Causal Inference (What If)

Ch11 Why models?

이종진

Seoul National University

ga0408@snu.ac.kr

August 05, 2021

Ch 11, Why model?

- Part II (ch11 $\sim$ 18) is about the parametric (model-based) estimators.
- This chapter motivates the need for models in data analysis
- "Data cannot always speak for themselves, rather we need to supplement the data with a model."

Necessity of model; example

- Example
  - 16 individuals infected with HIV.
  - A continuous outcome Y(CD4 cell count)
  - Wish to estimate the mean of Y among individuals with treatment level $A = a$
  - Conditional mean, $E(Y|A = a)$
- Treatment A could be
  - a dichotomous variables with two possible values
  - a polytomous variables with 4 possible values
  - a integer values from 0 to 100mg

Necessity of model; example

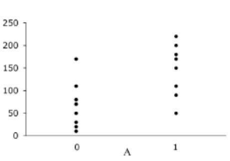

Figure 11.1

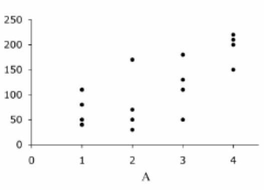

Figure 11.2

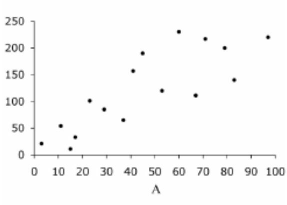

Figure 11.3

- # of possible values of $A >$ # of observed individuals
- How can we estimate the mean of outcome Y among individuals with treatment level $A = 90$ which not observed?

Necessity of model; example

- We often need to supplement the data with a model.
- With a priori restrictions, model can compensate for the lack of sufficient information in the data.
- $E(Y|A = a) = \theta_0 + \theta_1 A$

Nonparametric & parametric

$$E(Y|A) = \theta_0 + \theta_1 A$$

- Nonparametric estimators
  - Treatment A is a dichotomous; 2 quantities and 2 parameters
  - Without any a priori restrictions,
  - Standardization, IP weighting, stratification and matching.
- Parametric estimators
  - Treatment A takes integer values; 101 quantities and 2 parameters
  - Impose a priori restrictions on conditional mean (linear)

Smoothing & The bias-variance trade-off

- Smoothing
  - Linear model can be more flexible, introducing $A^2 \ldots, A^{15}$
  - The curve generally becomes more "wiggly", or less smooth, as the number of paramters increase.
- The bias-variance trade-off
  - The larger the number of paramters in model, the more protection afforded against bias from the model misspecification.
  - Although less smooth models may yield a less biased estimate, they also result in larger variance.