Introduction and Review
○○○○○○○○○○○○

Multivariate Regression
○○○○○○○○○○○○

Matching
○○○○○○○○○○○○○○○○○○○○○○

# Causal Inference: Selection on Observables, Multiple Regression and Matching

Ken Stiller

16th May 2022

Introduction and Review
●○○○○○○○○○○○

Multivariate Regression
○○○○○○○○○○○○

Matching
○○○○○○○○○○○○○○○○○○○○

# Reminders

▶ First assignments are due next Friday 12pm.

▶ Next week's drop-in session will be taking place online.

## Last week

- ▶ How can randomization solve the selection bias issue

- ▶ Important challenges to RCTs

  1. Translating your research question to a treatment and measurable outcome
  2. Randomization of the treatment
  3. Power of the experiment
  4. Validity of the design
  5. Correct treatment and mechanisms
  6. Non-Compliance (in field experiments)

## This week

- ▶ Non Compliance: Does information increase turnout?

- ▶ What happens if we cannot randomize?

- ▶ Selection on observables and the prospects for causal inference

- ▶ Multivariate Regression (less)

- ▶ Matching (more)

*Let's begin with a recap!*

# A Challenge in Field Experimentation

- ▶ In field experiments we have little control over our units/subjects

- ▶ Often times, we assign units to the treatment group, but they are not eventually treated

**We call this Non-Compliance**

- ▶ Often times, our inability to treat units can correlate with the outcome or the treatment

- ▶ This can have serious consequences for the estimation of the treatment effects

- ▶ Discarding those not treated although assigned to the treatment group: compliance is often self-selected

Typical Example: Canvassing and Turnout!

# Calculating Effects in Field Experiments

## Complier Average Causal Effect

$$CACE = \frac{ITT}{E[D_{1i}]}$$

i.e. the ATE among compliers

Example from GG(2012), pp 150-1:

|                                   | T              | C              |
|-----------------------------------|----------------|----------------|
| Turnout by those contacted        | 54.43 (395)    |                |
| Turnout among those not contacted | 36.48 (1,050)  | 37.54 (5,645)  |
| Overall Turnout                   | 41.38 (1,445)  | 37.54 (5,645)  |

Treatment Status (the CACE denominator):

$$D_{1i} = \frac{395}{1445} - \frac{0}{5,645} = 0.273$$

i.e. 27,3% actually treated from those assigned to treatment

## Calculating Effects in Field Experiments

|                                    | T             | C             |
| ---------------------------------- | ------------- | ------------- |
| Turnout by those contacted         | 54.43(395)    |               |
| Turnout among those not contacted  | 36.48(1,050)  | 37.54 (5,645) |
| Overall Turnout                    | 41.38 (1,445) | 37.54 (5,645) |

Treatment Status (the CACE denominator):

$$D_{1i} = \frac{395}{1445} - \frac{0}{5,645} = 0.273$$

The ITT (i.e. the CACE numerator) can be calculated as

$$41.38 - 37.54 = 3.84\%$$

and the effect among compliers as

$$CACE = \frac{3.84}{0.273} = 14.1\%$$

# Table of Contents

Introduction and Review

Multivariate Regression

Matching

# Equivalence of Randomization to Selection on Observables

**Q1:** What is our <u>empirical</u> expectation when we randomize?
**Q2:** How could we <u>know</u>?

**Example:** Balance Table

| Covariates | Control | Treatment | Difference | Sig. |
|---|---|---|---|---|
| Age (years) | 41.6 | 41.2 | .40 | .40 |
| % Democrat | .76 | .78 | .02 | .33 |
| % Republican | .04 | .03 | .01 | .21 |
| % Female | .68 | .70 | .02 | .58 |
| % Living in Des Moines | .53 | .55 | .02 | .46 |
| % Voted 2008 primary | .32 | .32 | .00 | .93 |
| % Voted 2008 general | .93 | .93 | .00 | .94 |
| % Voted 2010 primary | .22 | .22 | .00 | .87 |

Randomization Check (control vs treatment groups *in pre-treatment outcomes*)

# Equivalence of Randomization to Selection on Observables

▶ The randomization check is a good indication that there is `balance` in the control and treatment group

▶ In other words, the distribution of pre-treatment variables (let's call them $X$) in the control and treatment group is very similar.

