Causal Inference Notes (Rubin Causal Model)

1) SETUP

Assume we have a sample of individuals where some receive a treatment (e.g. college, job training, minimum wage increase, etc.) and the others do not. We denote a particular individual with the subscript *i.* Let's define two concepts with the following notation….

Actual Treatment Status of individual *i*:

$$D_i = \begin{cases} 1 \; if \; treated \\ 0 \; if \; \; NOT \; treated \end{cases}$$

Potential Outcomes of individual *i*:

Y_{i1} -- The outcome for individual *i* if she is treated

Y_{i0} -- The outcome for individual *i* if she is NOT treated

Individual *i's* potential outcomes are just that, potential. **They DO NOT depend on *i's* ACTUAL treatment status!**

Think of it this way ---

- Susie has one future salary if she is "treated" with job training $Y_{susie,1}$ and one if she is not treated, $Y_{susie,0}$.[1]
- These two future salaries are *set in stone prior* to her receiving (or not receiving) the treatment.
- When the treatment actually occurs, this only **REVEALS** the relevant potential outcome: $Y_{susie,1}$ if $D_{susie} = 1$ OR $Y_{susie,0}$ if $D_{susie} = 0$
- Of course only one of the potential outcomes is revealed because we do not get to see Susie's outcome if she is treated AND if she is not treated.

2) GOAL

We are after the causal effect of the treatment.

- The average (causal) treatment effect (ATE) is the average of the difference between individuals potential outcomes with and without the treatment. It is given by:

$$\tau = E[Y_{i1} - Y_{i0}]$$

- Why is it possible to estimate the average difference between potential outcomes if the individual difference between potential outcomes is unobserved? Because we will use other individuals (or the same individual at a different point in time) as proxies for unobserved counterfactuals!

[1] I am playing with notation a little for clarity, really we define $i \in \{1, \dots, N\}$, where N is the number of observations. Susie might be i= 364.

3) WHAT WE CAN AND CAN'T OBSERVE

With observational data, we can observe the DIFFERENCE between

A) $E[Y_{i1}|D_i = 1]$ -- The average of the potential outcomes if treated (Y_{i1}) for those who had this outcome revealed by actually receiving the treatment ($D_i = 1$), AND

B) $E[Y_{i0}|D_i = 0]$ -- The average of the potential outcomes if not treated ((Y_{i0}) for those who had this outcome revealed by not receiving the treatment ($D_i = 0$),.

Putting these together, we observe: $E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0]$

Ex: Those with a college degree earn 57% more than those without a college degree.

We CAN NOT observe the counterfactuals:

A) $E[Y_{i0}|D_i = 1]$ -- The average of the potential outcomes if not treated (Y_{i0}) for those who actually received the treatment ($D_i = 1$)

Ex: The average wage that those with a college degree would have earned if they had not graduated from college.

B) $E[Y_{i1}|D_i = 0]$ -- The average of the potential outcomes if treated (Y_{i1}) for those who did not receive the treatment ($D_i = 0$)

Ex: The average wage that those without a college degree would have earned if they had graduated from college.

4) AVERAGE (CAUSAL) TREATMENT EFFECT ON TREATED (ATET)

Add and subtract the counterfactual for the treated group, $E[Y_i(0)|D_i = 1]$, to the observed mean difference in outcomes:

$E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0]$

$$= \underbrace{E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1]}_{\text{ATET: } E[Y_{i1} - Y_{i0}|D_i = 1]} + \underbrace{E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]}_{\text{SELECTION BIAS (1)}}$$

- Average Treatment Effect on Treated (ATT) measures the effect of the treatment on those that were treated. $E[Y_{i1} - Y_{i0}|D_i = 1]$

  Ex: The causal effect of a college degree on those who actually went to college.

- Selection Bias measures the degree to which the non-treated group can serve as a counterfactual for the treated group. Specifically, here it is the difference between the actually treated and non-treated groups in the average *potential* outcome if not treated. $E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]$

  Ex: The difference in wage between the group that graduated from college and the group that didn't, HAD THEY BOTH NOT GRADUATED.

