the FSATT Mahalanobis Frontier

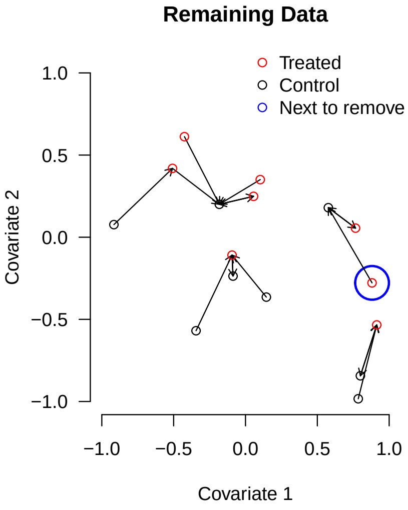


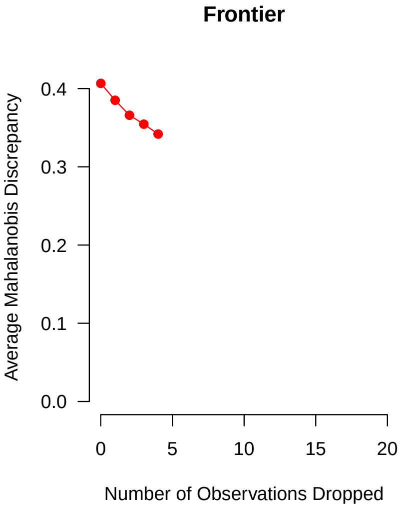


Constructing the FSATT Mahalanobis Frontier

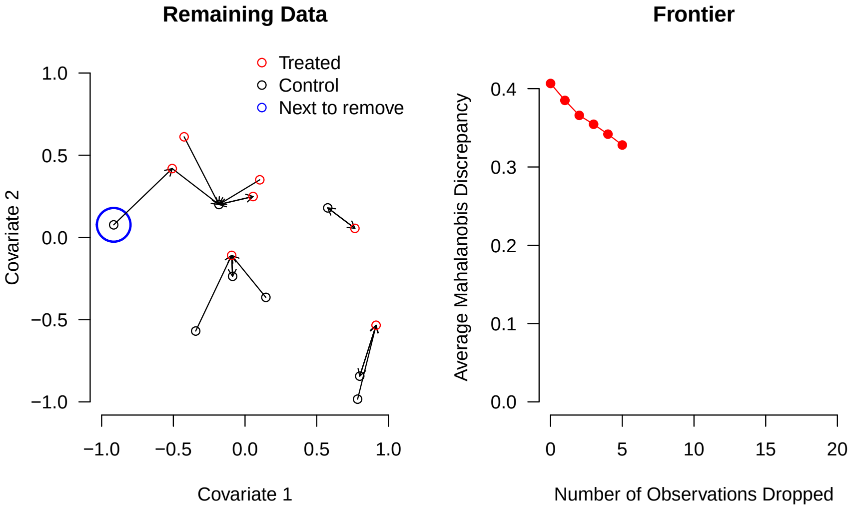


Constructing the FSATT Mahalanobis Frontier

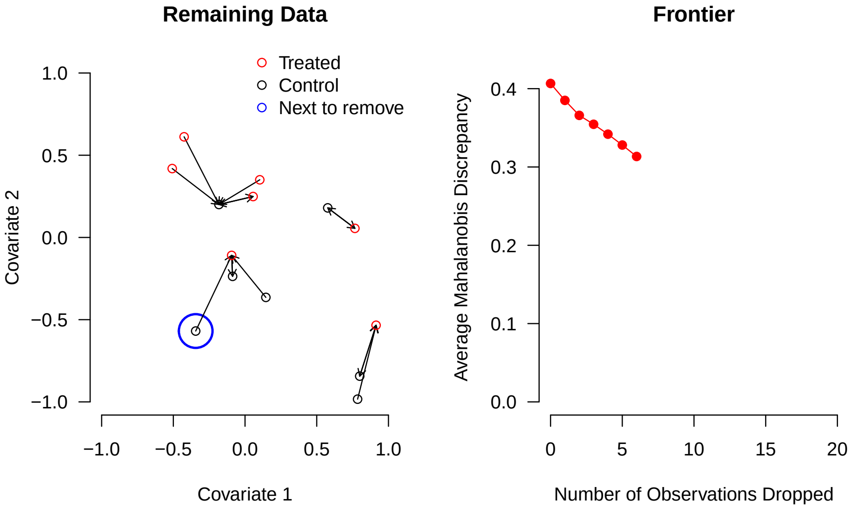




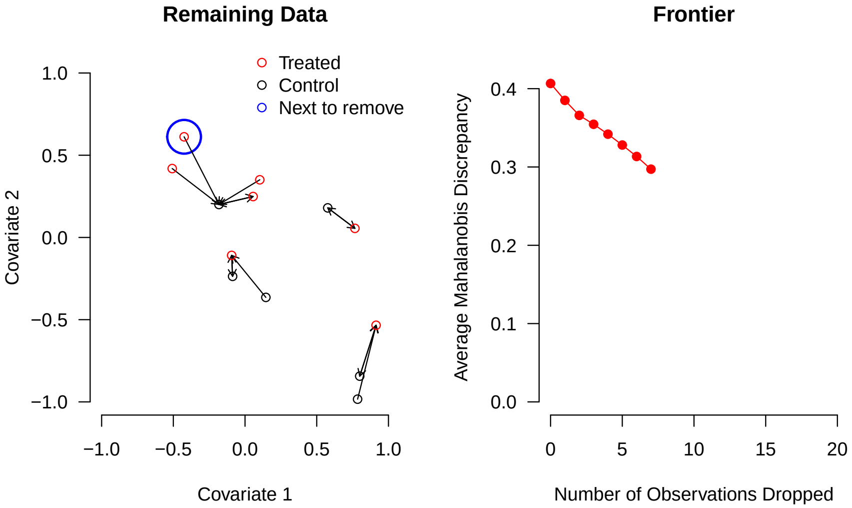
