EDS241: Assignment 3

Joe DeCesaro

02/18/2022

1 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:

- birthwgt=birth weight of infant in grams
- tobacco=indicator for maternal smoking

The control variables:

- 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)

```
# Reading in the data
birth_data <- read.csv("data/SMOKING_EDS241.csv")
```

1.1 (a) 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.

```
model1 <- lm_robust(formula = birthwgt ~ tobacco, data = birth_data)
huxreg(model1)
```

	(1)
(Intercept)	3430.286 ***
	(1.781)
tobacco	-244.539 ***
	(4.150)
N	94173
R2	0.037

*** p < 0.001; ** p < 0.01; * p < 0.05.

The unadjusted mean difference in birth weight of infants with smoking and non-smoking mothers is 245 grams. This difference corresponds with the average treatment effect under the assumption that a mother's smoking is the only thing that affects birth weight of a newborn and that smoking is randomly assigned in a population.

```
model1.1 <- lm_robust(formula = tobacco ~ meduc, data = birth_data)
huxreg(model1.1)
```

	(1)
(Intercept)	0.789 ***
	(0.008)
meduc	-0.046 ***
	(0.001)
N	94173
R2	0.061

*** p < 0.001; ** p < 0.01; * p < 0.05.

From the model1.1, where we regress tobacco on mother's education, there is a decrease in likelihood of smoking for every year increase that is statistically significant and therefore there is an effect of years of education on likelihood of smoking. This means that the assumption that smoking is randomly assigned is incorrect because of this correlation.

1.2 (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.

```
model2 <- lm_robust(formula = birthwgt ~ tobacco + mage + meduc + mblack + alcohol + first + diabete + a
                    data = birth_data)

summary(model2)
```

```
##
## Call:
## lm_robust(formula = birthwgt ~ tobacco + mage + meduc + mblack +
##     alcohol + first + diabete + anemia, data = birth_data)
##
## Standard error type:  HC2
##
## Coefficients:
##             Estimate Std. Error  t value   Pr(>|t|) CI Lower   CI Upper    DF
## (Intercept) 3362.258    12.0765 278.4133  0.000e+00 3338.588 3385.92805 94164
## tobacco     -228.073     4.2768 -53.3282  0.000e+00 -236.456 -219.69063 94164
## mage          -0.694     0.3682  -1.8849  5.944e-02   -1.416    0.02764 94164
## meduc         11.688     0.8618  13.5630  7.262e-42    9.999   13.37742 94164
## mblack      -240.030     5.3478 -44.8842  0.000e+00 -250.512 -229.54873 94164
## alcohol      -77.350    14.0392  -5.5096  3.607e-08 -104.866  -49.83312 94164
## first        -96.944     3.4880 -27.7934 2.528e-169 -103.781  -90.10763 94164
## diabete       73.228    13.2355   5.5327  3.162e-08   47.286   99.16895 94164
## anemia        -4.796    17.8739  -0.2683  7.884e-01  -39.829   30.23630 94164
##
## Multiple R-squared:  0.0717 ,    Adjusted R-squared:  0.07162
## F-statistic: 877.6 on 8 and 94164 DF,  p-value: < 2.2e-16
```

The estimated coefficient on tobacco is -228.07

1.3 (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 222*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).

Create indicators

1.4 (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.

Interpret the signs Age is the only one that is quadratic (squared) because it is only of the only things that you can

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

Weighted refression is not that different from unweighted, why do we want to do this? Not having a big difference somewhat proves that this is reasonable

Two methods show that you re controlling for this OVB or if you are not.