▶ In cases where the causal factor is not randomly assigned, we seek to limit our analysis to exchangeable units. i.e. units that are very similar in the distributions of the $X$s in the control and treatment group.

▶ Under -strict- assumptions, we remove bias by conditioning

Introduction and Review
○○○○○○○○○●○○

Multivariate Regression
○○○○○○○○○○○○

Matching
○○○○○○○○○○○○○○○○○○○○○○○○○

## Definitions

### Covariates

A covariate is a variable that is predetermined with respect to treatment $D_i$: $X_0 = X_1$, i.e. its value does not depend on the value of $D_i$.

▶ Does not imply that $X$ and $D$ are independent
▶ Predetermined variables are often time invariant (sex, race, etc.), but time invariance is not necessary

### Outcomes

Those variables,$Y$, that are (possibly) not predetermined are called outcomes (for some individual $i$ , $Y_{0,i} \neq Y_{1i}$)

In general, one should not condition on outcomes, because this may induce `post-treatment bias`.

Introduction and Review
○○○○○○○○○○○●○

Multivariate Regression
○○○○○○○○○○○○

Matching
○○○○○○○○○○○○○○○○○○○○○○○○○

# Removing Bias by Conditioning

We need to satisfy: $P(D_i = 1) \perp\!\!\!\perp Y_0, Y_1$

Experiments

▶ Randomization ensures unconfoundedness without selection on observables: $P(D_i = 1) \perp\!\!\!\perp Y_0, Y_1$ which translates into:

▶ $E(Y_1|D = 1) = E(Y_1|D = 0)$ &

▶ $E(Y_0|D = 1) = E(Y_0|D = 0)$

Typical Observational Studies

▶ Unconfoundedness can be assumed to hold only after conditioning on a set of pre-treatment variables: $P(D_i = 1|X_i) \perp\!\!\!\perp Y_0, Y_1$ which translates into:

▶ $E(Y_1|D = 1, X) = E(Y_1|D = 0, X)$ &

▶ $E(Y_0|D = 1, X) = E(Y_0|D = 0, X)$

# Conditioning on Observables

► Regression

► Matching

## Balancing

► All studies have a common goal: to **balance** the distributions of covariates for units which are treated and units untreated.

► They differ in how they try to achieve **balance**. In some instances estimation of causal effects happens (seemingly) simultaneously with the attempt to maximise balance on the observables (e.g. regression).

► In others, these two steps are clearly distinguished, with the design stage being the first stage in which balance is attempted and estimation follows after balance is achieved (e.g. matching).

Introduction and Review
○○○○○○○○○○○○

**Multivariate Regression**
●○○○○○○○○○○○

Matching
○○○○○○○○○○○○○○○○○○○○○○○○○

# Multivariate Regression and Causality

Recall the bivariate regression:

$$y_i = \alpha + \beta D_i + \epsilon_i$$

▶ If we believe that the treatment of $D_i$ is as good as randomly assigned, the $\beta$ coefficient will have a causal interpretation of the regression

Let's imagine we are interested in the effects of reading the Daily Mail on political preferences (voting Conservative):

$$\text{Conservative}_i = \alpha + \beta \text{Daily Mail}_i + \epsilon_i$$

▶ Is reading the Daily Mail as good as randomly assigned?
▶ What other variables affect both reading the Daily Mail and voting Conservative for an individual $i$?

