Distribution-Free Assessment of Population Overlap in Observational Studies

Lihua Lei

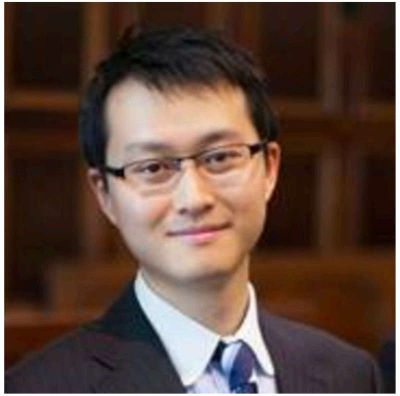
How similar are treatment and control groups?

Lihua Lei

Collaborators

Alexander D'Amour

(Google Brain)



Peng Ding

(UC Berkeley)

Avi Feller

(UC Berkeley)

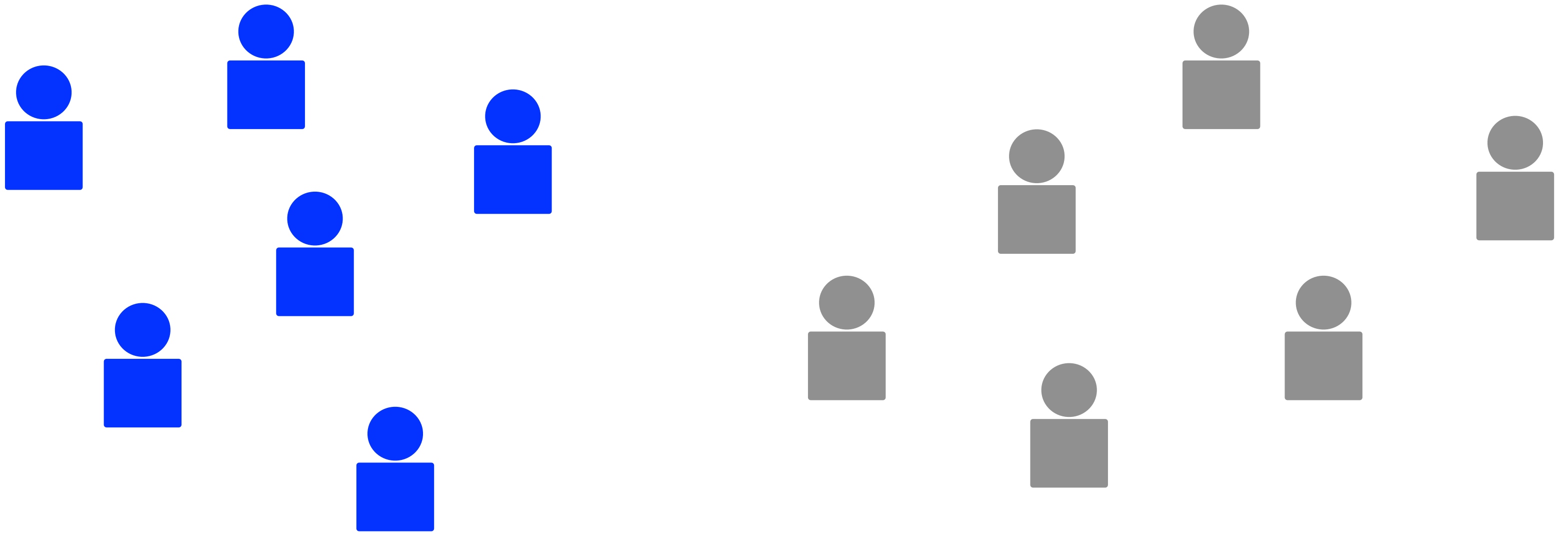
Jasjeet Sekhon

(Yale)

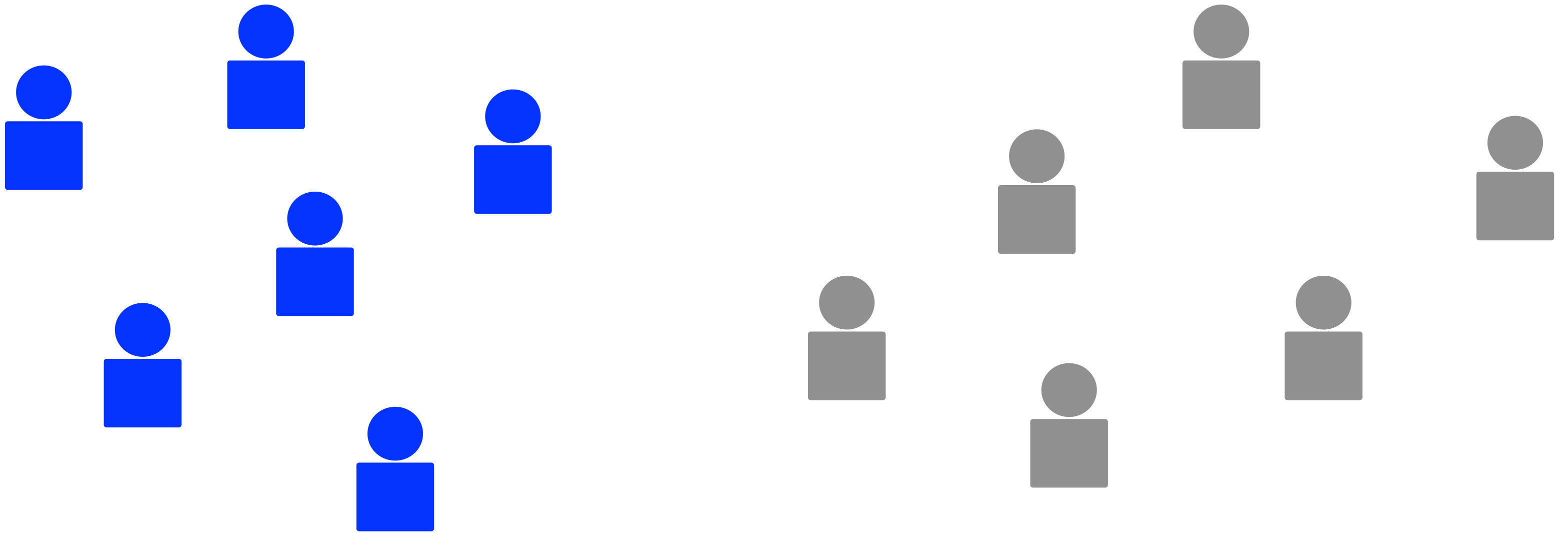
Motivation & background

“Robust” causal inference requires treatment & control groups to be “similar”

“Robust” causal inference requires treatment & control groups to be “similar”

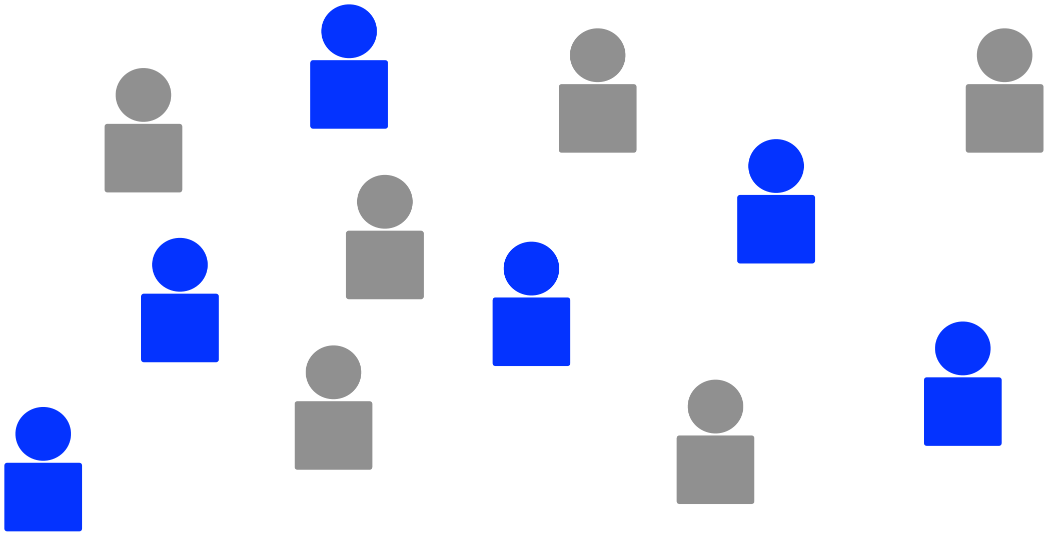


“Robust” causal inference requires treatment & control groups to be “similar”

