University of California, Santa Barbara   **Olivier Deschenes**
Bren School of Env. Science & Management   **Winter 2022**

EDS 241
Assignment 3
Due on 2/20/22 at 17:00

Turn in your Markdown pdf on Gauchospace in the "Assignment Turn in Area"

Question 1: Application of estimators based on treatment ignorability

This exercise asks you to implement some of the techniques presented in Lectures 6-7. The goal is to estimate the causal effect of maternal smoking during pregnancy on infant birth weight using the treatment ignorability assumptions. The data are taken from the National Natality Detail Files, and the extract "SMOKING_EDS241.csv"' is a random sample of all births in Pennsylvania during 1989-1991. Each observation is a mother-infant pair. The key variables are:

The outcome and treatment variables are:
birthwgt=birth weight of infant in grams
tobacco=indicator for maternal smoking

The control variables are:
mage (mother's age), meduc (mother's education), mblack (=1 if mother black), alcohol (=1 if consumed alcohol during pregnancy), first (=1 if first child), diabete (=1 if mother diabetic), anemia (=1 if mother anemic)

(a) What is the unadjusted mean difference in birth weight of infants with smoking and non-smoking mothers? Under what hypothesis does this correspond to the average treatment effect of maternal smoking during pregnancy on infant birth weight? Provide some simple empirical evidence for or against this hypothesis.

(b) Assume that maternal smoking is randomly assigned conditional on the observable covariates listed above. Estimate the effect of maternal smoking on birth weight using a linear regression. Report the estimated coefficient on tobacco and its standard error.

(c) Use the exact matching estimator to estimate the effect of maternal smoking on birth weight. For simplicity, consider the following covariates in your matching estimator: create a 0-1 indicator for mother's age (=1 if mage>=34), and a 0-1 indicator for mother's education (1 if meduc>=16), mother's race (mblack), and alcohol consumption indicator (alcohol). These 4 covariates will create 2*2*2*2 = 16 cells. Report the estimated average treatment effect of smoking on birthweight using the exact matching estimator and its linear regression analogue (Lecture 6, slides 12-14).

(d) Estimate the propensity score for maternal smoking using a logit estimator and based on the following specification: mother's age, mother's age squared, mother's education, and indicators for mother's race, and alcohol consumption.

(e) Use the propensity score weighted regression (WLS) to estimate the effect of maternal smoking on birth weight (Lecture 7, slide 12).

Note: This homework is a simple examination of these data. More research would be needed to obtain a more definitive assessment of the causal effect of smoking on infant health outcomes. Further, for this homework, you can ignore the adjustments to the standard errors that are necessary to reflect the fact that the propensity score is estimated. Just use heteroskedasticity robust standard errors in R. If you are interested, you can read Imbens and Wooldridge (2009) and Imbens (2014) for discussions of various approaches and issues with standard error estimations in models based on the propensity score.