Simplifying Matching Methods for Causal Inference

Gary King[1]

Institute for Quantitative Social Science
Harvard University

(Talk at Princeton University, Center for Statistics and Machine Learning, 2/6/2015)

[1] GaryKing.org

3 Problems, 3 Solutions

3 Problems, 3 Solutions

- Current practice, matching as preprocessing:

3 Problems, 3 Solutions

- Current practice, matching as preprocessing: violates current statistical theory.

3 Problems, 3 Solutions

- Current practice, matching as preprocessing: violates current statistical theory. So let's change the theory:

3 Problems, 3 Solutions

- Current practice, matching as preprocessing: 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)

3 Problems, 3 Solutions

- Current practice, matching as preprocessing: 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 49,600 articles!) sounds magical:

3 Problems, 3 Solutions

- Current practice, matching as preprocessing: 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 49,600 articles!) sounds magical:
  - ⇝ "Why Propensity Scores Should Not Be Used for Matching" (Gary King, Richard Nielsen)

3 Problems, 3 Solutions

- Current practice, matching as preprocessing: 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 49,600 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':

3 Problems, 3 Solutions

- Current practice, matching as preprocessing: 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 49,600 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)

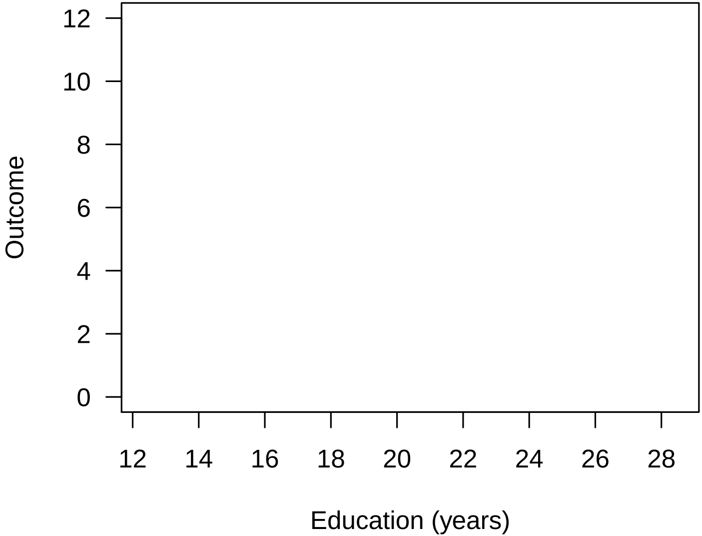
Matching to Reduce Model Dependence

Matching to Reduce Model Dependence

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

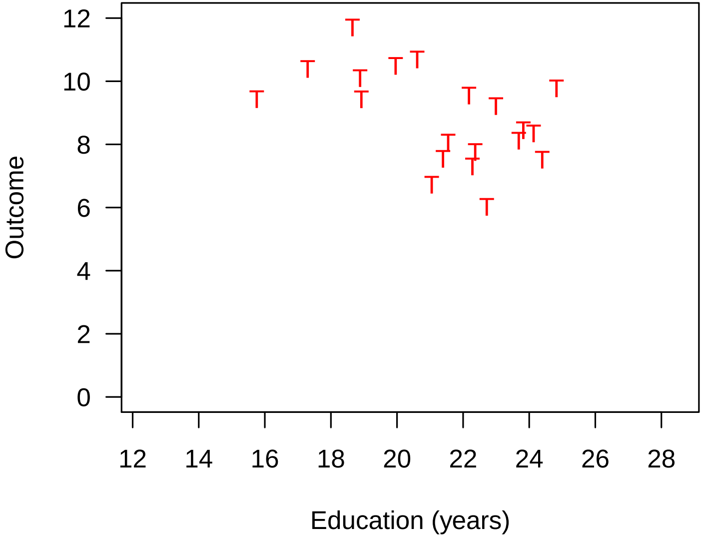
Matching to Reduce Model Dependence

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



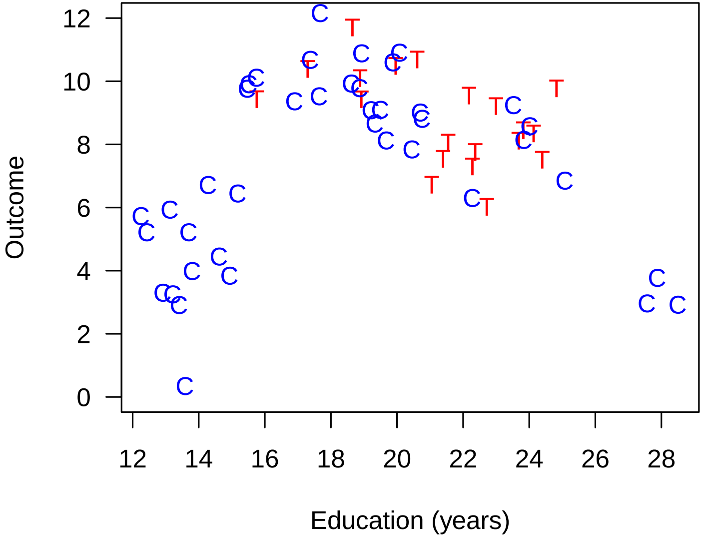
Matching to Reduce Model Dependence

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



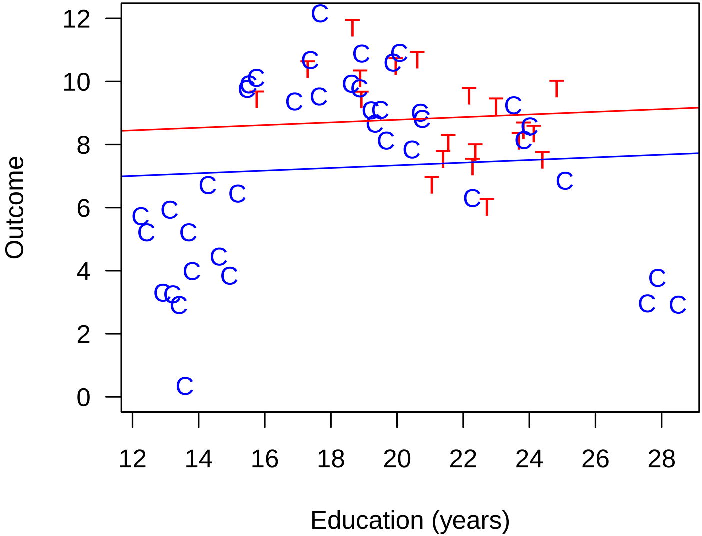
Matching to Reduce Model Dependence

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



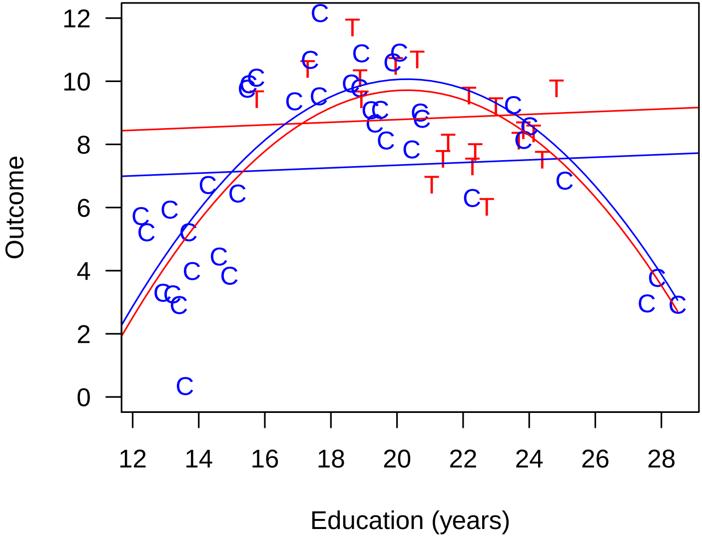
Matching to Reduce Model Dependence

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



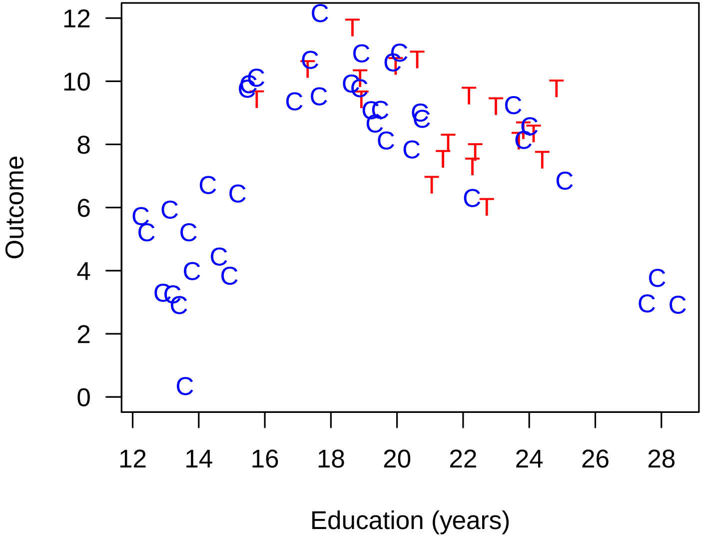
Matching to Reduce Model Dependence

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



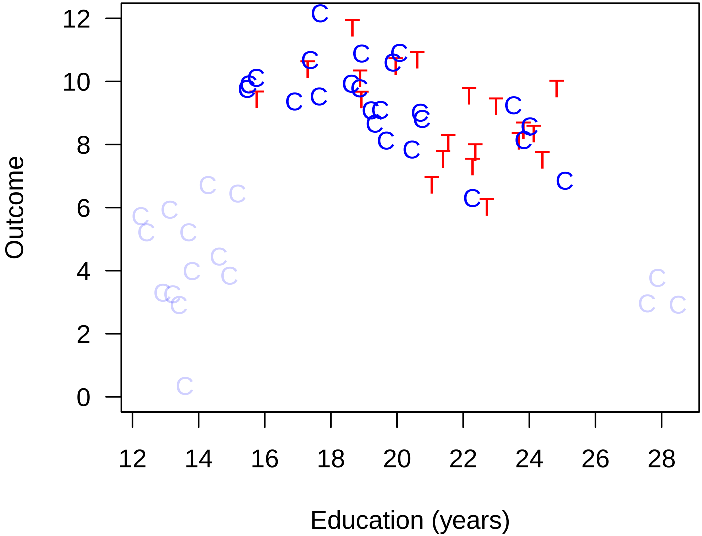
Matching to Reduce Model Dependence

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



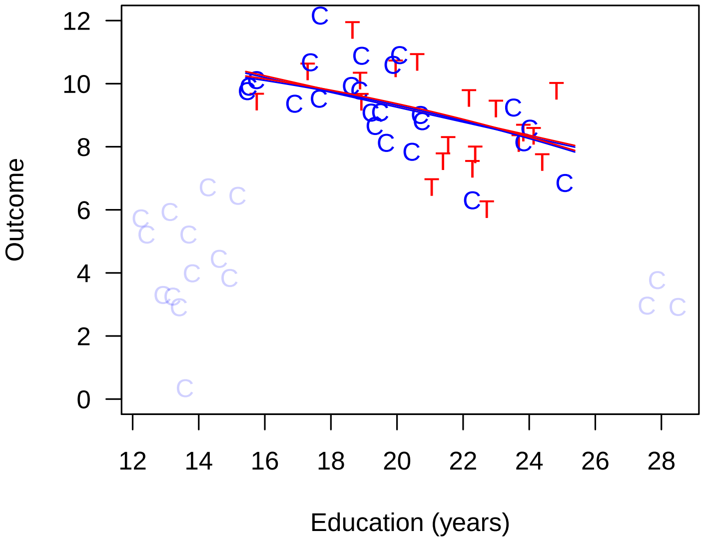
Matching to Reduce Model Dependence

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



Matching to Reduce Model Dependence

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



The Advantage of Matching

Without Matching:

The Advantage of Matching

Without Matching:

Imbalance

The Advantage of Matching

Without Matching:

Imbalance $\rightsquigarrow$ Model Dependence

The Advantage of Matching

Without Matching:

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

The Advantage of Matching

Without Matching:

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

The Advantage of Matching

With~~out~~ Matching:

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

The Advantage of Matching

Without Matching:

~~Im~~balance ⇝ ~~Model Dependence~~ ⇝ Researcher discretion ⇝ Bias

The Advantage of Matching

Without Matching:

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

The Advantage of Matching

Without ~~out~~ Matching:

~~Im~~balance ⇝ ~~Model Dependence~~ ⇝ ~~Researcher discretion~~ ⇝ ~~Bias~~

Current Practice: Matching as Preprocessing

Current Practice: Matching as Preprocessing

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

Current Practice: Matching as Preprocessing

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

Current Practice: Matching as Preprocessing

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

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

Current Practice: Matching as Preprocessing

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

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

Current Practice: Matching as Preprocessing

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

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

Current Practice: Matching as Preprocessing

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

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

- Quantities of Interest:

Current Practice: Matching as Preprocessing

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

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

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

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

Current Practice: Matching as Preprocessing

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

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

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

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

  2. FSATT: Feasible Average Treatment effect on the Treated

Current Practice: Matching as Preprocessing

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

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

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

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

  2. FSATT: Feasible Average Treatment effect on the Treated
- Estimate $Y_i(0)$ with Y_j from matched ($X_i \approx X_j$) control

Current Practice: Matching as Preprocessing

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

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

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

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

  2. FSATT: Feasible Average Treatment effect on the Treated
- Estimate $Y_i(0)$ with Y_j from matched ($X_i \approx X_j$) control
- Prune nonmatches: reduces imbalance & model dependence

Current Practice: Matching as Preprocessing

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

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

- Quantities of Interest:
  1. SATT: Sample Average Treatment effect on the Treated:

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

  2. FSATT: Feasible Average Treatment effect on the Treated
- Estimate $Y_i(0)$ with Y_j from matched ($X_i \approx X_j$) control
- Prune nonmatches: reduces imbalance & model dependence
- Big convenience: Follow preprocessing with whatever statistical method you'd have used without matching

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)

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)
- 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)
- 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)
- 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)
- 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)
- 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)
- 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

Approximating Randomized Experiments

Approximating Randomized Experiments

- Types of experiments:

Approximating Randomized Experiments

- Types of experiments:
    1. Compete Randomization: Treatment assignment by coin flips

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample

Approximating Randomized Experiments

- Types of experiments:
    1. Compete Randomization: Treatment assignment by coin flips
        - $\rightsquigarrow$ Balance on X: only on average
        - $\rightsquigarrow$ Balance on unmeasured vars: only on average
    2. Fully Blocked: Match pairs on X (exactly), then flip coins
        - $\rightsquigarrow$ Balance on X: perfect in sample
        - $\rightsquigarrow$ Balance on unmeasured vars: only on average

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for:

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance,

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - ⇝ Balance on X: only on average
     - ⇝ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - ⇝ Balance on X: perfect in sample
     - ⇝ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence,

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - ⇝ Balance on X: only on average
     - ⇝ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - ⇝ Balance on X: perfect in sample
     - ⇝ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power,

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency,

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - ⇝ Balance on X: only on average
     - ⇝ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - ⇝ Balance on X: perfect in sample
     - ⇝ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias,

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs,

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - ⇝ Balance on X: only on average
     - ⇝ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - ⇝ Balance on X: perfect in sample
     - ⇝ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - ⇝ Balance on X: only on average
     - ⇝ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - ⇝ Balance on X: perfect in sample
     - ⇝ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.
- Matching methods approximate which experiment?

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.
- Matching methods approximate which experiment?
  - PSM: complete randomization

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - ⇝ Balance on X: only on average
     - ⇝ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - ⇝ Balance on X: perfect in sample
     - ⇝ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.
- Matching methods approximate which experiment?
  - PSM: complete randomization
  - Other methods: fully blocked

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.
- Matching methods approximate which experiment?
  - PSM: complete randomization
  - Other methods: fully blocked
- $\Longrightarrow$ As we show, other methods usually dominate PSM

Approximating Randomized Experiments

- Types of experiments:
  1. Compete Randomization: Treatment assignment by coin flips
     - $\rightsquigarrow$ Balance on X: only on average
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
  2. Fully Blocked: Match pairs on X (exactly), then flip coins
     - $\rightsquigarrow$ Balance on X: perfect in sample
     - $\rightsquigarrow$ Balance on unmeasured vars: only on average
- Fully blocked dominates complete randomization for: imbalance, model dependence, power, efficiency, bias, research costs, and robustness.
- Matching methods approximate which experiment?
  - PSM: complete randomization
  - Other methods: fully blocked
- $\Longrightarrow$ As we show, other methods usually dominate PSM (but wait, it gets worse for PSM)

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
3. Checking Measure imbalance, tweak, repeat, . . .

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - $\text{Distance}(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 1: Mahalanobis Distance Matching

(Approximates Fully Blocked Experiment)

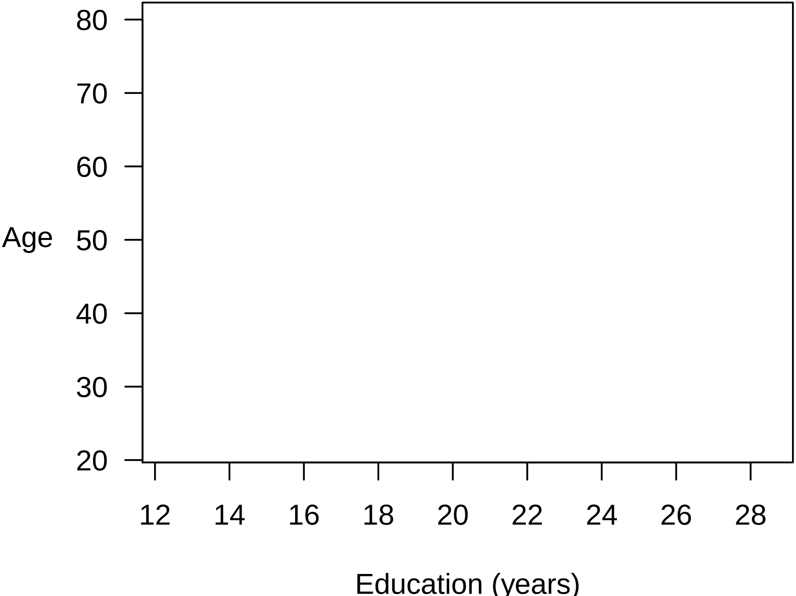
1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - Match each treated unit to the nearest control unit
   - Control units: not reused; pruned if unused
   - Prune matches if Distance$>$*caliper*
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 1: Mahalanobis Distance Matching

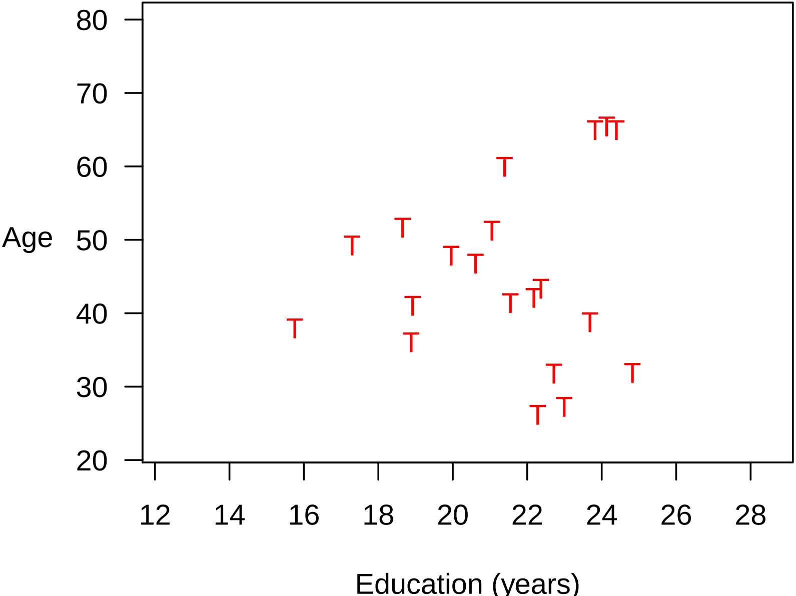
(Approximates Fully Blocked Experiment)

1. Preprocess (Matching)
   - Distance$(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
   - 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
3. Checking Measure imbalance, tweak, repeat, . . .

