Econometrics

TA Session 6

Lucia Sauer

2025-11-07

Planification — 2° Part

TA6 — RCT, Matching   Hand in Assignment 6 by 17th Nov

TA7 — IV, RDD   Hand in Assignment 7 by 25th Nov

TA8 — Difference-in-Differences   Hand in Assignment 8 by 2nd Dec

TA9 — Review for the Exam   Week of Dec 8

Overview

- Introduction
- Birthweight and smokers
- Randomized Controlled Trials
- Matching
- Propensity Score Matching

Introduction to Treatment Effects

- Allow us to estimate the causal effect of a treatment on an outcome.

Some examples:

Treatment	Outcome
New drug	Blood Pressure
Surgery	Mobility
Job training	Employment Status
Ad Campaign	Sales

Potential Outcomes

- $Y_i(0)$ outcome for individual i without treatment
- $Y_i(1)$ outcome for individual i with treatment
- D_i treatment indicator (1 if treated, 0 if not)
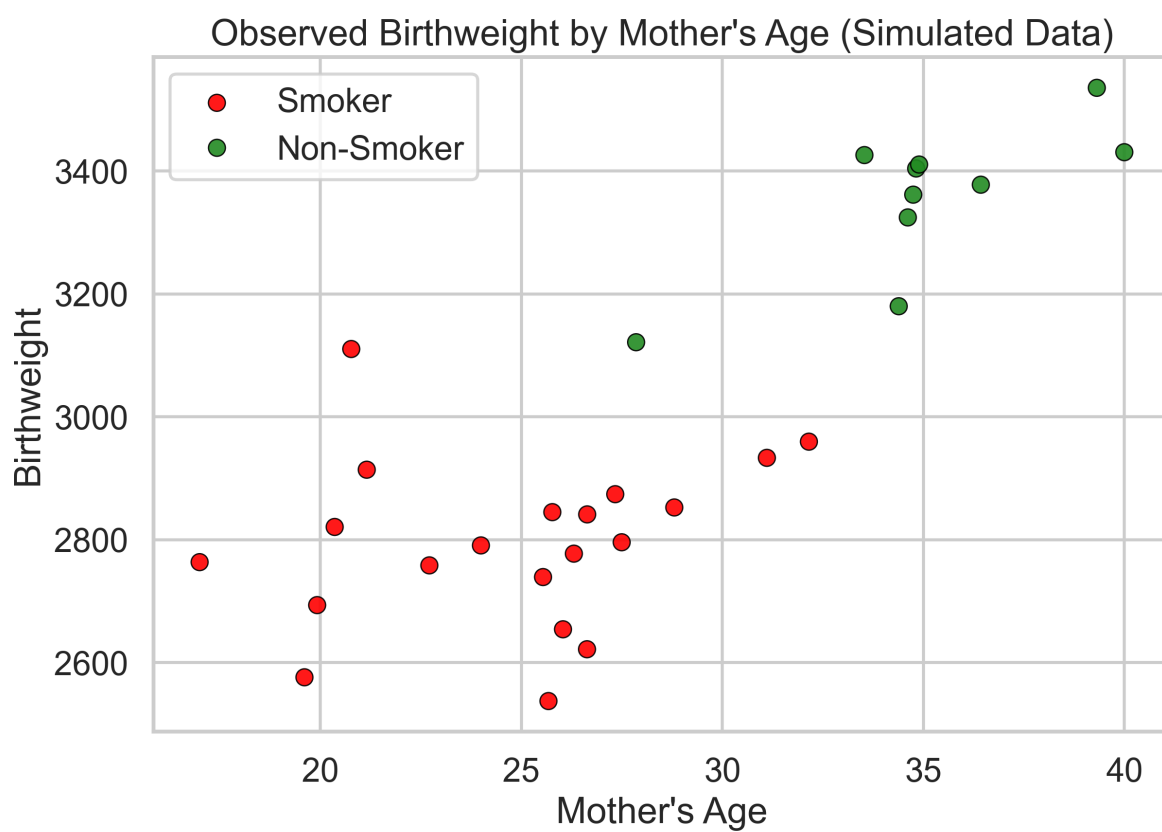- Observed outcome:
$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$
- Causal effect for individual i:
$$\tau_i = Y_i(1) - Y_i(0)$$

