# Matching (cont.)

observed : $y_i , D_i , x_i$
for $i$

$$y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$ = observed outcome for $i$

Assumptions to estimate $\Delta^{ATE}$

$$Y(0), Y(1) \perp D \mid x$$

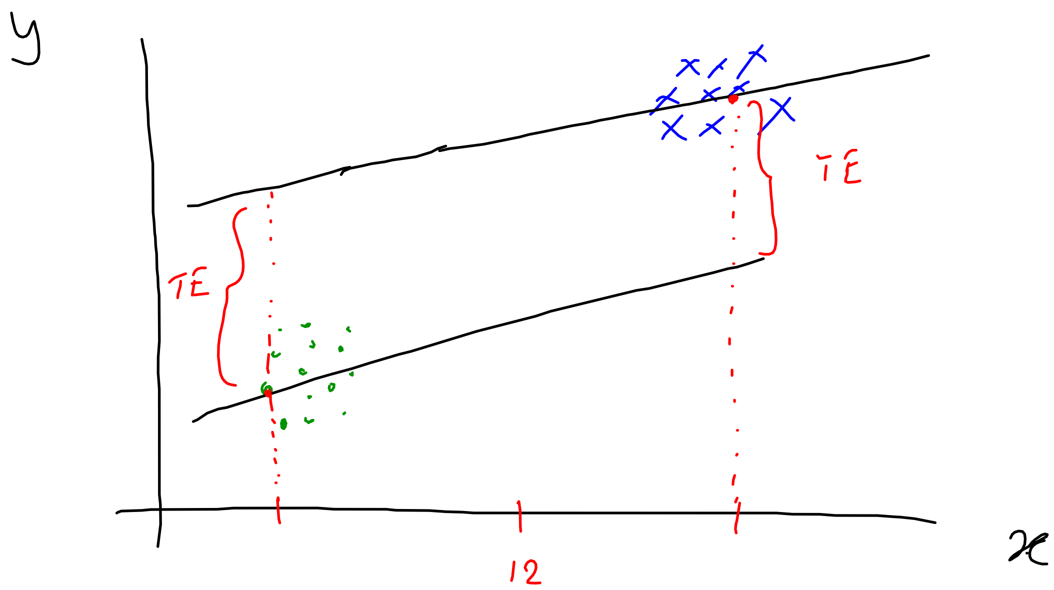
↳ independent of

$$0 < P(D = 1 \mid x) < 1$$

Random sample $\{ y, D, x \}$

Matching $\longrightarrow$ when comparing treated & control obs. more weight given to similar obs.

No functional form assumptions.



$$\hat{\Delta}^{ATE} = \frac{1}{N} \sum_i [Y_i(1) - Y_i(0)]$$

$\downarrow$

infeasible $\longrightarrow$ missing data / counterfactual

$Y_i(1)$ observed if $D_i = 1$ ; $Y_i(0)$ unobserved if $D_i = 1$

Replace $Y_i(0)$ with a close match to $i$

but with $D = 0$ (e.g. $Y_j(0)$ where $D_j = 0$)

Various matching approaches to determine what weight to attach to observations from the <u>opposite</u> group.

One approach relies on propensity score, $p(x)$ :

$$p(x) = Pr(D = 1 \mid x)$$

$\hookrightarrow$ typically estimated using logit / probit

$$Y(0), Y(1) \perp D \mid x \implies$$

$$Y(0), Y(1) \perp D \mid p(x)$$

Single nearest neighbor matching

If $D_i = 1$ & $Y_i(0)$ is unobserved use $Y_j(0) = y_j$

where $D_j = 0$ & $p_j$ is closest to $p_i$

prop. score for $j$

prop. score for $i$

" $D_i = 0$ & $Y_i(1)$ . . . . use $Y_j(0) = y_j$

where $D_j = 1$ & $p_j$ closest to $p_i$

| obs. | D | Y(1) | Y(0) | PS $P(D=1 \mid X)$ | Matched Y | TE |
|---|---|---|---|---|---|---|
| 1 | 1 | 18 | ? | 0.6 ✓ | 12 | 6 |
| 2 | 1 | 24 | ? | 0.7 ✓ | 12 | 12 |
| 3 | 0 | ? | 15 | 0.3 ✓ | 16 | 1 |
| 4 | 0 | ? | 12 | 0.4 ✓✓ | 16 | 4 |
| 5 | 1 | 16 | ? | 0.5 ✓✓ | 12 | 4 |
| | | | | | | 27 |

$$ATE = \frac{27}{5} = 5.4$$

Treated group

24 yrs

Family inc. = $100,000

only child

high SAT

Control

| D | age | family income | . . . |
|---|---|---|---|
| 0 | | | |
| 0 | | | |
| 1 | | | |
| 1 | | | |
| 1 | | | |
| 0 | | | |