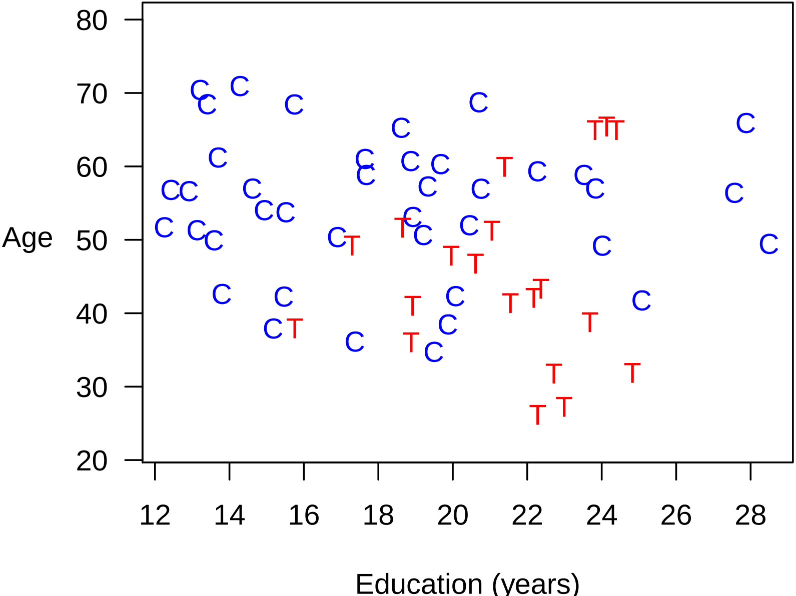
Mahalanobis Distance Matching



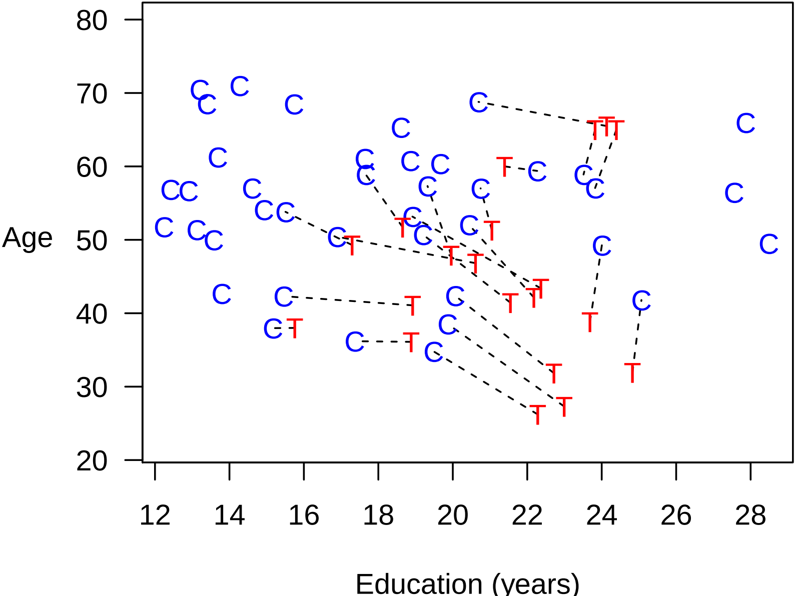
Mahalanobis Distance Matching



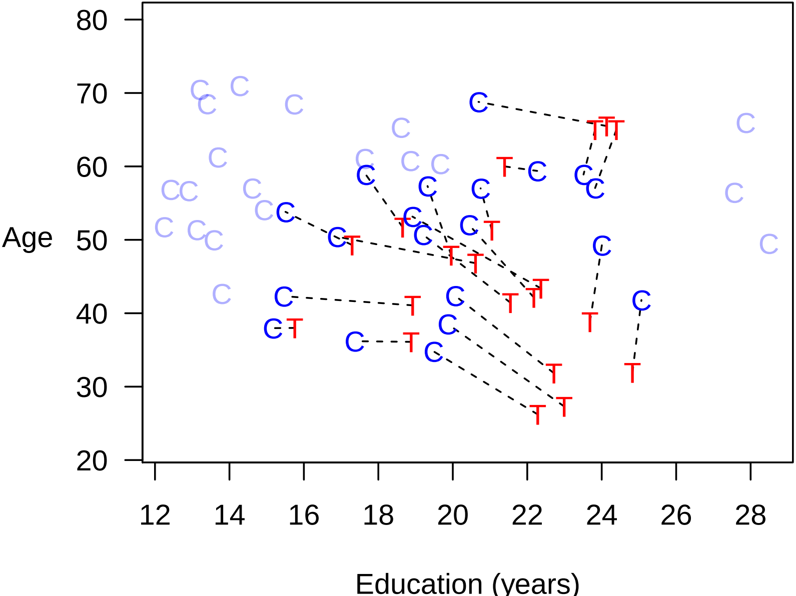
Mahalanobis Distance Matching



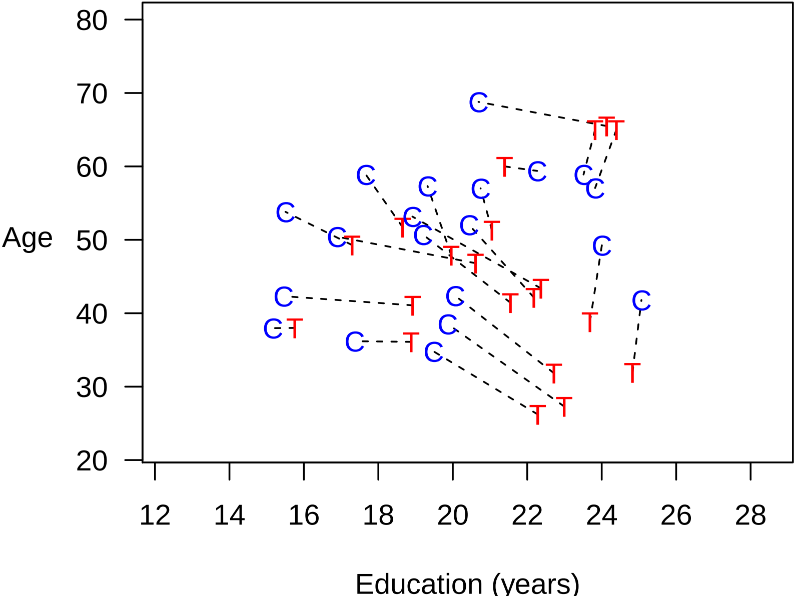
Mahalanobis Distance Matching



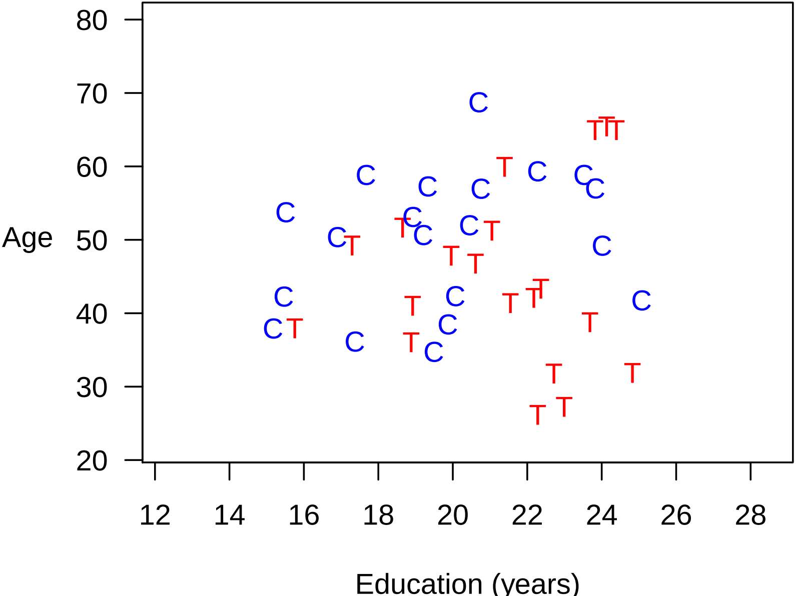
Mahalanobis Distance Matching



Mahalanobis Distance Matching



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

3. Checking Determine matched sample size, tweak, repeat, ...

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

3. Checking Determine matched sample size, tweak, repeat, ...

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
3. Checking Determine matched sample size, tweak, repeat, . . .

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

3. Checking Determine matched sample size, tweak, repeat, ...

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
3. Checking Determine matched sample size, tweak, repeat, ...

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
3. Checking Determine matched sample size, tweak, repeat, . . .

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
3. Checking Determine matched sample size, tweak, repeat, ...

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
   - Need to weight controls in each stratum to equal treateds
3. Checking Determine matched sample size, tweak, repeat, . . .

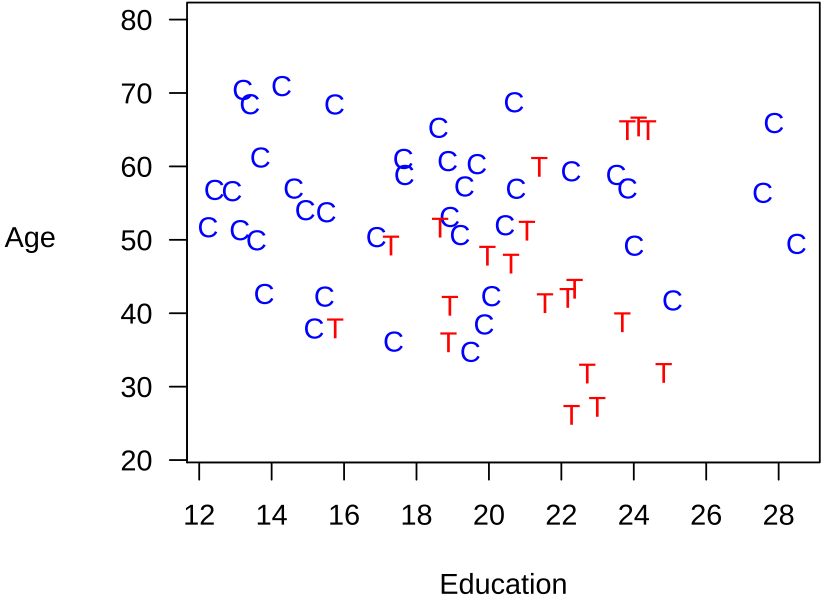
Method 2: Coarsened Exact Matching

(Approximates Fully Blocked Experiment)

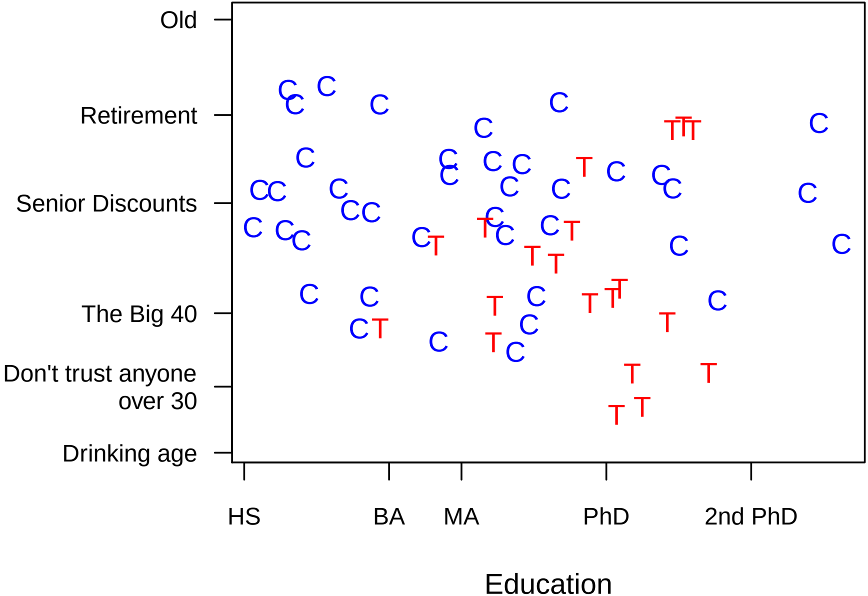
1. Preprocess (Matching)
   - Temporarily coarsen X as much as you're willing
     - e.g., Education (grade school, high school, college, graduate)
   - Apply exact matching to the coarsened X, $C(X)$
     - Sort observations into strata, each with unique values of $C(X)$
     - Prune any stratum with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