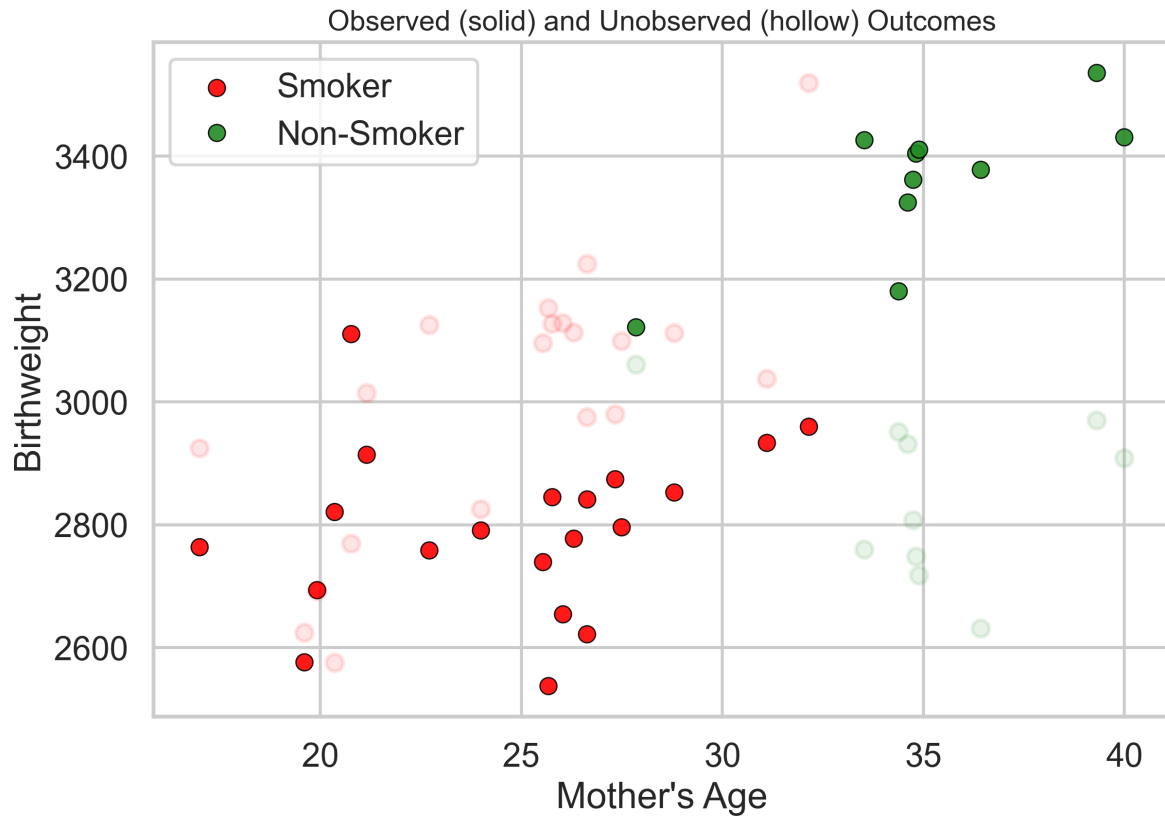But we can never observe both $Y_i(1)$ and $Y_i(0)$ for the same individual!

Example: Birthweight and Smokers

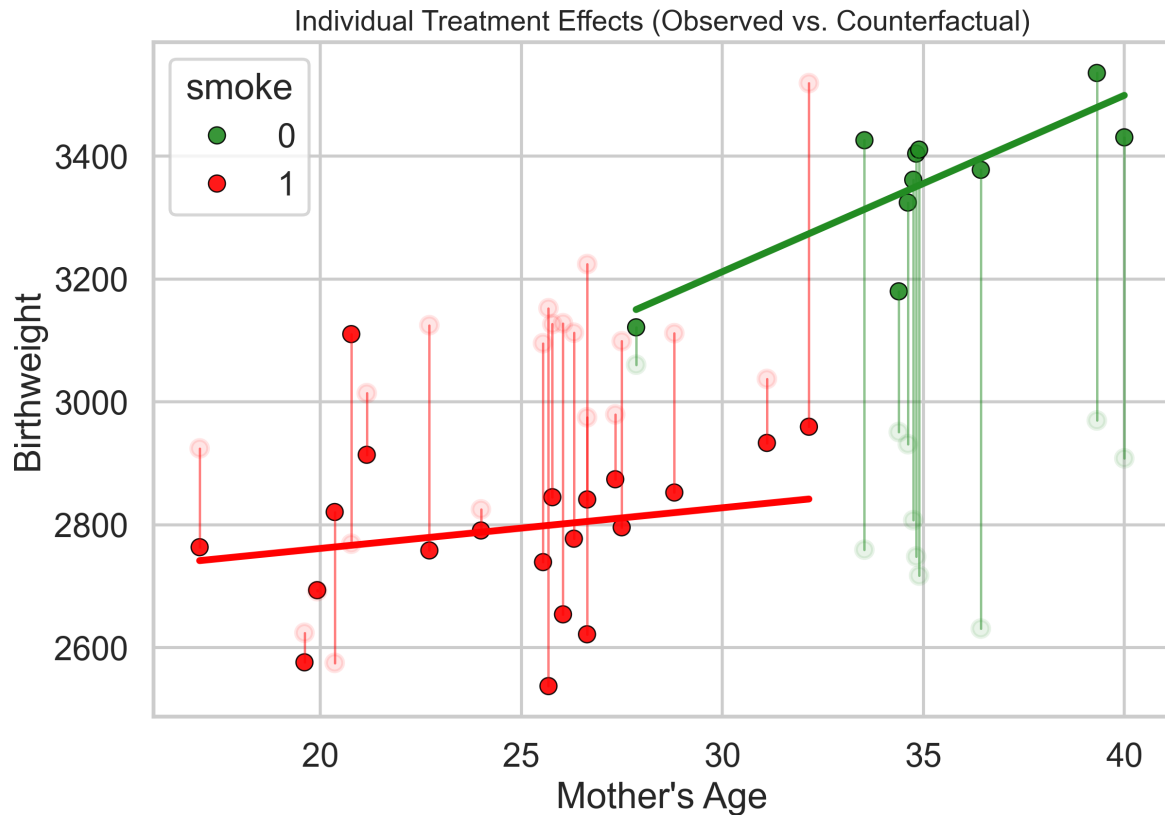*Does smoking during pregnancy affect birthweight?*

Observed Birthweight by Mother's Age (Simulated Data)

Potential outcomes:

- $Y_i(1)$: Birthweight if mother smokes
- $Y_i(0)$: Birthweight if mother does not smoke

Observed (solid) and Unobserved (hollow) Outcomes

- Causal effect: $\tau_i = Y_i(1) - Y_i(0)$
- We can only observe one of these outcomes for each mother.

