EDS241: Assignment 3

Scou Leonard

02/16/2022

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 if infant in grams
- `tobacco` = indicator for maternal smoking

The control variables are:

- `mage`: mother's age
- `meduc`: mother's education
- `mblack`: = 1 if mother is Black
- `alcohol`: =1 if consumed alcohol during pregnancy
- `first`: =1 if first child
- `diabete`: = 1 if mother is diabetic
- `anemia`: =1 if mother anemic

1 Load and Clean Data

```
#read in the data
smoking_df <- read.csv(here("data", "SMOKING_EDS241.csv"))
```

2 Homework Questions

2.1 Part 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.

```
mod_a1 <- lm_robust(birthwgt ~ tobacco, data = smoking_df)

huxtable::huxreg(mod_a1)
```

	(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 -244.5 grams.

This tells us about the effect of smoking on infant birth weight assuming that mothers who smoke and mothers who do are statistically different.

```
mod_a2 <- lm_robust(meduc ~ tobacco, data = smoking_df)

huxtable::huxreg(mod_a2)
```

	(1)
(Intercept)	13.239 ***
	(0.008)
tobacco	-1.318 ***
	(0.014)
N	94173
R2	0.061

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

Evidence against this hypothesis is shown in the code chunk above; when we regress tobacco usage on education, the results tell us that there is a significantly significant difference between the education levels of mothers who use tobacco during pregancy and those who do not. From this results, we know that mother's education is correlated with tobacco usage, but it is not included in our prediction of birthweights of infants of mothers who use tobacco above. This means there might be omitted variables bias, and the there may be additional variables interacting with the condition of tobacco use which are excluded.

2.2 Part 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.

The code chunk below estimates the effect of of maternal smoking on birth weight using a linear regression.

```
#linear regression with birthweight conditional on all of the variables
mod_B <- lm_robust(birthwgt ~ .,
                   data = smoking_df)

#standard error
mod_B[[2]][[2]]
```

```
## [1] 17.87392
```

The average treatment effect of maternal smoking on birth weight when all other covariants are held equal is -228.07 grams (on average, infants of mothers who use tobacco weigh 236.46 grams less than the infants of mothers who do not), with a standard error of 4.28.

2.3 Part 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).

The code chunk below creates 0-1 indicators for mother's education and mother's age.

```
#create indicators for mage and meduc
smoking_df <- smoking_df %>%
  mutate(mage_indicator = case_when(mage >= 34 ~ 1,
                                    mage < 34 ~ 0)) %>%
  mutate(meduc_indicator = case_when(meduc >= 16 ~ 1,
                                     meduc < 16 ~ 0)) %>%
  mutate(g = paste0(as.factor(mage_indicator),
                    as.factor(meduc_indicator),
                    as.factor(mblack),
                    as.factor(alcohol)))
```

The code chunk below generates the linear regression analogue to estimate the effect of smoking on birth weight.

```
#LINEAR REG ANALOGUE
mod_C <- lm_robust(birthwgt ~  tobacco +
                     as.factor(mage_indicator) +
                     as.factor(meduc_indicator) +
                     as.factor(mblack) +
                     as.factor(alcohol) +
                     as.factor(mage_indicator):as.factor(meduc_indicator) +
                     as.factor(mage_indicator):as.factor(mblack) +
                     as.factor(mage_indicator):as.factor(alcohol) +
                     as.factor(meduc_indicator):as.factor(mblack) +
                     as.factor(meduc_indicator):as.factor(alcohol) +
                     as.factor(mblack):as.factor(alcohol) +
                     as.factor(mage_indicator):as.factor(meduc_indicator):as.factor(mblack) +
                     as.factor(mage_indicator):as.factor(meduc_indicator):as.factor(alcohol) +
                     as.factor(meduc_indicator):as.factor(mblack):as.factor(alcohol) +
                     as.factor(mage_indicator):as.factor(meduc_indicator):as.factor(mblack):as.factor(a
                   data = smoking_df)
```

```
#generate table of coefficients for mod_C
huxtable::huxreg(mod_C)
```

