[PPHA34600] Probelm Evaluation PS3

Lee Kyung Ko

2022-11-03

```
#setwd("C:/Users/LEE KO/Desktop/Program Eval/PS/PS3")

library(tidyverse)
library(dplyr)
library(haven)
library(corrplot)
library(AER)
library(stargazer)
library(knitr)
library(estimatr)
library(Rcpp)
library(broom)
library(tidyr)

#Loading data and merging

data <- read_dta("almond_etal_2008.dta")
```

Q1.

Descriptive statistics of birth weight in the sample. What is the mean, standard deviation, minimum, and maximum?

- mean: 1511.6

- standard deviation: 89.0
- minimum: 1350.0
- maximum: 1650.0

```
bweight_col <- data %>% select(bweight)

#multiple function to one column
#source: https://www.projectpro.io/recipes/apply-multiple-functions-on-list-r-one-go
multiple_function <- function(x){
  c(mean = mean(x), sd = sd(x), min = min(x), max= max(x))
}

descr_stat <- bweight_col %>% lapply(multiple_function)

print(descr_stat)
```

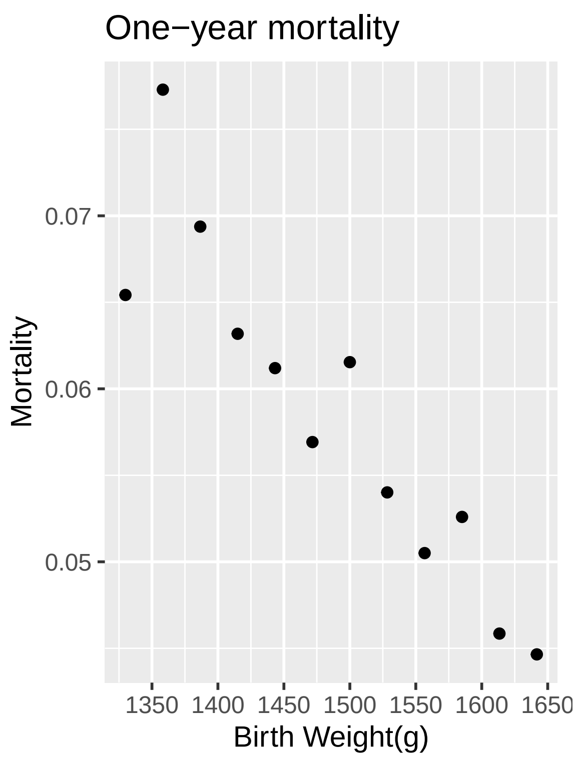
```
## $bweight
##       mean         sd        min        max
## 1511.57565   89.01614 1350.00000 1650.00000
```

Q2.