Individual Treatment Effects (Observed vs. Counterfactual)

- The challenge: how can we estimate an average effect if one of the two outcomes is always missing?

RCTs

Randomly assign individuals to treatment $D_i = 1$ or control $D_i = 0$.

$$Y_i(1), Y_i(0) \perp\!\!\!\perp D_i$$

- Randomization guarantees that, on average:
  - Treated and control groups are **identical in all characteristics**, except for the treatment.
  - Any difference in outcomes can be **causally attributed** to the treatment.

Ham and LaLonde (1996)

Does participation in the job training program lead to higher earnings?

- Outcome: Earnings in 1978 (post-treatment)
- Treatment: Participation in job training program

	train	age	educ	black	hisp	married	nodegree	mosinex	re74	re75	re78	unem74	unem
0	1	37	11	1	0	1	1	13	0.0	0.0	9.93005	1	1
1	1	22	9	0	1	0	1	13	0.0	0.0	3.59589	1	1
2	1	30	12	1	0	0	0	13	0.0	0.0	24.90950	1	1
3	1	27	11	1	0	0	1	13	0.0	0.0	7.50615	1	1
4	1	33	8	1	0	0	1	13	0.0	0.0	0.28979	1	1

Proposed Analysis

> Exercise
>
> 1. What are the average earnings for participants versus non-participants?

```
tabulate train, summarize(re78) means standard
```

Group	Average Earnings	SD of Earnings
Treated	4.55	5.48
Controls	6.35	7.86

> Exercise
>
> 2. Is the difference in mean earnings statistically significant?

$$H_0 : \mu_{treat} - \mu_{control} = 0 \quad H_a : not\ H_0$$

Test statistic using **different** variances:

$$t = \frac{(\bar{Y}_{treat} - \bar{Y}_{control}) - 0}{\sqrt{\frac{s^2_{treat}}{n_{treat}} + \frac{s^2_{control}}{n_{control}}}} \overset{\sim}{\underset{Under~H_0}{}} t_{df}$$

Two-sample t-test: Difference in Means

	Treatment mean	Control Mean	Difference
earnings	6.35	4.55	-1.79**
N	185	260	445

> **Exercise**
>
> 3. Assuming that randomization was properly done, can we estimate the effect of training on earnings using regression?

If randomization was properly done, we can estimate the effect of training on earnings using a simple regression model:

$$\texttt{earnings}_i = \beta_0 + \beta_1 \texttt{D}_i + \epsilon_i$$

> **Exercise**
>
> 4. How does adding control variables affect the estimated treatment effect?

If randomization is successful, adding covariates should not substantially change the ATE.

ATE Comparison: Simple vs. +Controls

	Simple	Simple + Controls Mean
treat_coeff	1.794**	1.625*
se	(0.633)	(0.640)
r2	0.0178	0.0490

Regression Adjustment (RA): Intuition

1. **Fit two regressions**

   - One for treated $(D = 1)$ and one for controls $(D = 0)$:

$$\hat{E}[Y|D = d, X] = \alpha_d + X'\beta_d$$

2. **Predict potential outcomes**

$$\widehat{Y_i(1)}, \ \widehat{Y_i(0)}$$

3. **Compute individual effects**

$$\hat{\tau}_i = \widehat{Y_i(1)} - \widehat{Y_i(0)}$$

4. **Average**

   - , ate: average over **all individuals** $\rightarrow$ ATE
   - , atet: average over **treated only** $\rightarrow$ ATT

Exercise

5. Are the baseline characteristics balanced between treated and control groups?

Covariate	Treated Mean	Control Mean	Difference
re74	2.107	2.096	+0.011
re75	1.267	1.532	−0.265
age	25.054	25.816	−0.762
agesq	677.315	717.395	−40.079
nodegree	0.835	0.708	**+0.127***
married	0.154	0.189	−0.035
black	0.827	0.843	−0.016
hisp	0.108	0.059	**+0.048***

Matching

After examining the covariate balance table...

- Treated and control groups **differ systematically** in key characteristics (e.g., `nodegree`, `hisp`).
- These differences create **confounding** — bias in estimating the treatment effect.
- We need to make the groups **comparable**, as if treatment were randomly assigned.

  Idea: Find in the control group individuals who look similar (in X's) to the treated ones.

Why Not Just Control for Covariates?

- In a regression, we "control for" covariates **by modeling**:

$$Y_i = \alpha + \tau D_i + X_i'\beta + \varepsilon_i$$

  But if treated and control units differ too much in X, regression must **extrapolate** into regions with no comparable controls.

- This leads to **model dependence** and possible **bias** if the functional form is misspecified.

  Matching compares treated and control units **with similar X**, ensuring overlap — we only compare *apples with apples*.

Matching and the Counterfactual

- Recall the potential outcomes framework:

$$Y_i = D_i Y_i(1) + (1 - D_i)Y_i(0)$$

- For treated units $(D_i = 1)$, we observe $Y_i(1)$ but not $Y_i(0)$.

