Project Presentation

Alex Luedtke, Lucia Petito, Steven Pollack

PHC252D

Outline

- Background
- Specify SCM (and DAG)
- Specify counterfactuals and target causal quantity
- Introduce data and commit to a statistical model
- Discuss identifiability and estimand
- Get our hands dirty (estimation procedures)
- Results
- Interpretation

Background

We know that sleep affects weight, but does trying to lose weight affect sleep?

- We used National Health and Nutrition Examination Survey (NHANES) data – from the National Center for Health Statistics (NCHS) – a multistage survey of U.S. population
  - Stage 1: Counties
  - Stage 2: Segments
  - Stage 3: Households
  - Stage 4: Individuals
- Survey aims to study wide range of topics such as Cardiovascular disease, Obesity, Physical fitness and physical functioning, Reproductive history and sexual behavior, etc.

Background (Cont'd)

Notes about NHANES data:

- Individuals were subjected to interviews as well as physical examinations.
  - categorical as well as numerical data
  - some questions had a lot of valid responses, but made positivity questionable.
  - No shortage of missing data (either "I don't know"'s or unanswered questions).
- The sample for the survey is selected to represent the U.S. population of all ages. To produce reliable statistics, NHANES over-samples persons 60 and older, African Americans, and Hispanics.

W: Baseline Covariates

- Gender
- Age in months (300-959 months, 25-79 years)
- Race/Ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other)
- Education Level (less than high school, high school/GED, some college, college and above)
- Marital Status (never married, married/living with partner, divorced/separated)
- Annual Household Income (less than or greater than \$20k)
- Body Mass Index (continuous from 15-50)

A: Exposure Variable

- The subject's response to the question: "During the past 12 months, have you tried to lose weight?"
- Note that this does not restrict to dieting

Y: Response Variable

- The subject's response to the question: "How much sleep do you usually get at night on weekdays or workdays?
- Both A and Y sampled simultaneously, so temporal ordering is only assumed

SCM and DAG

Our observational data structure is $O = (W, A, Y) \sim P_0$. With mild temporal assumptions, one SCM is:

$$W = f_W(U_W)$$
$$A = f_A(W, U_A)$$
$$Y = f_Y(W, A, U_Y)$$

Note: no assumptions made on functional forms of W, A, or Y.

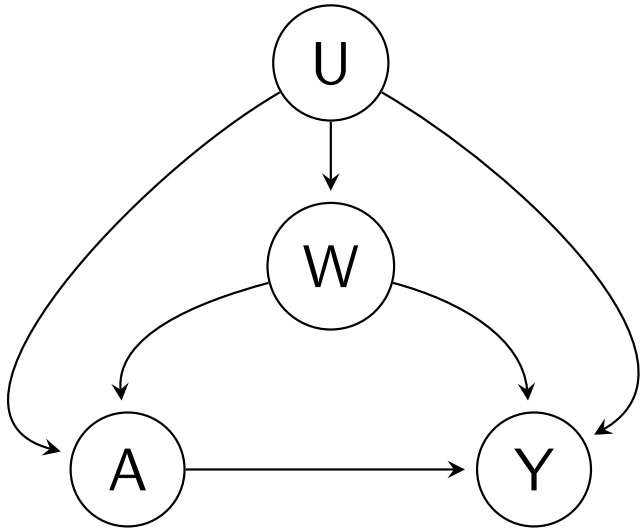

Figure : Simplified DAG – no independence assumptions on U's.

W: Baseline Covariates

- Gender
- Age in months (300-959 months, 25-79 years)
- Race/Ethnicity (Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, Other)
- Education Level (less than high school, high school/GED, some college, college and above)
- Marital Status (never married, married/living with partner, divorced/separated)
- Annual Household Income (less than or greater than \$20k)
- Body Mass Index (continuous from 15-50) – as of prior year

A: Exposure Variable

Y: Response Variable

Specify SCM (and DAG)