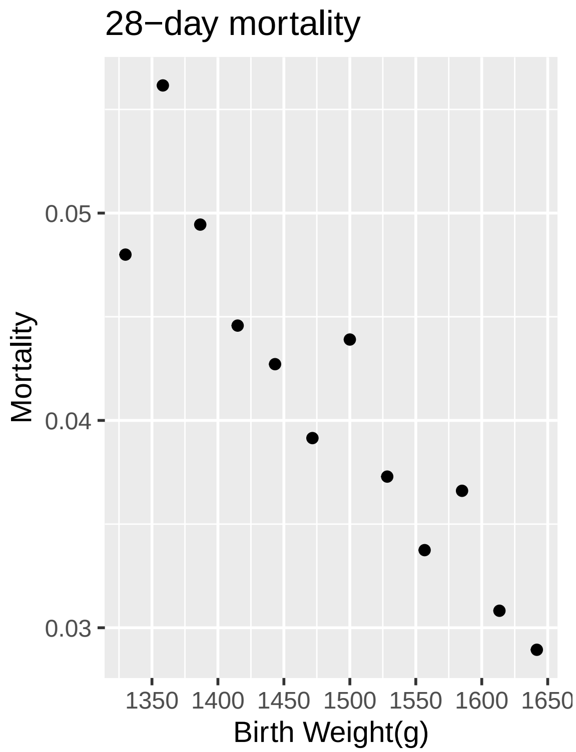
Plot one year and 28-day mortality rates against our running variable, birth weight. To do so, make bins of one ounce (28.35 grams) around the 1500 grams threshold and get the mean mortality rate in each bin. Make a separate graph for each outcome. Describe the relationship

```r
bin <- 28.35

#Cut function source
#https://statisticsglobe.com/group-data-frame-rows-range-r
#source: not scientifica notation: https://stackoverflow.com/questions/15497694/cut-function-in-r-label
data_bin <- data %>%
  mutate(bin_one_ounce = cut(bweight, seq(1500-6*bin, 1500+6*bin, bin), right=FALSE, dig.lab = 10))



data_bin_mean_mortality <- data_bin %>%
  group_by(bin_one_ounce) %>%
  summarise(obs_count=n(),
            mortality_one_year = mean(agedth5),
         mortality_28 = mean(agedth4))


#extract number from interval
#https://stackoverflow.com/questions/50669853/extract-value-from-an-interval-from-cut
#extract numbers from the interval to have integer x-scale
data_bin_mean_mortality <- data_bin_mean_mortality %>%
   mutate(bin_min= lapply(str_extract_all(bin_one_ounce, "-?[0-9.]+"), function(x) min(as.numeric(x))),
          bin_max =lapply(str_extract_all(bin_one_ounce, "-?[0-9.]+"), function(x) max(as.numeric(x))))

#apply function to multiple columns
#https://stackoverflow.com/questions/22772279/converting-multiple-
#columns-from-character-to-numeric-format-in-r
cols_name <- c("bin_min", "bin_max")

data_bin_mean_mortality[cols_name] <- lapply(data_bin_mean_mortality[cols_name], as.numeric)

data_bin_mean_mortality <- data_bin_mean_mortality %>%
  mutate(bin_center = (bin_min + bin_max)/2)


mortality_by_bin_one_year <- data_bin_mean_mortality %>%
  ggplot(aes(x=bin_min, y=mortality_one_year)) +
  geom_point() +
  scale_x_continuous(breaks=seq(1350, 1650, 50)) +
  labs(x="Birth Weight(g)", y="Mortality", title="One-year mortality")


mortality_by_bin_28_day <- data_bin_mean_mortality %>%
```

```
  ggplot(aes(x=bin_min, y=mortality_28)) +
  geom_point() +
  scale_x_continuous(breaks=seq(1350, 1650, 50)) +
  labs(x="Birth Weight(g)", y="Mortality", title="28-day mortality")

mortality_by_bin_one_year
```




```
mortality_by_bin_28_day
```




```
data_bin_mean_mortality %>%
  select(bin_one_ounce, obs_count)
```

```
## # A tibble: 12 x 2
##    bin_one_ounce    obs_count
##    <fct>                <int>
##  1 [1329.9,1358.25)      3271
##  2 [1358.25,1386.6)     30236
##  3 [1386.6,1414.95)     29002
##  4 [1414.95,1443.3)     31386
##  5 [1443.3,1471.65)     31653
##  6 [1471.65,1500)       32185
##  7 [1500,1528.35)       35830
##  8 [1528.35,1556.7)     35344
##  9 [1556.7,1585.05)     35680
## 10 [1585.05,1613.4)     39966
## 11 [1613.4,1641.75)     39649
## 12 [1641.75,1670.1)     32206
```

For both one year and 28-day mortality rates, mortality rate generally decreases as birth weight increases. However, in the bin that includes 1500g birth weight [1500,1528.35), mortality rate shows an increase thus suggesting a discontinuity at the VLBW threshold. A deviation from the general decrease is also observed in the bin right below 1600 grams. The shape of the plot for the one-year and 28-day mortality rate looks similar in general. The more observations are included in each bin, more bias and less variance is expected for mean estimates.

To note, it seems that the number of observations in each bin are generally similar except for the bin with the smallest birth weight value. The group shows discontinuous lower mortality rate, but it might needs to be considered separately from the general sample population with the very low birth weight.

Q3.