Matching allows us to **reconstruct the missing counterfactual** $Y_i(0)$ using similar individuals from the control group.

Under the **Conditional Independence Assumption (CIA):**

$$(Y_i(1), Y_i(0)) \perp D_i \mid X_i$$

treatment is "as good as random" given (X).

---

1. Nearest Neighbor Matching
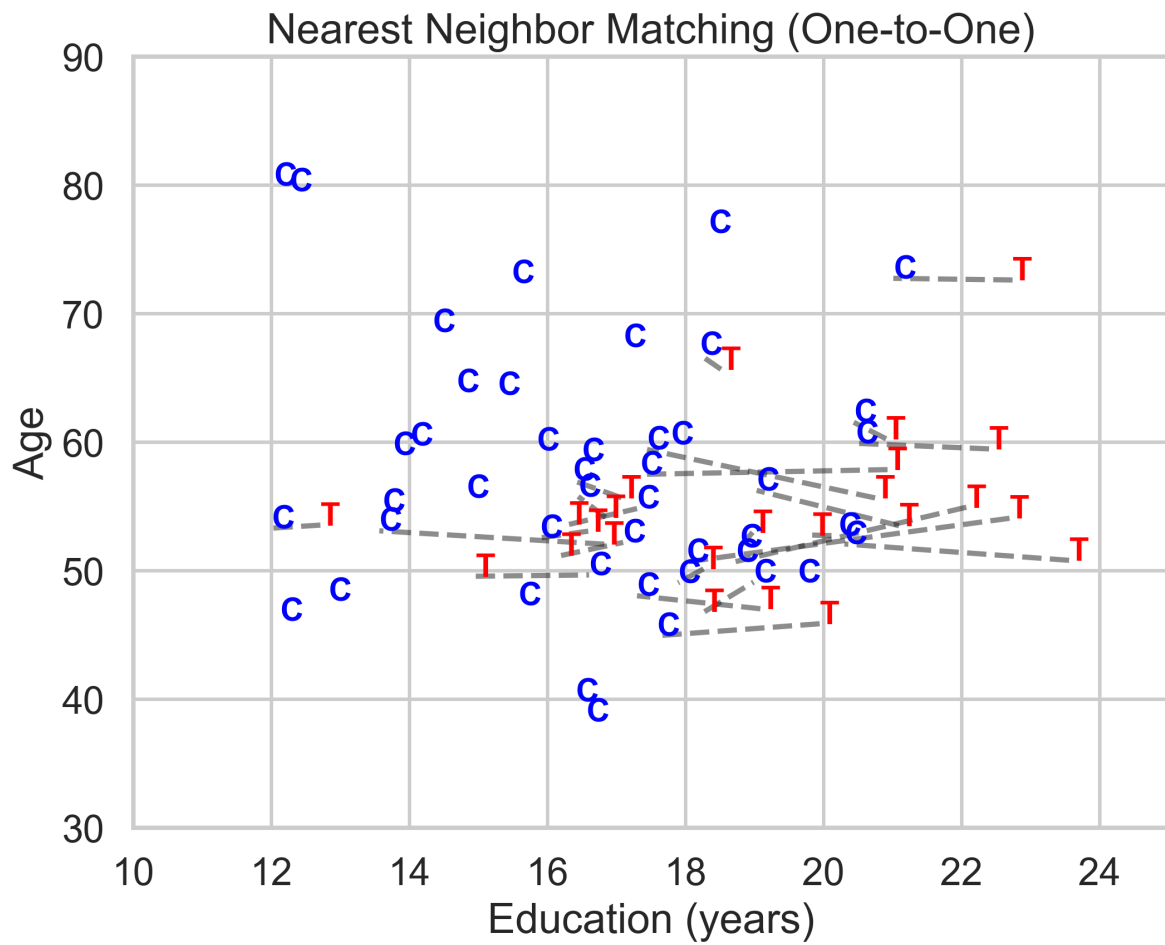
> Procedure
>
> 1. Preprocess:
>   - Distance $(X_c, X_t) = \sqrt{(X_c - X_t)' S^{-1} (X_c - X_t)}$
>   - Match each treated unit to nearest control unit
>   - Control units: not reused; prune if unused
>   - Prune matches if distance > caliper
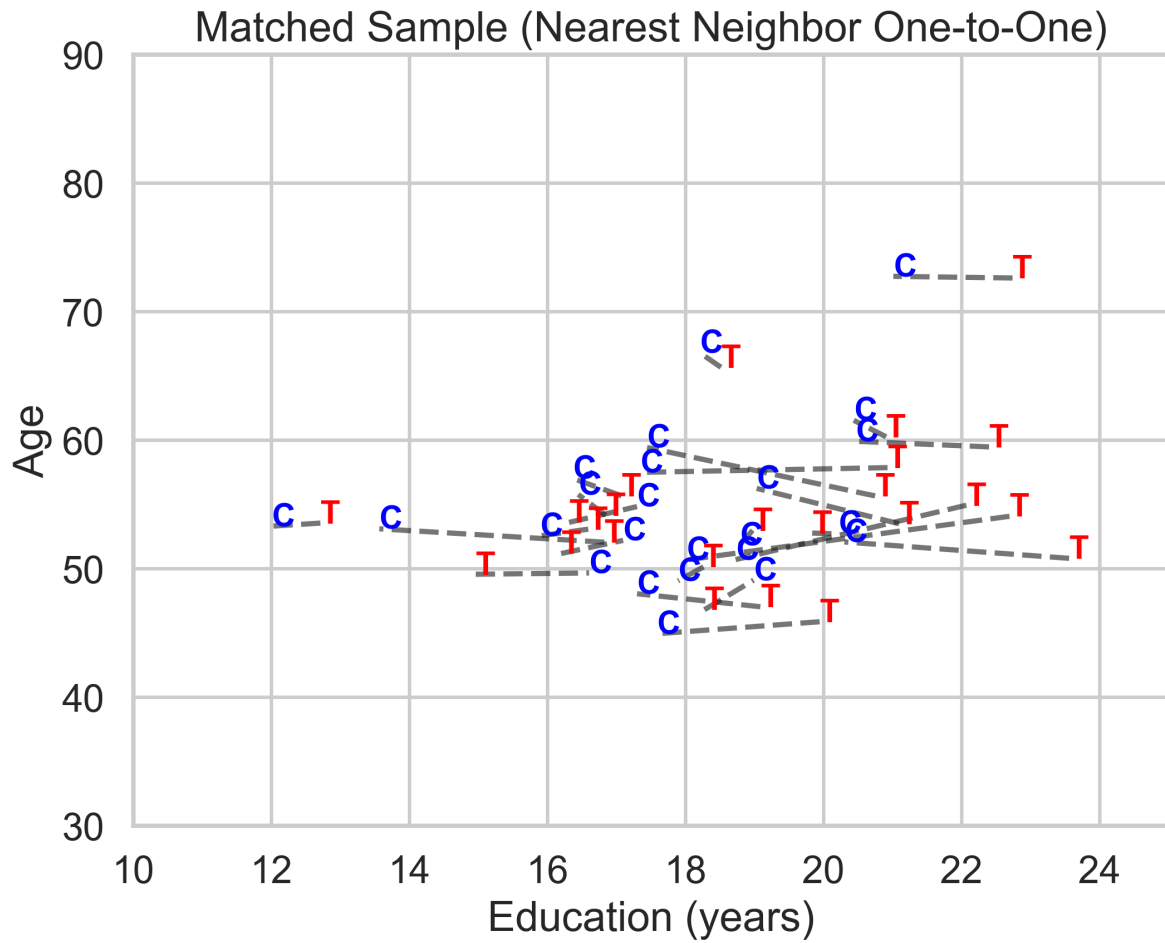>
> 2. Estimate treatment effect on matched sample