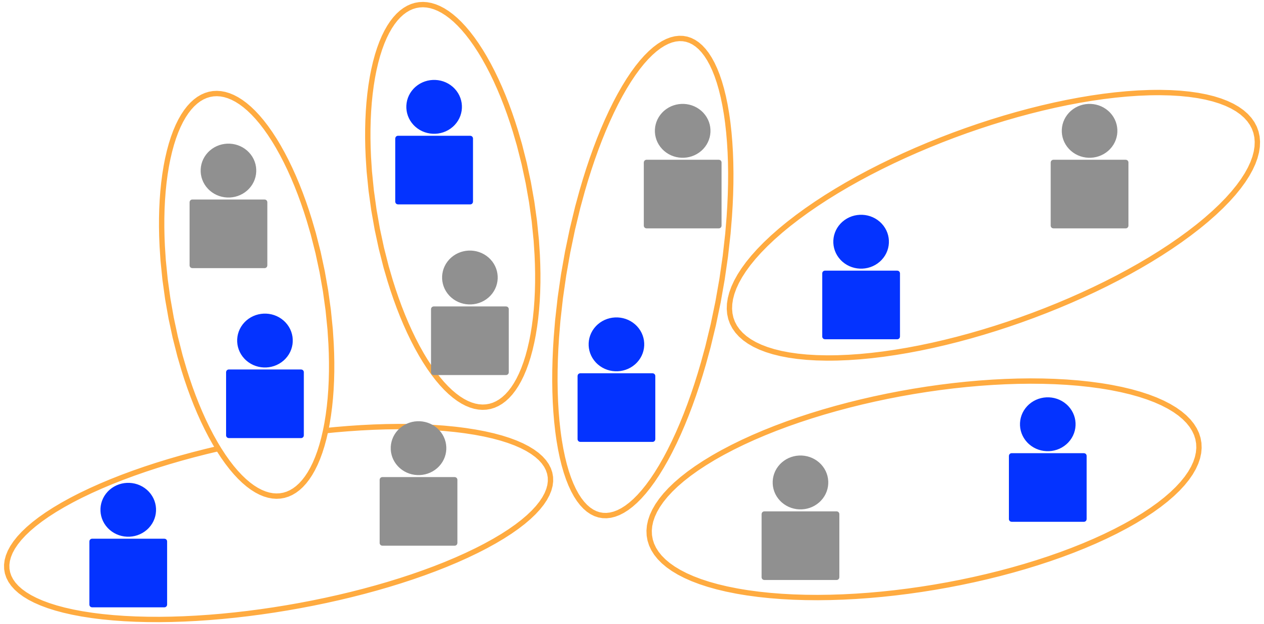


Need extrapolation

“Robust” causal inference requires treatment & control groups to be “similar”



“Robust” causal inference requires treatment & control groups to be “similar”



Infer counterfactuals from similar units

Assessing similarity via overlap

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^{n} \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Assessing similarity via overlap

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition:

$\mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0$, (e.g., $\mathcal{O}_0 = 0.1$)

One of the most fundamental conditions!

Overlap measures similarity of treatment/control groups

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition:

$\mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0$, (e.g., $\mathcal{O}_0 = 0.1$)

One of the most fundamental conditions!

$P_{X|T=1} = P_{X|T=0} \Leftrightarrow e(X) \equiv e$ (RCT)

$\Rightarrow$strict overlap condition $\not\Rightarrow P_{X|T=1} = P_{X|T=0}$

Overlap measures similarity of treatment/control groups

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition:

$\mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0$, (e.g., $\mathcal{O}_0 = 0.1$)

One of the most fundamental conditions!

$P_{X|T=1} = P_{X|T=0} \Leftrightarrow e(X) \equiv e$ (RCT)

$\Rightarrow$strict overlap condition $\not\Rightarrow P_{X|T=1} = P_{X|T=0}$

$$\mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0$$

$$\Leftrightarrow \frac{\mathcal{O}_0}{1-\mathcal{O}_0} \leq \frac{P(T=1 \mid X)}{P(T=0 \mid X)} \leq \frac{1-\mathcal{O}_0}{\mathcal{O}_0}$$

$$\Leftrightarrow \frac{\mathcal{O}_0}{1-\mathcal{O}_0} \leq \frac{P(X \mid T=1)P(T=1)}{P(X \mid T=0)P(T=0)} \leq \frac{1-\mathcal{O}_0}{\mathcal{O}_0}$$

$$\Leftrightarrow \frac{\mathcal{O}_0}{1-\mathcal{O}_0}\frac{P(T=0)}{P(T=1)} \leq \frac{P(X \mid T=1)}{P(X \mid T=0)} \leq \frac{1-\mathcal{O}_0}{\mathcal{O}_0}\frac{P(T=0)}{P(T=1)}$$

$\Rightarrow$ Density ratio of covariate distributions is bounded

A summary measure of overlap

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition:

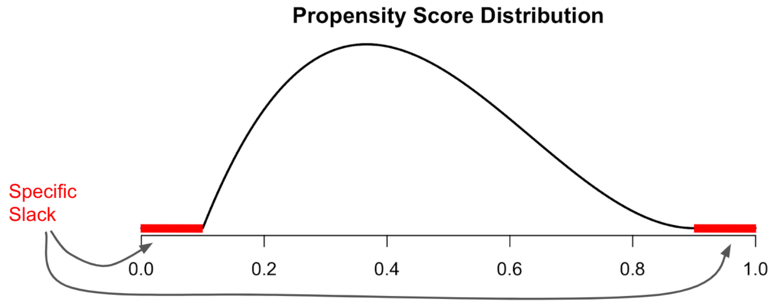
$$\mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0, \text{ (e.g., } \mathcal{O}_0 = 0.1)$$

One of the most fundamental conditions!

Definition (population overlap slack)

$$\mathcal{O}^* \triangleq \min_x \min\{e(x), 1 - e(x)\}$$

Strict overlap condition $\iff \mathcal{O}^* \geq \mathcal{O}_0$



A summary measure of overlap

Setting

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition:

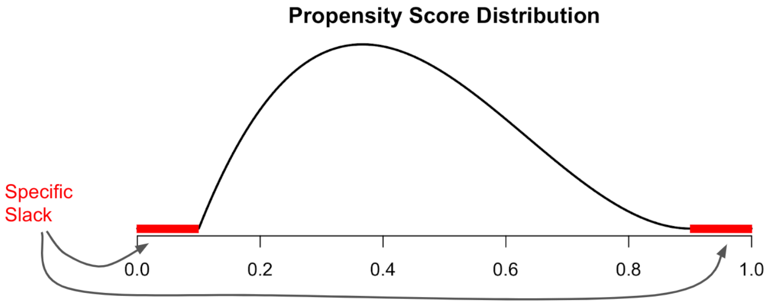
$\mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0$, (e.g., $\mathcal{O}_0 = 0.1$)

One of the most fundamental conditions!

Definition (population overlap slack)

$$\mathcal{O}^* \triangleq \min_x \min\{e(x), 1 - e(x)\}$$

Strict overlap condition $\iff \mathcal{O}^* \geq \mathcal{O}_0$



$n\mathcal{O}^*$ is the **effective sample size** for **ATE** w/o outcome restrictions (Hong, Leung, Li. '20)

Common practice: estimated propensity scores

Setting

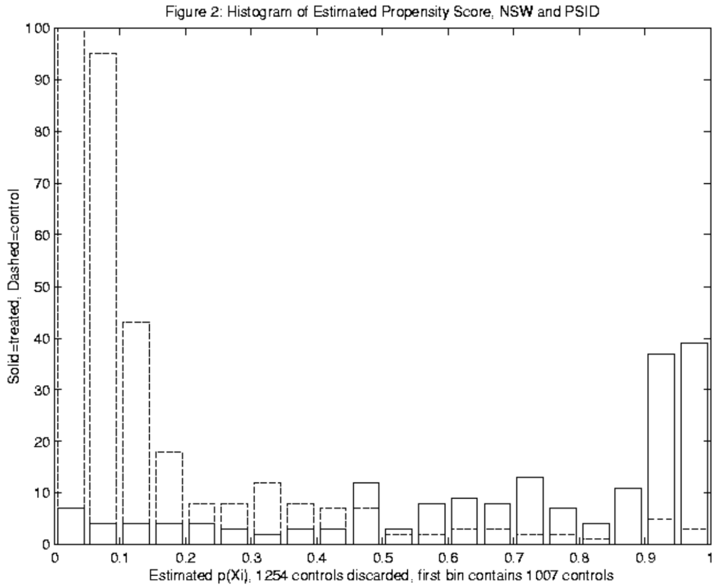
- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition:

$$\mathcal{O}_0 \le e(X) \le 1 - \mathcal{O}_0, \text{ (e.g., } \mathcal{O}_0 = 0.1)$$

One of the most fundamental conditions!

