EDS241: Assignment 3

Wylie Hampson

02/15/2022

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 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.

```
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

model_a <- lm(formula = meduc ~ tobacco, smoking)
```

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

broom::tidy(model_a)
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)    13.2    0.00743    1782.        0
## 2 tobacco        -1.32   0.0169      -77.9       0
```

The unadjusted mean difference in birth weight between babies from smoking and nonsmoking mothers is 244.5393875 grams. This assumes that smoking is a randomly assigned treatment which we can see by regressing tobacco use on mothers education, is not true. By regressing tobacco use on mothers education we can see that on average mothers that use tobacco would be expected to have an education level lower by 1.32 with a very low p-value. This shows that there is a relationship between those covariates, meaning tobacco use is not randomly assigned.

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 ~ ., 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)	3362.2582445	12.0764983	0.0000000
anemia	-4.7963916	17.8739216	0.7884338
diabete	73.2275309	13.2354917	0.0000000
tobacco	-228.0730765	4.2767834	0.0000000
alcohol	-77.3497487	14.0391720	0.0000000
mblack	-240.0303000	5.3477693	0.0000000
first	-96.9441154	3.4880224	0.0000000
mage	-0.6940244	0.3681995	0.0594445
meduc	11.6883416	0.8617788	0.0000000

The coefficient on tobacco is -244.54 with a standard error of 4.15.

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),
         g = paste0(age_ind, edu_ind, mblack, alcohol))

model_2 <- lm(formula = birthwgt ~
                tobacco +
                age_ind + edu_ind + mblack + alcohol +
                factor(g),
              data = smoking)

model_2_table <- broom::tidy(model_2) %>%
  dplyr::select(term, estimate, std.error, p.value) %>%
  knitr::kable()

model_2_table
```

term	estimate	std.error	p.value
(Intercept)	3445.87305	2.245419	0.0000000
tobacco	-226.24503	4.142761	0.0000000
age_ind	-191.81183	697.398486	0.7832868
edu_ind	258.11128	354.718261	0.4668287
mblack	-212.48783	349.057690	0.5426936
alcohol	-39.56216	346.504796	0.9090994
factor(g)0001	-23.56199	347.023921	0.9458675
factor(g)0010	-29.35116	349.101565	0.9329960
factor(g)0011	-131.95648	692.801352	0.8489428
factor(g)0100	-220.30216	354.746217	0.5345923
factor(g)0101	-130.03786	696.168645	0.8518251
factor(g)0110	-166.39891	696.621344	0.8112103
factor(g)0111	-225.25932	1047.129826	0.8296737
factor(g)1000	202.17095	697.423351	0.7719066
factor(g)1001	128.52100	362.462701	0.7229072
factor(g)1010	152.61320	357.351580	0.6693312
factor(g)1011	NA	NA	NA
factor(g)1100	-25.47460	351.170292	0.9421707
factor(g)1101	NA	NA	NA
factor(g)1110	NA	NA	NA
factor(g)1111	NA	NA	NA

```
TIA_table <- smoking %>%
  group_by(g, tobacco)%>%
  summarise(n_obs = n(),
            birthwgt_mean= mean(birthwgt, na.rm = T))%>% #Calculate number of observations and birthwgt
  gather(variables, values, n_obs:birthwgt_mean)%>% #Reshape data
  mutate(variables = paste0(variables,"_",tobacco, sep=""))%>% #Combine the treatment and variables for
  pivot_wider(id_cols = g, names_from = variables,values_from = values)%>% #Reshape data by treatment a
  ungroup()%>%  #Ungroup from g values
  mutate(birthwgt_diff = birthwgt_mean_1 - birthwgt_mean_0, #calculate birthwgt_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
## --------------------------------------------------------------------------------
```

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

```
## [1] -224.2583
```

```
ATT=sum((TIA_table$w_ATT)*(TIA_table$birthwgt_diff))
ATT
```

```
## [1] -222.589
```

The average treatment effect of smoking on birthweight using the exact matching estimator is -224.26 grams.

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.

```
smoking <- smoking %>%
  mutate(mage_sq = mage ^ 2)

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

```
##
## Call:
## glm(formula = tobacco ~ mage + mage_sq + meduc + mblack + alcohol,
##     family = binomial(), data = smoking)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5482  -0.7182  -0.5461  -0.3214   2.6709
##
## Coefficients:
##              Estimate Std. Error z value             Pr(>|z|)
## (Intercept)  1.929611   0.191814  10.060 < 0.0000000000000002 ***
## mage         0.077636   0.014915   5.205     0.00000019355476 ***
## mage_sq     -0.001941   0.000278  -6.983     0.00000000000288 ***
## meduc       -0.321597   0.005144 -62.520 < 0.0000000000000002 ***
## mblack      -0.059525   0.026506  -2.246               0.0247 *
## alcohol      2.022696   0.060358  33.511 < 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: 84825  on 94167  degrees of freedom
## AIC: 84837
##
## 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)

wls1_table <- broom::tidy(wls1) %>%
  dplyr::select(term, estimate, std.error, p.value) %>%
  knitr::kable()

wls1_table
```

term	estimate	std.error	p.value
(Intercept)	3425.9937	2.288148	0
tobacco	-225.4748	3.262564	0