Selection into Different Outcomes

- People select into different outcomes
- Simple difference in mean:

$$
\begin{aligned}
& E[Y|D=1] - E[Y|D=0] \\
& = (E[Y_1|D=1] - E[Y_0|D=1]) + (E[Y_0|D=1] - E[Y_0|D=0]) \\
& = (\text{ATT}) + (\text{Selection Bias})
\end{aligned}
$$

- Under random assignment, there will be no selection bias (why?)

Unconfoundedness

- Assumption for unconfoundedness: $\{Y_d\}_{d \in \mathcal{D}} \perp D|X$
- Intuition: conditional on X, treatment is as good as randomly assigned
- Also known as "Conditional Independence Assumption (CIA)"
- Similar identification results can be shown
- Thought experiment
  - Fix an $X = x$
  - Find treatment and control with the same x
  - Compare their outcomes

Bad Controls and Propensity Score

- X should be pre-determined
- Too much control might be more biased (Not the more the better!)
- Conditional on the whole X is hard (Curse of dimensionality)
- Rosenbaum and Rubin (1983): $(Y_1, Y_0) \perp D|X \Rightarrow (Y_1, Y_0) \perp D|p(X)$
- where $p(x) \equiv P(D = 1|X = x)$ is the propensity score

Rosenbaum and Rubin (1983)

$$\begin{aligned}
P[D = 1|Y_0, Y_1, P] &= E(P[D = 1|Y_0, Y_1, P, X]|Y_0, Y_1, P) \\
&= E(P[D = 1|Y_0, Y_1, X]|Y_0, Y_1, P), (\because P \text{ is a function of } X) \\
&= E(P[D = 1|X]|Y_0, Y_1, P), (\because \text{CIA}) \\
&= E(p(X)|Y_0, Y_1, P) \\
&= E(P|Y_0, Y_1, P) \\
&= P
\end{aligned}$$

- Note that the last term P doesn't depend on (Y_0, Y_1)
- So $P[D = 1|Y_0, Y_1, P]$ equals to that does not condition on (Y_0, Y_1)

Implementation Methods

- Many ways: matching, block matching
- A whole book is dedicated to this
- Re-weigh data so that
  - the treatment and control groups have a similar distribution of propensity
- e.g., the following estimator

$$\begin{aligned}\hat{\tau}^{\mathrm{IP}} &= \frac{1}{n}\sum_{i:D_i=1}\frac{Y_i}{\hat{p}(X_i)} - \frac{1}{n}\sum_{i:D_i=0}\frac{Y_i}{1-\hat{p}(X_i)} \\ &= \frac{1}{n}\sum_{i}^{n}\frac{D_i Y_i}{\hat{p}(X_i)} - \frac{1}{n}\sum_{i}^{n}\frac{(1-D_i)Y_i}{1-\hat{p}(X_i)}.\end{aligned}$$

Why would this work?

$$E\left[\frac{YD}{p(X)}\right] = E[Y_1], \quad \text{and} \quad E\left[\frac{Y(1-D)}{1-p(X)}\right] = E[Y_0]$$

$$\begin{aligned}
E\left[\frac{Y\ D}{p(X)}\right] &= E\left[E\left[\frac{YD}{p(X)}\middle| X\right]\right] \\
&= E\left[E\left[\frac{Y_1\ D}{p(X)}\middle| X\right]\right] \\
&= E\left[\frac{E[Y_1|X]\ E[D|X]}{p(X)}\right] \\
&= E\left[E[Y_1|X]\right] = E\left[Y_1\right]
\end{aligned}$$

Remarks

- In practice, estimate the propensity by logit.
- There are also ATT versions of this.
- Adding linear control is a special case of these.