Check the distribution of estimated propensity scores



Dehejia & Wahba ('98)

- Plug-in estimate: $\hat{\mathcal{O}} = \min \min\{\hat{e}(x), 1 - \hat{e}(x)\}$

Common practice: estimated propensity scores

Setting

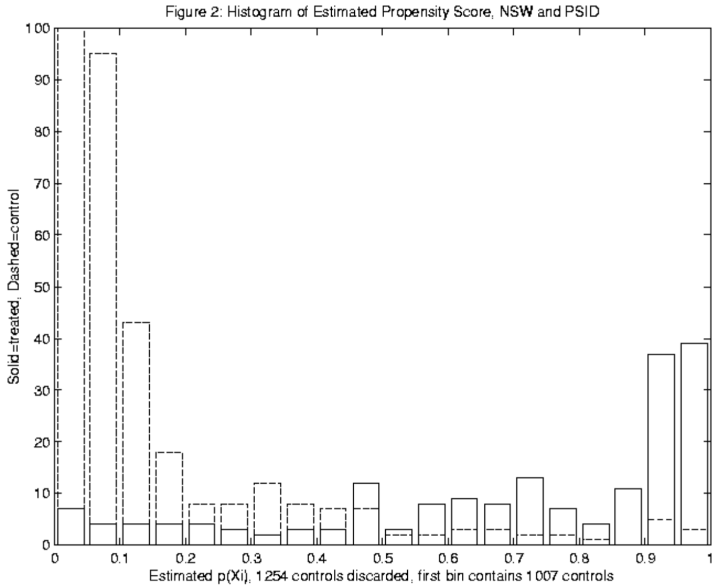
- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^n \overset{i.i.d.}{\sim} (T, X)$ (the only assumption!)
- **Propensity score**: $e(x) \triangleq P(T = 1 \mid X = x)$

Strict overlap condition:

$\mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0$, (e.g., $\mathcal{O}_0 = 0.1$)

One of the most fundamental conditions!

Check the distribution of estimated propensity scores



Dehejia & Wahba ('98)

- Plug-in estimate: $\hat{\mathcal{O}} = \min \min\{\hat{e}(x), 1 - \hat{e}(x)\}$
- Misspecified propensity score model
- $\mathcal{O}^*$ is hard to estimate; irregular parameter

Our contribution: O-value

We propose **O-values** as **upper confidence bounds** of $\mathcal{O}^*$, denoted by $\hat{\mathcal{O}}$, that

- **guarantees coverage**, i.e., $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$ (e.g., $\alpha = 0.05$)
- **in finite samples** (no asymptotics!)
- **without any assumption other than i.i.d.** (uniform inference)
- is able to wrap around **any black-box** algorithm to estimate $e(X)$

Analogous to p-value:

- small $\hat{\mathcal{O}} \Rightarrow$ strong evidence against overlap
- large $\hat{\mathcal{O}} \not\Rightarrow$ sufficient overlap
- conservative but reliable assessment of overlap (Armstrong & Kolesar, '18, '21; Armstrong, Kolesar & Kwon, '20)

What can we do with O-values? Testing overlap condition

O-value is an upper confidence bound of $\mathcal{O}^*$: $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$

- Test the strict overlap condition (treated as a composite null hypothesis)

$$H_0 : \mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0 \text{ or } H_0 : \mathcal{O}^* \geq \mathcal{O}_0$$

Reject H_0 if $\hat{\mathcal{O}} < \mathcal{O}_0$; valid in finite samples with size at most α

What can we do with O-values? Testing overlap condition

O-value is an upper confidence bound of $\mathcal{O}^*$: $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$

- Test the strict overlap condition (treated as a composite null hypothesis)

$$H_0 : \mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0 \text{ or } H_0 : \mathcal{O}^* \geq \mathcal{O}_0$$

  Reject H_0 if $\hat{\mathcal{O}} < \mathcal{O}_0$; valid in finite samples with size at most α

- Run simulations to determine a "safe" $\mathcal{O}_0$ for the estimator to be used and test it

What can we do with O-values? Testing overlap condition

O-value is an upper confidence bound of $\mathcal{O}^*$: $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$

- Test the strict overlap condition (treated as a composite null hypothesis)

$$H_0 : \mathcal{O}_0 \leq e(X) \leq 1 - \mathcal{O}_0 \text{ or } H_0 : \mathcal{O}^* \geq \mathcal{O}_0$$

Reject H_0 if $\hat{\mathcal{O}} < \mathcal{O}_0$; valid in finite samples with size at most α

- Run simulations to determine a "safe" $\mathcal{O}_0$ for the estimator to be used and test it

We evaluate coverage of confidence intervals in the "many-cluster" setting for different choices of β, n, and p; results are given in Table 6. Coverage is generally better with more overlap ($\eta = 0.25$) rather than less ($\eta = 0.1$), and with sparser choices of β. Moreover, coverage rates appear to improve as n increases, suggesting that we are in a regime where the asymptotics from Corollary 6 are beginning to apply.

Athey, Imbens, Wager ('17)

What can we do with O-values? Assessing efficiency loss

O-value is an upper confidence bound of $\mathcal{O}^*$: $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$

- Recall that $n\mathcal{O}^*$ is the effective sample size (Hong, Leung, Li. '20)
- Given the group sizes n_1, n_0, the maximum effective sample size is given by $\min\{n_1, n_0\}$

What can we do with O-values? Assessing efficiency loss

O-value is an upper confidence bound of $\mathcal{O}^*$: $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$

- Recall that $n\mathcal{O}^*$ is the effective sample size (Hong, Leung, Li. '20)
- Given the group sizes n_1, n_0, the maximum effective sample size is given by $\min\{n_1, n_0\}$
- $L = 1 - n\mathcal{O}^* / \min\{n_1, n_0\}$ measures the efficiency loss due to imbalance
- Estimate L by $\hat{L} = 1 - n\hat{\mathcal{O}} / \min\{n_1, n_0\}$; optimistic assessment of efficiency loss

Construction of O-values

General idea

- Standard econometric pipeline

Identification $\Rightarrow$ Estimation $\Rightarrow$ Uncertainty quantification (tests or confidence intervals)

General idea

- Standard econometric pipeline

Identification ⇒ Estimation ⇒ Uncertainty quantification (tests or confidence intervals)

- Strictly speaking, $e(x)$ is identifiable; so is $\mathcal{O}^* = \min_x \min\{e(x), 1 - e(x)\}$

- However, $\mathcal{O}^*$ is not "estimable" (irregularity; model misspecification)

General idea

- Standard econometric pipeline

$$\text{Identification} \Rightarrow \text{Estimation} \Rightarrow \text{Uncertainty quantification (tests or confidence intervals)}$$

- Strictly speaking, $e(x)$ is identifiable; so is $\mathcal{O}^* = \min_x \min\{e(x), 1 - e(x)\}$
- However, $\mathcal{O}^*$ is not "estimable" (irregularity; model misspecification)
- Surprisingly, there are non-trivial estimable upper bounds of $\mathcal{O}^*$ that is robust to misspecification!
- If $\tilde{\mathcal{O}}^* \geq \mathcal{O}^*$ and $\mathbb{P}(\hat{\mathcal{O}} \geq \tilde{\mathcal{O}}^*) \geq 1 - \alpha$, then $\hat{\mathcal{O}}$ is a valid O-value:

$$\mathbb{P}(\hat{\mathcal{O}} \geq \mathcal{O}^*) \geq \mathbb{P}(\hat{\mathcal{O}} \geq \tilde{\mathcal{O}}^*) \geq 1 - \alpha$$

General idea

- Standard econometric pipeline

Identification ⇒ Estimation ⇒ Uncertainty quantification (tests or confidence intervals)

- Strictly speaking, $e(x)$ is identifiable; so is $\mathcal{O}^* = \min_x \min\{e(x), 1 - e(x)\}$

- However, $\mathcal{O}^*$ is not "estimable" (irregularity; model misspecification)

- Surprisingly, there are non-trivial estimable upper bounds of $\mathcal{O}^*$ that is robust to misspecification!

