EDS241: Assignment 3

Peter Menzies

02/19/2022

Reading in data

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

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

```
mean_wt_smoking <- mean(df$birthwgt[df$tobacco == 1])
mean_wt_nonsmoking <- mean(df$birthwgt[df$tobacco == 0])

wt_diff <- mean_wt_smoking - mean_wt_nonsmoking
```

```
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 between infants with mothers who smoked during pregnancy and those with mothers who did not 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 relevant baseline characteristics.

The mean difference in alcohol consumption during pregnancy 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 all control variables

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

```
mdl_b %>%
  tidy() %>%
  xtable()
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	3362.26	12.08	278.41	0.00	3338.59	3385.93	94164.00	birthwgt
2	anemia	-4.80	17.87	-0.27	0.79	-39.83	30.24	94164.00	birthwgt
3	diabete	73.23	13.24	5.53	0.00	47.29	99.17	94164.00	birthwgt
4	tobacco	-228.07	4.28	-53.33	0.00	-236.46	-219.69	94164.00	birthwgt
5	alcohol	-77.35	14.04	-5.51	0.00	-104.87	-49.83	94164.00	birthwgt
6	mblack	-240.03	5.35	-44.88	0.00	-250.51	-229.55	94164.00	birthwgt
7	first	-96.94	3.49	-27.79	0.00	-103.78	-90.11	94164.00	birthwgt
8	mage	-0.69	0.37	-1.88	0.06	-1.42	0.03	94164.00	birthwgt
9	meduc	11.69	0.86	13.56	0.00	10.00	13.38	94164.00	birthwgt

The estimated coefficient on tobacco in this model is -228.0730765 $\pm$ 4.2767834 grams.

(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 = as.integer(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)
```

```
xtable(tia, digits = c(0, 0, 0, 0 , 2, 2, 2, 2, 2))
```

	g	n_obs_0	n_obs_1	mean_wt_0	mean_wt_1	wt_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	0.10	0.11
4	0011	71	226	3120.07	2817.34	-302.73	0.00	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	0.00	0.00
7	0110	625	61	3319.22	3159.05	-160.17	0.01	0.00
8	0111	4	10	2983.50	3097.70	114.20	0.00	0.00
9	1000	5115	976	3467.41	3171.42	-295.98	0.06	0.05
10	1001	56	45	3358.32	3097.73	-260.59	0.00	0.00
11	1010	396	135	3185.08	2994.67	-190.41	0.01	0.01
12	1011	7	26	2739.71	2846.38	106.67	0.00	0.00
13	1100	4492	201	3487.19	3249.45	-237.74	0.05	0.01
14	1101	57	17	3534.91	3037.47	-497.44	0.00	0.00
15	1110	147	19	3328.29	2852.16	-476.13	0.00	0.00
16	1111	1	1	3459.00	2835.00	-624.00	0.00	0.00

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

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

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

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

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

```
lm_c %>%
  tidy() %>%
  xtable()
```

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

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	3445.87	2.23	1543.60	0.00	3441.50	3450.25	94156.00	birthwgt
2	tobacco	-226.25	4.22	-53.61	0.00	-234.52	-217.97	94156.00	birthwgt
3	g0001	-63.12	20.43	-3.09	0.00	-103.17	-23.08	94156.00	birthwgt
4	g0010	-241.84	5.74	-42.12	0.00	-253.09	-230.58	94156.00	birthwgt
5	g0011	-384.01	29.87	-12.86	0.00	-442.55	-325.46	94156.00	birthwgt
6	g0100	37.81	4.53	8.34	0.00	28.92	46.70	94156.00	birthwgt
7	g0101	88.51	38.41	2.30	0.02	13.22	163.80	94156.00	birthwgt
8	g0110	-120.78	18.98	-6.36	0.00	-157.97	-83.58	94156.00	birthwgt
9	g0111	-219.20	127.34	-1.72	0.09	-468.79	30.40	94156.00	birthwgt
10	g1000	10.36	6.82	1.52	0.13	-3.01	23.72	94156.00	birthwgt
11	g1001	-102.85	45.14	-2.28	0.02	-191.33	-14.37	94156.00	birthwgt
12	g1010	-251.69	24.11	-10.44	0.00	-298.93	-204.44	94156.00	birthwgt
13	g1011	-443.86	79.41	-5.59	0.00	-599.51	-288.21	94156.00	birthwgt
14	g1100	40.82	7.40	5.51	0.00	26.31	55.34	94156.00	birthwgt
15	g1101	26.74	55.25	0.48	0.63	-81.56	135.03	94156.00	birthwgt
16	g1110	-146.19	38.55	-3.79	0.00	-221.76	-70.62	94156.00	birthwgt
17	g1111	-185.75	198.89	-0.93	0.35	-575.58	204.08	94156.00	birthwgt

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

```
df_d <- df %>%
  mutate(mage2 = mage ** 2)
```

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

```
logit_mdl %>%
  tidy() %>%
  xtable()
```

	term	estimate	std.error	statistic	p.value
1	(Intercept)	1.93	0.19	10.06	0.00
2	mage	0.08	0.01	5.21	0.00
3	mage2	-0.00	0.00	-6.98	0.00
4	meduc	-0.32	0.01	-62.52	0.00
5	mblack	-0.06	0.03	-2.25	0.02
6	alcohol	2.02	0.06	33.51	0.00

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

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

(e) Using 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 %>%
  tidy() %>%
  xtable()
```

	term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
1	(Intercept)	3425.99	1.85	1847.79	0.00	3422.36	3429.63	94171.00	birthwgt
2	tobacco	-225.47	5.03	-44.87	0.00	-235.32	-215.63	94171.00	birthwgt

The estimated effect of maternal smoking on birth weight using propensity score weighted regression is -225.4748287 grams.