Constructing the FSATT Mahalanobis Frontier



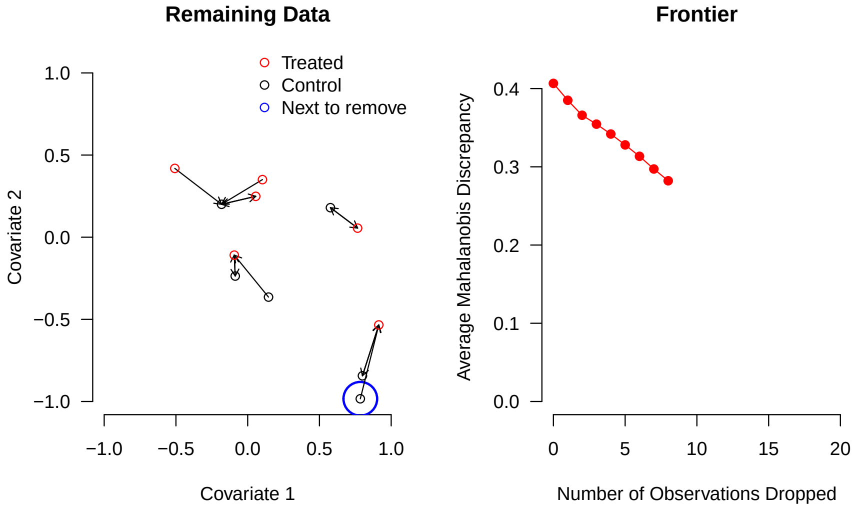
Constructing the FSATT Mahalanobis Frontier



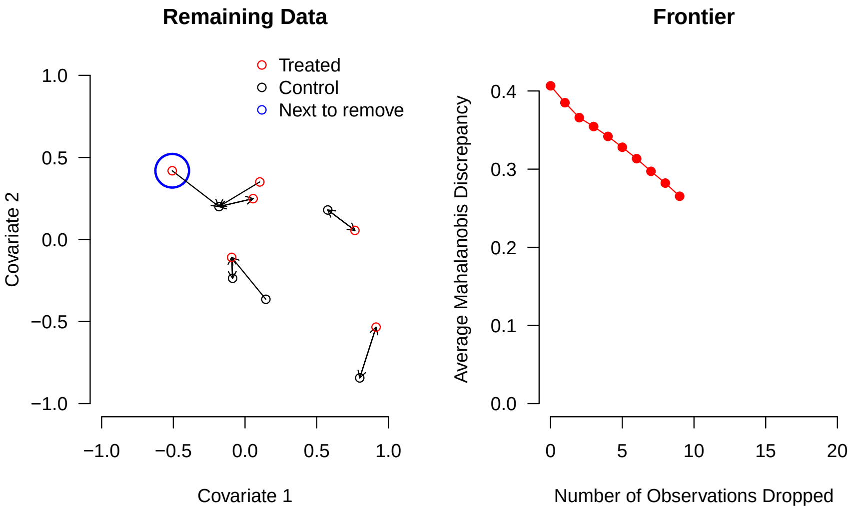
Constructing the FSATT Mahalanobis Frontier



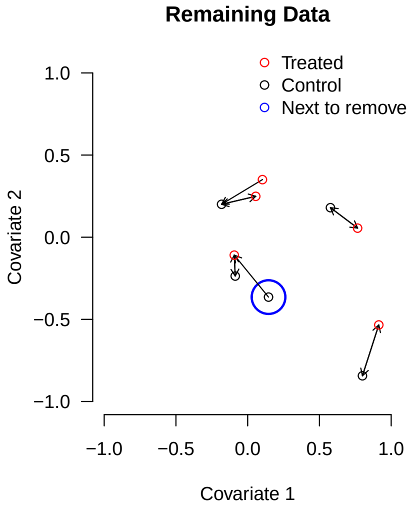
Constructing the FSATT Mahalanobis Frontier