A key assumption for an RDD to provide a causal estimate is that individuals are not able to sort according to the running variable, i.e., they should not be able to manipulate its value. Discuss in your own words whether this is a reasonable assumption in this case.

- Yes, it is a reasonable assumption. Parents and babies are unlikely and hardly able to know and try to manipulate the weight and manipulate it to exceed or fall short of the VLBW threshold.

Q4.

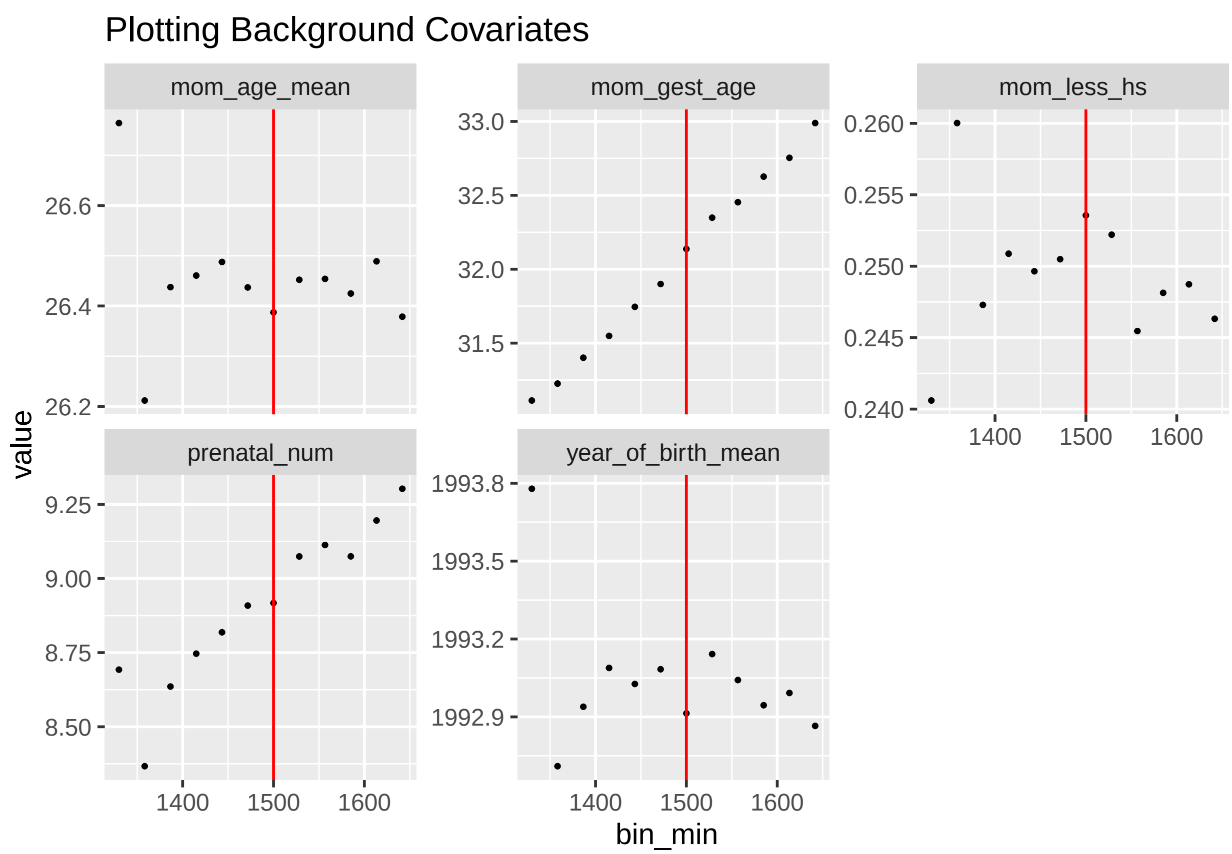
Now plot background covariates including mom_age, mom_ed1, gest, nprenatal, yob against birth weight

```r
#similarly repeat the previous steps, but in more systematic way
data_bin_one_ounce <- data_bin %>%
   mutate(bin_min = lapply(str_extract_all(bin_one_ounce, "-?[0-9.]+"), function(x) min(as.numeric(x))),
          bin_max =lapply(str_extract_all(bin_one_ounce, "-?[0-9.]+"), function(x) max(as.numeric(x))))


cols_name <- c("bin_min", "bin_max")

data_bin_one_ounce[cols_name] <- lapply(data_bin_one_ounce[cols_name], as.numeric)


#plot x-axis:
data_bin_covariates <- data_bin_one_ounce %>%
    group_by(bin_min) %>%
    summarise(
            obs_count=n(),
            mortality_one_year = mean(agedth5),
            mortality_28 = mean(agedth4),
            mom_age_mean = mean(mom_age),
            mom_less_hs = mean(mom_ed1),
            mom_gest_age = mean(gest, na.rm=T),
            prenatal_num = mean(nprenatal, na.rm=T),
            year_of_birth_mean = mean(yob))


data_bin_covariates
```

```