Coarsened Exact Matching

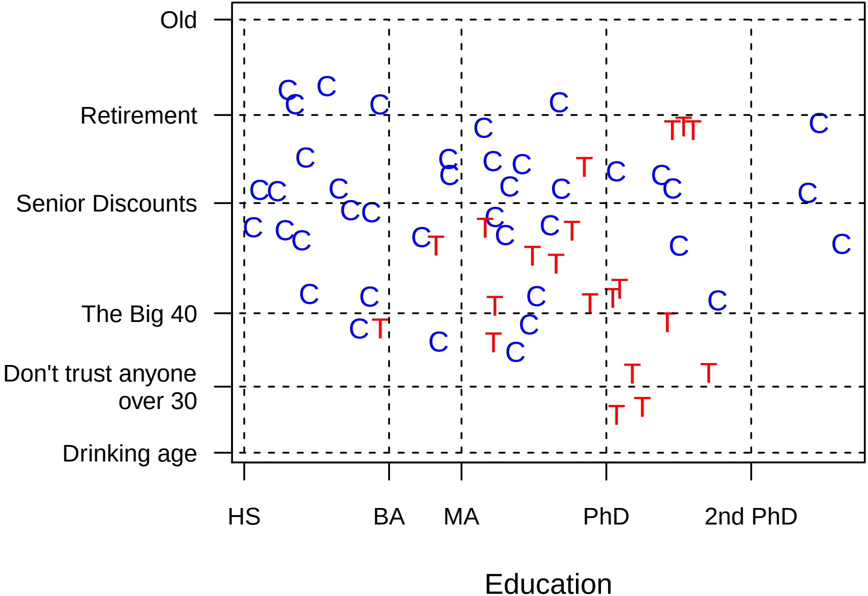
Coarsened Exact Matching



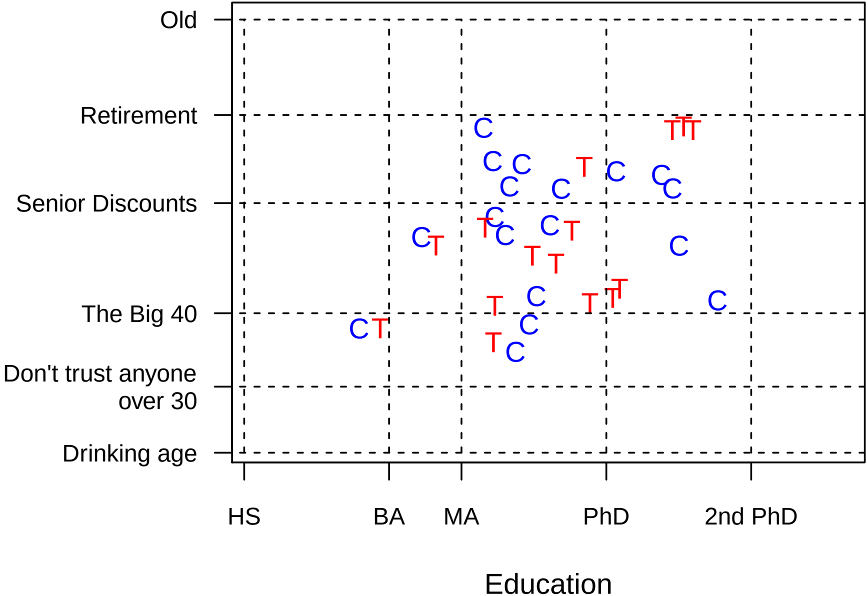
Coarsened Exact Matching



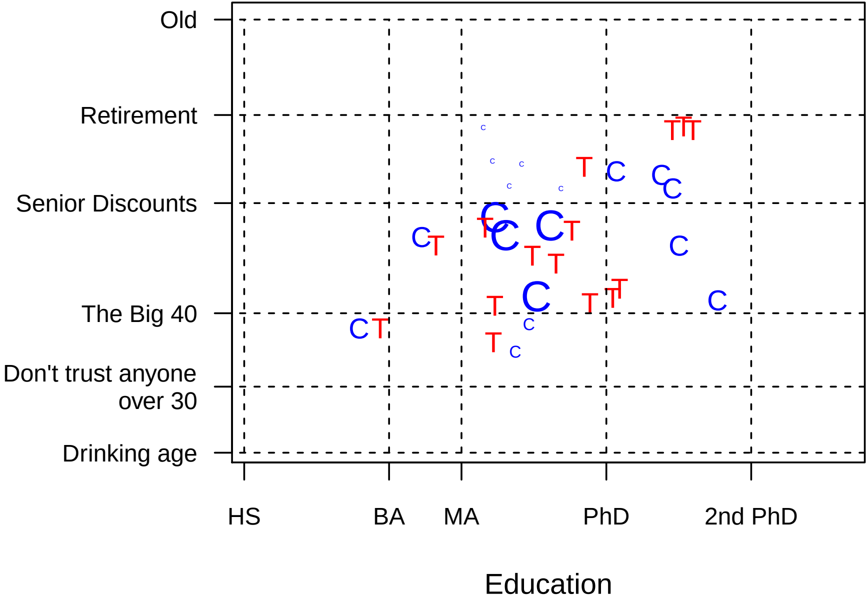
Coarsened Exact Matching



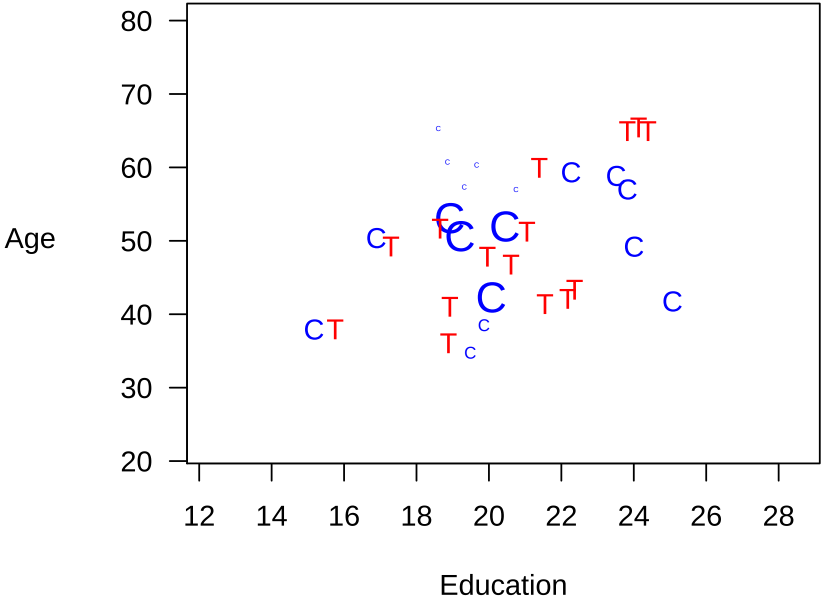
Coarsened Exact Matching



Coarsened Exact Matching



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
3. Checking Measure imbalance, tweak, repeat, . . .

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
3. Checking Measure imbalance, tweak, repeat, ...

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$(X_i, X_j) = |\pi_i - \pi_j|$
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

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

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
    - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
    - $\text{Distance}(X_i, X_j) = |\pi_i - \pi_j|$
    - Match each treated unit to the nearest control unit
    - Control units: not reused; pruned if unused
2. Estimation Difference in means or a model
3. Checking Measure imbalance, tweak, repeat, ...

Method 3: Propensity Score Matching

(Approximates Completely Randomized Experiment)

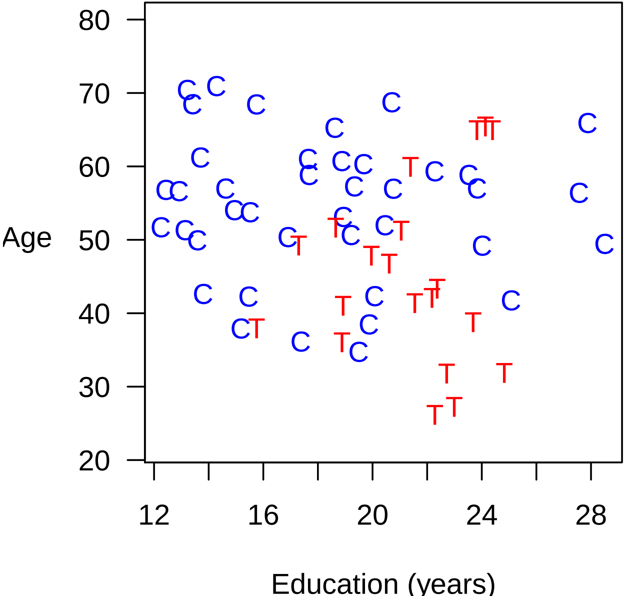
1. Preprocess (Matching)
    - Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
    - $\text{Distance}(X_i, X_j) = |\pi_i - \pi_j|$
    - Match each treated unit to the nearest control unit
    - Control units: not reused; pruned if unused
    - Prune matches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

Method 3: Propensity Score Matching

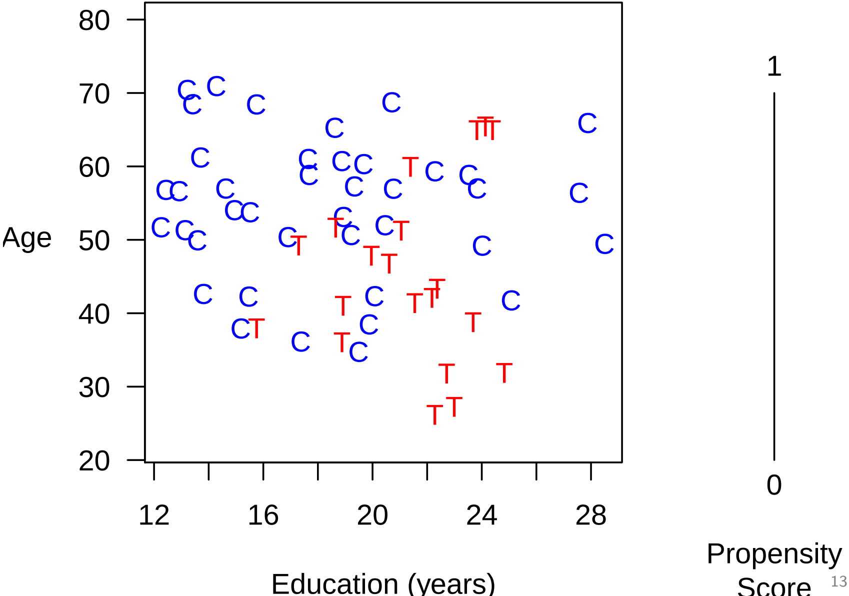
(Approximates Completely Randomized Experiment)

1. Preprocess (Matching)
    - Reduce k elements of X to scalar
      $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
    - $\text{Distance}(X_i, X_j) = |\pi_i - \pi_j|$
    - 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
3. Checking Measure imbalance, tweak, repeat, . . .