- If $\tilde{\mathcal{O}}^* \geq \mathcal{O}^*$ and $\mathbb{P}(\hat{\mathcal{O}} \geq \tilde{\mathcal{O}}^*) \geq 1 - \alpha$, then $\hat{\mathcal{O}}$ is a valid O-value:

$$\mathbb{P}(\hat{\mathcal{O}} \geq \mathcal{O}^*) \geq \mathbb{P}(\hat{\mathcal{O}} \geq \tilde{\mathcal{O}}^*) \geq 1 - \alpha$$

- **Partial identification** for identifiable parameters (to gain robustness)

Two challenges

- Arbitrary covariates

Two challenges

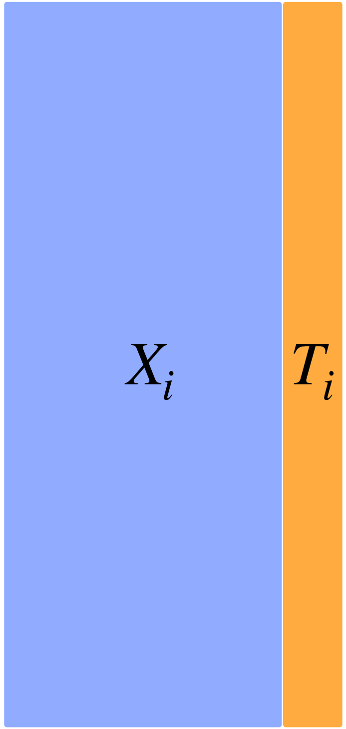
- Arbitrary covariates
  - Standardize X into estimated propensity scores
  - Use data splitting to avoid overfitting

Two challenges

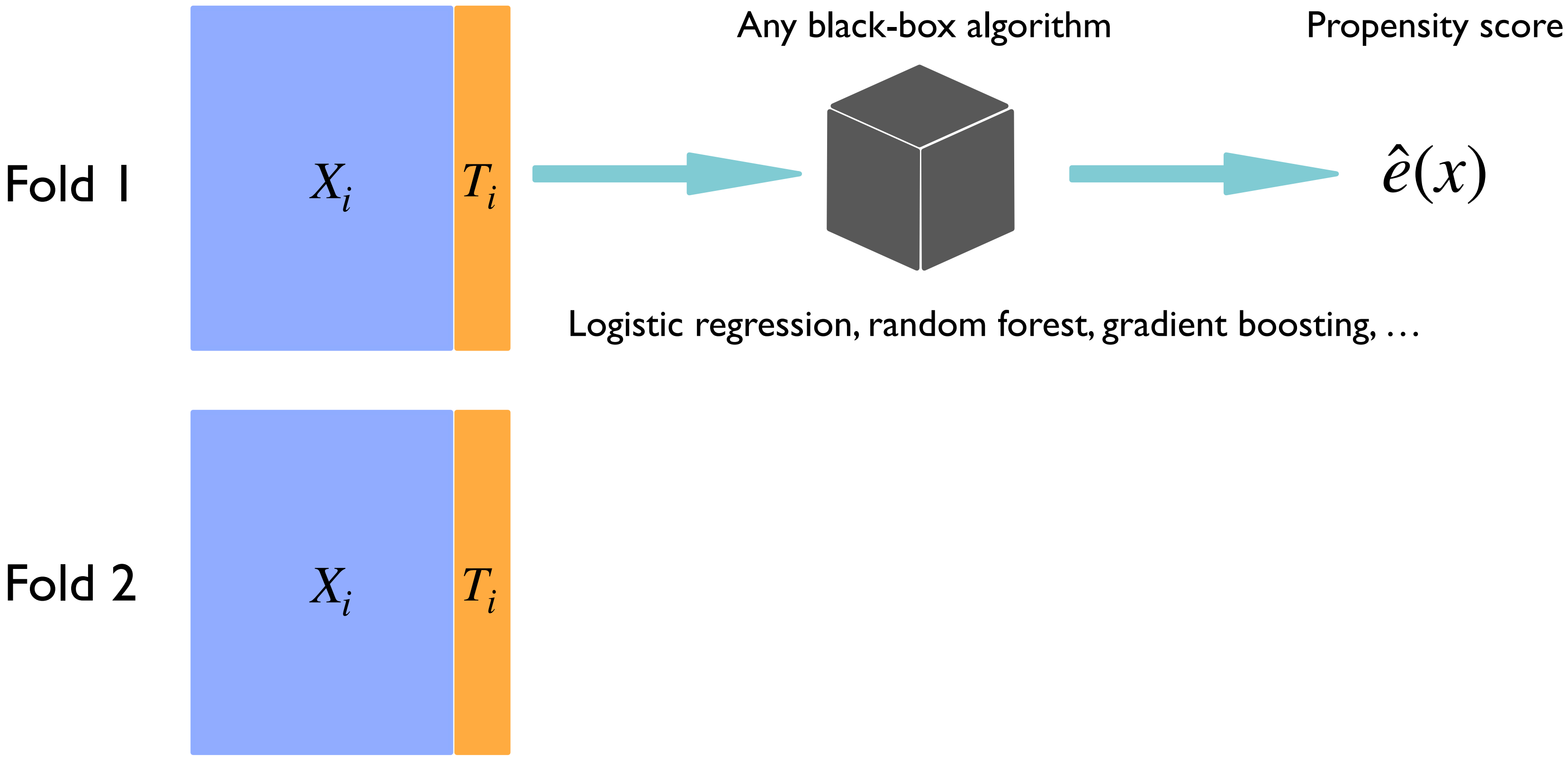
- Arbitrary covariates
  - Standardize X into estimated propensity scores
  - Use data splitting to avoid overfitting
- Pass the estimated propensity scores into careful balance checks (instead of using them directly)
  - Difference-in-Means (DiM) O-value
  - Difference-in-Tails (DiT) O-value
  - Difference-in-Ranks (DiR) O-value
  - Classification-Error (CE) O-value

