EDS 241 Assignment 2

Mia Forsline

2/20/2022

1 Set Up

2 Introduction

2.1 Goal: to estimate the causal effect of maternal smoking during pregnancy on infant birth weight using the treatment ignorability assumptions

- data come from the Child & Family Data Archive's National Natality Detail Files
- the data we will be using is a random sample of all births in Pennsylvania during 1989 - 1991
- each observation is a mother-infant pair

2.2 Variables

- Outcome variable: `birthwgt` = birth weight of infant in grams

- Treatment variable: `tobacco` = indicator for maternal smoking

- Control variables:

    - `mage` = mother's age
    - `meduc` = mother's education
    - `mblack` = 1 if the mother is Black
    - 'alcohol = 1 if the mother consumed alcohol during pregnancy
    - `first` = 1 if this is the mother's first child
    - `diabete` = 1 if the mother is diabetic
    - `anemia` = 1 if the mother is anemic

2.3 Note

This homework is a simple examination of these data. More research would be needed to obtain a more definitive assessment of the causal effect of smoking on infant health outcomes. Further, for this homework, you can ignore the adjustments to the standard errors that are necessary to reflect the fact that the propensity score is estimated. Just use heteroskedasticity robust standard errors in R. If you are interested, you can read Imbens and Wooldridge (2009) and Imbens (2014) for discussions of various approaches and issues with standard error estimations in models based on the propensity score

```
data <- read_csv(here("data", "SMOKING_EDS241.csv"))
```

3 (a) What is the unadjusted mean difference in birth weight of infants with smoking and nonsmoking mothers?

```
data_smoke <- subset(data, tobacco == 1)
smoke_mean <- round(mean(data_smoke$birthwgt), digits = 2)

data_no_smoke <- subset(data, tobacco == 0)
no_smoke_mean <- round(mean(data_no_smoke$birthwgt), digits = 2)

diff <- no_smoke_mean - smoke_mean
```

On average, smoking mothers give birth to babies that weight 3185.75g. On average, non-smoking mothers give birth to babies that weight 3430.29g. Thus, the unadjusted mean difference in infant birth weights between smoking and nonsmoking mothers is 244.54g.

3.1 (a) Under what hypothesis 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.

```
y1 <- round(mean(data_smoke$birthwgt), digits = 2)
y0 <- round(mean(data_no_smoke$birthwgt), digits = 2)

ate <- y1 - y0
```

The average treatment effect (ATE) is the expected value of $Y(1)$ - $Y(0)$ within the population of interest. Thus, we would hypothesize that for mothers giving birth in Pennsylvania during 1989 - 1991, smoking during pregnancy influences infant birth weight. Specifically, we hypothesize that smoking during pregnancy results in lower infant birth weight.

H_0: Maternal smoking during pregnancy does not influence infant birth weight ($\mu = 0$)

H_A: Maternal smoking during pregnancy does influence infant birth weight ($\mu \neq 0$)

In this case, $Y(1) = 3185.75$ and $Y(0) = 3430.29$, so the ATE = -244.54.

4   (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

5   (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. 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).

6   (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.

7   (e) Use the propensity score weighted regression (WLS) to estimate the effect of maternal smoking on birth weight (Lecture 7, slide 12).