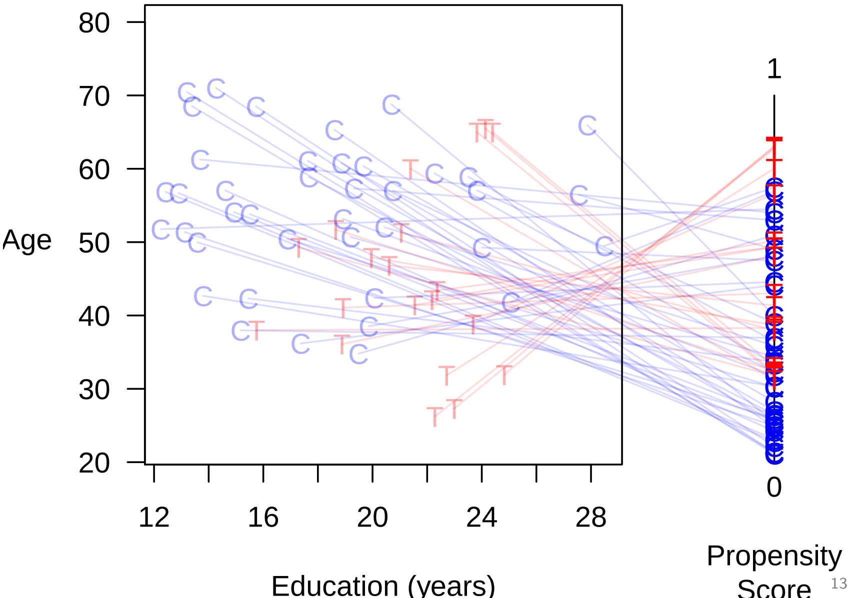
Propensity Score Matching



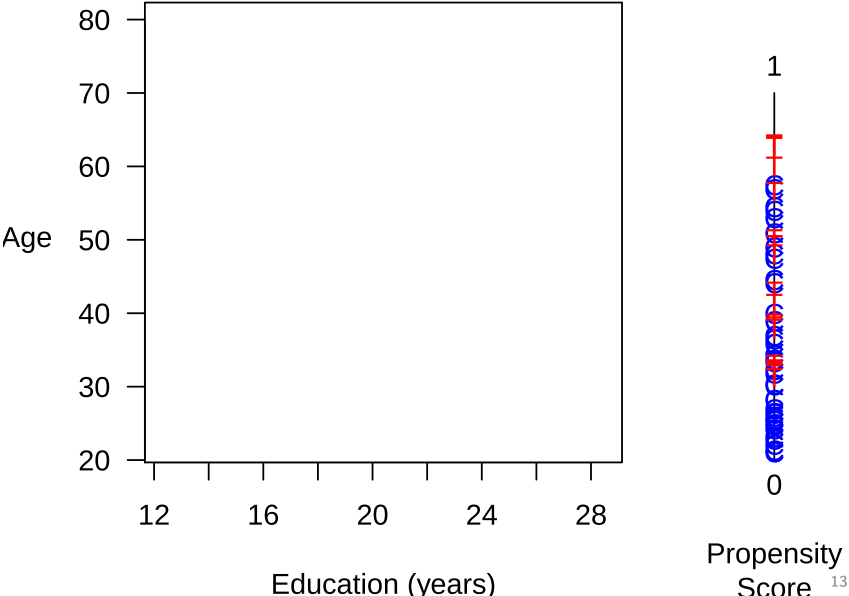
Propensity Score Matching



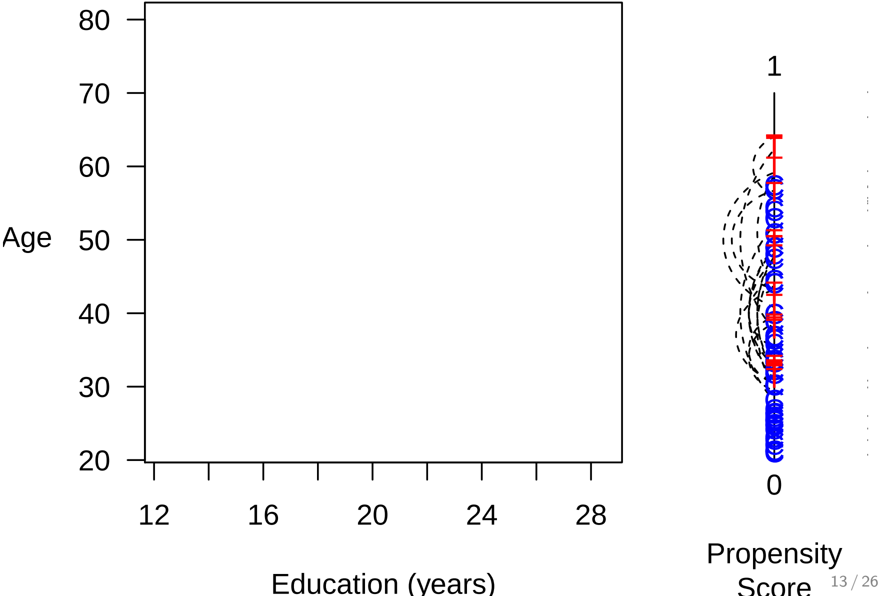
Propensity Score Matching



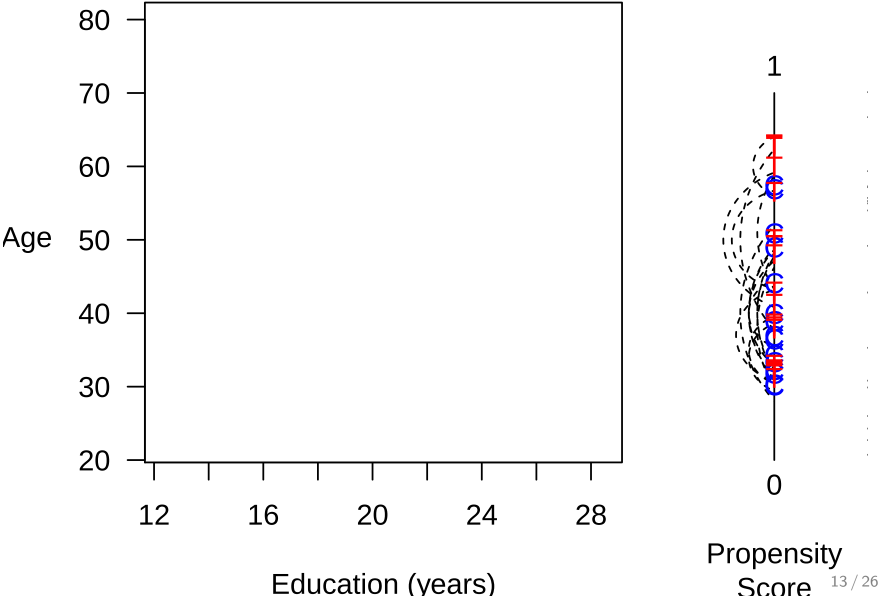
Propensity Score Matching



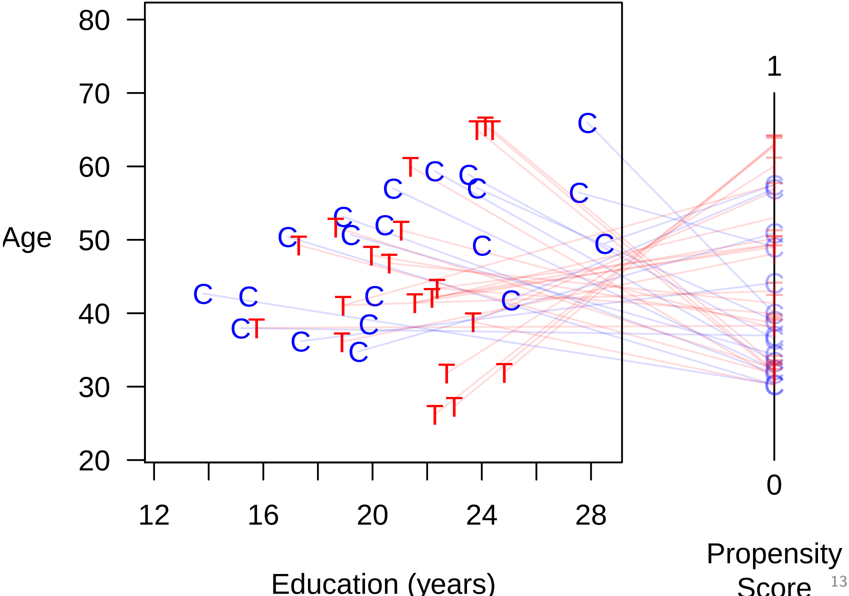
Propensity Score Matching



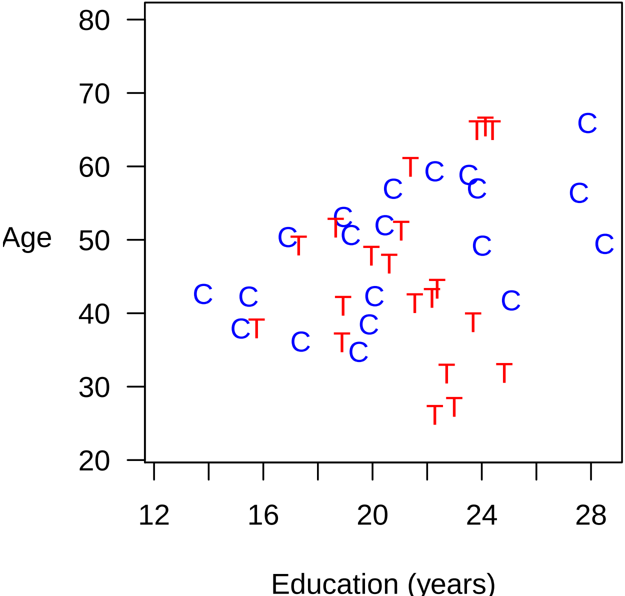
Propensity Score Matching



Propensity Score Matching



Propensity Score Matching



PSM's Statistical Properties

PSM's Statistical Properties

- PSM is Inefficient:

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random
  - Random matching increases imbalance!

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\rightsquigarrow$ higher imbalance $\rightsquigarrow$ more inefficiency

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\rightsquigarrow$ higher imbalance $\rightsquigarrow$ more inefficiency
  - If the data have no good matches, the paradox won't be a problem but you're cooked anyway

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\rightsquigarrow$ higher imbalance $\rightsquigarrow$ more inefficiency
  - If the data have no good matches, the paradox won't be a problem but you're cooked anyway
- PSM is Biased:

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\rightsquigarrow$ higher imbalance $\rightsquigarrow$ more inefficiency
  - If the data have no good matches, the paradox won't be a problem but you're cooked anyway
- PSM is Biased:
  - Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias

PSM's Statistical Properties

- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\rightsquigarrow$ higher imbalance $\rightsquigarrow$ more inefficiency
  - If the data have no good matches, the paradox won't be a problem but you're cooked anyway
- PSM is Biased:
  - Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias
- Curse of Dimensionality Problems:

PSM's Statistical Properties

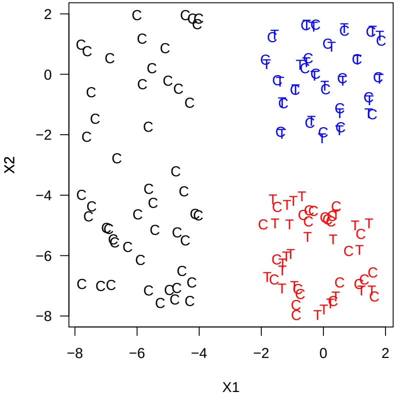
- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\rightsquigarrow$ higher imbalance $\rightsquigarrow$ more inefficiency
  - If the data have no good matches, the paradox won't be a problem but you're cooked anyway
- PSM is Biased:
  - Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias
- Curse of Dimensionality Problems:
  - The Promise: avoid it by balancing on π rather than X

PSM's Statistical Properties

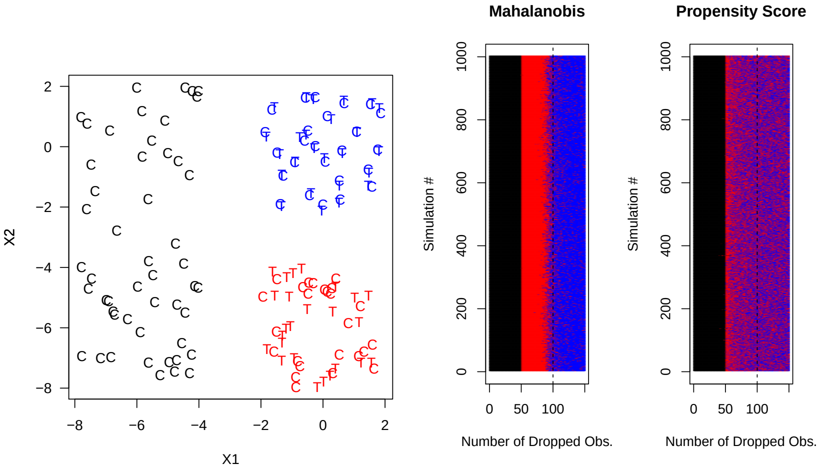
- PSM is Inefficient:
  - Efficient relative to complete randomization, but
  - Inefficient relative to full blocking (Imai, King, and Nall: up to 600% difference in SEs in experiments)
- The PSM Paradox:
  - If data are balanced to begin with, or after some pruning, $\hat{\pi} \approx 0.5$ (or constant within strata) $\rightsquigarrow$ matching is at random
  - Random matching increases imbalance!
  - Approximating complete randomization (by pruning) $\rightsquigarrow$ higher imbalance $\rightsquigarrow$ more inefficiency
  - If the data have no good matches, the paradox won't be a problem but you're cooked anyway
- PSM is Biased:
  - Imbalance $\rightsquigarrow$ Inefficency $\rightsquigarrow$ Model dependence $\rightsquigarrow$ Bias
- Curse of Dimensionality Problems:
  - The Promise: avoid it by balancing on π rather than X
  - The Reality: The PSM Paradox is bigger with more covariates

PSM is Blind Where Other Methods Can See

PSM is Blind Where Other Methods Can See

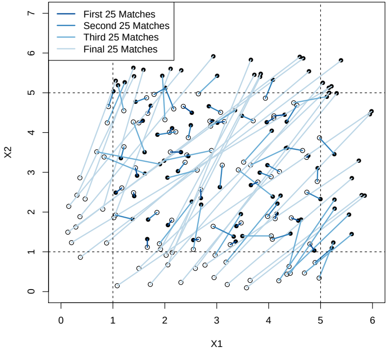


PSM is Blind Where Other Methods Can See

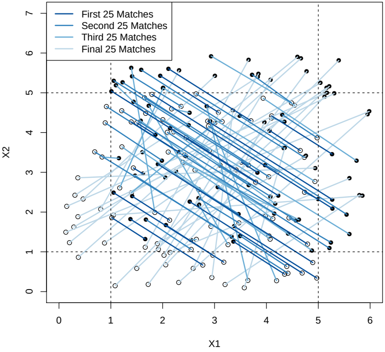


What Does PSM Match?

MDM Matches

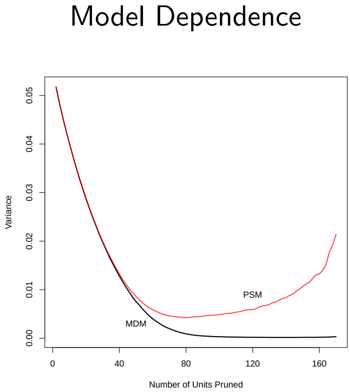


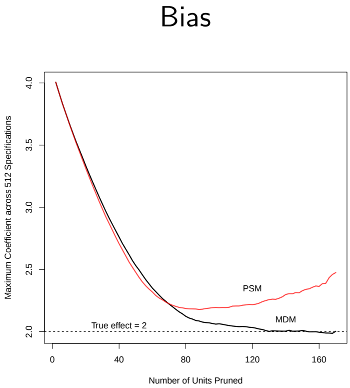
PSM Matches



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

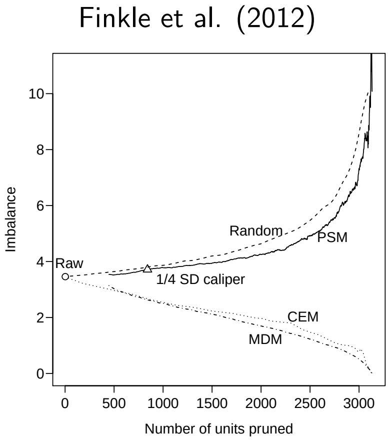
PSM Increases Model Dependence & Bias

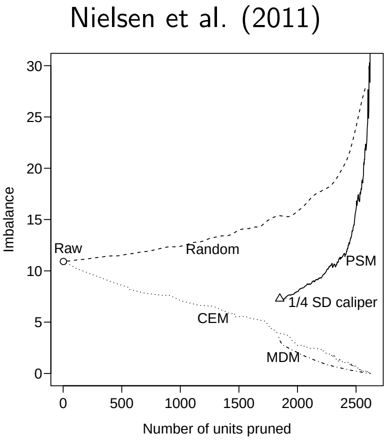




$$Y_i = 2T_i + X_{1i} + X_{2i} + \epsilon_i$$
$$\epsilon_i \sim N(0,1)$$

The Propensity Score Paradox





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
- (Maybe we can beat MDM/CEM for a given #pruned?)

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. 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
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. 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
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. 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
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. 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
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - No need to choose or use a matching method

The Matching Frontier

- Bias-Variance trade off $\rightsquigarrow$ Imbalance-n Trade Off
  Frontier = matched dataset with lowest imbalance for each n
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - No need to choose or use a matching method
  - 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
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - No need to choose or use a matching method
  - 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
- (Maybe we can beat MDM/CEM for a given #pruned?)
- To use, make 2 choices:
  1. Quantity of interest: SATT (prune Cs only) or FSATT
  2. Fixed- or variable-ratio matching
- Result:
  - Simple to use
  - No need to choose or use a matching method
  - All solutions are optimal
  - No iteration or diagnostics required
  - No cherry picking possible; you see everything optimal

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 all $\binom{N}{n}$ 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 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"

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

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!

