

EDS241: Assignment 03 - National Natality Detail

Julia Parish

2022/02/20

1 EDS241 Environmental Policy Evaluation Assignment 03

This statistical analysis was completed as an assignment for the course, Environmental Data Science 241: Environmental Policy Evaluation. It is an application of estimators based on treatment ignorability. The goal of this assignment was 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 data files for this assignment is a random sample of all births in Pennsylvania during 1989-1991. Each observation is a mother-infant pair.

The outcome and treatment variables are:

birthwgt = birth weight of infant in grams

tobacco = indicator for maternal smoking

Note: This exercise asks you to implement some of the techniques presented in Lectures 6-7. 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.

```
data <- read_csv(here("hw03/data/smoking.csv"))
```

Question 1A

What is the unadjusted mean difference in birth weight of infants with smoking and non-smoking mothers? Under what assumption 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.

Question 1B

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.

lm_robust - include all variables all in (age, education are numeric). The way to get cond't avg tx effect, have to have all factors - here add them linearly. If you add age, education as factors creates a number of bins. How do we control these as observables, the easiest way it to include them linearly.

```
mod2 <- lm_robust(birthwgt ~ ., data = data)
```

Question 1 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 age>=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). *see TIA Table #ydiff = delta for x, w_ATE # of obs for rows, w_ATT = weights*

```
# mod3 <- lm_robust(data = data,  
#                   birthwgt ~ tobacco + as.factor(alcobol) + as.factor()  
#  
#  
#                   )
```

2 linear analogue - saturated model, exact matching estimator: calculating weights

-224

- (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.

3 create new column for mage squared

- (e) Use the propensity score weighted regression (WLS) to estimate the effect of maternal smoking on birth weight (Lecture 7, slide 12). *# see cgl.R, glm is a logit*

4