Step 1: covariate standardization

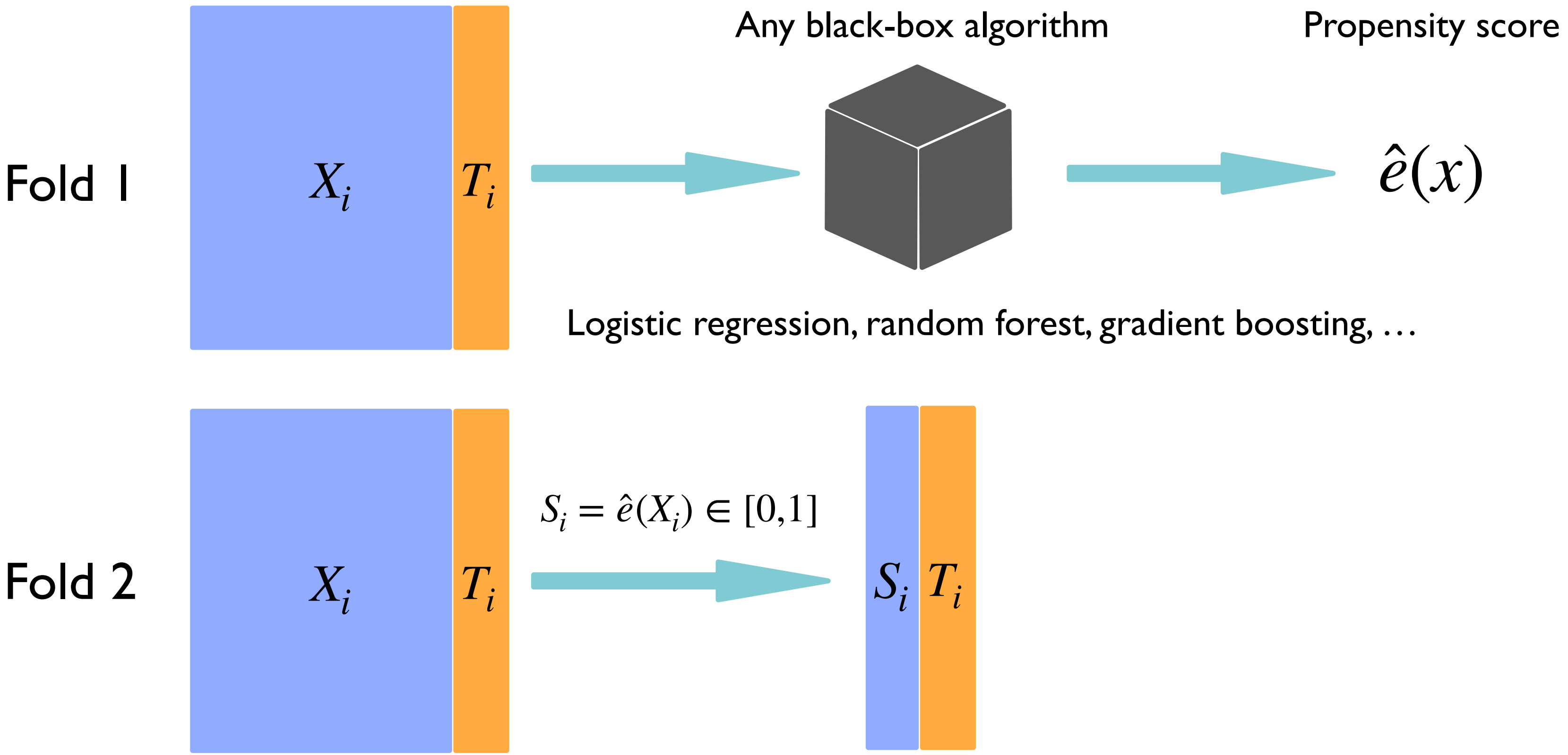
Data



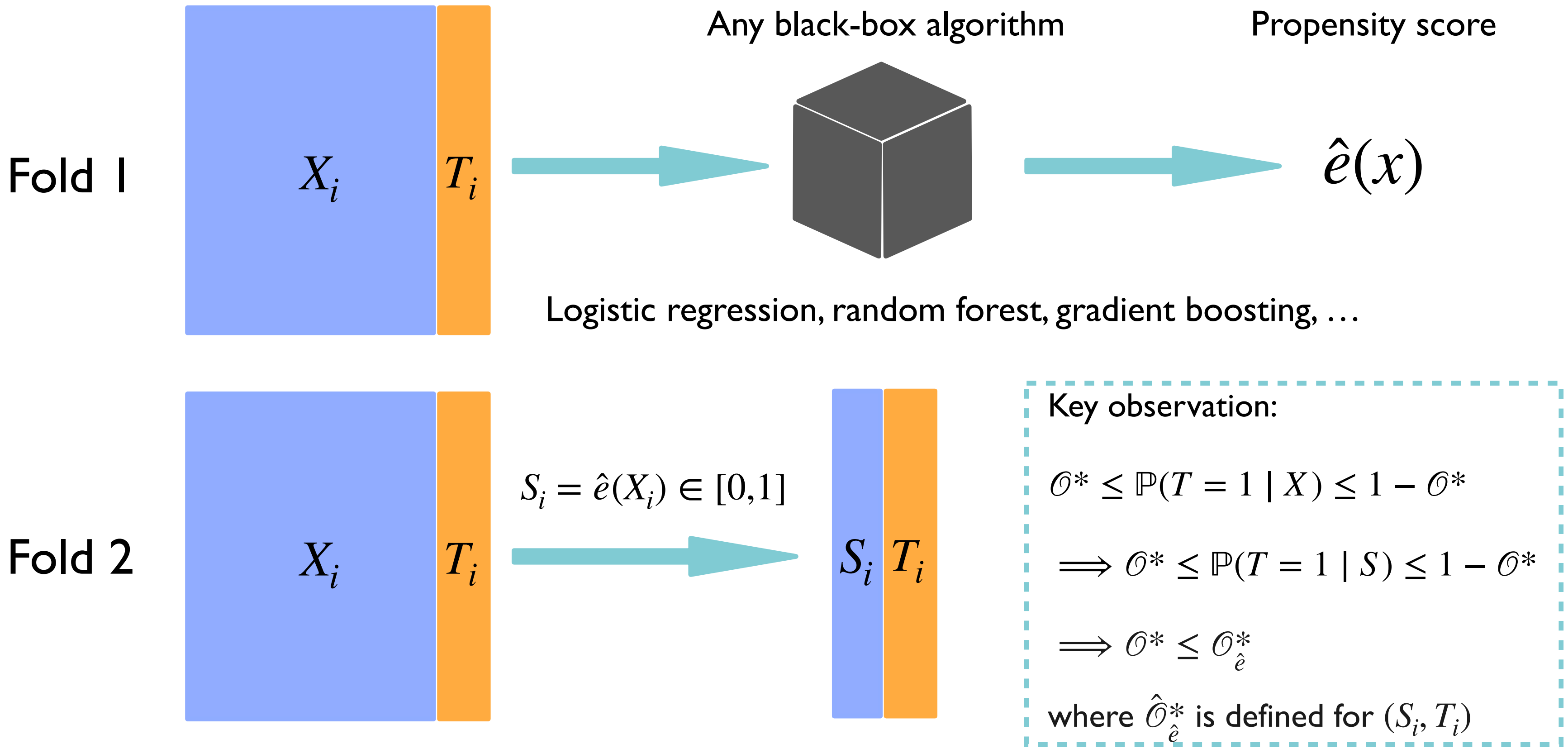
Step 1: covariate standardization



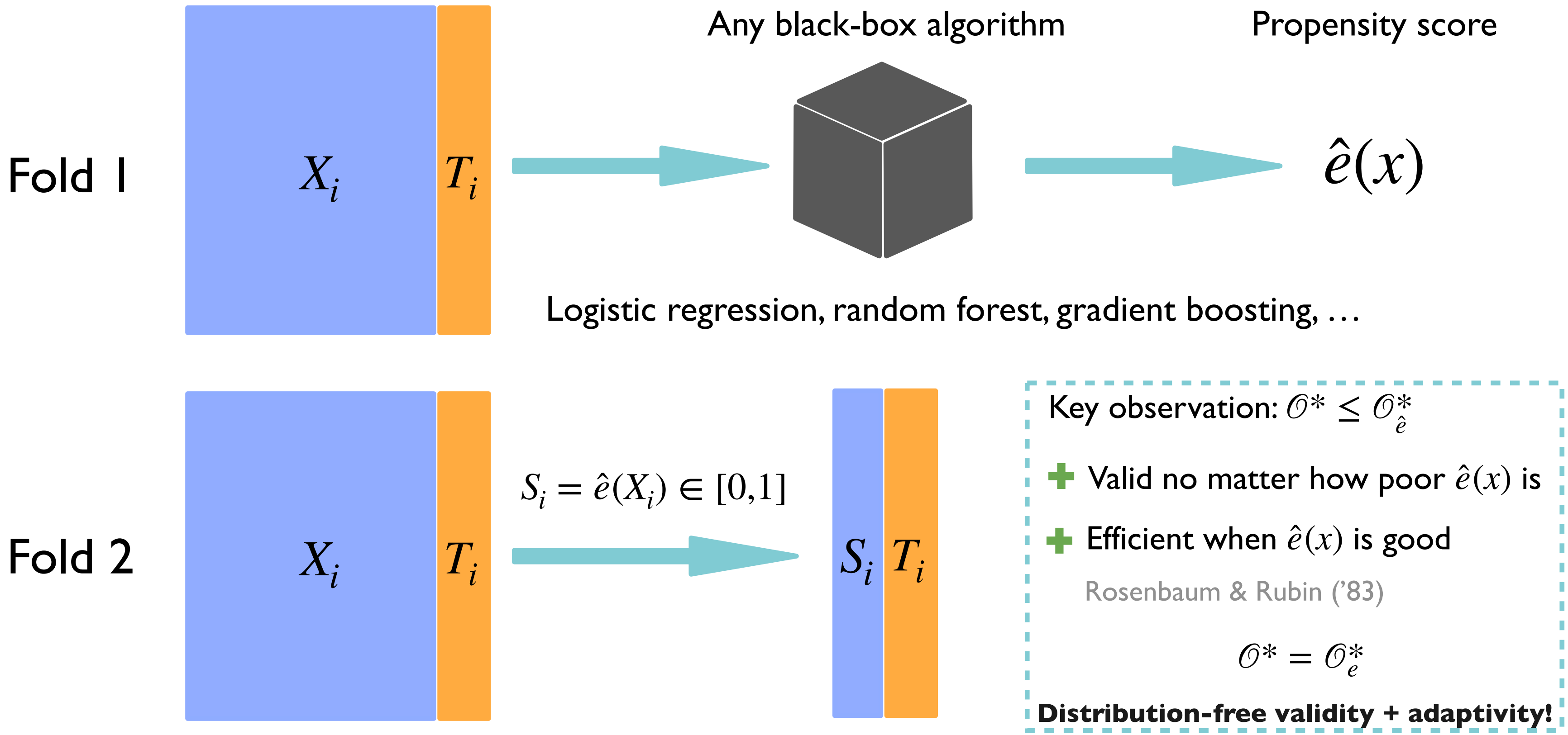
Step 1: covariate standardization



Step I: covariate standardization



Step 1: covariate standardization



Step I: covariate standardization

Before standardization:

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^{n} \overset{i.i.d.}{\sim} (T, X)$
- $e(x) \triangleq P(T = 1 \mid X = x)$
- $\mathcal{O}^* = \min_x \min\{e(x), 1 - e(x)\}$

After standardization:

- Binary treatment $T \in \{0,1\}$
- Covariate $S = \hat{e}(X)$: one-dimensional; $S \in [0,1]$
- $(S_i, T_i)_{i=1}^{n_2} \overset{i.i.d.}{\sim} (S, T)$ (on the second fold)
- $\tilde{e}(s) \triangleq P(T = 1 \mid S = s)$
- $\tilde{\mathcal{O}}^* = \min_s \min\{\tilde{e}(s), 1 - \tilde{e}(s)\}$

Step I: covariate standardization

Before standardization:

- Binary treatment $T \in \{0,1\}$
- Covariates X: no constraint
- $(T_i, X_i)_{i=1}^{n} \overset{i.i.d.}{\sim} (T, X)$
- $e(x) \triangleq P(T = 1 \mid X = x)$
- $\mathcal{O}^* = \min_x \min\{e(x), 1 - e(x)\}$