5) <u>AVERAGE (CAUSAL) TREATMENT EFFECT ON NON-TREATED (ATEN)</u>

Add and subtract the counterfactual for the non-treated group, $E[Y_i(1)|D_i = 0]$, to the observed mean difference in outcomes:

$E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$

$$= \underbrace{E[Y_{i1}|D_i = 0] - E[Y_{i0}|D_i = 0]}_{\text{ATN: } E[Y_{i1} - Y_{i0}|D_i = 0]} + \underbrace{E[Y_{i1}|D_i = 1] - E[Y_{i1}|D_i = 0]}_{\text{SELECTION BIAS (2)}}$$

- Average Treatment Effect on Non-Treated (ATEN) measures the effect of the treatment on those that were treated. $E[Y_{i1} - Y_{i0}|D_i = 0]$

  Ex: The causal effect of a college degree on those who actually went to college.

- Again, the Selection Bias term measures the degree to which the non-treated group and treated group are comparable. But, here it is the difference between the actually treated and non-treated groups in the average *potential* outcome if treated. $E[Y_{i1}|D_i = 1] - E[Y_{i1}|D_i = 0]$

  Ex: The difference in wage between the group that graduated from college and the group that didn't, HAD THEY BOTH GRADUATED.

6) <u>AVERAGE (CAUSAL) TREATMENT EFFECT (ATE)</u>

Recall from the Section 2 above that ATE is defined as, $\tau = E[\tau_i] = E[Y_{i1} - Y_{i0}]$. We can derive the following decomposition.

$$\tau = E[Y_{i1} - Y_{i0}]$$

$$\tau = E[Y_{i1} - Y_{i0}|\, D_i = 1]\Pr(D_i = 1) + E[Y_{i1} - Y_{i0}|\, D_i = 0]\Pr(D_i = 0)$$

$$\tau = \Big[E\big[Y_{i1}|D_i = 1] - E[Y_{i0} \;\; |\, D_i = 1\big]\Big](1 - \Pr(D_i = 0))$$

$$+ \Big[E\big[Y_{i1}|D_i = 0] - E[Y_{i0} \;\; |\, D_i = 0\big]\Big](1 - \Pr(D_i = 1))$$

$$\tau = E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 0]$$

$$-E[Y_{i1}|\, D_i = 1]\Pr(D_i = 0) - E[Y_{i0}|\, D_i = 1]\Pr(D_i = 1))$$

$$+E[Y_{i0}|\,D_i=0]\,\Pr(D_i=1)+\;E[Y_{i1}|\,D_i=0]\Pr\,(D_i=0)$$

Observed Mean Difference

$$\rightarrow E[Y_{i1}|D_i=1]-E[Y_{i0}|D_i=0]$$

$$=\;\tau+\Big[E\big[Y_{i1}|D_i=1]-E[Y_{i1}\;\;|\,D_i=0\big]\Big]\Pr(D_i=0)+\Big[E\big[Y_{i0}|D_i=1]-E[Y_{i0}\;\;|\,D_i=0\big]\Big]\Pr\,(D_i=1)$$

ATE SELECTION BIAS (2) SELECTION BIAS (1)

The observed average difference is only equal to the average treatment effect if both flavors of selection bias are zero.

7) RANDOMIZED CONTROL TRIALS

In a randomized experiment the assignment to either the treatment or non-treatment group is random. Therefore, the treatment status, D_i, and potential outcomes, Y_{i1} and Y_{i0} , are independent. This implies,

$E[Y_{i1}|D_i=1]=E[Y_{i1}]$ and $E[Y_{i0}|D_i=0]=E[Y_{i0}]$

The intuition here is that assignment to the treatment or control group is random and therefore can not be correlated with any characteristic of the individual (observed or unobserved). In other words, we are choosing which potential outcome to reveal by chance, it is not up to the individual to choose which to reveal.

This implies that both flavors of selection bias are zero:

$E[Y_{i0}|D_i=1]-\;E[Y_{i0}|D_i=0]$ = $E[Y_{i0}]-\;E[Y_{i0}]$ = 0

$E[Y_{i1}|D_i=1]-\;E[Y_{i1}|D_i=0]$ = $E[Y_{i1}]-\;E[Y_{i1}]$ = 0

Therefore, ATE = ATET = ATEN