EDS241: Assignment 3

Peter Menzies

02/16/2022

Reading in data

```
df <- read_csv(here("data", "SMOKING_EDS241.csv")) %>%
  clean_names()
```

(a) Mean difference in birth weight of infants with smoking and nonsmoking mothers

```
df_wt_diff <- df %>%
  group_by(tobacco) %>%
  summarize(mean_wt = mean(birthwgt))

wt_diff <- df_wt_diff$mean_wt[df_wt_diff$tobacco == 0] - df_wt_diff$mean_wt[df_wt_diff$tobacco == 1]
```

```
t.test(df$alcohol[df$tobacco == 1], df$alcohol[df$tobacco == 0])
```

```
##
##  Welch Two Sample t-test
##
## data:  df$alcohol[df$tobacco == 1] and df$alcohol[df$tobacco == 0]
## t = 23.838, df = 19620, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.03403037 0.04012795
## sample estimates:
##  mean of x  mean of y
## 0.04418246 0.00710330
```

The unadjusted mean difference in birth weight of infants with smoking and nonsmoking mothers is 244.5393875 grams.

This would correspond to the average treatment effect of maternal smoking during pregnancy on infant birth weight under the assumption that the treatment (mother smoking during pregnancy) is statistically independent of all control variables.

The mean difference in alcohol consumption during pregnancy (one of the control variables) between smoking and non-smoking mothers is statistically different from zero (shown above with a t-test), and thus we can't make the previous assumption regarding ATE.

(b) Multiple linear regression of birth weight on smoking and control variables

```
mdl_b <- lm_robust(birthwgt ~ ., df)
```

```
mdl_b %>%
  starprep() %>%
  stargazer(header = FALSE)
```

Table 1:

(Intercept)	anemia	diabete	tobacco	alcohol	mblack	first	mage	meduc
12.076	17.874	13.235	4.277	14.039	5.348	3.488	0.368	0.862

The estimated coefficient on tobacco in this model is -228.0730765 +/- 4.2767834

(c) Estimating the effect of maternal smoking on birth weight using exact matching estimator

```
df_c <- df %>%
  mutate(mage = case_when(mage >= 34 ~ 1,
                          mage < 34 ~ 0)) %>%
  mutate(meduc = case_when(meduc >= 16 ~ 1,
                           meduc < 16 ~ 0)) %>%
  mutate(g = as.factor(paste0(mage, meduc, mblack, alcohol)))
```

```
tia <- df_c %>%
  group_by(g, tobacco) %>%
  summarise(n_obs = n(),
            mean_wt = mean(birthwgt, na.rm = T)) %>%
  gather(variables, values, n_obs:mean_wt) %>%
  mutate(variables = paste0(variables, "_", tobacco)) %>%
  pivot_wider(id_cols = g, names_from = variables, values_from = values) %>%
  ungroup() %>%
  mutate(wt_diff = mean_wt_1 - mean_wt_0,
         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)) %>%
  mutate_if(is.numeric, round, 2)
```

```
stargazer(tia, summary = FALSE, digits = 2, header = FALSE)
```

```
ate = sum((tia$w_ate) * (tia$wt_diff))
ate
```

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

Table 2:

	g	n_obs_0	n_obs_1	mean_wt_0	mean_wt_1	wt_diff	w_ate	w_att
1	1	44274	13443	3445.69	3220.25	-225.44	0.61	0.74
2	2	214	448	3450.28	3124.25	-326.03	0.01	0.02
3	3	7007	1980	3195.97	3006.31	-189.66	0.1	0.11
4	4	71	226	3120.07	2817.34	-302.73	0	0.01
5	5	13425	535	3483.02	3273.94	-209.08	0.15	0.03
6	6	130	29	3510.95	3413.21	-97.74	0	0
7	7	625	61	3319.22	3159.05	-160.17	0.01	0
8	8	4	10	2983.5	3097.7	114.2	0	0
9	9	5115	976	3467.41	3171.42	-295.98	0.06	0.05
10	10	56	45	3358.32	3097.73	-260.59	0	0
11	11	396	135	3185.08	2994.67	-190.41	0.01	0.01
12	12	7	26	2739.71	2846.38	106.67	0	0
13	13	4492	201	3487.19	3249.45	-237.74	0.05	0.01
14	14	57	17	3534.91	3037.47	-497.44	0	0
15	15	147	19	3328.29	2852.16	-476.13	0	0
16	16	1	1	3459	2835	-624	0	0

```
att = sum((tia$w_att) * (tia$wt_diff))
att
```

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

```
lm_c <- lm_robust(birthwgt ~ tobacco + g, df_c)
```

```
lm_c$coefficients %>%
  stargazer(header = FALSE, digits = 2, column.sep.width = "0.5pt")
```

Table 3:

(Intercept)	tobacco	g0001	g0010	g0011	g0100	g0101	g0110	g0111
3,445.87	-226.25	-63.12	-241.84	-384.01	37.81	88.51	-120.78	-219.20

g1000	g1001	g1010	g1011	g1100	g1101	g1110	g1111
10.36	-102.85	-251.69	-443.86	40.82	26.74	-146.19	-185.75

The estimated average treatment effect of smoking on birth weight using the exact matching estimator is -224.2583 compared with -226.2450329 for the analagous linear regression.

(d) Estimating the propensity score for maternal smoking using a logit estimator

```
df_d <- df %>%
  mutate(mage2 = mage ^ 2,
         wt_scaled = scale(birthwgt))
```

```
logit_mdl <- glm(tobacco ~ mage + mage2 + meduc + mblack + alcohol, family = binomial(link = "logit"), d
```

```
logit_mdl %>% stargazer(header = FALSE)
```

Table 4:

	Dependent variable:
	tobacco
mage	0.078***
	(0.015)
mage2	−0.002***
	(0.0003)
meduc	−0.322***
	(0.005)
mblack	−0.060**
	(0.027)
alcohol	2.023***
	(0.060)
Constant	1.930***
	(0.192)
Observations	94,173
Log Likelihood	−42,412.660
Akaike Inf. Crit.	84,837.330
Note:	*p<0.1; **p<0.05; ***p<0.01

```
eps <- predict(logit_mdl, type = "response")

ps_wgt <- (df_d$tobacco / eps) + ((1 - df_d$tobacco) / (1 - eps))
```

(e) Using the propensity score weighted regression to estimate the effect of maternal smoking on birth weight

```
wls_mdl <- lm_robust(birthwgt ~ tobacco, df, weights = ps_wgt)
```

```
wls_mdl %>%
  starprep() %>%
  stargazer(header = FALSE)
```

Table 5:

(Intercept)	tobacco
1.854	5.025