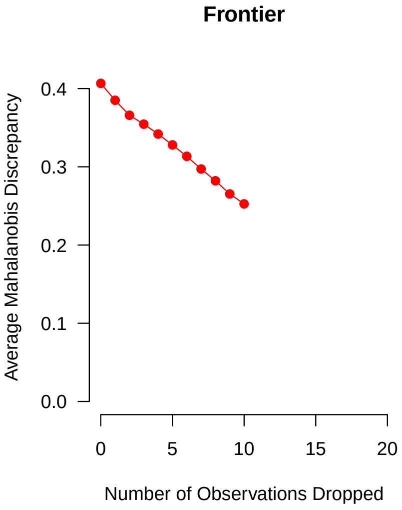


Constructing the FSATT Mahalanobis Frontier

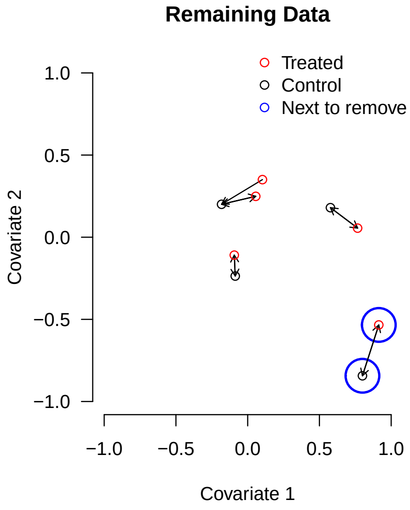


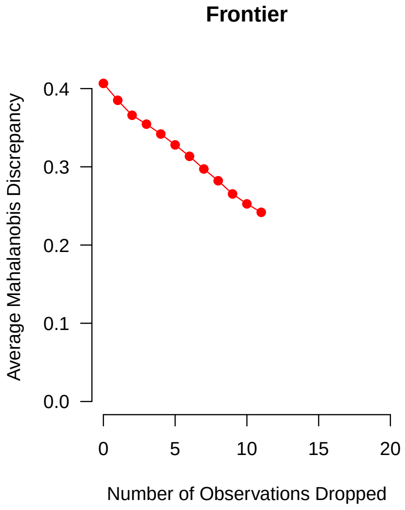
Constructing the FSATT Mahalanobis Frontier