The linear regression analogue tells us that the the infants of mothers who smoke during pregnancy weigh 226.25 grams less than the infants of mothers who do not smoke during pregnancy.

```
#EXACT MATCHING ESTIMATOR
TIA_table <- smoking_df %>%
  group_by(g,tobacco) %>%
  summarise(n_obs = n(),
            birthwgt_mean= mean(birthwgt, na.rm = TRUE )) %>%
  gather(variables, values, n_obs:birthwgt_mean) %>% #Reshape data
  mutate(variables = paste0(variables,"_",tobacco, sep="")) %>%
  pivot_wider(id_cols = g, names_from = variables,values_from = values) %>%
```

```r
  ungroup() %>%  #Ungroup from X values
  mutate(birthwgt_diff = birthwgt_mean_1 - birthwgt_mean_0, #calculate Y_diff
         w_ATE = (n_obs_0+n_obs_1)/(sum(n_obs_0)+sum(n_obs_1)),
         w_ATT = n_obs_1/sum(n_obs_1)) %>% #calculate weights
  mutate_if(is.numeric, round, 2) #Round data


stargazer(TIA_table, type= "text", summary = FALSE, digits = 2)
```

```
##
## =================================================================================
##      g   n_obs_0 n_obs_1 birthwgt_mean_0 birthwgt_mean_1 birthwgt_diff w_ATE w_ATT
## ---------------------------------------------------------------------------------
## 1  0000  44274   13443      3445.69         3220.25        -225.44    0.61  0.74
## 2  0001   214     448       3450.28         3124.25        -326.03    0.01  0.02
## 3  0010  7007    1980       3195.97         3006.31        -189.66     0.1  0.11
## 4  0011   71      226       3120.07         2817.34        -302.73      0   0.01
## 5  0100  13425    535       3483.02         3273.94        -209.08    0.15  0.03
## 6  0101   130     29        3510.95         3413.21        -97.74       0     0
## 7  0110   625     61        3319.22         3159.05        -160.17    0.01    0
## 8  0111    4      10        2983.5          3097.7          114.2       0     0
## 9  1000  5115     976       3467.41         3171.42        -295.98    0.06  0.05
## 10 1001   56      45        3358.32         3097.73        -260.59      0     0
## 11 1010   396     135       3185.08         2994.67        -190.41    0.01  0.01
## 12 1011    7      26        2739.71         2846.38        106.67       0     0
## 13 1100  4492     201       3487.19         3249.45        -237.74    0.05  0.01
## 14 1101   57      17        3534.91         3037.47        -497.44      0     0
## 15 1110   147     19        3328.29         2852.16        -476.13      0     0
## 16 1111    1       1          3459            2835          -624        0     0
## ---------------------------------------------------------------------------------
```

```r
# MULTIVARIATE MATCHING ESTIMATES OF ATE AND ATT
ATE=sum((TIA_table$w_ATE)*(TIA_table$Y_diff))
ATE
```

```
## [1] 0
```

2.4 Part 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.

```
# BASIC PROPENSITY SCORE --- THIS IS A TOY MODEL
# ESTIMATE PROPENSITY SCORE MODEL AND PREDICT (EPS)
ps_model <- glm(tobacco ~ mage + mage^2 + meduc + mblack + alcohol, family = binomial(), data = smoking

summary(ps_model)

EPS <- predict(ps_model, type = "response")
```

2.5 Part E

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

	(1)
(Intercept)	3445.873 ***
	(2.232)
tobacco	-226.245 ***
	(4.220)
as.factor(mage_indicator)1	10.359
	(6.819)
as.factor(meduc_indicator)1	37.809 ***
	(4.535)
as.factor(mblack)1	-241.839 ***
	(5.742)
as.factor(alcohol)1	-63.124 **
	(20.431)
as.factor(mage_indicator)1:as.factor(meduc_indicator)1	-7.343
	(10.601)
as.factor(mage_indicator)1:as.factor(mblack)1	-20.207
	(25.548)
as.factor(mage_indicator)1:as.factor(alcohol)1	-50.088
	(49.897)
as.factor(meduc_indicator)1:as.factor(mblack)1	83.254 ***
	(20.113)
as.factor(meduc_indicator)1:as.factor(alcohol)1	113.826 **
	(43.625)
as.factor(mblack)1:as.factor(alcohol)1	-79.043 *
	(36.419)
as.factor(mage_indicator)1:as.factor(meduc_indicator)1:as.factor(mblack)1	-8.222
	(50.558)
as.factor(mage_indicator)1:as.factor(meduc_indicator)1:as.factor(alcohol)1	-14.702
	(84.115)
as.factor(meduc_indicator)1:as.factor(mblack)1:as.factor(alcohol)1	-70.081
	(139.210)
as.factor(mage_indicator)1:as.factor(meduc_indicator)0:as.factor(mblack)1:as.factor(alcohol)1	0.080
	(101.358)
as.factor(mage_indicator)1:as.factor(meduc_indicator)1:as.factor(mblack)1:as.factor(alcohol)1	123.650