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 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

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

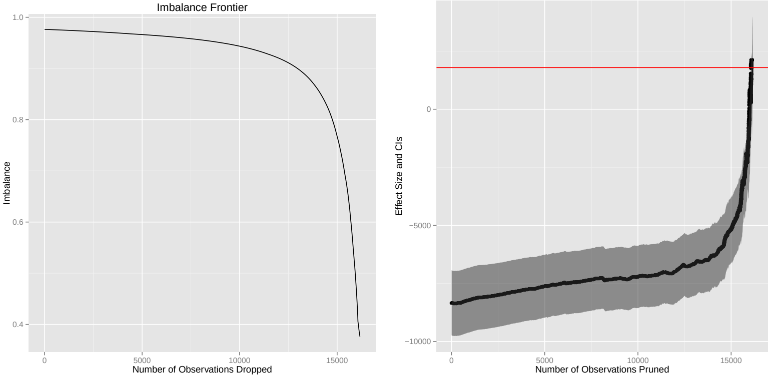
How hard is the frontier to calculate?

- Consider 1 point on the SATT frontier:
  - Start with matrix of N control units X_0
  - Calculate imbalance for 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
  - 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 $n = N, N-1, \ldots, 1$
  - 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)
  - $\rightsquigarrow$ It's **easy** to calculate!

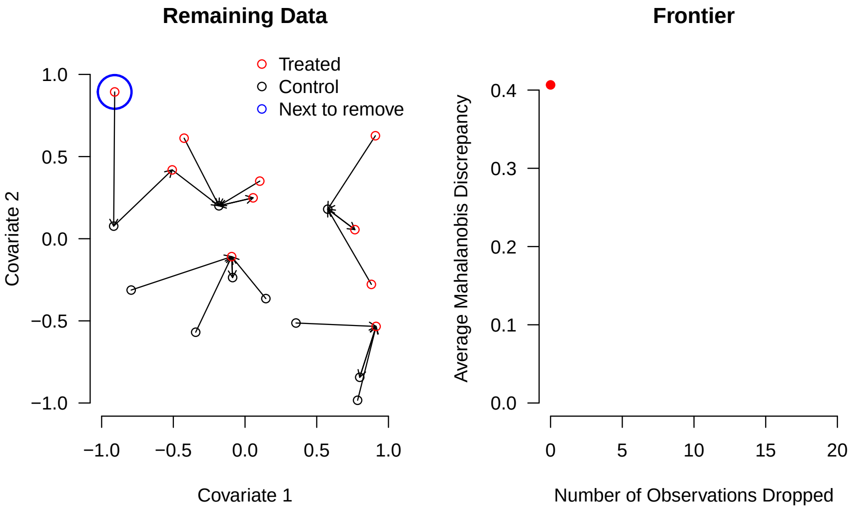
Job Training Data: Frontier and Causal Estimates



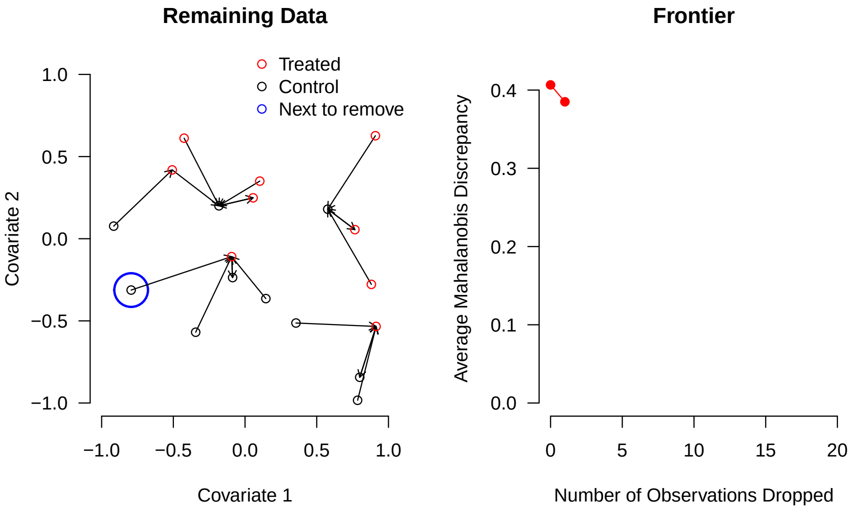
- 185 Ts; pruning most 16,252 Cs won't increase variance much
- Huge bias-variance trade-off after pruning most Cs
- Estimates converge to experiment after removing bias
- No mysteries: basis of inference clearly revealed

Constructing the FSATT Mahalanobis Frontier

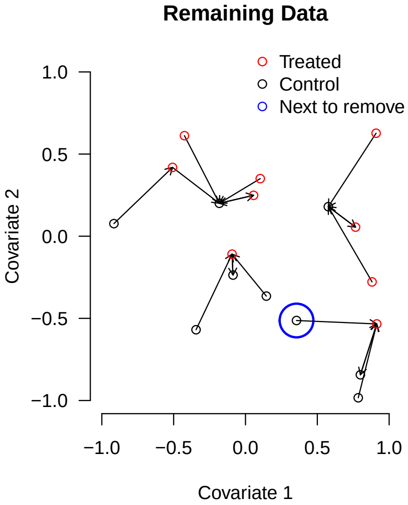
Constructing the FSATT Mahalanobis Frontier

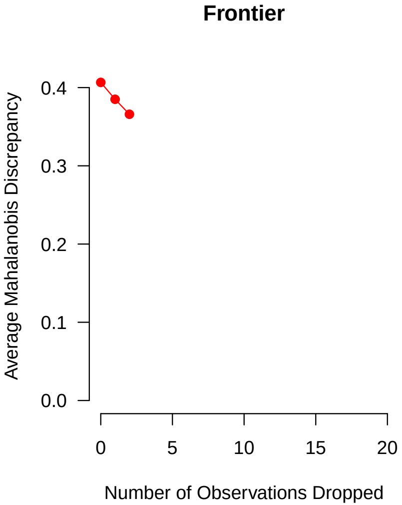


Constructing the FSATT Mahalanobis Frontier

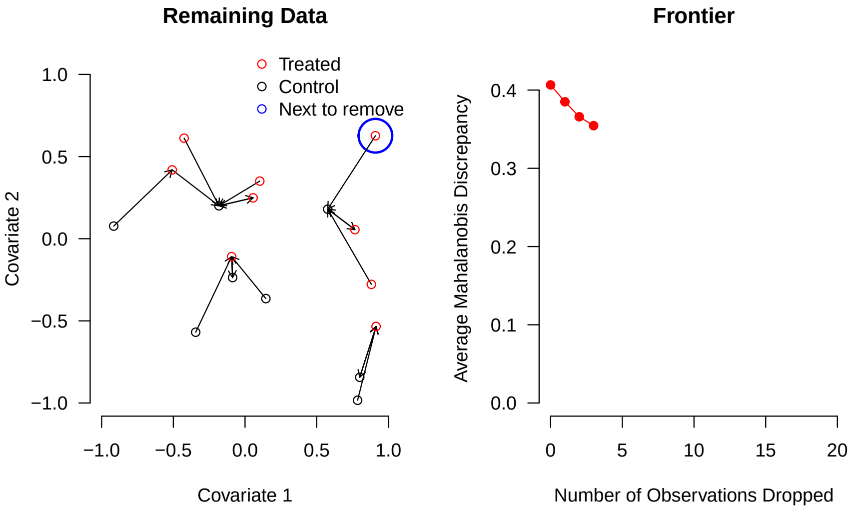


Constructing the FSATT Mahalanobis Frontier

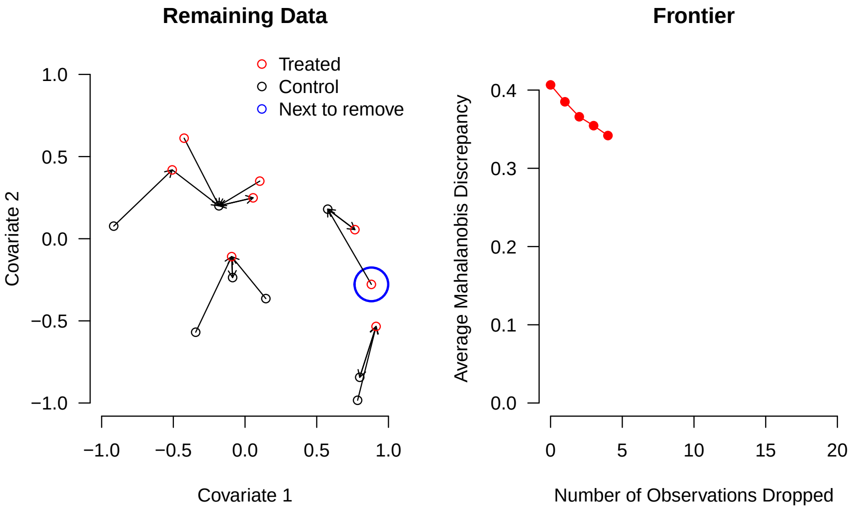




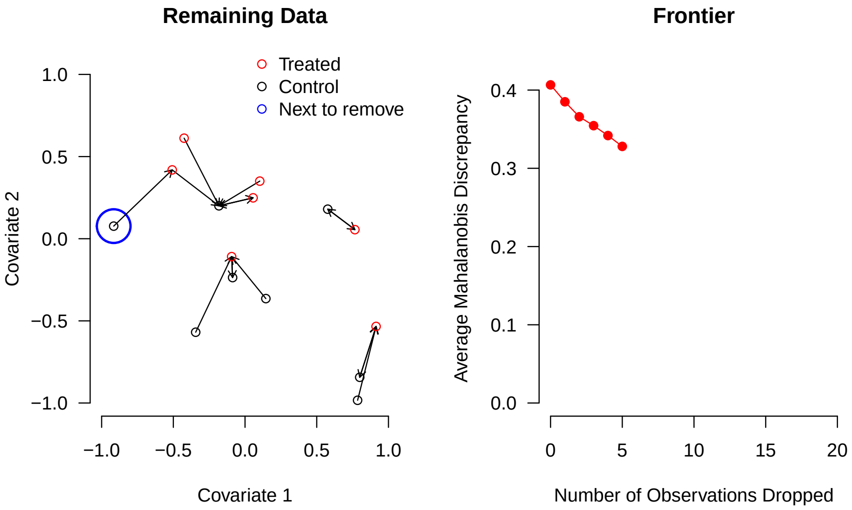
Constructing the FSATT Mahalanobis Frontier



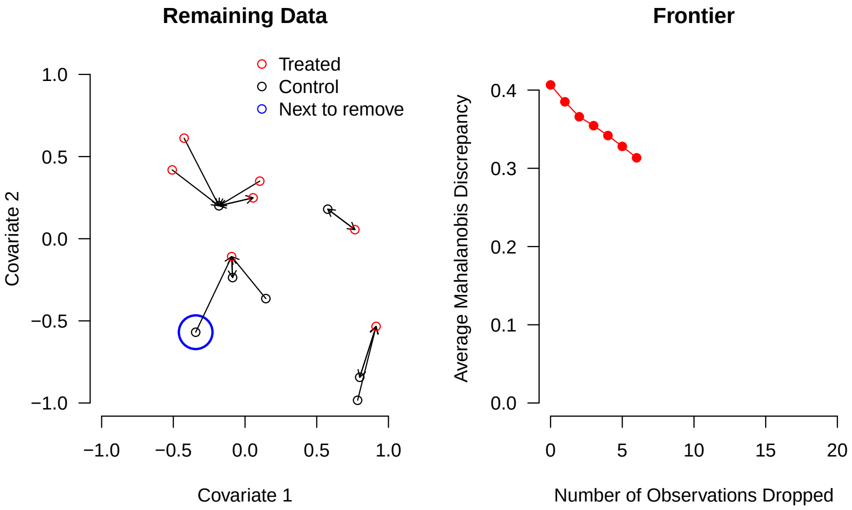
Constructing the FSATT Mahalanobis Frontier