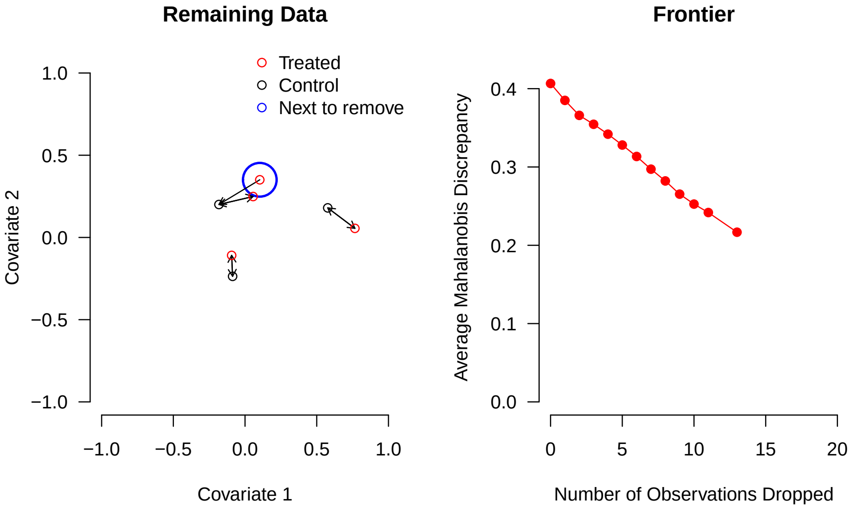


Constructing the FSATT Mahalanobis Frontier

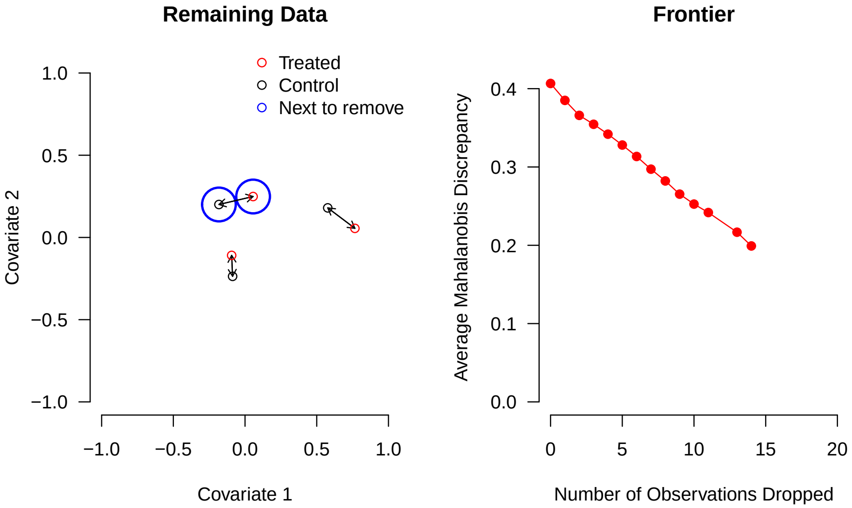




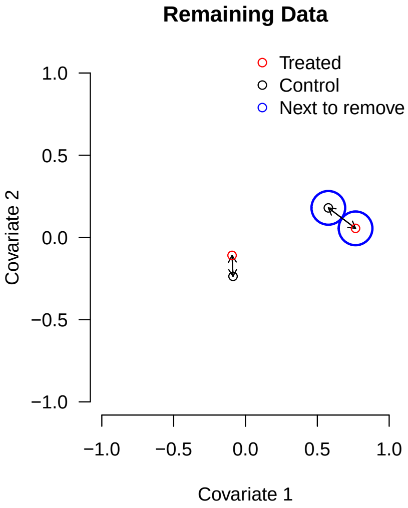
Constructing the FSATT Mahalanobis Frontier