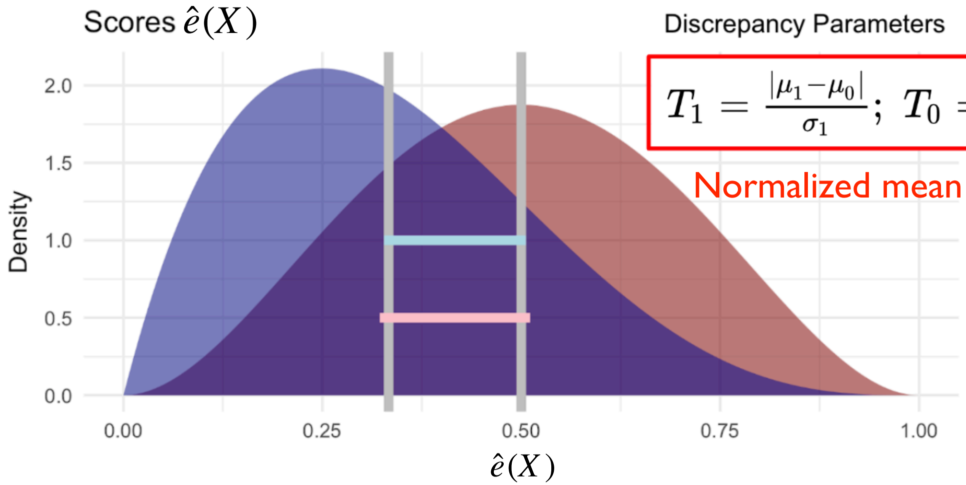
After standardization:

- Binary treatment $T \in \{0,1\}$
- Covariate $S = \hat{e}(X)$: one-dimensional; $S \in [0,1]$
- $(S_i, T_i)_{i=1}^{n_2} \overset{i.i.d.}{\sim} (S, T)$ (on the second fold)
- $\tilde{e}(s) \triangleq P(T = 1 \mid S = s)$
- $\tilde{\mathcal{O}}^* = \min_s \min\{\tilde{e}(s), 1 - \tilde{e}(s)\}$

$\tilde{\mathcal{O}}^* \geq \mathcal{O}^*$ regardless of model specification; $\tilde{\mathcal{O}}^* \approx \mathcal{O}^*$ if model is good ($\hat{e} \approx e$)

Leveraging the powerful ML algorithms without worrying about failure modes

Step 2: careful balance check (Difference-in-Means O-value)



Discrepancy Parameters

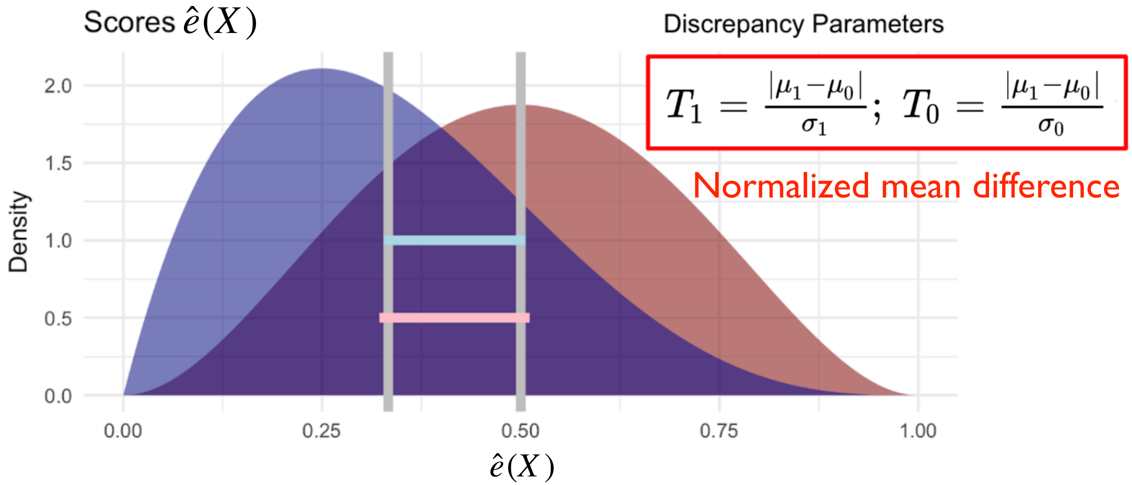
$$\boxed{T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1};\; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}}$$

Normalized mean difference

Intuition:

large $\mathcal{O}^*$ $\Longrightarrow$ smaller discrepancy between $S \mid T = 1$ and $S \mid T = 0$

Step 2: careful balance check (Difference-in-Means O-value)



Theorem

$\mu_1, \sigma_1 \leftarrow$ mean, sd of $S \mid T = 1$

$\mu_0, \sigma_0 \leftarrow$ mean, sd of $S \mid T = 0$

$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1}, \ T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

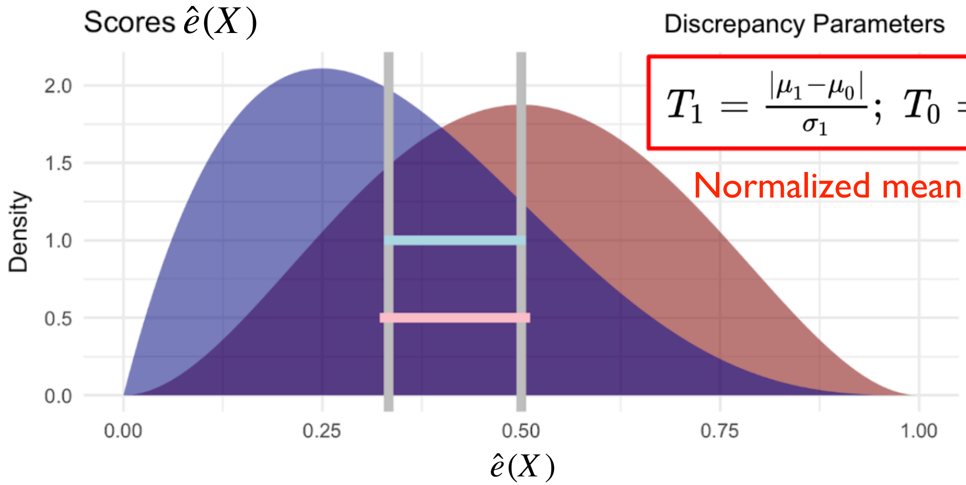
Then $\mathcal{O}^* \leq f(T_1, T_0)$ for a decreasing f

D'Amour, Ding, Feller, Lei, Sekhon ('21)

Intuition:

large $\mathcal{O}^* \implies$ smaller discrepancy between $S \mid T = 1$ and $S \mid T = 0$

Step 2: careful balance check (Difference-in-Means O-value)



Discrepancy Parameters

$$\boxed{T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1};\; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}}$$

Normalized mean difference

Theorem

$\mu_1, \sigma_1 \leftarrow$ mean, sd of $S \mid T = 1$

$\mu_0, \sigma_0 \leftarrow$ mean, sd of $S \mid T = 0$

$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1},\; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Then $\mathcal{O}^* \leq f(T_1, T_0)$ for a decreasing f

D'Amour, Ding, Feller, Lei, Sekhon ('21)

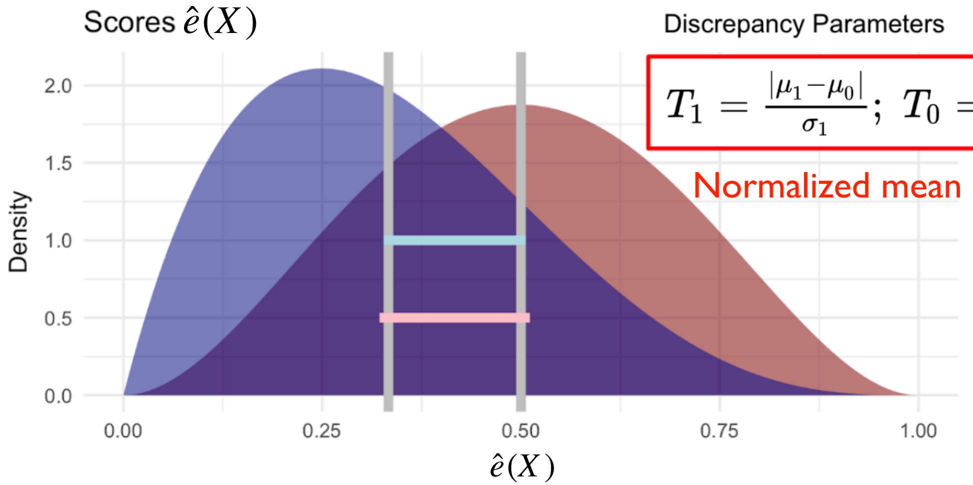
$$f(T_1, T_0) = \frac{1}{2} - \sqrt{\frac{1}{4} - \frac{\pi(1-\pi)}{\max\{\pi T_0, (1-\pi)T_1\}^2 + 1}}$$

