EDS241: Assignment 3

Scou Leonard

02/15/2022

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 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.

OH NOTES If we are not adjusting on anything, what is the assumption This is randomization of treatment, but unconditional randomization of treatment
Smoking status is randomly assigned between mothers - the assumption

statitsical significance fodifference in birthweights does not tell you anything about random assignment of smoking treatment -> we need random assignment of smoking status

Evidence: we need to show that any statistical difference between mothers that smoke and mothers that don't smoke -> run a model: income ~ tobacco, educ ~ tobacco, age ~ tobacco *are the means statitsically differnece for these?* provide some evidence that smoking is correlated with these variables . note the p values !

hypothesis: testing that treatment is independent of potential outcomes. what would need to be true for the difference in weights to be the average treatment effect? unconditional treatment ignorability *END*

```
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

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

```
mod_a2 <- lm_robust(birthwgt ~ mblack, data = smoking_df)

huxtable::huxreg(mod_a2)
```

The unadjusted mean difference in birth weight of infants with smoking and non- smoking mothers is -244.5 grams.

This corresponds to the alternative hypothesis that there is a difference between the infant birthweight between those with and without maternal smoking during pregnancy, holding all else equal regarding the mother's pregnancy and health.

Evidence against this hypothesis is that not all pregnancies and maternal health is equal; a mother living in wealth who smokes might have a heavier baby than a mother who lives in poverty and does not smoke, for example, if socioeconomic status has an impact on infant birth weight.

	(1)
(Intercept)	3412.271 ***
	(1.708)
mblack	-255.911 ***
	(5.307)
N	94173
R2	0.026

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

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.

OH NOTES

because it is conditional on the other variables, we include them all

in order to get the ate, they need to be as factors. but here, maybe we do them linearly.

getting towards matching and conditioning on observables.

the simplest way to start controlling for the covarites is to linearly include them

END

```r
# should we include the mage and meduc variables as.factor??
mod <- lm_robust(birthwgt ~ tobacco + mage + meduc + as.factor(mblack) + as.factor(alcohol) + as.factor

mod
```

```
##                          Estimate Std. Error     t value
## (Intercept)          3362.2582445 12.0764983 278.4133404
## tobacco              -228.0730765  4.2767834 -53.3281804
## mage                   -0.6940244  0.3681995  -1.8849143
## meduc                  11.6883416  0.8617788  13.5630420
## as.factor(mblack)1   -240.0303000  5.3477693 -44.8841915
## as.factor(alcohol)1   -77.3497487 14.0391720  -5.5095663
## as.factor(first)1     -96.9441154  3.4880224 -27.7934326
## as.factor(diabete)1    73.2275309 13.2354917   5.5326642
## as.factor(anemia)1     -4.7963916 17.8739216  -0.2683458
##
## (Intercept)          0.0000000000000000000000000000000000000000000000000000000000000000000000000000000
## tobacco              0.0000000000000000000000000000000000000000000000000000000000000000000000000000000
## mage                 0.0594444923473254546730792924336128635331988334655761718750000000000000000000000
## meduc                0.0000000000000000000000000000000000000000072624386997872499283709106250529202989
## as.factor(mblack)1   0.0000000000000000000000000000000000000000000000000000000000000000000000000000000
## as.factor(alcohol)1  0.0000000360657160655466344204580174653135271611859025142621248960494995117187500
## as.factor(first)1    0.0000000000000000000000000000000000000000000000000000000000000000000000000000000
## as.factor(diabete)1  0.0000000316239144957658212877400522467646881707992179144639521837234497070312500
```

```
## as.factor(anemia)1  0.788433752343045246568919992569135501980781555175781250000000000000000000000000
##                       CI Lower      CI Upper    DF
## (Intercept)          3338.588438 3385.92805061 94164
## tobacco              -236.455526 -219.69062730 94164
## mage                   -1.415691    0.02764255 94164
## meduc                   9.999265   13.37741865 94164
## as.factor(mblack)1  -250.511870 -229.54873009 94164
## as.factor(alcohol)1 -104.866374  -49.83312355 94164
## as.factor(first)1   -103.780602  -90.10762925 94164
## as.factor(diabete)1   47.286110   99.16895142 94164
## as.factor(anemia)1   -39.829084   30.23630133 94164
```

```r
#se
mod[[2]][[2]]
```

```
## [1] 4.276783
```

The average effect of maternal smoking on birth weight when all other covarients are held equal is -236.46 grams 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 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).

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))
```

```
#matching TIA pull from TIA table
```

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

```
#LINEAR REG ANALOGUE
mod_B <- 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)

huxtable::huxreg(mod_B)
```

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.

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.

```r
# 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)
```

```
##
## Call:
## glm(formula = tobacco ~ mage + mage^2 + meduc + mblack + alcohol,
##     family = binomial(), data = smoking_df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5257  -0.7017  -0.5466  -0.3300   2.5648
##
## Coefficients:
##              Estimate Std. Error z value             Pr(>|z|)
## (Intercept)  3.194297   0.064381  49.616 < 0.0000000000000002 ***
## mage        -0.025947   0.001759 -14.751 < 0.0000000000000002 ***
## meduc       -0.316296   0.005070 -62.385 < 0.0000000000000002 ***
## mblack      -0.082448   0.026357  -3.128              0.00176 **
## alcohol      2.022760   0.060089  33.663 < 0.0000000000000002 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 92325  on 94172  degrees of freedom
## Residual deviance: 84875  on 94168  degrees of freedom
## AIC: 84885
##
## Number of Fisher Scoring iterations: 5
```

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

(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