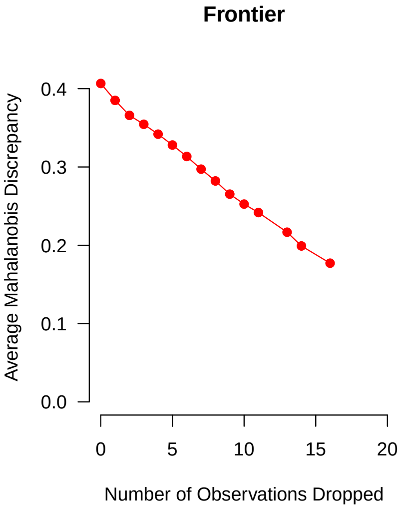


Constructing the FSATT Mahalanobis Frontier

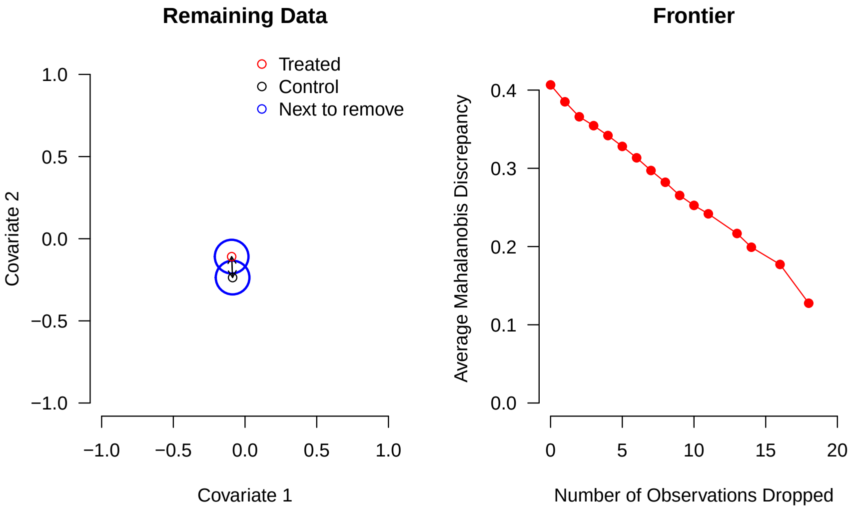


Constructing the FSATT Mahalanobis Frontier

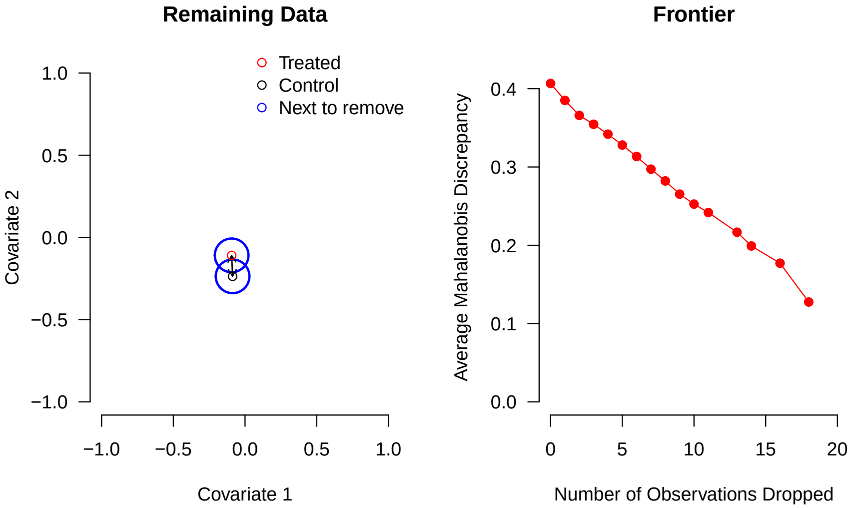




Constructing the FSATT Mahalanobis Frontier

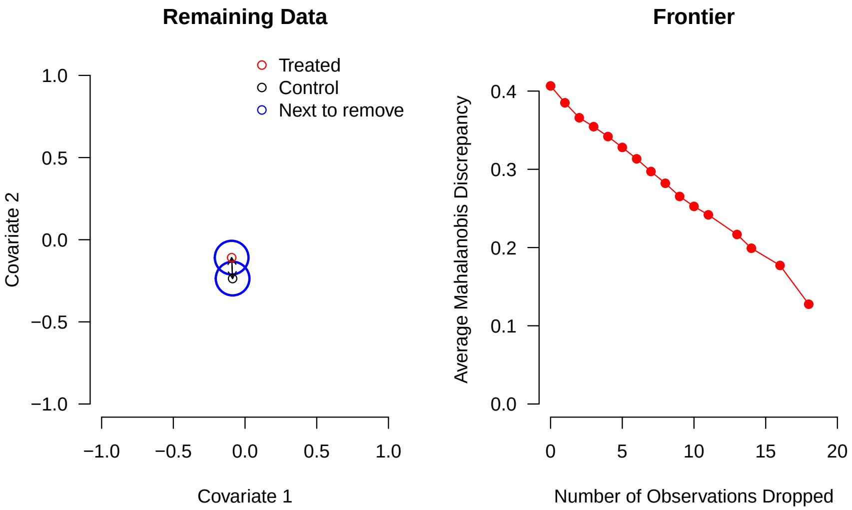


Constructing the FSATT Mahalanobis Frontier



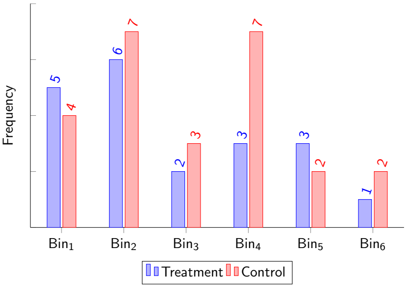
- Warning: figure omits some details!

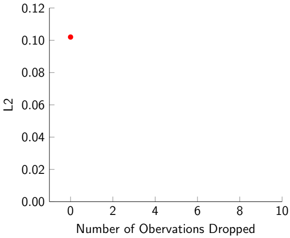
Constructing the FSATT Mahalanobis Frontier



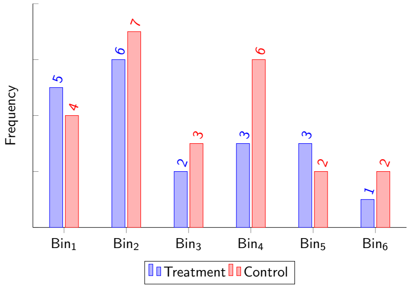
- Warning: figure omits some details!
- Very fast; works with any continuous imbalance metric