Building Intuition

- Consider 2 covariates, Education and Age. No dependent variable here.
- We have some Treated units and some Control units.

Treated (T) and Control (C) Units by Covariates

Matching treated and control units

- For every Treated unit we find the closest Control unit.
- Order in which your data are set up actually determines the match.
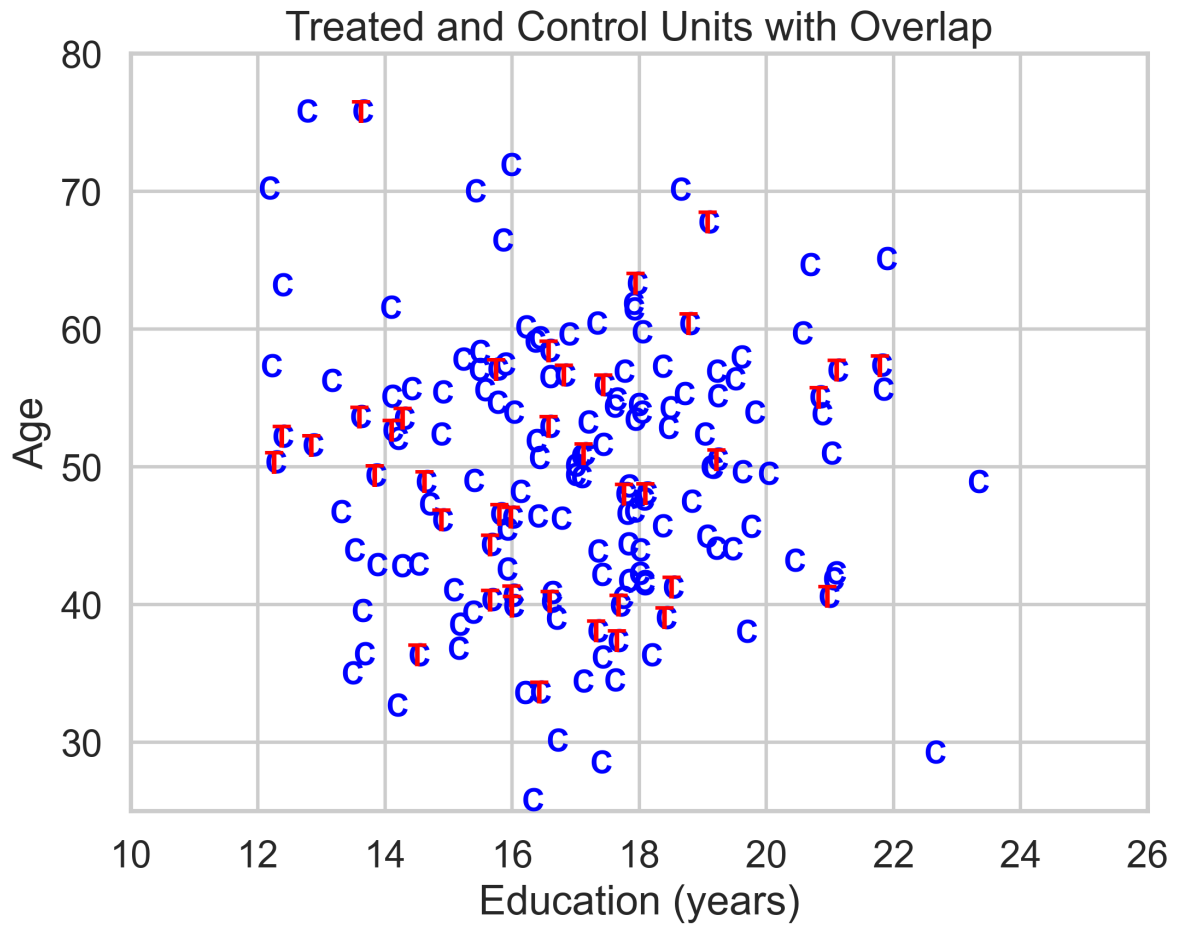
Nearest Neighbor Matching (One-to-One)

