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.

```r
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.

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

# summary(model2)
```

The estimated coefficient on tobacco is -228.07 and the standard error is 4.28.

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).

```r
# produce indicators
c_data <- birth_data %>%
  select("tobacco",
         "alcohol",
         "mblack",
         "mage",
         "meduc",
         "birthwgt") %>%
  mutate(mage_d = ifelse(mage >= 34, 1, 0),
         meduc_d = ifelse(meduc >= 16, 1, 0))

# make group variables
c_data <- c_data %>%
  select("tobacco",
         "alcohol",
         "mblack",
         "mage_d",
         "meduc_d",
         "birthwgt") %>%
  mutate(g = paste0(alcohol,
                    mblack,
                    mage_d,
                    meduc_d))

# make the model
model3 <- lm_robust(formula = birthwgt ~ tobacco + as.factor(g),
                    data = c_data)
```

```r
# huxreg("birthweight(g)" = model3)
```

```r
# Report the estimated ATE of smoking on birthweight using the exact matching estimator and its linear 
ATE_table <- c_data %>%
  group_by(g, tobacco)%>%
  # calculate number of observations
  summarise(n_obs = n(),
            # calculate birthwgt mean by X by treatment cells
            birthwgt_mean= mean(birthwgt, na.rm = T))%>%
  # Reshape dataframe
  gather(variables, values, n_obs:birthwgt_mean)%>%
  # Combine the treatment and variables for reshaping
  mutate(variables = paste0(variables,"_", tobacco, sep=""))%>%
  # Reshape data by treatment and X cell
  pivot_wider(id_cols = g, names_from = variables,values_from = values)%>%
  # Ungroup from X values
  ungroup()%>%
  # calculate birthwgt_diff
  mutate(birthwgt_diff = birthwgt_mean_1 - birthwgt_mean_0,
         w_ATE = (n_obs_0+n_obs_1)/(sum(n_obs_0)+sum(n_obs_1)),
         # calculate weights
         w_ATT = n_obs_1/sum(n_obs_1))%>%
  # Round data
  mutate_if(is.numeric, round, 2)

# huxtable(ATE_table)
```

```r
# Exact matching estimator ATE
ATE=sum((ATE_table$w_ATE)*(ATE_table$birthwgt_diff))
# ATE
```

Using the linear regression analogue, the average effect of a mother smoking on birth weight is a decrease of 226.25 grams. The average effect of a mother smoking on birth weight is a decrease of 224.26 grams using the the exact matching estimator.

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.

```r
# make mage^2
data_propensity <- birth_data %>%
  mutate(mage_sq = mage * mage)

# Estimate propensity score and predict (EPS)
ps_model1 <- glm(tobacco ~ mage + mage_sq + meduc + mblack + alcohol,
                 family = binomial(),
                 data = data_propensity)
# summary(ps_model1)
```

```r
# Make new response variable
EPS <- predict(ps_model1, type = "response")

# Propensity score
PS_WGT <- (data_propensity$tobacco/EPS) +
  ((1 - data_propensity$tobacco) / (1-EPS))
```

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

```r
# Use propensity score to weighted regression (WLS) to estimate maternal smoking effect on birth weight
wls <- lm(data = data_propensity,
          birthwgt ~ tobacco + mage + mage_sq + meduc + mblack + alcohol,
          weights = PS_WGT)

huxreg("Birth Weight" = wls)
```

Using the propensity score, the estimated effect of a mother smoking on birth weight is a decrease of 220.23.

	Birth Weight
(Intercept)	2971.444 ***
	(36.122)
tobacco	-220.233 ***
	(3.223)
mage	27.627 ***
	(2.693)
mage_sq	-0.478 ***
	(0.049)
meduc	7.472 ***
	(0.849)
mblack	-220.990 ***
	(4.994)
alcohol	-71.914 ***
	(13.709)
N	94173
R2	0.074
logLik	-728569.509
AIC	1457155.018

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