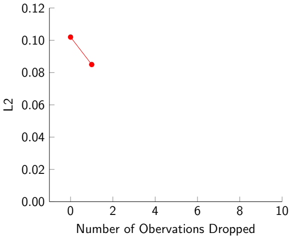
Constructing the L1/L2 SATT Frontier



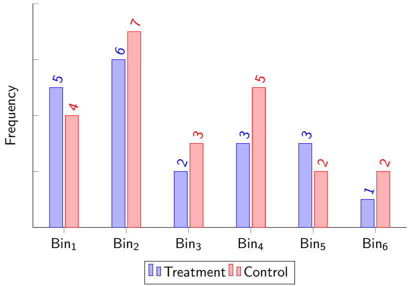


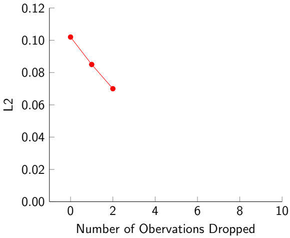
Constructing the L1/L2 SATT Frontier



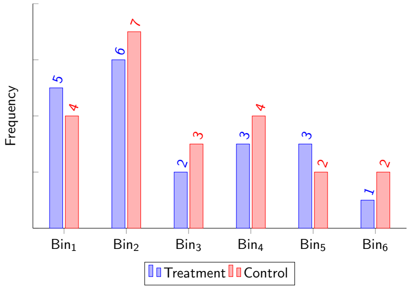


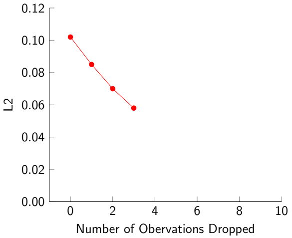
Constructing the L1/L2 SATT Frontier



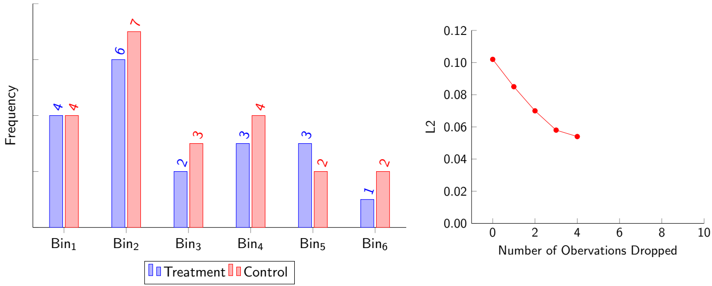


Constructing the L1/L2 SATT Frontier

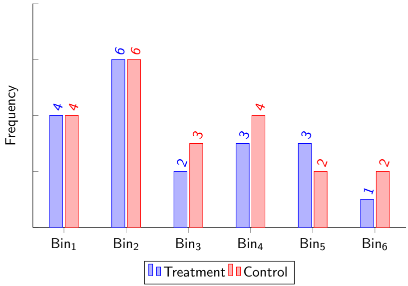


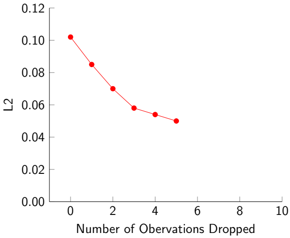


Constructing the L1/L2 SATT Frontier

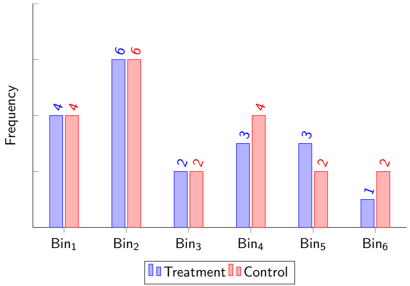


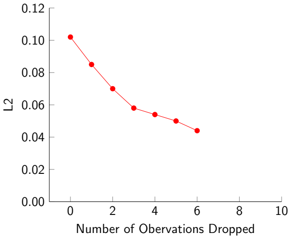
Constructing the L1/L2 SATT Frontier



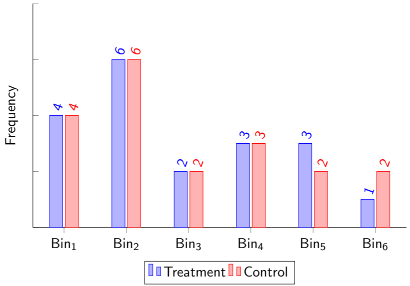


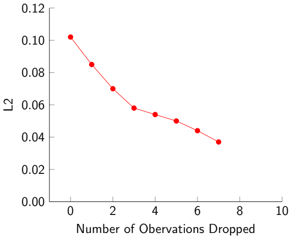
Constructing the L1/L2 SATT Frontier



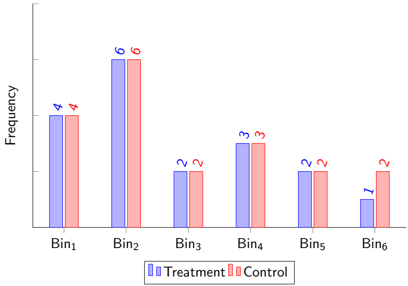


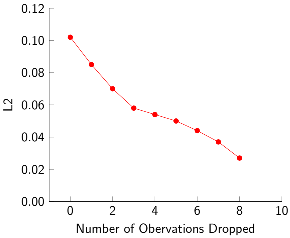
Constructing the L1/L2 SATT Frontier



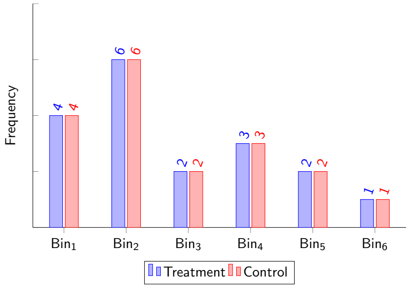


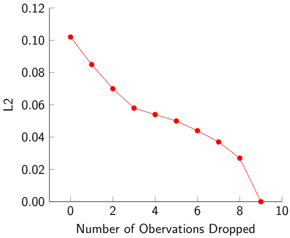
Constructing the L1/L2 SATT Frontier





Constructing the L1/L2 SATT Frontier





Constructing the L1/L2 SATT Frontier

Constructing the L1/L2 SATT Frontier

- Warning: This figure omits some technical details too!