Pruning Unmatched Units

- We can get rid of observations that did not achieve a match.
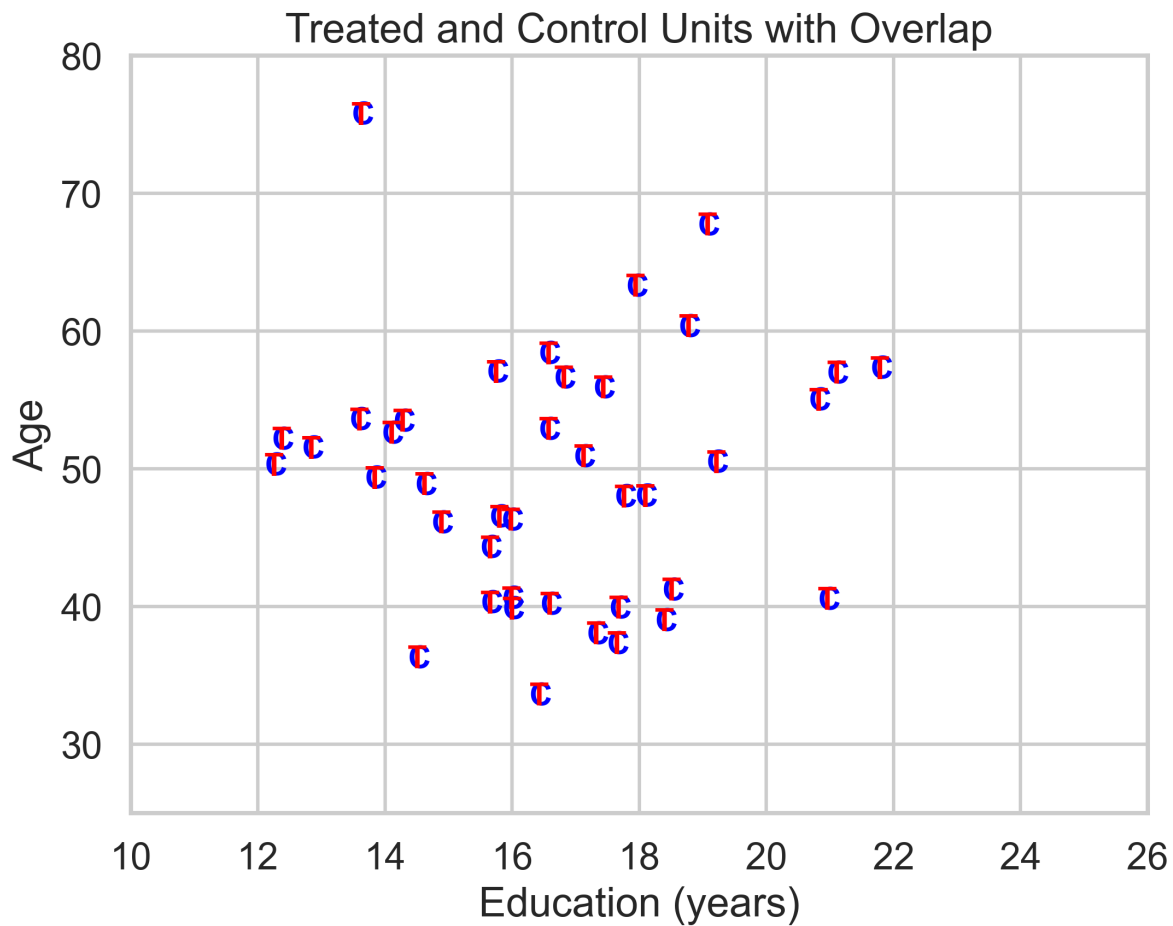- Now we have something closer to a randomized experiment.

Matched Sample (Nearest Neighbor One-to-One)

---

2. **Exact Matching**

- Such a beautiful dataset! Tons of control units.
- Every red T has a blue C on top.

Treated and Control Units with Overlap

- No matter what we do with Age and Education, it couldn't predict the difference between treatment and control. They can't close bias.
- We can look at the causal effect on the outcome variable by just taking the difference in means.