# Connecting Multivariate Regression to Potential Outcomes Framework

We would like to compare:

- ▶ An individual $i$ that reads the Daily Mail

- ▶ to the **same** individual $i$ that does not read the Daily Mail

Multivariate regression allows us to compare:

- ▶ An individual $i$ that reads the Daily Mail with income $I$, education levels $E$, occupation $O$, living in region $R$ and Age $A$

- ▶ to an individual $j$ with same income $I$, same education levels $E$, same occupation $O$, living in same region $R$ and same Age $A$ that does not read the Daily Mail

Introduction and Review
00000000000

Multivariate Regression
00●000000000

Matching
000000000000000000000

# Multivariate Regression: Other Variables Leading to Selection Bias?

Let's imagine we are interested in the effects of reading the Daily Mail on political preferences (voting Conservative):

$$\text{Conservative}_i = \alpha + \beta \text{Daily Mail}_i + \epsilon_i$$

Imagine though the true model specification is:

$$\text{Conservative}_i = \alpha + \beta \text{Daily Mail}_i + \gamma Age_i + u_i$$

▶ Adding another variable we suspect may correlate with both the treatment and the outcome is called 'conditioning on $X_i$' in non-experimental data ('controlling for').

▶ The idea is 'holding $X_i$ constant' - we eliminate the effect of $X_i$ on the treatment before comparing treated and control outcomes

Introduction and Review
0000000000000

Multivariate Regression
0000000000000

Matching
0000000000000000000000000

## Notations

Recall the bivariate regression:

$$y_i = \alpha + \beta^S D_i + \epsilon_i \qquad (1)$$

Imagine though the true model specification is:

$$y_i = \alpha + \beta^L D_i + \gamma X_i + \epsilon_i \qquad (2)$$

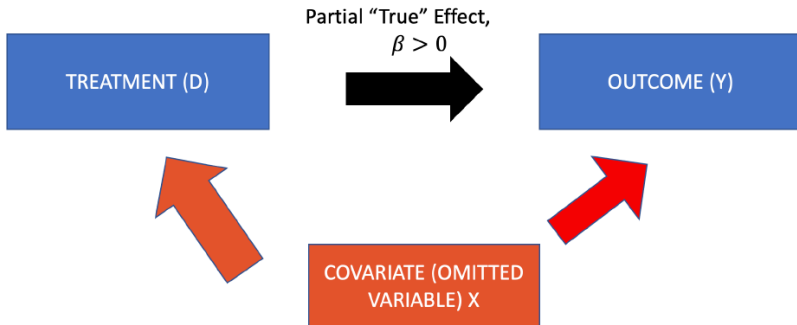Will the estimated $\hat{\beta^S}$ from Equation 1 be biased if we don't condition on $X$, i.e. Age?

$$\beta^S = \frac{cov(Y_i, D_i)}{var(D_i)} = \beta^L + \gamma \times \delta(D_i, X_i)$$

where $\delta(D_i, X_i)$ is the correlation between the treatment and covariate (or the regression coefficient from a regression of $X_i$ on $D_i$).
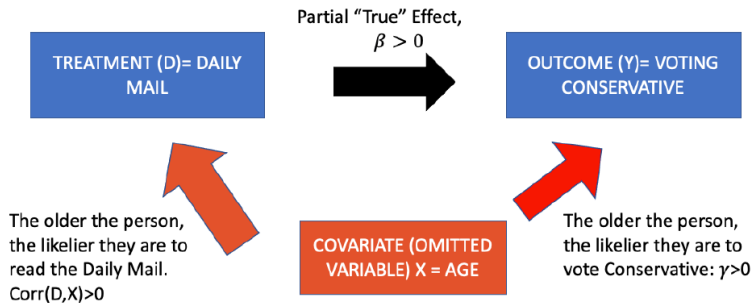
Introduction and Review
○○○○○○○○○○○

Multivariate Regression
○○○○○●○○○○○○

Matching
○○○○○○○○○○○○○○○○○○○○

# The more the covariates the better? Can we remove every omitted variable bias?

▶ If $\gamma \neq 0$ but $\delta(D_i, X_i) = 0$, we do not need to include the covariate for OVB purposes. What about precision of the estimate?

