EDS241: Assignment 3

Clarissa Boyajian

02/16/2022

In this assignment we estimate the causal effect of maternal smoking during pregnancy on infant birth weight using the treatment ignorability assumptions. Data come from the National Natality Detail Files and are a random sample of all births in Pennsylvania during 1989-1991, with each observation being a mother-infant pair.

1 Load and clean data

The following code loads and cleans the data.

```
data_smoking <- read.csv("data/SMOKING_EDS241.csv") %>%
  clean_names()
```

2 Unadjusted mean difference of birth weights

The code chunk below runs a simple linear regression of birth weights on the binary variable of whether the mother was a smoker or nonsmoker.

```
model_1.1 <- lm_robust(data = data_smoking,
                       birthwgt ~ tobacco)

huxreg("Birth Weight" = model_1.1)
```

	Birth Weight
(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 code chunk below runs a simple linear regression of mothers' education on the binary variable of whether the mother was a smoker or nonsmoker.

```
model_1.2 <- lm_robust(data = data_smoking,
                       meduc ~ tobacco)

model_1.3 <- lm_robust(data = data_smoking,
                       mage ~ tobacco)

huxreg("Education" = model_1.2, "Age" = model_1.3)
```

	Education	Age
(Intercept)	13.239 ***	27.453 ***
	(0.008)	(0.019)
tobacco	-1.318 ***	-1.915 ***
	(0.014)	(0.043)
N	94173	94173
R2	0.061	0.020

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

a. The average un-adjusted mean difference in birth weight of infants with smoking and nonsmoking mothers is a decrease of 244.54 grams. This estimate only holds if smoking status is random assigned between mothers unconditional on the other covariates. This is likely not true, as we know that whether someone smokes or not is impacted by their environment. To check whether tobacco usage is correlated with a mother's level of education or her age, I ran two simple linear regressions. The first regressed mother's education on tobacco, and showed a statistically significant different of -1.32. The second regressed mother's age on tobacco, and showed a statistically significant different of -1.91.

3 Ddjusted mean difference of birth weights

3.1 Linear regression

The code chunk below runs a simple linear regression of birth weights on the binary variable of whether the mother was a smoker or nonsmoker, holding other covariates constant.

```
model_2 <-
  lm_robust(data = data_smoking,
            birthwgt ~ tobacco + anemia + diabete +
              alcohol + mblack + first + mage + meduc)

huxtable_model_2 <- huxreg("Birthweight" = model_2)
```

	Birthweight		Birthweight		Birthweight
(Intercept)	3362.258 ***	alcohol	-77.350 ***	meduc	11.688 ***
	(12.076)		(14.039)		(0.862)
tobacco	-228.073 ***	mblack	-240.030 ***	N	94173
	(4.277)		(5.348)	R2	0.072
anemia	-4.796	first	-96.944 ***	*** p < 0.001; ** p < 0.01; * p < 0.05.	
	(17.874)		(3.488)		
diabete	73.228 ***	mage	-0.694		
	(13.235)		(0.368)		

b. The average effect of maternal smoking on birth weight when all other covariates are held equal is a decrease of 228.07 grams. The robust standard error is 4.28.

3.2 Exact matching estimator

The code chunk below creates a new dataframe that transforms into binary indicators mother's age (=1 if `mage` >= 34) and level of education (=1 if `meduc` >= 16). This code chunk also creates a column tracking which of the four dummy variables (age, education, alcohol, and race) are turned on for each observation.

```
data_smoking_binary <-
  data_smoking %>%
  mutate(mage = case_when(mage >= 34 ~ 1,
                          mage < 34 ~ 0),
         meduc = case_when(meduc >= 16 ~ 1,
                           meduc < 16 ~ 0),
         covariates = paste0(alcohol, mblack, mage, meduc))
```

The code chunk below calculates the ATE of smoking on birth weight using the exact matching estimator.

```
ATE_table <- data_smoking_binary %>%
  group_by(covariates, tobacco) %>%
  # calculate number of observations
  summarise(n_obs = n(),
            # calculate birthwgt mean by X by treatment cells
            birthwgt_mean = mean(birthwgt, na.rm = TRUE)) %>%
  # reshape the 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 = covariates,
              names_from = variables,
              values_from = values) %>%
  ungroup() %>%
  # calculate birthwgt_diff
```

```
  mutate(birthwgt_diff = birthwgt_mean_1 - birthwgt_mean_0,
         # calculate ATE weights
         w_ATE = (n_obs_0 + n_obs_1) / (sum(n_obs_0) + sum(n_obs_1)),
         # calculate ATT weights
         w_ATT = n_obs_1 / sum(n_obs_1)) %>%
  mutate_if(is.numeric, round, 2)

huxtable(ATE_table)
```