## # A tibble: 12 x 9
##    bin_min obs_count mortality~1 morta~2 mom_a~3 mom_l~4 mom_g~5 prena~6 year_~7
##      <dbl>     <int>       <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1   1330.      3271      0.0654  0.0480    26.8   0.241    31.1    8.69   1994.
##  2   1358.     30236      0.0773  0.0562    26.2   0.260    31.2    8.37   1993.
##  3   1387.     29002      0.0694  0.0494    26.4   0.247    31.4    8.64   1993.
##  4   1415.     31386      0.0632  0.0446    26.5   0.251    31.5    8.75   1993.
##  5   1443.     31653      0.0612  0.0427    26.5   0.250    31.7    8.82   1993.
##  6   1472.     32185      0.0569  0.0391    26.4   0.250    31.9    8.91   1993.
##  7   1500      35830      0.0615  0.0439    26.4   0.254    32.1    8.92   1993.
##  8   1528.     35344      0.0540  0.0373    26.5   0.252    32.3    9.07   1993.
##  9   1557.     35680      0.0505  0.0337    26.5   0.245    32.5    9.11   1993.
## 10   1585.     39966      0.0526  0.0366    26.4   0.248    32.6    9.07   1993.
```

```
## 11   1613.     39649      0.0459  0.0308    26.5   0.249    32.8    9.20   1993.
## 12   1642.     32206      0.0447  0.0289    26.4   0.246    33.0    9.30   1993.
## # ... with abbreviated variable names 1: mortality_one_year, 2: mortality_28,
## #   3: mom_age_mean, 4: mom_less_hs, 5: mom_gest_age, 6: prenatal_num,
## #   7: year_of_birth_mean
```

```
##GGPLOT for each COLUMN
#Source: drawing multiple plots for each column
#https://stackoverflow.com/questions/58818300/ggplot-for-each-column-in-a-data
data_bin_covariates %>%
  select(-c(obs_count, mortality_one_year, mortality_28)) %>%
  tidyr::gather(variable, value, -bin_min) %>%
  ggplot(aes(x = bin_min, y = value)) +
  geom_point(size = 0.5) +
  geom_vline(xintercept = 1500, color="red")+
  facet_wrap(~ variable, scales = "free_y", nrow=2) +
  labs(title = "Plotting Background Covariates")
```