▶ We should not include covariates that are an outcome of the treatment themselves (bad controls), as they reintroduce selection bias.

▶ Thinking about correlation across different covariates

▶ Thinking about the direction of omitted variable bias and changes in estimates as we introduce more covariates

# Good Control/Covariate

# Good Control/Covariate: Example

Partial "True" Effect,
$\beta > 0$

TREATMENT (D)= DAILY MAIL

OUTCOME (Y)= VOTING CONSERVATIVE

The older the person, the likelier they are to read the Daily Mail. Corr(D,X)>0

COVARIATE (OMITTED VARIABLE) X = AGE

The older the person, the likelier they are to vote Conservative: $\gamma > 0$

# Bad Control/Covariate

# Bad Control/Covariate

Introduction and Review
○○○○○○○○○○○○

Multivariate Regression
○○○○○○○○○●○○

Matching
○○○○○○○○○○○○○○○○○○○○○○○
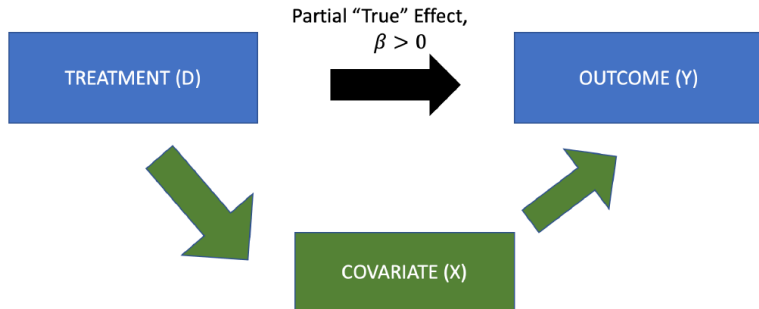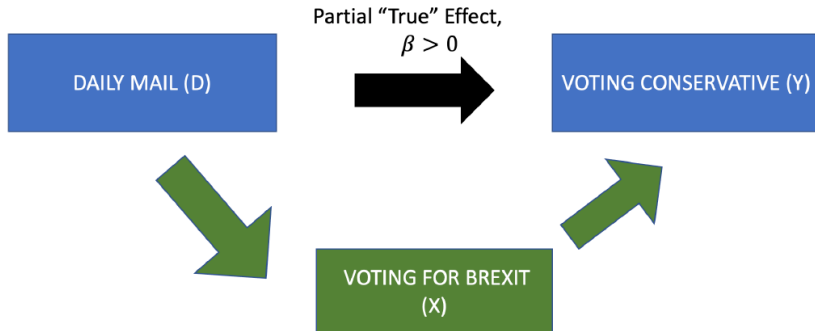
# Intuition: Bad Controls Reintroduce Selection Bias

- ▶ Imagine that indeed there was an RCT that randomly assigned Daily Mail readership to individuals.
- ▶ However, when we estimate the effects of Daily Mail readership, we add 'Voting for Brexit' as a covariate as we suspect it could predict voting for the Conservative party.
- ▶ Issue: voting for Brexit is itself an outcome of Daily Mail readership
- ▶ That means we are comparing two things: i) a person that does read the Daily Mail and voted for Brexit to a person who does not read the Daily Mail and voted for Brexit, and ii) a person that does read the Daily Mail and did not vote for Brexit to a person who does not read the Daily Mail and did not vote for Brexit
- ▶ But in this comparison, these are definitely not the right counterfactuals

# Goal of Conditioning

**Aim**: Once we condition on $X_i$, we believe that the treatment of $D_i$ becomes as good as randomly assigned. Then the $\beta$ coefficient will have a causal interpretation of the regression.
*Is this feasible with conditioning on a few covariates?*

Imagine a model specification:

$$\text{Political Preferences}_i = \alpha + \beta Daughters_i + u_i$$

Can we interpret $\beta$ as a causal estimate? Is there anything we can condition on to make **the number of daughters** as good as randomly assigned?