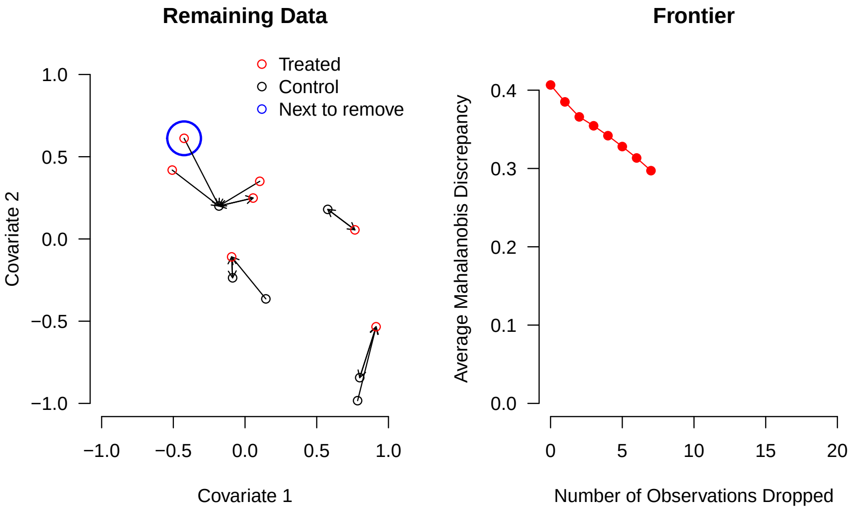
Constructing the FSATT Mahalanobis Frontier



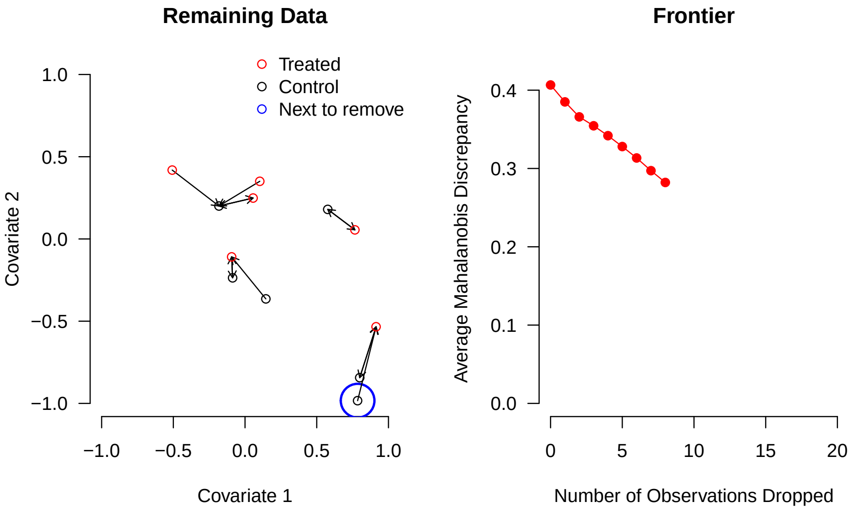
Constructing the FSATT Mahalanobis Frontier



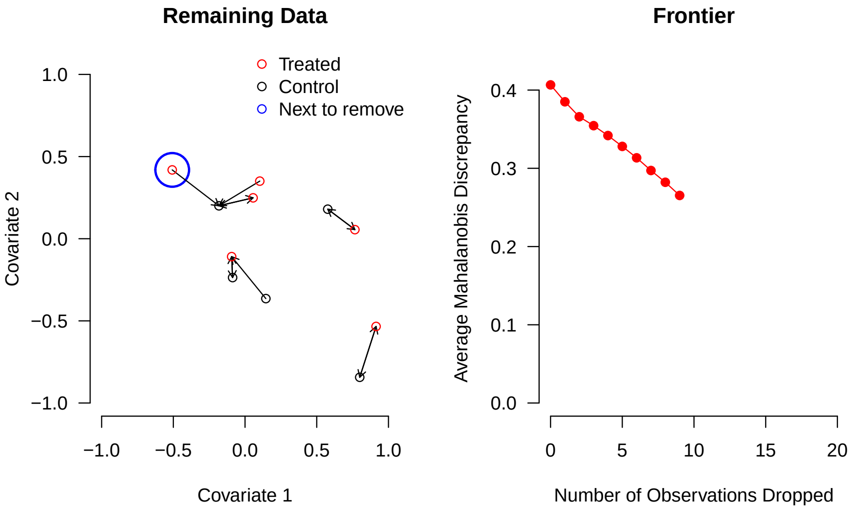
Constructing the FSATT Mahalanobis Frontier



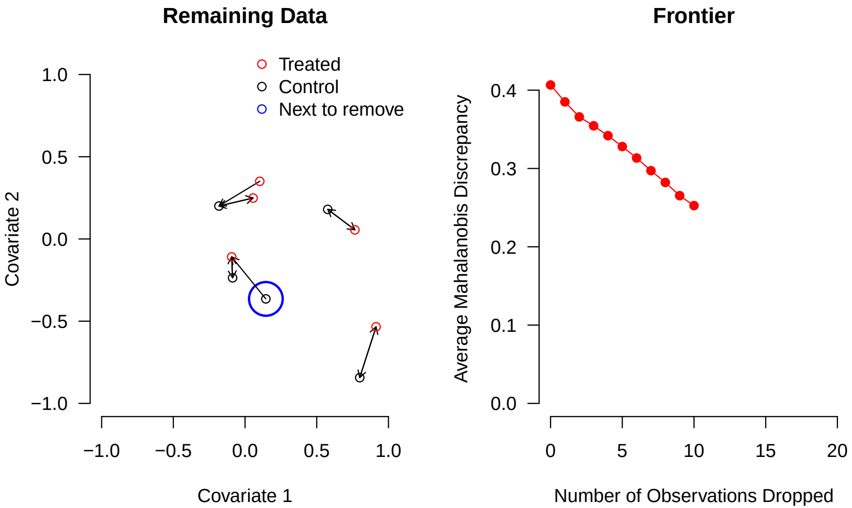
Constructing the FSATT Mahalanobis Frontier



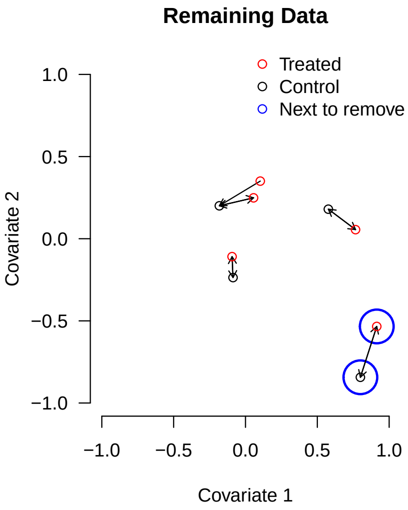
Constructing the FSATT Mahalanobis Frontier

