

P8122 Homework 3

Due: October 14 2022 at 5:00pm

Instructions

- Upload a single pdf file for your homework on Canvas.
- The file should include the code in the appendix. Comment your code.
- You may discuss these problems with each other verbally, but must write up the answers on your own, and may not share or show your answers to anyone else.
- Short and clear answers, please.
- No late homework are allowed.

Inspired by the study on the effect that light at night has on weight gain and other variables in mice, we created data by simulating new attributes and exposure assignment for the mice, representing what we would expect to see in an observational study of this relationship. The code to simulate the data is provided. We will proceed using the g-formula to estimate the sample average causal effect.

- a) (30 points) We have provided R-code to simulate data of 16 mice. Obesity, a binary covariate (C), is simulated using a Bernoulli distribution by assuming that $1/5$ (p) of the mice are obese at baseline. Exposure to light variable (A) is simulated as a Bernoulli distribution conditional on the obesity covariate C. Here we assume that obese mice are less likely to get exposed to light because they are more sedentary ($\theta_1 = -1/5$). Finally, we generate a glucose outcome variable (Y), a normally distributed variable whose mean (μ_Y) is both a function of obesity at baseline and light ($\beta_0 + \beta_1 * \text{obesity} + \beta_2 * \text{light}$, $\beta_0 = 110$, $\beta_1 = 20$, $\beta_2 = -5$). Interpret all parameters (p , θ_0 , θ_1 , β_0 , β_1 , β_2).
- b) (10 points) Write the marginal and conditional PACE. Under which assumptions marginal and conditional PACE are identified?
- c) (10 points) Show the g-formula for the randomized study we considered in the previous homework in which we studied the effect of light (dark=DL Vs bright=LL) on obesity. Compare the g-formula of the current simulated observational study.
- d) (20 points) Provide the estimate of $E[Y|A = 1] - E[Y|A = 0]$ in your simulated data from part (a). Interpret the results.
- e) (20 points) Provide the estimate of $E[Y_1] - E[Y_0]$ in your simulated data from part (a) using the g-formula. Interpret the results. Explain the differences between the inferences obtained in (d) and (e).
- f) (10 points) Consider now a hypothetical observational study with 10 continuous covariates. Under which assumptions can you estimate $E[Y_1] - E[Y_0]$ using linear regression.