Constructing the L1/L2 SATT Frontier

- Warning: This figure omits some technical details too!
- Works very fast, even with very large data sets

Formal Notation and Assumptions for Causal Inference

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ is the (potential) value of outcome variable if $T_i = t$ (for $t = 0, 1$)

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ is the (potential) value of outcome variable if $T_i = t$ (for $t = 0, 1$)
- Treatment effect: $\mathrm{TE}_i = Y_i(1) - Y_i(0)$

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ is the (potential) value of outcome variable if $T_i = t$ (for $t = 0, 1$)
- Treatment effect: $\text{TE}_i = Y_i(1) - Y_i(0)$
- Fundamental problem of causal inference: $Y_i(0)$ and $Y_i(1)$ are never both observed for any i

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ is the (potential) value of outcome variable if $T_i = t$ (for $t = 0, 1$)
- Treatment effect: $\mathrm{TE}_i = Y_i(1) - Y_i(0)$
- Fundamental problem of causal inference: $Y_i(0)$ and $Y_i(1)$ are never both observed for any i
- Observed outcome: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \ldots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ is the (potential) value of outcome variable if $T_i = t$ (for $t = 0, 1$)
- Treatment effect: $\mathrm{TE}_i = Y_i(1) - Y_i(0)$
- Fundamental problem of causal inference: $Y_i(0)$ and $Y_i(1)$ are never both observed for any i
- Observed outcome: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$
- Pretreatment control variables: X_i (k-vector)

Formal Notation and Assumptions for Causal Inference

- Units: $i = 1, \dots, n$
- Treatment variable: $T_i = 1$ for treateds; 0 for controls
- Potential outcomes: $Y_i(t)$ is the (potential) value of outcome variable if $T_i = t$ (for $t = 0, 1$)
- Treatment effect: $\mathrm{TE}_i = Y_i(1) - Y_i(0)$
- Fundamental problem of causal inference: $Y_i(0)$ and $Y_i(1)$ are never both observed for any i
- Observed outcome: $Y_i = T_i Y_i(1) + (1 - T_i) Y_i(0)$
- Pretreatment control variables: X_i (k-vector)
- Simplification we will use: focus on TE for treated units; so $Y_i(0)$ is unobserved; $Y_i(1) = Y_i$

Implied Notation Assumptions

Implied Notation Assumptions

- Overlap ("common support"): $\Pr(T_i = 1|X) < 1 \ \forall$ treated i

Implied Notation Assumptions

- Overlap ("common support"): $\Pr(T_i = 1|X) < 1 \; \forall$ treated i
- If no overlap: TE_i does not logically exist

Implied Notation Assumptions

- Overlap ("common support"): $\Pr(T_i = 1|X) < 1 \;\forall$ treated i
- If no overlap: TE_i does not logically exist
- Stable Unit Treatment Value (SUTVA): logical consistency, so potential values are fixed if T changes (or "no interference"; no "versions of treatment").

Implied Notation Assumptions

- Overlap ("common support"): $\Pr(T_i = 1|X) < 1 \ \forall$ treated i
- If no overlap: TE_i does not logically exist
- Stable Unit Treatment Value (SUTVA): logical consistency, so potential values are fixed if T changes (or "no interference"; no "versions of treatment").
- Example of no SUTA: $Y_i(0) = 5$ if $T_i = 1$ but $Y_i(0) = 8$ if it were the case that $T_i = 0$

Implied Notation Assumptions

- Overlap ("common support"): $\Pr(T_i = 1|X) < 1 \; \forall$ treated i
- If no overlap: TE_i does not logically exist
- Stable Unit Treatment Value (SUTVA): logical consistency, so potential values are fixed if T changes (or "no interference"; no "versions of treatment").
- Example of no SUTA: $Y_i(0) = 5$ if $T_i = 1$ but $Y_i(0) = 8$ if it were the case that $T_i = 0$
- If overlap, but no SUTVA: $Y_i(0)$ and TE_i exist but are not fixed QOIs

Implied Notation Assumptions

- Overlap ("common support"): $\Pr(T_i = 1|X) < 1 \ \forall$ treated i
- If no overlap: TE_i does not logically exist
- Stable Unit Treatment Value (SUTVA): logical consistency, so potential values are fixed if T changes (or "no interference"; no "versions of treatment").
- Example of no SUTA: $Y_i(0) = 5$ if $T_i = 1$ but $Y_i(0) = 8$ if it were the case that $T_i = 0$
- If overlap, but no SUTVA: $Y_i(0)$ and TE_i exist but are not fixed QOIs
- Example of SUTVA violation: use entire data set to define Y or T by cluster analysis. Then when T changes from 0 to 1, the values of Y may change, but the meaning of the categories will as well.