Intuition:

large $\mathcal{O}^* \implies$ smaller discrepancy between $S \mid T = 1$ and $S \mid T = 0$

Using information theory Rukhin ('93)

Step 2: careful balance check (Difference-in-Means O-value)



$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1}; \; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Normalized mean difference

Theorem

$\mu_1, \sigma_1 \leftarrow$ mean, sd of $S \mid T = 1$

$\mu_0, \sigma_0 \leftarrow$ mean, sd of $S \mid T = 0$

$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1}, \; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Then $\mathcal{O}^* \leq f(T_1, T_0)$ for a decreasing f

Confidence region $\mathscr{C}$ of $(\mu_1, \sigma_1, \mu_0, \sigma_0)$

$\Longrightarrow$ Upper confidence bound on $f(T_1, T_0)$

$$\hat{\mathcal{O}} = \max_{T_1, T_0 \in \mathscr{C}} f(T_1, T_0)$$

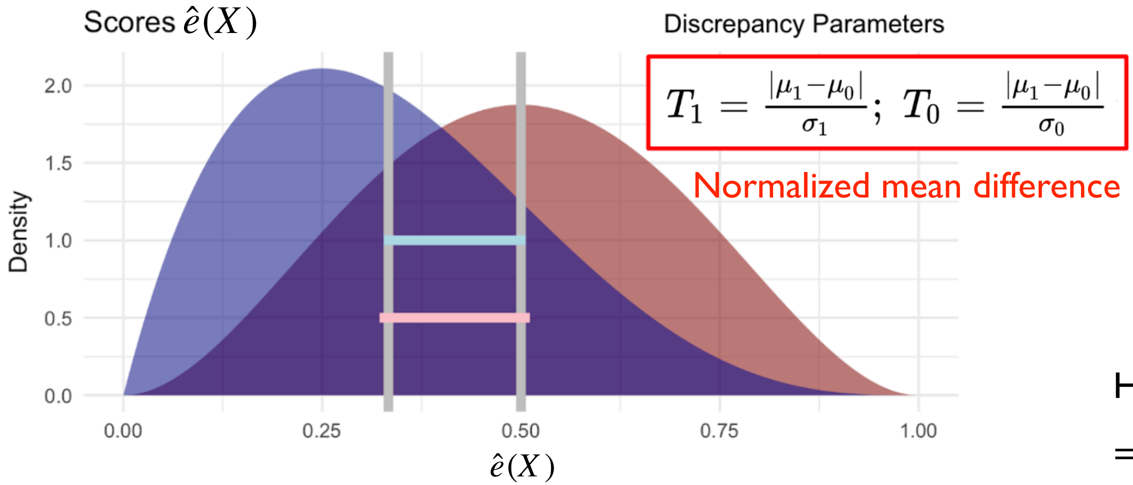
$\Longrightarrow$ DiM O-value

$$\mathbb{P}(\hat{\mathcal{O}} \geq \mathcal{O}^*) \geq \mathbb{P}((\mu_1, \sigma_1, \mu_0, \sigma_0) \in \mathscr{C}) \geq 1 - \alpha$$

Intuition:

large $\mathcal{O}^* \Longrightarrow$ smaller discrepancy between $S \mid T = 1$ and $S \mid T = 0$

Step 2: careful balance check (Difference-in-Means O-value)



Theorem

$\mu_1, \sigma_1 \leftarrow$ mean, sd of $S \mid T = 1$

$\mu_0, \sigma_0 \leftarrow$ mean, sd of $S \mid T = 0$

$$T_1 = \frac{|\mu_1 - \mu_0|}{\sigma_1}, \; T_0 = \frac{|\mu_1 - \mu_0|}{\sigma_0}$$

Then $\mathcal{O}^* \leq f(T_1, T_0)$ for a decreasing f

Hedged-capital bound + Hoeffding-Bentkus

$\Longrightarrow$ Confidence region $\mathscr{C}$ of $(\mu_1, \sigma_1, \mu_0, \sigma_0)$

$\Longrightarrow$ Upper confidence bound on $f(T_1, T_0)$

$$\hat{\mathcal{O}} = \max_{T_1, T_0 \in \mathscr{C}} f(T_1, T_0)$$

$\Longrightarrow$ DiM O-value

$\mathbb{P}(\hat{\mathcal{O}} \geq \mathcal{O}^*) \geq \mathbb{P}((\mu_1, \sigma_1, \mu_0, \sigma_0) \in \mathscr{C}) \geq 1 - \alpha$

Intuition:

large $\mathcal{O}^* \Longrightarrow$ smaller discrepancy between $S \mid T = 1$ and $S \mid T = 0$

Step 2: careful balance check (general scheme)

- Let $P_1 = \mathbb{P}(S \mid T = 1), P_0 = \mathbb{P}(S \mid T = 0)$
- Find an estimable "discrepancy" measure $\Delta(P_0, P_1)$ and a bound $B_\Delta(\mathcal{O}) \downarrow \mathcal{O}$ such that

$$\Delta(P_0, P_1) \leq B_\Delta(\mathcal{O}^*)$$ (population property)

Step 2: careful balance check (general scheme)

- Let $P_1 = \mathbb{P}(S \mid T = 1), P_0 = \mathbb{P}(S \mid T = 0)$
- Find an estimable "discrepancy" measure $\Delta(P_0, P_1)$ and a bound $B_\Delta(\mathcal{O}) \downarrow \mathcal{O}$ such that

$$\Delta(P_0, P_1) \leq B_\Delta(\mathcal{O}^*) \text{ (population property)}$$

- Compute a lower confidence bound for $\Delta(P_0, P_1)$

$$\mathbb{P}(\hat{\Delta}^- \leq \Delta(P_0, P_1)) \geq 1 - \alpha \text{ (sample property)}$$

Step 2: careful balance check (general scheme)

- Let $P_1 = \mathbb{P}(S \mid T = 1), P_0 = \mathbb{P}(S \mid T = 0)$

- Find an estimable "discrepancy" measure $\Delta(P_0, P_1)$ and a bound $B_\Delta(\mathcal{O}) \downarrow \mathcal{O}$ such that

$$\Delta(P_0, P_1) \leq B_\Delta(\mathcal{O}^*) \text{ (population property)}$$

- Compute a lower confidence bound for $\Delta(P_0, P_1)$

$$\mathbb{P}(\hat{\Delta}^- \leq \Delta(P_0, P_1)) \geq 1 - \alpha \text{ (sample property)}$$

- $\hat{\mathcal{O}} = B_\Delta^{-1}(\hat{\Delta}^-)$ is a valid O-value

$$\mathbb{P}(\hat{\mathcal{O}} \geq \mathcal{O}^*) = \mathbb{P}(\hat{\Delta}^- \leq B_\Delta(\mathcal{O}^*)) \geq \mathbb{P}(\hat{\Delta}^- \leq \Delta(P_0, P_1)) \geq 1 - \alpha$$

Step 2: careful balance check (DiM/DiT/DiR/CE O-values)

	Δ	$B_\Delta(\mathcal{O}^*)$	$\hat{\Delta}_-$
DiM	T-stat.	χ^2-divergence	Hedged capital bound Waudby-Smith-Ramdas ('20)
DiT	LR	Trivial algebra	Line-crossing probability Dempster ('59) Generalized Simes' inequality Sarkar ('98)
DiR	AUC	Generalized Neyman-Pearson	Hybrid bound for U-statistics Bentkus ('04), Maurer ('06)
CE	class. error	Formula of Bayes risk	Self-normalized Vapnik bound (Anthony-Shaw-Taylor, '93) Improved Devroye bound (Devroye, '82)

A sketch of DiT O-value

- "Discrepancy" measures:

$$\nu_1 = \sup_{A \in \mathscr{A}} \frac{P_1(A)}{P_0(A)}, \quad \nu_0 = \sup_{A \in \mathscr{A}} \frac{P_0(A)}{P_1(A)}, \quad \mathscr{A} = \{[0,x] : x \in [0,1]\} \cup \{[x,1] : x \in [0,1]\}$$

- Population property:

$$\nu_1 \le \frac{1-\pi}{\pi} \frac{1-\mathcal{O}^*}{\mathcal{O}^*}, \quad \nu_0 \le \frac{\pi}{1-\pi} \frac{1-\mathcal{O}^*}{\mathcal{O}^*}$$

- Induced upper bound on $\mathcal{O}^*$:

$$\mathcal{O}^* \le \min \left\{ \frac{\pi}{\pi + (1-\pi)\nu_0}, \frac{1-\pi}{1-\pi+\pi\nu_1} \right\}$$