## Goal of Conditioning

**Aim**: Once we condition on $X_i$, we believe that the treatment of $D_i$ becomes as good as randomly assigned. Then the $\beta$ coefficient will have a causal interpretation of the regression.
*Is this feasible with conditioning on a few covariates?*

Is there anything we can condition on to make **the number of daughters** as good as randomly assigned?
Imagine a model specification:

$$\text{Political Preferences}_i = \alpha + \beta_1 Daughters_i + \beta_2 Children + \epsilon_i$$

Can we interpret $\beta_1$ as a causal estimate?

# Table of Contents

# Matching

### The Underlying Logic

Think of matching as a way to address the missing data problem, by "imputing" missing observations for potential outcomes, using observed outcomes from units chosen on the basis of information about a set of $X's$, which—we believe—drive subjects into their treatment status.

So, if $X$ denotes a set of pre-treatment characteristics for subjects, matching is based on the following assumption:

### Unconfoundedness

$Y_1, Y_0 \perp\!\!\!\perp D|X$

# Matching

▶ We seek to find treated and untreated units that are exchangeable

▶ In Matching, we search across "covariates" for similar units that only differ in terms of $D_i$.

▶ Recall JS Mill's idea that across all potential causes we observe similar patterns, while only the true cause is dissimilar.

▶ Imagine a dataset with characteristics such as height, hair colour, and instrument choice: You are looking for the tall, blonde units who plays the guitar and was treated and the tall blonde guitarist who was untreated.

▶ Setting aside other possible combinations of variables, you could estimate the effect $D_i$ on $Y_i$

▶ A better example might serve us well!

Introduction and Review
○○○○○○○○○○○

Multivariate Regression
○○○○○○○○○○○

Matching
○○○●○○○○○○○○○○○○○○○○

# Matching: A Running Example,MPs for Sale?

### Research Question

What is the effect of serving in Parliament on politicians' wealth?

### Definition: Treatment

$D_i$ : Indicator of treatment status for politician $i$

$$D_i = \left\{ \begin{array}{ll} 1 & \text{if } i \text{ was elected into Parliament} \\ 0 & \text{if } i \text{ was not elected into Parliament.} \end{array} \right.$$

### Definition: Observed Outcome

$Y_i$ : Observed wealth at death for politician $i$

### Definition: Potential Outcomes

$Y_{0i}$ and $Y_{1i}$: Potential Outcomes for politician $i$

# What we Observe

An Example with 10 candidates:

| | $D_i$=Won? |
|---|---|
| Candidate 1 | Yes=1 |
| Candidate 2 | Yes=1 |
| Candidate 3 | No=0 |
| Candidate 4 | No=0 |
| Candidate 5 | No=0 |
| Candidate 6 | Yes=1 |
| Candidate 7 | No=0 |
| Candidate 8 | No=0 |
| Candidate 9 | Yes=1 |
| Candidate 10 | Yes=1 |

# Selection bias?

(Eggers & Hainmueller 2009) Are there returns to wealth in politics? For which ideology?

**TABLE 1.  Gross Wealth at Death (Real 2007 GBP) for Competitive Candidates Who Ran for House of Commons Between 1950 and 1970 (Estimation Sample)**

|  | Mean | Min. | 1st Qtr. | Median | 3rd Qtr. | Max. | Obs. |
|---|---|---|---|---|---|---|---|
| **Both Parties** | | | | | | | |
| All candidates | 599,385 | 4,597 | 186,311 | 257,948 | 487,857 | 12,133,626 | 427 |
| Winning candidates | 828,379 | 12,111 | 236,118 | 315,089 | 722,944 | 12,133,626 | 165 |
| Losing candidates | 455,172 | 4,597 | 179,200 | 249,808 | 329,103 | 8,338,986 | 262 |
| **Conservative Party** | | | | | | | |
| All candidates | 836,934 | 4,597 | 192,387 | 301,386 | 743,342 | 12,133,626 | 223 |
| Winning candidates | 1,126,307 | 34,861 | 252,825 | 483,448 | 1,150,453 | 12,133,626 | 104 |
| Losing candidates | 584,037 | 4,597 | 179,259 | 250,699 | 485,832 | 8,338,986 | 119 |
| **Labour Party** | | | | | | | |
| All candidates | 339,712 | 12,111 | 179,288 | 250,329 | 298,817 | 7,926,246 | 204 |
| Winning candidates | 320,437 | 12,111 | 193,421 | 254,763 | 340,313 | 1,036,062 | 61 |
| Losing candidates | 347,934 | 40,604 | 177,203 | 243,526 | 295,953 | 7,926,246 | 143 |