Causal Quantities of Interest

Causal Quantities of Interest

- Sample Average Treatment Effect on the Treated

$$\text{SATT} = \text{mean}_{i \in \{T=1\}}(\text{TE}_i)$$

Causal Quantities of Interest

- Sample Average Treatment Effect on the Treated

$$\text{SATT} = \text{mean}_{i \in \{T=1\}}(\text{TE}_i)$$

- Feasible SATT

FSATT: TE_i averaged over well-matched treated units

Causal Quantities of Interest

- Sample Average Treatment Effect on the Treated

$$\text{SATT} = \text{mean}_{i \in \{T=1\}}(\text{TE}_i)$$

- Feasible SATT

FSATT: TE_i averaged over well-matched treated units

- Other QOIs: PATT, FPATT; SATE, FSATE; PATE, FPATE

Causal Estimation Assumptions

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
  - Goal: Identification (we can learn the truth if $n \to \infty$)

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0) | X$ for all treateds

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0) | X$ for all treateds
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0) | X$ for all treateds
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)
  - Goal: $\widehat{\text{SATT}} = \text{SATT}$

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0) | X$ for all treateds
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)
  - Goal: $\widehat{\text{SATT}} = \text{SATT}$
  - Sufficient (not necess.): $(Y_i(0) | T_i = 1, X) = (Y_j(0) | T_j = 0, X)$ $\forall$ treateds i & matching controls j

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0) | X$ for all treateds
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)
  - Goal: $\widehat{\text{SATT}} = \text{SATT}$
  - Sufficient (not necess.): $(Y_i(0)|T_i = 1, X) = (Y_j(0)|T_j = 0, X)$ $\forall$ treateds i & matching controls j
  - Necessary: $\text{mean}(Y_i(0)|T = 1, X) = \text{mean}(Y_j(0)|T = 0, X)$ (average within strata of X are equal)

Causal Estimation Assumptions

- Ignorable Treatment Assignment (ITA)
  - aka: "selection on observables," "no selection on Y," "unconfoundedness," "conditional independence"; special cases: "exogeneity," "no omitted variable bias"
  - Goal: Identification (we can learn the truth if $n \to \infty$)
  - The mechanism that produces treatment assignment (T_i) is independent of the potential outcomes
  - Formally: $T_i \perp Y_i(0)|X$ for all treateds
- Uncorrelated Treatment Assignment (UTA) (ITA Simplified)
  - Goal: $\widehat{\text{SATT}} = \text{SATT}$
  - Sufficient (not necess.): $(Y_i(0)|T_i = 1, X) = (Y_j(0)|T_j = 0, X)$ $\forall$ treateds i & matching controls j
  - Necessary: $\text{mean}(Y_i(0)|T = 1, X) = \text{mean}(Y_j(0)|T = 0, X)$ (average within strata of X are equal)
  - Simpler special case under 1-to-1 matching on X: $\text{mean}(Y_i(0)|T_i = 1) = \text{mean}(Y_j|T_j = 0)$

Assumptions to Justify Current Practice

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
    - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
    - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints, like CEM)

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints, like CEM)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints, like CEM)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints, like CEM)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints, like CEM)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$
  - *Set-wide Common support*: $\Pr(T = 1|A) < 1$

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T=1|X)<1$ ($T=0,1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints, like CEM)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$
  - *Set-wide Common support*: $\Pr(T=1|A)<1$
- Fits all common matching methods & practices; no asymptotics

Assumptions to Justify Current Practice

Existing Theory of Inference: Stop What You're Doing!

- Framework: simple random sampling from a population
- Exact matching: Rarely possible; but would make estimation easy
- Assumptions:
  - *Unconfoundedness*: $T \perp Y(0) \mid X$ (Healthy & unhealthy get meds)
  - *Common support*: $\Pr(T = 1|X) < 1$ ($T = 0, 1$ are both possible)
- Approximate matching (bias correction, new variance estimation): common, but all current practices would have to change

Alternative Theory of Inference: It's Gonna be OK!

- Framework: stratified random sampling from a population
- Define A: a stratum in a partition of the product space of X ("continuous" variables have natural breakpoints, like CEM)
- We already know and use these procedures: Group strong and weak partisans; Don't match college dropout with 1st year grad student
- Assumptions:
  - *Set-wide Unconfoundedness*: $T \perp Y(0) \mid A$
  - *Set-wide Common support*: $\Pr(T = 1|A) < 1$
- Fits all common matching methods & practices; no asymptotics
- Easy extensions for: multi-level, continuous, & mismeasured treatments; A too wide, n too small