A sketch of DiT O-value

- Applying DKWM inequality (not good; just for illustration)

$$\mathbb{P}\left(\sup_{A\in\mathscr{A}} |\hat{P}_t(A) - P_t(A)| \leq \sqrt{\frac{\log(2/\delta)}{2n_t}}\right) \geq 1 - \delta$$

- Deducing lower confidence bounds for ν_1 and ν_0 (Bonferroni on two groups)

$$\hat{\nu}_1^- = \sup_{A\in\mathscr{A}} \frac{\hat{P}_1(A) - \sqrt{\log(4/\alpha)/2n_1}}{\hat{P}_0(A) + \sqrt{\log(4/\alpha)/2n_0}}, \quad \hat{\nu}_0^- = \sup_{A\in\mathscr{A}} \frac{\hat{P}_0(A) - \sqrt{\log(4/\alpha)/2n_0}}{\hat{P}_1(A) + \sqrt{\log(4/\alpha)/2n_1}}$$

- DiT O-value (with known π)

$$\hat{\mathscr{O}}_{\mathrm{DiT}} = \min\left\{\frac{\pi}{\pi + (1-\pi)\hat{\nu}_0^-}, \frac{1-\pi}{1-\pi+\pi\hat{\nu}_1^-}\right\}$$

Application: O-values for Lalonde data

- National Supported Work Demonstration program (Lalonde '86)
- Treatment group has 185 units (Dehejia & Wahba '98)
- 7 control groups: 6 observational and 1 experimental
- DiT O-value; Gradient boosting to estimate propensity scores

	CPS			PSID			RCT		
$(\alpha = .05)$	n_0	$n\hat{\mathcal{O}}$	loss	n_0	$n\hat{\mathcal{O}}$	loss	n_0	$n\hat{\mathcal{O}}$	loss
Raw	15992	42	77%	2490	47	75%	260	222	0%
V2 (Trimmed)	2369	65	65%	253	106	43%			
V3 (Trimmed)	429	118	36%	128	97	47%			

Efficiency loss due to imbalance : $\hat{L} = 1 - \underbrace{n\hat{\mathcal{O}}}_{\text{effective sample size}} \;/\; \underbrace{\min\{n_1, n_0\}}_{\text{effective sample size in an RCT}}$

Extensions

Other measures of overlap

- $\mathcal{O}^* = \min_x \min\{e(x), 1 - e(x)\}$ is motivated by the strict overlap condition for ATE

Other measures of overlap

- $\mathcal{O}^* = \min_x \min\{e(x), 1 - e(x)\}$ is motivated by the strict overlap condition for ATE

- For ATT, the strict overlap condition is weaker:

$$e(X) \leq 1 - \mathcal{O}_0$$

  motivating a modified population overlap slack $\mathcal{O}^* = \min_x \{1 - e(x)\}$

- Partial identification bounds can still be derived

Other measures of overlap

$$\mathcal{O}^* = \min_x \{1 - e(x)\}$$

	CPS		PSID		RCT	
$(\alpha = .05)$	$n\hat{\mathcal{O}}$	loss	$n\hat{\mathcal{O}}$	loss	$n\hat{\mathcal{O}}$	loss
Raw	13235	0%	1060	0%	248	0%
V2 (Trimmed)	1443	0%	77	58%		
V3 (Trimmed)	192	0%	41	78%		

Other measures of overlap

- $\mathcal{O}^* = \min_x \min\{e(x), 1 - e(x)\}$ is motivated by the strict overlap condition for ATE

- For ATT, the strict overlap condition is weaker:

$$e(X) \leq 1 - \mathcal{O}_0$$

  motivating a modified population overlap slack $\mathcal{O}^* = \min_x \{1 - e(x)\}$

- Partial identification bounds can still be derived, though looser

- Can also consider quantile population overlap (much harder to derive bounds):

$$\mathcal{O}_\eta^* = \text{quantile}_\eta(\min\{e(X), 1 - e(X)\})$$

Summary

O-value assesses population overlap with distribution-free guarantees in finite samples

$$\text{Data} \xrightarrow{\text{Machine learning}} \hat{e}(x) \xrightarrow[\text{Computable concentration inequalities}]{\text{Partial identification}} \text{O-value}$$

Thank you!

The paper is available at https://lihualei71.github.io/ovalue.pdf

What can we do with O-values? Model check for trimming

O-value is an upper confidence bound of $\mathcal{O}^*$: $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$

- Common practice: trim samples with extreme **estimated** propensity scores (Crump et al. ’09)
- Want to keep units with $e(x) \in [0.1, 0.9]$. What if the propensity score model is misspecified?

What can we do with O-values? Model check for trimming

O-value is an upper confidence bound of $\mathcal{O}^*$: $\mathbb{P}(\mathcal{O}^* \leq \hat{\mathcal{O}}) \geq 1 - \alpha$

- Common practice: trim samples with extreme **estimated** propensity scores (Crump et al. '09)
- Want to keep units with $e(x) \in [0.1, 0.9]$. What if the propensity score model is misspecified?
- Check if $\hat{\mathcal{O}} \geq 0.1$ for the trimmed samples!

Common practice: marginal balance check

TABLE 1—THE SAMPLE MEANS AND STANDARD DEVIATIONS OF PRE-TRAINING EARNINGS AND OTHER CHARACTERISTICS FOR THE NSW AFDC AND MALE PARTICIPANTS

	Full National Supported Work Sample			
	AFDC Participants		Male Participants	
Variable	Treatments	Controls	Treatments	Controls
Age	33.37	33.63	24.49	23.99
	(7.43)	(7.18)	(6.58)	(6.54)
Years of School	10.30	10.27	10.17	10.17
	(1.92)	(2.00)	(1.75)	(1.76)
Proportion	.70	.69	.79	.80
High School Dropouts	(.46)	(.46)	(.41)	(.40)
Proportion Married	.02	.04	.14	.13
	(.15)	(.20)	(.35)	(.35)
Proportion Black	.84	.82	.76	.75
	(.37)	(.39)	(.43)	(.43)
Proportion Hispanic	.12	.13	.12	.14
	(.32)	(.33)	(.33)	(.35)
Real Earnings	$393	$395	1472	1558
1 year Before	(1,203)	(1,149)	(2656)	(2961)
Training	[43]	[41]	[58]	[63]
Real Earnings	$854	$894	2860	3030
2 years Before	(2,087)	(2,240)	(4729)	(5293)
Training	[74]	[79]	[104]	[113]
Hours Worked	90	92	278	274
1 year Before	(251)	(253)	(466)	(458)
Training	[9]	[9]	[10]	[10]
Hours Worked	186	188	458	469
2 years Before	(434)	(450)	(654)	(689)
Training	[15]	[16]	[14]	[15]
Month of Assignment	−12.26	−12.30	−16.08	−15.91
(Jan. 78 = 0)	(4.30)	(4.23)	(5.97)	(5.89)
Number of Observations	800	802	2083	2193

Lalonde ('86)

Common practice: marginal balance check

- Test for equality?
- Choice of cutoffs?
- Adjust for multiplicity?
- Account for interactions?
- Lack of a summary measure!

TABLE 1—THE SAMPLE MEANS AND STANDARD DEVIATIONS OF PRE-TRAINING EARNINGS AND OTHER CHARACTERISTICS FOR THE NSW AFDC AND MALE PARTICIPANTS

	Full National Supported Work Sample			
	AFDC Participants		Male Participants	
Variable	Treatments	Controls	Treatments	Controls
Age	33.37 (7.43)	33.63 (7.18)	24.49 (6.58)	23.99 (6.54)
Years of School	10.30 (1.92)	10.27 (2.00)	10.17 (1.75)	10.17 (1.76)
Proportion High School Dropouts	.70 (.46)	.69 (.46)	.79 (.41)	.80 (.40)
Proportion Married	.02 (.15)	.04 (.20)	.14 (.35)	.13 (.35)
Proportion Black	.84 (.37)	.82 (.39)	.76 (.43)	.75 (.43)
Proportion Hispanic	.12 (.32)	.13 (.33)	.12 (.33)	.14 (.35)
Real Earnings 1 year Before Training	$393 (1,203) [43]	$395 (1,149) [41]	1472 (2656) [58]	1558 (2961) [63]
Real Earnings 2 years Before Training	$854 (2,087) [74]	$894 (2,240) [79]	2860 (4729) [104]	3030 (5293) [113]
Hours Worked 1 year Before Training	90 (251) [9]	92 (253) [9]	278 (466) [10]	274 (458) [10]
Hours Worked 2 years Before Training	186 (434) [15]	188 (450) [16]	458 (654) [14]	469 (689) [15]
Month of Assignment (Jan. 78 = 0)	−12.26 (4.30)	−12.30 (4.23)	−16.08 (5.97)	−15.91 (5.89)
Number of Observations	800	802	2083	2193

Lalonde ('86)