- Yes, plots seem to be smooth around the threshold of 1500. Although it seems that they are not on perfectly linear lines, the plots show smoothness considering the scale on the y-axis. Thus at least regarding the observable variables plotted against birth weight, babies have similar background characteristics.

(Note) Here I set the x-axis of each bin to the smallest number in each interval.As in (2), the data in the interval with the smallest birth weight deviates from general trend. It is understandable, as those babies in the lowest recorded birth weight might present especially different health conditions and relevant factors.

Q5.

Now formalize the evidence on smoothness of the covariates by estimating regressions

```
#make columns that match the regression terms
data_for_reg <- data %>%
  mutate(vlbw_dummy=ifelse(bweight<1500, 1, 0),
         bweight_centralized = bweight - 1500,
         under_vlbw_bweight_centralized = vlbw_dummy * bweight_centralized,
         over_vlbw_bweight_centralized = (1-vlbw_dummy) * bweight_centralized
         )

#limit sample to above/below the threshold by caliper
caliper_sample <- function(caliper, df){
 df %>%
  filter(bweight>=1500-caliper & bweight <=1500+caliper)
  }

data_caliper_85 <- caliper_sample(85, data_for_reg)



# Regression variables
covariates<-c("mom_age", "mom_ed1", "gest", "nprenatal", "yob")

#saved models to reg
regs <- sapply(covariates, function(x)
  lm(as.formula(paste0(x, " ~ vlbw_dummy + over_vlbw_bweight_centralized + under_vlbw_bweight_centraliz

output <- lapply(regs, function(x) summary(x)$coefficient)


covariates
```

[1] "mom_age" "mom_ed1" "gest" "nprenatal" "yob"

```
stargazer(regs, type = "text", column.labels = c("mom's age", "mom's edu less than HS", "gestational ag
```

==
Dependent variable:

——-

paste0(x, " ~ vlbw_dummy + over_vlbw_bweight_centralized + under_vlbw_bweight_centralized")
mom's age mom's edu less than HS gestational age prenatal care vist # year of birth
(1) (2) (3) (4) (5)

——

——————— vlbw_dummy 0.241*** (0.060) -0.003 (0.004) -0.128*** (0.031) 0.089 (0.055) 0.590*** (0.057)
over_vlbw_bweight_centralized 0.003*** (0.001) -0.0001** (0.0001) 0.005*** (0.0004) 0.004*** (0.001) 0.008*** (0.001)
under_vlbw_bweight_centralized 0.002** (0.001) 0.00003 (0.0001) 0.006*** (0.0005) 0.003*** (0.001) 0.005*** (0.001)
Constant 26.310*** (0.034) 0.255*** (0.002) 32.152*** (0.018) 8.904*** (0.031) 1,992.738*** (0.033)

——

——————— Observations 202,078 202,078 185,293 182,582 202,078

R2 0.0001 0.00003 0.010 0.001 0.001
Adjusted R2 0.0001 0.00001 0.010 0.001 0.001
Residual Std. Error 6.505 (df = 202074) 0.433 (df = 202074) 3.196 (df = 185289) 5.622 (df = 182578) 6.171 (df = 202074)
F Statistic 7.910*** (df = 3; 202074) 1.848 (df = 3; 202074) 601.631*** (df = 3; 185289) 35.500*** (df = 3; 182578) 53.891*** (df = 3; 202074) ===
Note: *p<0.1;* ***p<0.05;*** p<0.01

```
#Another form: but name not showing
#stargazer(output, object.names=TRUE, model.names = TRUE, type="text")
```

- The coefficient for VLBW (alpha 1) needs to be zero to show continuity around the threshold. The printed estimates are for the covariates ("mom_age", "mom_ed1", "gest", "nprenatal", "yob") respectively. When taking a look at p-value, mom_age (p_value 0.0001), gestational age(p_value 0.00003), and year of birth (p_value 0) are significantly different from zero. These are evidence of discontinuities around the threshold. If not controlled, these covariates will affect RDD estimates.

Looking at each estimate, mother's age (estimate: 0.241) and year of birth (0.59) seems to have positive correlation and it means that RDD estimate could be overestimated if these two variables are not controlled. Gestational age (-0.128) has negative correlation and it means that an RDD estimate could be underestimated if not controlled of this variable.

Q6.

