EDS 241: Assignment 1

Javier Patrón

2023-03-11

Question 1:

Application of estimators based on the "treatment ignorability" assumption. The goal is to estimate the causal effect of maternal smoking during pregnancy on infant birth weight using the treatment ignorability assumptions (Lecture 6 & 7).

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)

Load in the libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(stargazer)
```

```
##
## Please cite as:
##
##  Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
##  R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```
library(estimatr)
```

Load in the data

```
smoke <- read_csv(here::here("SMOKING_EDS241.csv"))
```

```
## Rows: 94173 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (9): anemia, diabete, tobacco, alcohol, mblack, first, mage, meduc, birt...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

(a) What is the unadjusted mean difference in birth weight of infants with smoking and nonsmoking 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 assumption.

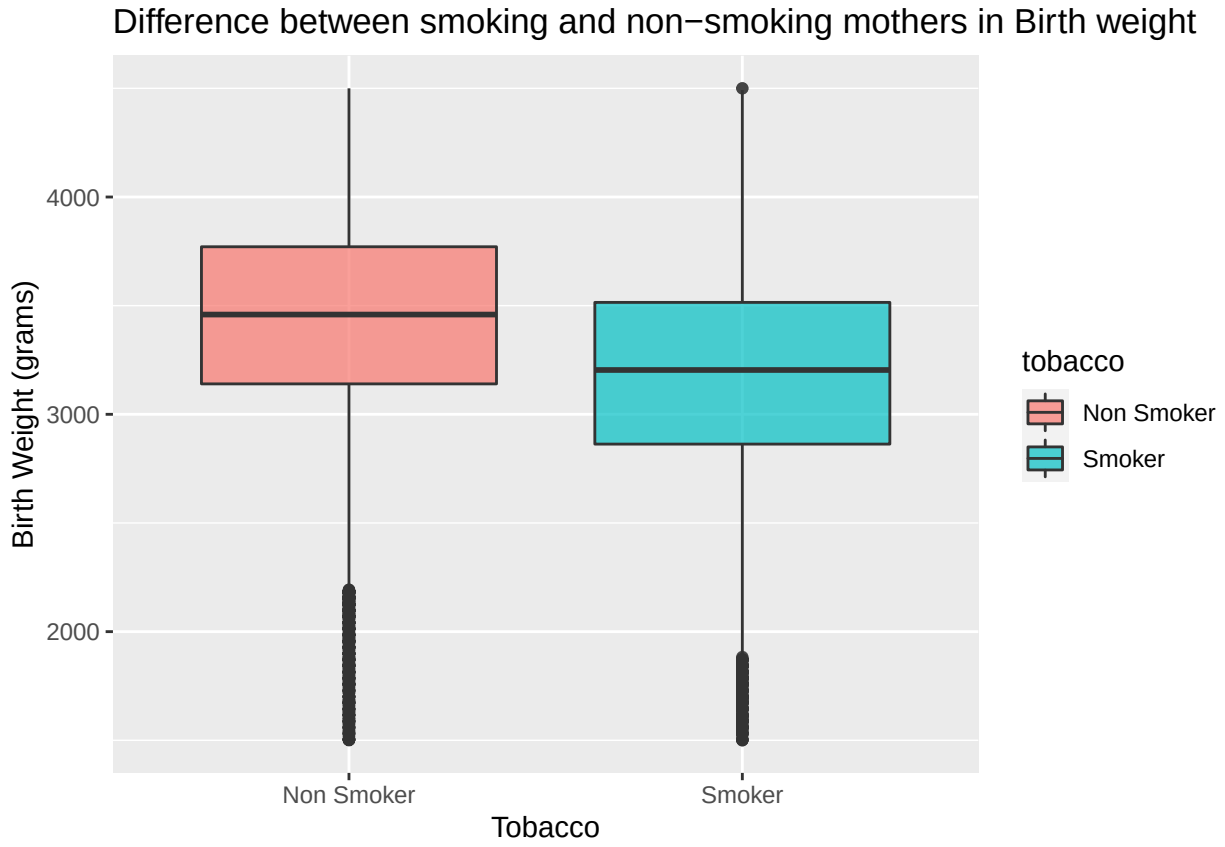
```
#Question (a)

#What is the mean difference between smoking and non smoking

weight_means <- smoke |>
  group_by(tobacco) |>
  summarise(birthwgt_mean = mean(birthwgt))

weight_means$birthwgt_mean[1]
```

```
## [1] 3430.286
```

```
print(paste0("The difference in birth weight means from smoking and non smoking is: ", round(weight_mea
```

```
## [1] "The difference in birth weight means from smoking and non smoking is: 244.54(grams) meaning tha
```

```
smoke |>
  mutate(tobacco = case_when(tobacco == 0 ~ "Non Smoker",
                             tobacco == 1 ~ "Smoker")) |>
  ggplot(aes(x= as.factor(tobacco), y = birthwgt, fill = tobacco)) +
  geom_boxplot(alpha = 0.7) +
  labs(x = "Tobacco",
       y = "Birth Weight (grams)",
       title = "Difference between smoking and non-smoking mothers in Birth weight ")
```



print(paste0("The difference in birth weight means from smoking and non smoking is:", 244.54,"(grams)", " meaning that the weight of the babies from smoker mothers are 244.54 lighter"))

Another way of seeing this is by the lm() function

```
lm(data = smoke, birthwgt ~ tobacco)
```

```
##
## Call:
## lm(formula = birthwgt ~ tobacco, data = smoke)
##
## Coefficients:
## (Intercept)      tobacco
##      3430.3       -244.5
```

Thanks to the results above with lm() we can see the same result as above. The "slope" tell us that the difference between weight babies between smokers and nonsmokers is -244.5 grams. For every "unit" which is YES or NO, on average a baby will be 244.5 grams lighter **omitting all of the other variables**.

(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 an OLS regression with linear controls for the covariates. Report the estimated coefficient on tobacco and its standard error.

```
# Question (b)
```

```r
# Fit the linear regression model with all the variables
all <- lm(data = smoke, birthwgt ~ .)

# Create a pretty table of the results
x <- starprep(all,
              stat = c("std.error"),
              se_type = "HC1",
              alpha = 0.05)


stargazer(all,
           se = x,
           type="text")
```

```
##
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                                birthwgt
## ------------------------------------------------
## anemia                          -4.796
##                                (17.864)
##
## diabete                         73.228***
##                                (13.232)
##
## tobacco                        -228.073***
##                                 (4.277)
##
## alcohol                        -77.350***
##                                (14.034)
##
## mblack                         -240.030***
##                                 (5.348)
##
## first                          -96.944***
##                                 (3.488)
##
## mage                             -0.694*
##                                 (0.368)
##
## meduc                           11.688***
##                                 (0.862)
##
## Constant                      3,362.258***
##                                (12.076)
##
## ------------------------------------------------
## Observations                     94,173
## R2                                0.072
## Adjusted R2                       0.072
## Residual Std. Error    484.733 (df = 94164)
## F Statistic         909.176*** (df = 8; 94164)
```

```
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

As you can see in the table above the tobacco impact in birth weight has moved from 228.073, meaning that tobacco is not the only variable that is affecting the decrease in weight for new borns. There are other factors and variables meaning that the initial assumption is not right.

(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 2 * 2 * 2 * 2 = 16 cells.

```
# Question (c)


#In this case you need to create a new column with the ceros and ones and then you apply the lm(weight


#and you say and predict what is the new factor
```

Report the estimated average treatment effect of smoking on birthweight using the exact matching estimator and its linear regression analogue.