Additional Covariates:

▶ Education

▶ Aristocrat

▶ Gender

▶ Schooling

# Selection bias?

(Eggers & Hainmueller 2009) Are there returns to wealth in politics? For which ideology?

Additional Covariates:

▶ Education

▶ Aristocrat

▶ Gender

▶ Schooling

## Back to the Problem

$E(Y_{1i}|D=1) - E(Y_{0i}|D=0)$ leads us to the problem of selection bias that we have already seen.

Randomization would solve the problem, but you cannot randomize who gets elected and who does not.

Instead, we try to find direct comparisons: `matches` for each treated unit.

# Exact Matching

An Example with 10 candidates:

|              | Observed Outcome: Wealth at Death | $D_i$ | Male? |
|--------------|-----------------------------------|-------|-------|
| Candidate 1  | 855,557                           | 1     | 1     |
| Candidate 2  | 912,331                           | 1     | 1     |
| Candidate 3  | 566,271                           | 0     | 1     |
| Candidate 4  | 319,838                           | 0     | 1     |
| Candidate 5  | 612,233                           | 0     | 0     |
| Candidate 6  | 601,222                           | 1     | 0     |
| Candidate 7  | 485,709                           | 0     | 1     |
| Candidate 8  | 102,509                           | 0     | 1     |
| Candidate 9  | 991,511                           | 1     | 1     |
| Candidate 10 | 757,972                           | 1     | 1     |

What do we do next?

# Exact Matching: Counterfactuals for Observed Outcome, Ordered According to Covariate Values

|  | Potential Outcome Under Treatment | Potential Outcome Under Control | $D_i$ | Male? |
|---|---|---|---|---|
| Candidate 1 | 855,557 | ? | 1 | 1 |
| Candidate 2 | 912,331 | ? | 1 | 1 |
| Candidate 9 | 991,511 | ? | 1 | 1 |
| Candidate 10 | 757,972 | ? | 1 | 1 |
| Candidate 3 | ? | 566,271 | 0 | 1 |
| Candidate 4 | ? | 319,838 | 0 | 1 |
| Candidate 7 | ? | 485,709 | 0 | 1 |
| Candidate 8 | ? | 102,509 | 0 | 1 |
|  | Potential Outcome Under Treatment | Potential Outcome Under Control | $D_i$ | Male? |
| Candidate 5 | ? | 612,233 | 0 | 0 |
| Candidate 6 | 601,222 | ? | 1 | 0 |

# Counterfactuals for Observed Outcome

|  | Potential Outcome Under Treatment | Potential Outcome Under Control | $D_i$ | Male? |
|---|---|---|---|---|
| Candidate 1 | 855,557 | ? | 1 | 1 |
| Candidate 2 | 912,331 | ? | 1 | 1 |
| Candidate 9 | 991,511 | ? | 1 | 1 |
| Candidate 10 | 757,972 | ? | 1 | 1 |
| Candidate 3 | ? | 566,271 | 0 | 1 |
| Candidate 4 | ? | 319,838 | 0 | 1 |
| Candidate 7 | ? | 485,709 | 0 | 1 |
| Candidate 8 | ? | 102,509 | 0 | 1 |
|  | Potential Outcome Under Treatment | Potential Outcome Under Control | $D_i$ | Male? |
| Candidate 5 | 601,222 | 612,233 | 0 | 0 |
| Candidate 6 | 601,222 | 612,233 | 1 | 0 |