Get an estimate of the size of discontinuity in one year & 28-day mortality, using a caliper of 85 grams. Interpret the coefficients.

```
#28-day
OLS_no_control_28_day <- lm(agedth4 ~ vlbw_dummy + under_vlbw_bweight_centralized + over_vlbw_bweight_c

#summary(OLS_no_control_28_day)

#one-year
OLS_no_control_one_year<- lm(agedth5 ~ vlbw_dummy + under_vlbw_bweight_centralized + over_vlbw_bweight_

#summary(OLS_no_control_28_day)

stargazer(OLS_no_control_28_day, OLS_no_control_one_year, type="text", column.labels = c("28-day mortali
```

```
##
## ======================================================================
##                                          Dependent variable:
##                                    -----------------------------------
##                                        agedth4          agedth5
##                                    28-day mortality one-year mortality
##                                          (1)              (2)
## ----------------------------------------------------------------------
## vlbw_dummy                            -0.009***        -0.010***
##                                        (0.002)          (0.002)
##
## under_vlbw_bweight_centralized        -0.0001***       -0.0001***
##                                       (0.00003)        (0.00003)
```

```
##
## over_vlbw_bweight_centralized        -0.0002***        -0.0002***
##                                      (0.00002)         (0.00003)
##
## Constant                              0.045***          0.063***
##                                      (0.001)           (0.001)
##
## ----------------------------------------------------------------------
## Observations                          202,078           202,078
## R2                                     0.001             0.001
## Adjusted R2                            0.0005            0.0005
## Residual Std. Error (df = 202074)      0.196             0.233
## F Statistic (df = 3; 202074)         34.513***         33.817***
## ======================================================================
## Note:                                      *p<0.1; **p<0.05; ***p<0.01
```

- For 28-day mortality rate,

a1 is approximately -0.009, a2 is approximately -0.0001, and a3 is approximately -0.0002. That all coefficients being negative means negative relationship between the birth weight and 28-day mortality within a caliper of 85 grams.

a1 indicates the effect of treatment at the threshold 1500grams. That is, the medical treatment (expense) that starts at the birth weight threshold of 1500 grams (treatment effect) is relevant with approximately 0.9 percentage point of decrease in 28-day mortality rate. a2 indicates how birth weight is connected with 28-day mortality rate when a newborn weighs less than 1500 grams (one gram of birth weight increase is associated with 0.01 percentage point decrease in 28-day mortality rate). a3 indicates the relationship between the birth weight and the 28-day mortality weight when a newborn weights more than 1500 grams (one gram birth weight increases is associated with 0.02 percentage point decrease in the rate) The negative connection is more strong (slope is steeper) above the threshold.

- Similarly for one-year mortality rate, there is approximately 1 percentage point decrease in mortality rate at the birth weight threshold of 1500 grams. Under the threshold, one gram increase in birth weight is associated with 0.01 percentage point mortality rate decrease and similarly 0.02 percentage point over the threshold. The negative relationship is more strong (slope steeper) over the threshold likewise.

Q7.