14

Treated and Control Units with Overlap

The Curse of Dimensionality

When there are many covariates $(X_1, X_2, ..., X_k)$:

- Hard to find exact matches in all dimensions.
- Matching becomes sparse or impossible.

Solution: Summarize all covariates into a single number:

$$\pi(X) = P(D = 1|X)$$

the **propensity score**.

Then match treated and controls with similar $\pi(X)$ values instead of the full vector X.

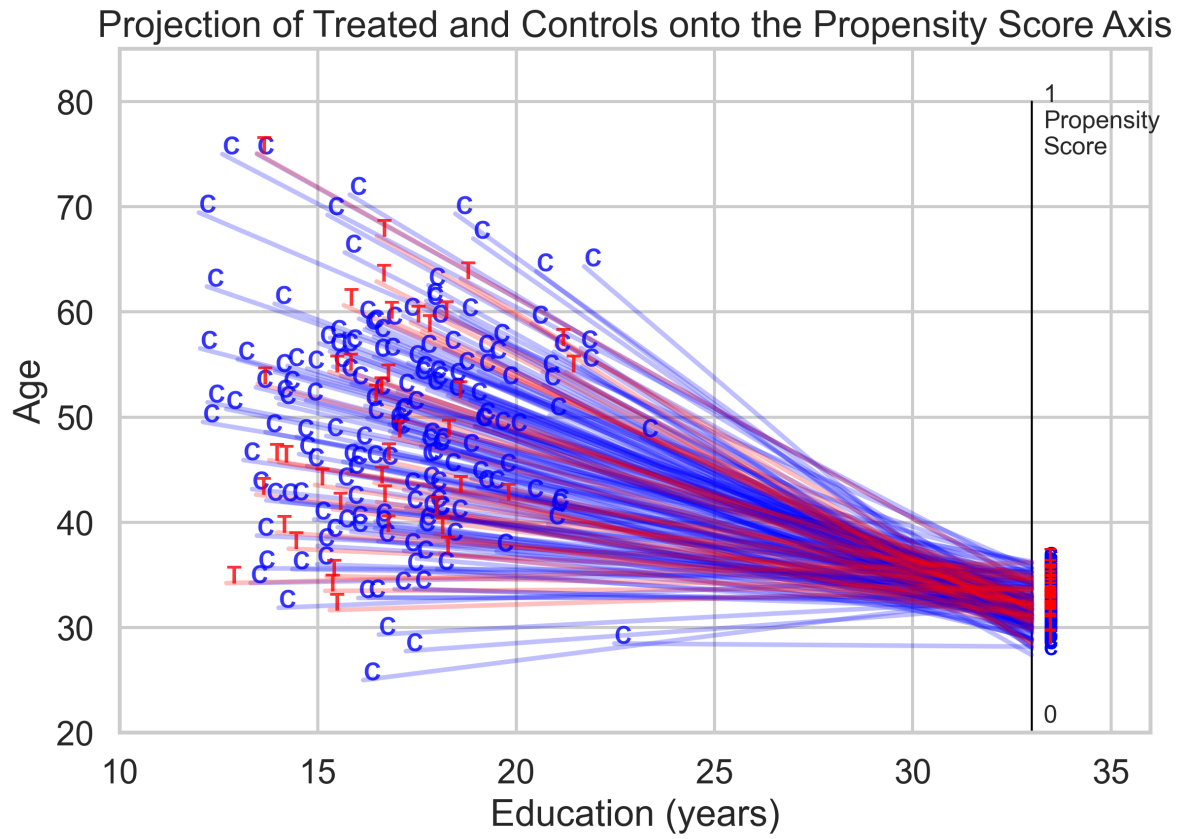### 3. Propensity Score Matching

---

Procedure

1. Preprocess:

   - Reduce k elements of X to scalar $\pi_i = P(T_i = 1|X) = \frac{1}{1+e^{-X_i'\beta}}$
   - Compute distance $(X_c, X_t) = |\pi_c - \pi_t|$
   - Match each treated unit to the nearest control unit.
   - Control units: not reused; prune if unused
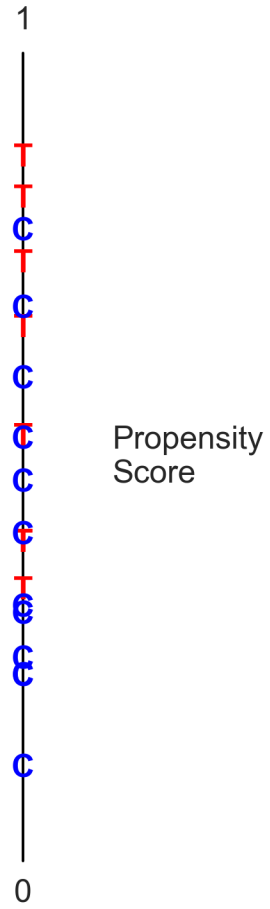   - Prune matches if distance $>$ caliper