Introduction and Review
○○○○○○○○○○○○

Multivariate Regression
○○○○○○○○○○○○

**Matching**
○○○○○○○○○○○○●○○○○○○○○

# Imputing the Missing Outcomes

|  | Potential Outcome Under Treatment | Potential Outcome Under Control | $D_i$ | Male? |
|---|---|---|---|---|
| Candidate 1 | 855,557 | 368,581.75 | 1 | 1 |
| Candidate 2 | 912,331 | 368,581.75 | 1 | 1 |
| Candidate 9 | 991,511 | 368,581.75 | 1 | 1 |
| Candidate 10 | 757,972 | 368,581.75 | 1 | 1 |
| Candidate 3 | 879,342.75 | 566,271 | 0 | 1 |
| Candidate 4 | 879,342.75 | 319,838 | 0 | 1 |
| Candidate 7 | 879,342.75 | 485,709 | 0 | 1 |
| Candidate 8 | 8793,42.75 | 102,509 | 0 | 1 |
|  |  |  |  |  |
|  | Potential Outcome Under Treatment | Potential Outcome Under Control | $D_i$ | Male? |
| Candidate 5 | 601,222 | 612,233 | 0 | 0 |
| Candidate 6 | 601,222 | 612,233 | 1 | 0 |

Introduction and Review
○○○○○○○○○○○○

Multivariate Regression
○○○○○○○○○○○○

Matching
○○○○○○○○○○○○●○○○○○○○

## Estimate Treatment Effects

ATT

- ▶ $E[Y_{1i} - Y_{0i}|D = 1, X = 1]$ for each $i = 1, 2, 9, 10$
- ▶ $E[Y_{1,6} - Y_{0,6}|D = 1, X = 0]$
- ▶ And take weighted average

ATC

- ▶ $E[Y_{1i} - Y_{0i}|D = 0, X = 1]$ for each $i = 3, 4, 7, 8$
- ▶ $E[Y_{1,5} - Y_{0,5}|D = 0, X = 0]$ &
- ▶ And take the weighted average.

ATE

- ▶ $E[Y_{1i} - Y_{0i}|X = 1]$ for each $i = 1, 2, 3, 4, 7, 8, 9, 10$
- ▶ $E[Y_{1i} - Y_{0i}|X = 0]$ for each $i = 5, 6$
- ▶ And take the weighted average.

# A Complication: Adding Covariates

|  | $D_i$ | Male? | Oxbridge? |
|---|---|---|---|
| Candidate 1 | 1 | 1 | 0 |
| Candidate 2 | 1 | 1 | 1 |
| Candidate 3 | 0 | 1 | 0 |
| Candidate 4 | 0 | 1 | 0 |
| Candidate 5 | 0 | 0 | 0 |
| Candidate 6 | 1 | 0 | 1 |
| Candidate 7 | 0 | 1 | 0 |
| Candidate 8 | 0 | 1 | 1 |
| Candidate 9 | 1 | 1 | 1 |
| Candidate 10 | 1 | 1 | 0 |

# Rearrange with respect to values of $X_1$

|  | $D_i$ | Male? | Oxbridge? |
|---|---|---|---|
| Candidate 1 | 1 | 1 | 0 |
| Candidate 2 | 1 | 1 | 1 |
| Candidate 3 | 0 | 1 | 0 |
| Candidate 4 | 0 | 1 | 0 |
| Candidate 7 | 0 | 1 | 0 |
| Candidate 8 | 0 | 1 | 1 |
| Candidate 9 | 1 | 1 | 1 |
| Candidate 10 | 1 | 1 | 0 |
|  | $D_i$ | Male? | Oxbridge? |
| Candidate 5 | 0 | 0 | 0 |
| Candidate 6 | 1 | 0 | 1 |

Is there a match for Candidate 6?