Add covariates to the model in 6. Compare with the estimates in the previuos question and explain the difference.

```
#exclude rows those missing values in control variables
#mom_ed4, gest_skw4, nprenatal_4
data_caliper_85_no_missing <- data_caliper_85 %>%
  filter(mom_ed5 != 1 & gest_wks4 != 1 & nprenatal_4 !=1)


#Didn't include last categories of dummies as we need include (n-1) dummies and one cateogy will be reg
#mom_ed1, gest_wks1, nprenatal_1
#and regard year of birth (yob)  and mom_race as factors
#source: https://bookdown.org/carillitony/bailey/chp6.html
OLS_control_85_28_day <- lm(agedth4 ~ vlbw_dummy + under_vlbw_bweight_centralized + over_vlbw_bweight_c
```

```r
                    mom_age + mom_ed2 + mom_ed3 + mom_ed4 + as.factor(mom_race) + as.factor(yob) + gest_
                    nprenatal_3 , data= data_caliper_85_no_missing)

OLS_control_85_one_year <- lm(agedth5 ~ vlbw_dummy + under_vlbw_bweight_centralized + over_vlbw_bweight_
                    mom_age + mom_ed2 + mom_ed3 + mom_ed4 + as.factor(mom_race) + as.factor(yob) + gest_
                    nprenatal_3 , data= data_caliper_85_no_missing)

stargazer(OLS_control_85_28_day, OLS_control_85_one_year, type="text", column.labels = c("28-day mortali
```

```
##
## ======================================================================
##                                          Dependent variable:
##                                     -----------------------------------
##                                         agedth4           agedth5
##                                     28-day mortality one-year mortality
##                                           (1)               (2)
## ----------------------------------------------------------------------
## vlbw_dummy                              -0.005**          -0.005**
##                                         (0.002)           (0.002)
##
## under_vlbw_bweight_centralized         -0.0001***        -0.0001***
##                                        (0.00003)         (0.00003)
##
## over_vlbw_bweight_centralized          -0.0001***        -0.0001***
##                                        (0.00003)         (0.00003)
##
## mom_age                                 -0.0001           -0.0002*
##                                         (0.0001)          (0.0001)
##
## mom_ed2                                 0.003***          -0.003*
##                                         (0.001)           (0.001)
##
## mom_ed3                                  -0.001          -0.009***
##                                         (0.001)           (0.002)
##
## mom_ed4                                  0.001           -0.009***
##                                         (0.002)           (0.002)
##
## as.factor(mom_race)2                   -0.019***         -0.015***
##                                         (0.001)           (0.001)
##
## as.factor(mom_race)3                     -0.003            0.001
##                                         (0.003)           (0.003)
##
## as.factor(yob)1984                       -0.002            -0.004
##                                         (0.003)           (0.004)
##
## as.factor(yob)1985                       0.003             0.005
##                                         (0.003)           (0.004)
##
## as.factor(yob)1986                       -0.003            -0.006
##                                         (0.003)           (0.004)
##
```

```
## as.factor(yob)1987                          -0.006*            -0.008**
##                                             (0.003)            (0.004)
##
## as.factor(yob)1988                          -0.006*             -0.005
##                                             (0.003)            (0.004)
##
## as.factor(yob)1989                         -0.009***          -0.013***
##                                             (0.003)            (0.004)
##
## as.factor(yob)1990                         -0.017***          -0.022***
##                                             (0.003)            (0.003)
##
## as.factor(yob)1991                         -0.019***          -0.023***
##                                             (0.003)            (0.003)
##
## as.factor(yob)1995                         -0.028***          -0.035***
##                                             (0.003)            (0.004)
##
## as.factor(yob)1996                         -0.029***          -0.036***
##                                             (0.003)            (0.004)
##
## as.factor(yob)1997                         -0.031***          -0.039***
##                                             (0.003)            (0.004)
##
## as.factor(yob)1998                         -0.029***          -0.038***
##                                             (0.003)            (0.004)
##
## as.factor(yob)1999                         -0.027***          -0.038***
##                                             (0.003)            (0.003)
##
## as.factor(yob)2000                         -0.030***          -0.042***
##                                             (0.003)            (0.003)
##
## as.factor(yob)2001                         -0.031***          -0.040***
##                                             (0.003)            (0.003)
##
## as.factor(yob)2002                         -0.035***          -0.046***
##                                             (0.003)            (0.003)
##
## gest_wks2                                   0.023***           0.030***
##                                             (0.002)            (0.002)
##
## gest_wks3                                    0.006              0.010
##                                             (0.006)            (0.008)
##
## nprenatal_2                                -0.007***          -0.009***
##                                             (0.001)            (0.001)
##
## nprenatal_3                                -0.011***          -0.012***
##                                             (0.002)            (0.002)
##
## Constant                                    0.067***           0.099***
##                                             (0.003)            (0.004)
##
```

```
## ----------------------------------------------------------------------
## Observations                             159,513          159,513
## R2                                        0.009            0.010
## Adjusted R2                               0.008            0.010
## Residual Std. Error (df = 159483)         0.189            0.225
## F Statistic (df = 29; 159483)           48.083***        54.055***
## ======================================================================
## Note:                                      *p<0.1; **p<0.05; ***p<0.01
```

- Estimates with controls

The estimates are all negative as expected (first three rows). The estimates didn't change significantly in general but the size of significance did changed at the threshold. The estimates with controls are identical for 28-day and one-year data.

Comparing 28-day mortality rate,

- a1: -0.005 (before control: -0.009) -> the size of treatment effect becomes smaller after control
- a2: -0.0001 (before control: -0.0001) -> the size of negative relationship below the threshold are identical
- a3: -0.0001 (before control: -0.0002) -> the size of negative relationship becomes smaller, but not much.