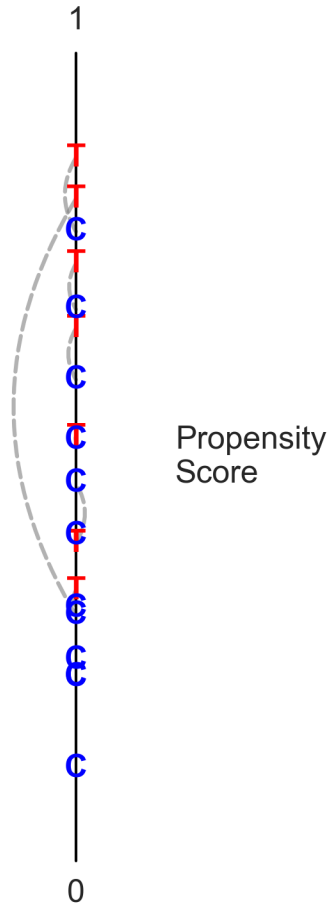2. Estimate treatment effect on matched sample

---

Building Intuition

- Create a Propensity Score π_i from 0 to 1.
- Project each of those observations into a one-dimensional space.

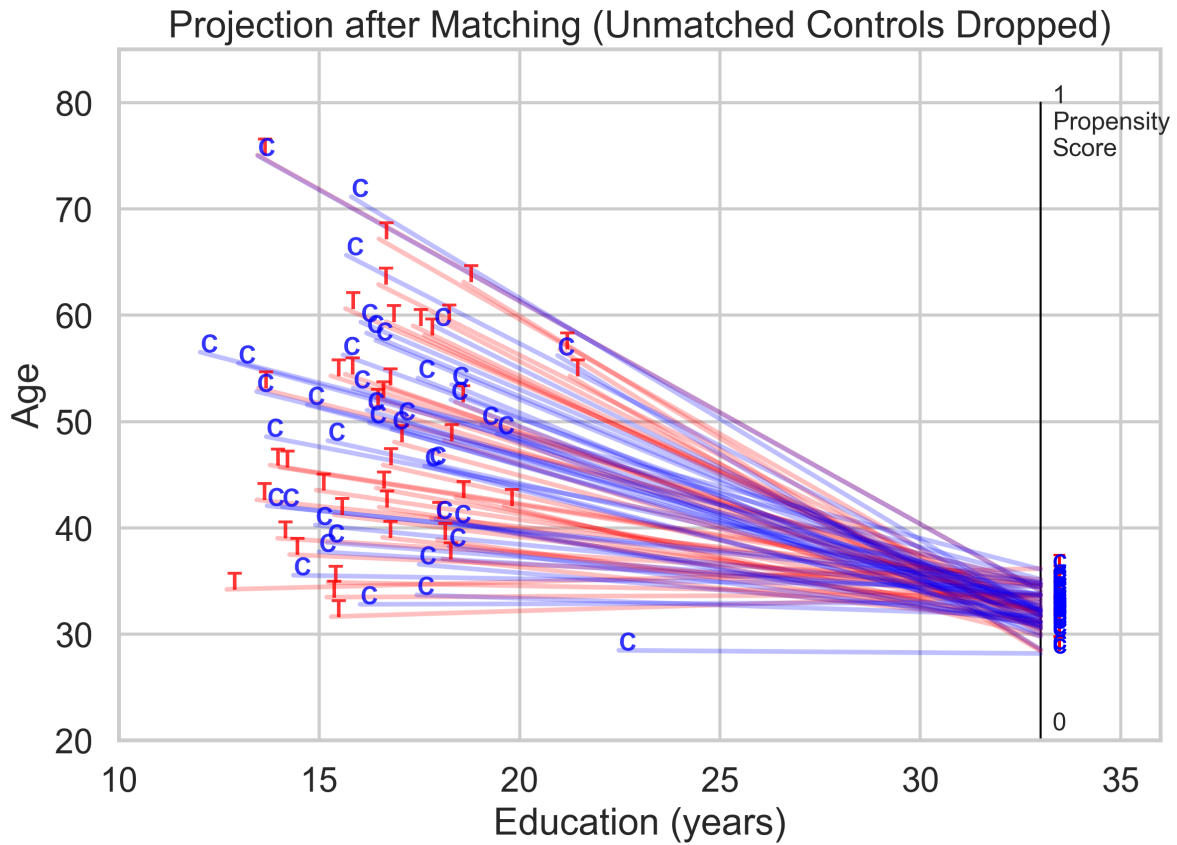Projection of Treated and Controls onto the Propensity Score Axis

Then we perform the matching on this one-dimension space.

Then we perform the matching on this one-dimension space.

Finally, we project them back to the multivariate space. This is the subset of the data that goes into the analysis.

Projection after Matching (Unmatched Controls Dropped)

Estimating the Propensity Score

1. **Estimate the Propensity Score**

$$P(D_i = 1|X_i) = \Phi(X_i'\beta)$$

   - Predicts each individual's probability of being treated (`myscore`).

2. **Describe the Distribution**

   - Reports min, max, mean, and percentiles of the estimated scores.
   - Identifies *common support*: where treated and controls overlap.

3. **Divide into Blocks (Stratification)**

   - Splits the sample into intervals of the propensity score (`myblock`).
   - In each block, treated and controls have similar scores on average.

4. **Check the Balancing Property**

   - Tests whether within each block, covariates X_i are balanced across treated and control units:
   $$D_i \perp X_i \mid p(X_i)$$
   - If balance is satisfied $\rightarrow$ the propensity score is valid.

- `myscore`: the estimated propensity score for each individual

- `myblock`: the block (interval) of the score where the unit falls