Is Candidate 5 good enough for any match?

## So, we are left with:

|              | $D_i$ | Male? | Oxbridge? |
|--------------|-------|-------|-----------|
| Candidate 1  | 1     | 1     | 0         |
| Candidate 10 | 1     | 1     | 0         |
| Candidate 3  | 0     | 1     | 0         |
| Candidate 4  | 0     | 1     | 0         |
| Candidate 7  | 0     | 1     | 0         |
|              | $D_i$ | Male? | Oxbridge? |
| Candidate 2  | 1     | 1     | 1         |
| Candidate 9  | 1     | 1     | 1         |
| Candidate 8  | 0     | 1     | 1         |

# So, we are left with:

|              | $D_i$ | Male? | Oxbridge? |              | $D_i$ | Male? | Oxbridge? |
|--------------|-------|-------|-----------|--------------|-------|-------|-----------|
| Candidate 1  | 1     | 1     | 0         |              |       |       |           |
| Candidate 3  | 0     | 1     | 0         | Candidate 2  | 1     | 1     | 1         |
| Candidate 4  | 0     | 1     | 0         | Candidate 8  | 0     | 1     | 1         |
| Candidate 7  | 0     | 1     | 0         | Candidate 9  | 1     | 1     | 1         |
| Candidate 10 | 1     | 1     | 0         |              |       |       |           |

<div align="center">Estimation</div>

1. Caclulate: $E(Y_i|D = 1, X_1 = Male, X_2 = NoOxbridge) - E(Y_i|D = 0, X_1 = Male, X_2 = NoOxbridge)$

2. Caclulate: $E(Y_i|D = 1, X_1 = Male, X_2 = Oxbridge) - E(Y_i|D = 0, X_1 = Male, X_2 = Oxbridge)$

3. Take weighted average.

4. What happened with Candidates 5 and 6?

# Even More Covariates

## What are we looking for?

▶ Units differing in their $D_i$ values while in the same time: Having the exact same values in all other columns ($X's$)

▶ Any chance?

| | $D_i$ | Male? | Oxbridge? | Aristocrat? | Public schooling? |
|---|---|---|---|---|---|
| Candidate 1 | 1 | 1 | 0 | 0 | 1 |
| Candidate 2 | 1 | 1 | 1 | 1 | 1 |
| Candidate 6 | 1 | 0 | 1 | 1 | 1 |
| Candidate 9 | 1 | 1 | 1 | 1 | 1 |
| Candidate 10 | 1 | 1 | 0 | 1 | 1 |
| Candidate 3 | 0 | 1 | 0 | 0 | 0 |
| Candidate 4 | 0 | 1 | 0 | 0 | 0 |
| Candidate 5 | 0 | 0 | 0 | 0 | 0 |
| Candidate 7 | 0 | 1 | 0 | 0 | 0 |
| Candidate 8 | 0 | 1 | 1 | 0 | 1 |

▶ **How about a multivariate regression here?**

## Extrapolations

▶ The "else equal" principle is often satisfied only through extrapolations beyond the range of the available data.

▶ Such extrapolations are in turn based on assumptions, which are typically untestable and 'invisible' within the regression framework.

▶ Matching, thus makes the stage of making units similar with regard to covariates more transparent.

▶ Imagine candidate 7 also went to a public school; thus, comparing Candidate 1 and 7 would provide an estimate of ATE.

▶ Again this is also the case with regression: extrapolations require attaching greater weight to most similar units. *Regressions?*

# Dimensionality

▶ In the original study, there are more than 400 observations available.

▶ But: many more covariates are taken into account

▶ As the number of covariates used to "match" units increases, it becomes exponentially more difficult to find perfect matches.

▶ Exact matching fails in finite samples if the dimensionality of $X$ is large: not enough information. Far too demanding for the vast majority of research questions and data available.

▶ With more than one continuous variable, it is also sub-optimal (Abadie & Imbens, 2006).