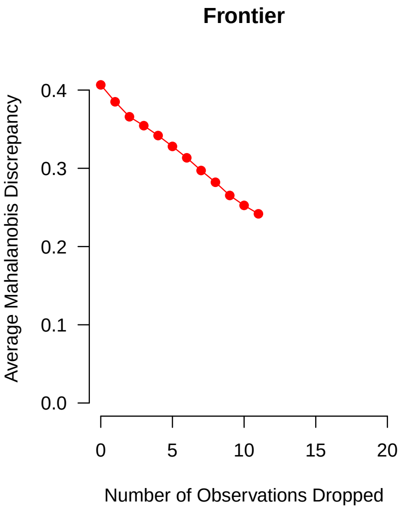


Constructing the FSATT Mahalanobis Frontier

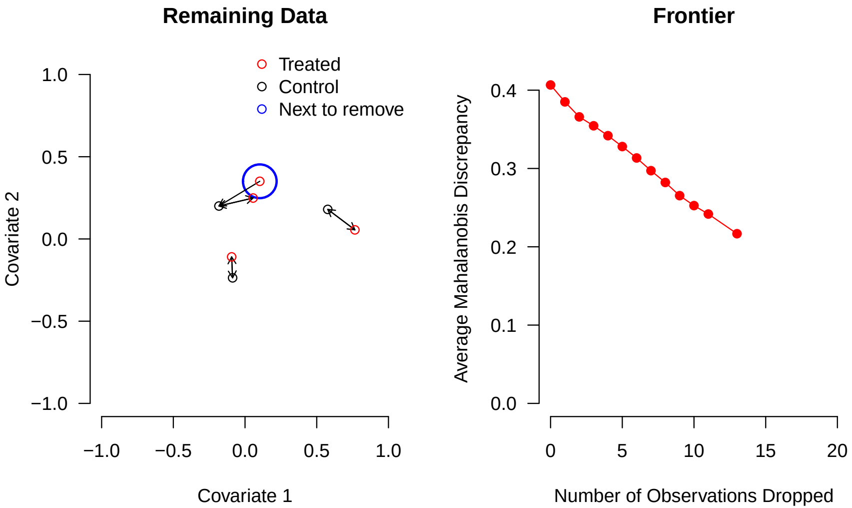


Constructing the FSATT Mahalanobis Frontier

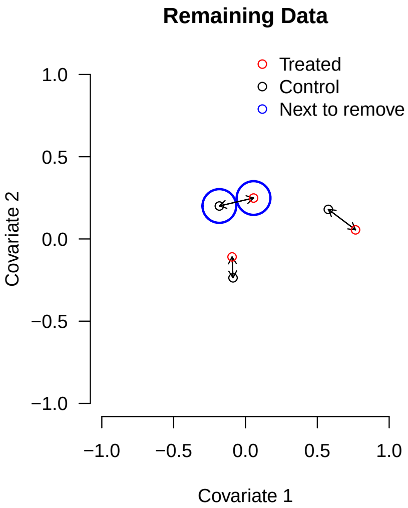


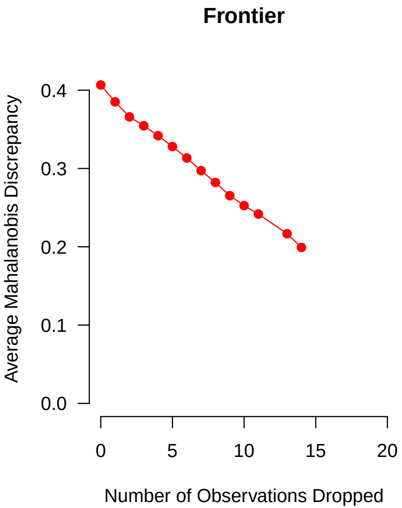


Constructing the FSATT Mahalanobis Frontier

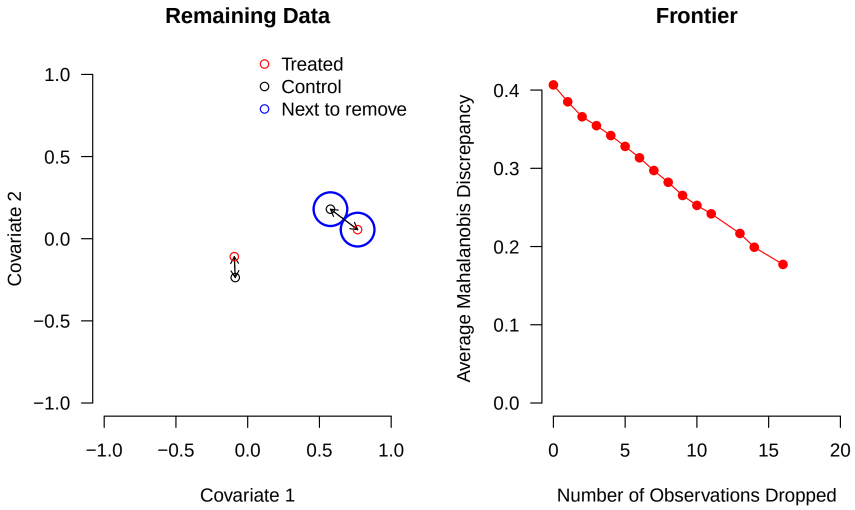


Constructing the FSATT Mahalanobis Frontier

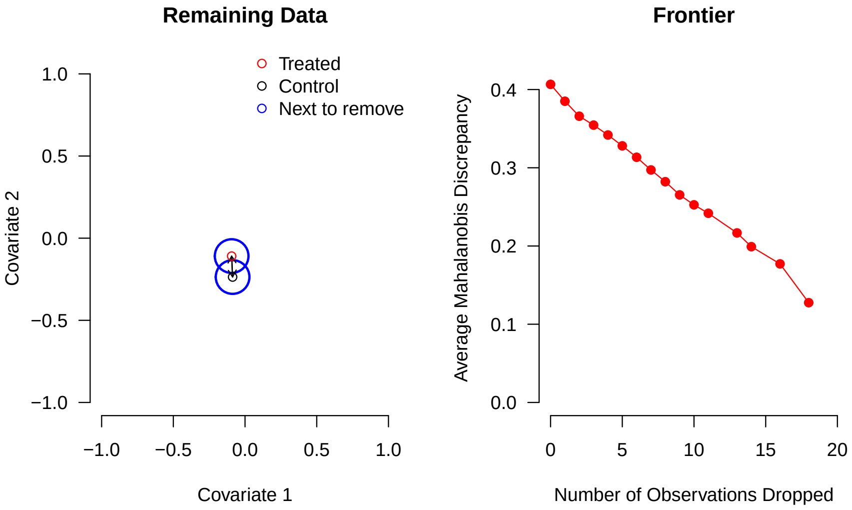




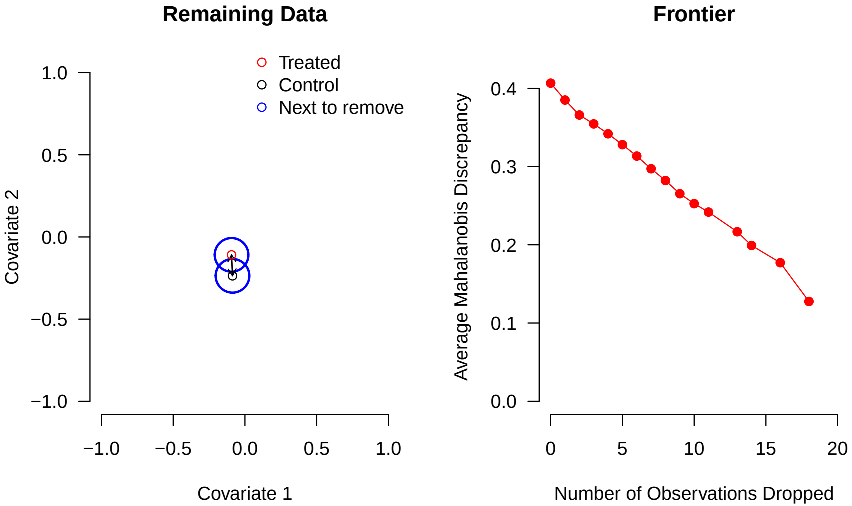
Constructing the FSATT Mahalanobis Frontier



Constructing the FSATT Mahalanobis Frontier

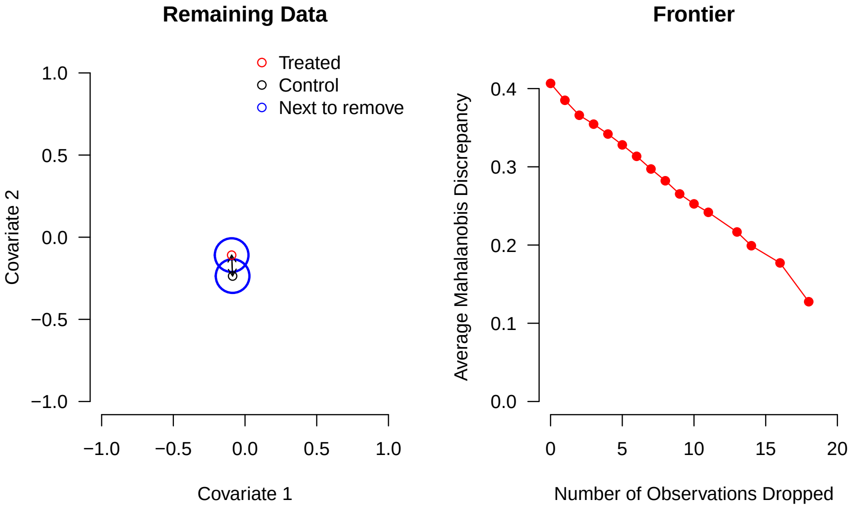


Constructing the FSATT Mahalanobis Frontier



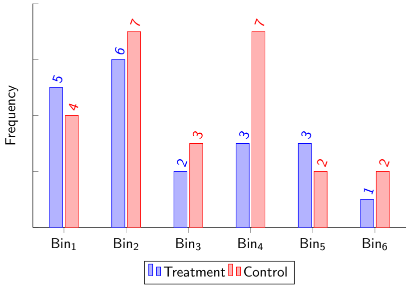
- Warning: figure omits details and the proof!

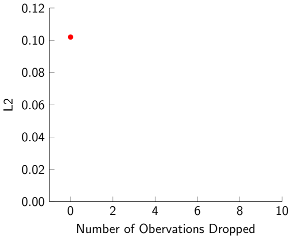
Constructing the FSATT Mahalanobis Frontier



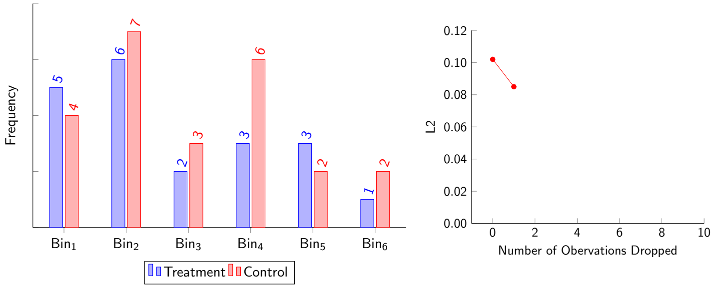
- Warning: figure omits details and the proof!
- Very fast; works with any continuous imbalance metric

Constructing the L1/L2 SATT Frontier

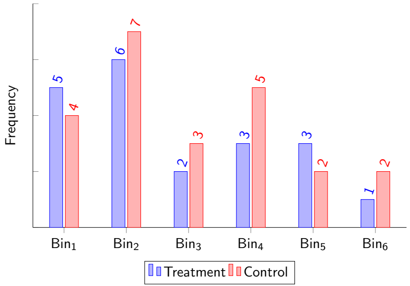


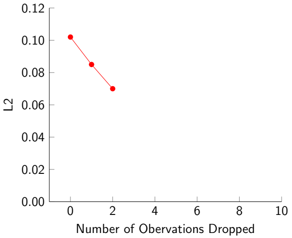


Constructing the L1/L2 SATT Frontier

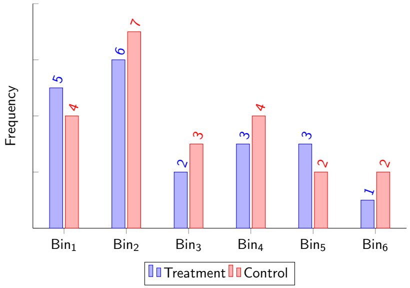


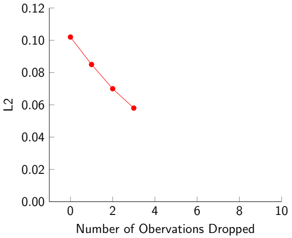
Constructing the L1/L2 SATT Frontier



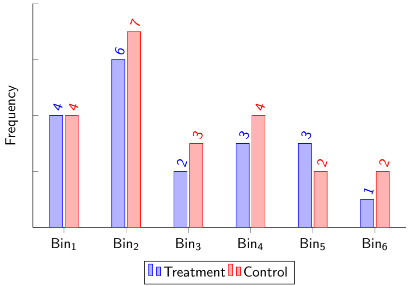


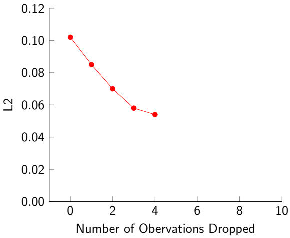
Constructing the L1/L2 SATT Frontier



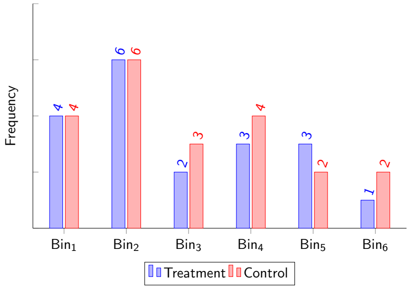


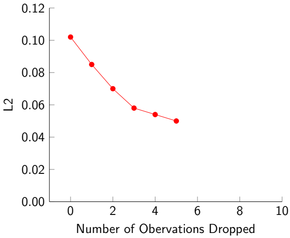
Constructing the L1/L2 SATT Frontier



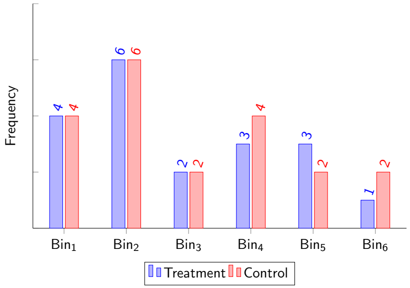


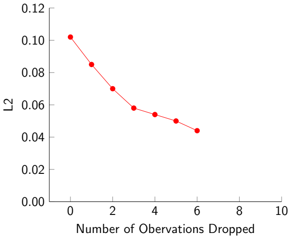
Constructing the L1/L2 SATT Frontier



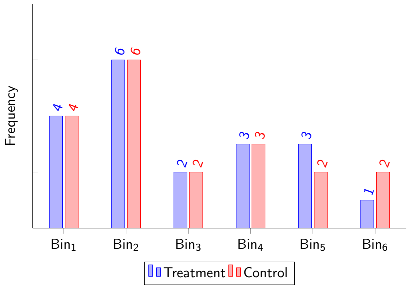


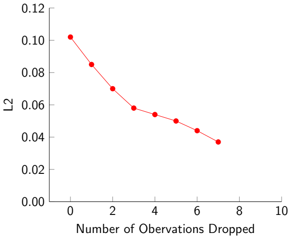
Constructing the L1/L2 SATT Frontier



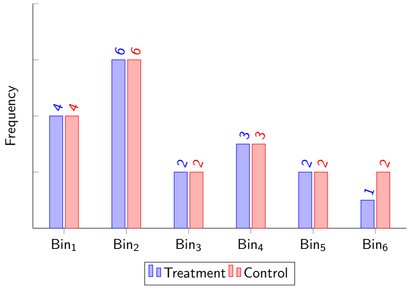


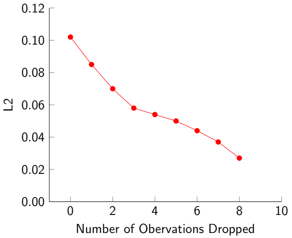
Constructing the L1/L2 SATT Frontier



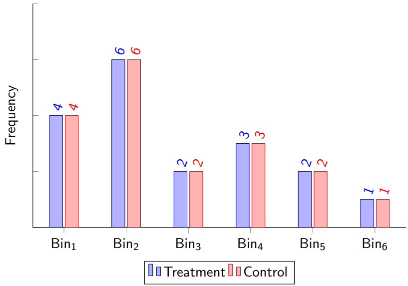


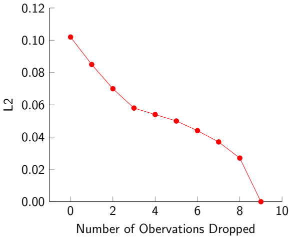
Constructing the L1/L2 SATT Frontier





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

Conclusions

Conclusions

- The Matching Frontier

Conclusions

- The Matching Frontier
    - Fast; easy; no iteratation; Software: MatchingFrontier

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM:

Conclusions

- The Matching Frontier
    - Fast; easy; no iteratation; Software: MatchingFrontier
    - No need to choose among matching methods
    - Optimal results from your choice of imbalance metric
- Propensity score matching:
    - Approximates complete, not fully blocked, experiments
    - Ignores information; exacerbates model dependence
    - Some mistakes with PSM: Controlling for irrelevant covariates;

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data;

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support;

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score;

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- Theory of Inference for Matching

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- Theory of Inference for Matching
  - Switch from simple to stratified random sampling

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- Theory of Inference for Matching
  - Switch from simple to stratified random sampling
  - Justifies current practices

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- Theory of Inference for Matching
  - Switch from simple to stratified random sampling
  - Justifies current practices
  - Clarifies how to improve inferences

Conclusions

- The Matching Frontier
  - Fast; easy; no iteratation; Software: MatchingFrontier
  - No need to choose among matching methods
  - Optimal results from your choice of imbalance metric
- Propensity score matching:
  - Approximates complete, not fully blocked, experiments
  - Ignores information; exacerbates model dependence
  - Some mistakes with PSM: Controlling for irrelevant covariates; Adjusting experimental data; Reestimating propensity score after eliminating noncommon support; 1/4 caliper on propensity score; Not switching to other methods.
- Theory of Inference for Matching
  - Switch from simple to stratified random sampling
  - Justifies current practices
  - Clarifies how to improve inferences
- $\rightsquigarrow$ Using more information is simpler and more powerful

For more information, papers, & software

GaryKing.org