EDS241: Assignment 3

Wylie Hampson

02/14/2022

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

The control variables are: 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)

Import the data:

```
smoking <- read.csv(here("data", "SMOKING_EDS241.csv"))
```

Question 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.

```
smoking_group <- smoking %>%
  group_by(tobacco) %>%
  summarize(mean_birthwgt = mean(birthwgt))

smoking_group
```

```
## # A tibble: 2 x 2
##   tobacco mean_birthwgt
##     <int>         <dbl>
## 1       0         3430.
## 2       1         3186.
```

```
non_smoke_mean_wgt <- smoking_group %>% filter(tobacco == 0)
smoke_mean_wgt <- smoking_group %>% filter(tobacco == 1)

mean_wgt_diff <- non_smoke_mean_wgt$mean_birthwgt - smoke_mean_wgt$mean_birthwgt
```

The unadjusted mean difference in birth weight between babies from smoking and nonsmoking mothers is 244.5393875 grams. This leads us to the hypothesis that on average maternal smoking causes a decrease in *birth weights. The simple empirical evidence is the simple fact that on average babies from mothers who smoked had lower birth weights than those that did not smoke.*

Question 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.

```
model_1 <- lm_robust(formula = birthwgt ~ tobacco, data = smoking)

model_1_table <- broom::tidy(model_1) %>%
  dplyr::select(term, estimate, std.error, p.value) %>%
  knitr::kable()

model_1_table
```

term	estimate	std.error	p.value
(Intercept)	3430.2863	1.780943	0
tobacco	-244.5394	4.149552	0

Question 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 2x2x2x2 = 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).

```
smoking <- smoking %>%
  mutate(age_ind = ifelse((mage >= 34), 1, 0),
         edu_ind = ifelse((meduc >= 16), 1, 0))
```

Question 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.

```
ps_model <- glm(tobacco ~ mage + (mage ^ 2) + meduc + mblack + alcohol, family = binomial(), data = smol
summary(ps_model)
```

```
##
## Call:
## glm(formula = tobacco ~ mage + (mage^2) + meduc + mblack + alcohol,
##     family = binomial(), data = smoking)
##
## 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
```

```
EPS <- predict(ps_model, type = "response")
PS_WGT <- (smoking$tobacco/EPS) + ((1-smoking$tobacco)/(1-EPS))
```

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

```
wls1 <- lm(formula = birthwgt ~ tobacco, data = smoking, weights = PS_WGT)

se_models = starprep(wls1, stat = c("std.error"), se_type = "HC2", alpha = 0.05)
stargazer(wls1, se = se_models, type="text", omit = "(tobacco)")
```

```
##
## ================================================
##                         Dependent variable:
##                     ----------------------------
##                               birthwgt
## ------------------------------------------------
## Constant                    3,425.729***
##                               (1.850)
##
## ------------------------------------------------
## Observations                   94,173
## R2                             0.048
## Adjusted R2                    0.048
## Residual Std. Error     702.146 (df = 94171)
## F Statistic         4,762.462*** (df = 1; 94171)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```