covariates	n_obs_0	n_obs_1	birthwgt_mean_0	birthwgt_mean_1	birthwgt_diff	w_ATE	w_ATT
0000	4.43e+04	1.34e+04	3.45e+03	3.22e+03	-225	0.61	0.74
0001	1.34e+04	535	3.48e+03	3.27e+03	-209	0.15	0.03
0010	5.12e+03	976	3.47e+03	3.17e+03	-296	0.06	0.05
0011	4.49e+03	201	3.49e+03	3.25e+03	-238	0.05	0.01
0100	7.01e+03	1.98e+03	3.2e+03	3.01e+03	-190	0.1	0.11
0101	625	61	3.32e+03	3.16e+03	-160	0.01	0
0110	396	135	3.19e+03	2.99e+03	-190	0.01	0.01
0111	147	19	3.33e+03	2.85e+03	-476	0	0
1000	214	448	3.45e+03	3.12e+03	-326	0.01	0.02
1001	130	29	3.51e+03	3.41e+03	-97.7	0	0
1010	56	45	3.36e+03	3.1e+03	-261	0	0
1011	57	17	3.53e+03	3.04e+03	-497	0	0
1100	71	226	3.12e+03	2.82e+03	-303	0	0.01
1101	4	10	2.98e+03	3.1e+03	114	0	0
1110	7	26	2.74e+03	2.85e+03	107	0	0
1111	1	1	3.46e+03	2.84e+03	-624	0	0

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

The code chunk below calculates the ATE of smoking on birth weight using the exact matching estimator's linear regression analogue.

```
## linear regression aspect
model_3 <-
  lm_robust(data = data_smoking_binary,
            birthwgt ~ tobacco + as.factor(covariates))

huxtable_model_3 <- huxreg("Birth weight" = model_3)
```

	Birth weight		Birth weight
(Intercept)	3445.873 ***	as.factor(covariates)1001	88.511 *
	(2.232)		(38.413)
tobacco	-226.245 ***	as.factor(covariates)1010	-102.853 *
	(4.220)		(45.144)
as.factor(covariates)0001	37.809 ***	as.factor(covariates)1011	26.737
	(4.535)		(55.254)
as.factor(covariates)0010	10.359	as.factor(covariates)1100	-384.006 ***
	(6.819)		(29.870)
as.factor(covariates)0011	40.825 ***	as.factor(covariates)1101	-219.198
	(7.404)		(127.345)
as.factor(covariates)0100	-241.839 ***	as.factor(covariates)1110	-443.862 ***
	(5.742)		(79.415)
as.factor(covariates)0101	-120.775 ***	as.factor(covariates)1111	-185.751
	(18.977)		(198.895)
as.factor(covariates)0110	-251.686 ***	N	94173
	(24.106)	R2	0.063
as.factor(covariates)0111	-146.188 ***	*** p < 0.001; ** p < 0.01; * p < 0.05.	
	(38.555)		
as.factor(covariates)1000	-63.124 **		
	(20.431)		

c. Using the exact matching estimator, the average effect of maternal smoking on birth weight is a decrease of 224.2583 grams. Using the linear regression analogue, the average effect of maternal smoking on birth weight is a decrease of 226.25 grams.

3.3 Propensity score

The code chunk below estimates the propensity score for maternal smoking using a logit estimator and includes the following covariables: mother's age, mother's age squared, mother's education, and indicators for mother's race, and alcohol consumption.

```
# calculate `mage^2`
data_smoking_propensity <-
  data_smoking %>%
  mutate(mage_2 = mage*mage)
```

```r
# ESTIMATE PROPENSITY SCORE MODEL AND PREDICT (EPS)
ps_mode_l <- glm(tobacco ~ mage + mage_2 + meduc + mblack + alcohol,
                 family = binomial(),
                 data = data_smoking_propensity)
# create new variable
EPS <- predict(ps_mode_l, type = "response")
# propensity-score
PS_WGT <- (data_smoking_propensity$tobacco / EPS) +
  ((1 - data_smoking_propensity$tobacco) / (1 - EPS))
```

d. The above code chunk calculates a table of estimate propensity scores that will be used below to run a weighted regression to estimate the

The code chunk below uses the propensity score weighted regression (WLS) to estimate the effect of maternal smoking on birth weight.

```r
wls1 <- lm(data = data_smoking_propensity,
           birthwgt ~ tobacco + mage + mage_2 + meduc + mblack + alcohol,
           weights = PS_WGT)

huxtable_wls1 <- huxreg("Birth Weight" = wls1)
```

e. The estimated effect of maternal smoking on birth weight using the propensity score is a decrease of 220.23 grams.

```r
restack_across(huxtable_wls1, 11)
```

	Birth Weight		Birth Weight
(Intercept)	2971.444 ***	mblack	-220.990 ***
	(36.122)		(4.994)
tobacco	-220.233 ***	alcohol	-71.914 ***
	(3.223)		(13.709)
mage	27.627 ***	N	94173
	(2.693)	R2	0.074
mage_2	-0.478 ***	logLik	-728569.509
	(0.049)	AIC	1457155.018
meduc	7.472 ***	*** p < 0.001; ** p < 0.01; * p < 0.05.	
	(0.849)		