Week 2 1-pager
Samuel Rowe
6/20/2022

1) The Takeaway
    a. Potential outcomes are a way to assess causal inference
        i. However, the problem is that some we only observe half of the outcomes, while the corresponding potential outcomes are unobserved
    b. The Simple differences in Mean Outcome estimate is biased because people make decisions about sorting in and out of treatment based up what they think is optimal
        i. This is selection bias and we cannot identify the causal effect when this happens with a simple difference in mean outcomes
    c. A well-designed and well-implemented randomized assignment is the best way to deal with selection bias
        i. When we have randomized assignment, a simple difference in mean outcomes will identify the causal effect of the treatment on the outcome of interest
    d. The goal of this course is to learn about the research designs that identify the causal effect by overcoming selection bias
        i. Randomized assignment is the gold standard, but we will talk about the strengths, weaknesses, and assumptions for research designs to identify the causal effect with observational data
    e. Randomized inference is a methodology to construct exact p-values when traditional methods might not be as appropriate
        i. What is the causal effect was a product of chance? What happens when we randomized treatment?
        ii. Randomized inference is helpful when large admin data sets instead of samples, not appealing to large n of an estimator, or utilizing placebo-based inference
2) Potential Outcomes
    a. The counterfactual is an important concept – what would the world be like if another outcome was chosen – but counterfactual are never observed in history because only one outcome occurs
    b. Counterfactual outcomes exist ex ante as a set of possibilities before one outcome is realized, but we will simplify into a binary outcome
    c. Potential outcomes are defined as Y_i^1 if unit *i* receives treatment and as Y_i^0 if unit *i* does not receive treatment
    d. Treatment effects can never be calculated, but they can be estimated
    e. Unit Specific Treatment Effects: $\delta_i = Y_i^1 - Y_i^0$
    f. Average Treatment Effect: $ATE = E[\delta_i] = E\left[Y_i^1 - Y_i^0\right] = E\left[Y_i^1\right] - E\left[Y_i^0\right]$
    g. Average Treatment on the Treated (ATT): $ATT = E\left[Y_i^1 | D_i = 1\right] - E\left[Y_i^0 | D_i = 1\right]$
    h. Average Treatment on the Untreated (ATU): $ATU = E\left[Y_i^1 | D_i = 0\right] - E\left[Y_i^0 | D_i = 0\right]$
    i. Simple difference in outcomes is an estimator for ATE
    j. Simple Difference in Outcome: $\frac{1}{N_T}\sum_{i=1}^{n}(y_i | d_i = 1) - \frac{1}{N_C}\sum_{i=1}^{n}(y_i | d_i = 0) = E[Y^1] - E[Y^0] + E[Y^0 | D = 1] - E[Y^0 | D = 0] + (1 - \pi)(ATT - ATU)$
        i. $SDO = ATE + SelectionBias + HeterogenousTreatmentBias$

k. Selection Bias is $E[Y^0|D=1] - E[Y^0|D=0] \neq 0$
l. Heterogenous Treatment Bias: $(1-\pi)(ATT - ATU)$
m. The SDO is bias because individual were ***optimally sorted*** (selected) into their best treatment option
n.

3) Experimental Research Design
   a. Randomized assignment (or experimental research) design is the gold standard in research methodologies, since they identify the causal effect of the treatment on the outcome of interest
   b. Strength
      i. Controls for selection bias on observed and unobserved confounders
   c. Weakness
      i. Expensive, logistical and implementation nightmares, and potential moral issues
   d. Assumptions
      i. Independence assumption: $(Y^0, Y^1) \perp D$
         1. Assignment of treatment is independent of potential outcomes
      ii. Stable Unit Treatment Value Assumption
         1. There are no spillovers in treatment and dosage is constant
   e. Tests
      i. Covariate Balance on observables
      ii. We actually cannot test covariate balance on unobservables
   f. When we use a regression for estimating causal effects in a randomized experiment and assume homogenous effects
      i. $Y_i = \alpha + \delta D_i + \varepsilon_i$
      ii. Where $\alpha = E[Y_i^0]$
      iii. Where $\delta = Y_i^1 - Y_i^0$ with homogenous effects
      iv. Where $\varepsilon_i = Y_i^0 - E[Y_i^0]$
   g. Covariates in randomized assignment is used for
      i. When we have conditional random assignment
      ii. Reduce the standard error around the estimate of causal effect and increase the precision