The difference in the estimates could be attributed to the significance of covariates. In other words, covariate variables have explanatory power thus controlling them affects estimates. Specifically, Mom's education, race, gestational age, prenatal care visit numbers are related with outcome variables by different size (although not all are significant). The most noticeable difference seems to be relevant to the year of birth. Especially when a baby is born in 2000s, it seems to have meaningfully lower (around 4 percent point) mortality rate compared to those born in 1983 (default group as constant).

- (Note) Reasoning for the ways I did filtering and adding controls: I filtered only to the observations without any missing values to mom's education, gestational age, and prenatal care visit numbers. This is because those observations with missing values (i.e., mom_ed1 = 1, gest_wks4 = 1, nprenatal_4 = 1) could not be considered as having common characteristics. It would be misleading to categorize different respondents in to one category (e.g. mom_ed5) and control for it (e.g. less than high school but missing education response vs more than college but missing education response)

Q8.

Bandwidth sensitivity: Use model in 7 to assess the sensitivity of the esimates to the use of different calipers. Use calipers of 65 and 105 grams. Are the estimates any different to those in 7? what is the tradeoff when we increase/decrease caliper?

```r
data_caliper_65 <- caliper_sample(65, data_for_reg)

data_caliper_65_no_missing <- data_caliper_65 %>%
  filter(mom_ed5 != 1 & gest_wks4 != 1 & nprenatal_4 !=1)


data_caliper_105 <- caliper_sample(105, data_for_reg)
```

```
data_caliper_105_no_missing <- data_caliper_105 %>%
  filter(mom_ed5 != 1 & gest_wks4 != 1 & nprenatal_4 !=1)


OLS_control_65_28_day <- lm(agedth4 ~ vlbw_dummy + under_vlbw_bweight_centralized + over_vlbw_bweight_ce
                    mom_age + as.factor(mom_race)  + mom_ed2 + mom_ed3 + mom_ed4 + as.factor(yob) + gest
                    nprenatal_3 , data= data_caliper_65_no_missing)


OLS_control_65_one_year <- lm(agedth5 ~ vlbw_dummy + under_vlbw_bweight_centralized + over_vlbw_bweight_
                    mom_age + as.factor(mom_race)  + mom_ed2 + mom_ed3 + mom_ed4 + as.factor(yob) + gest
                    nprenatal_3 , data= data_caliper_65_no_missing)



OLS_control_105_28_day <- lm(agedth4 ~ vlbw_dummy + under_vlbw_bweight_centralized + over_vlbw_bweight_c
                    mom_age + as.factor(mom_race)  + mom_ed2 + mom_ed3 + mom_ed4 + as.factor(yob) + gest
                    nprenatal_3 , data= data_caliper_105_no_missing)


OLS_control_105_one_year <- lm(agedth5 ~ vlbw_dummy + under_vlbw_bweight_centralized + over_vlbw_bweigh
                    mom_age + as.factor(mom_race)  + mom_ed2 + mom_ed3 + mom_ed4 + as.factor(yob) + gest
                    nprenatal_3 , data= data_caliper_105_no_missing)

#FOr 28-day, different calipers
stargazer(OLS_control_65_28_day, OLS_control_105_28_day, OLS_control_85_28_day,
          column.labels=c("caliper=65, 28-day", "caliper=105, 28-day", "caliper=85, 28-day"), type="tex
```

```
##
## ================================================================================================
##                                                                Dependent variable:
##                                  ---------------------------------------------------------------
##                                                                      agedth4
##                                     caliper=65, 28-day          caliper=105, 28-day         caliper=8
##                                             (1)                         (2)                        (
## ------------------------------------------------------------------------------------------------
## vlbw_dummy                               -0.005**                    -0.004**                   -0.0
##                                           (0.002)                     (0.002)                    (0.
##
## under_vlbw_bweight_centralized           -0.0001*                   -0.0001***                 -0.000
##                                          (0.0001)                   (0.00003)                  (0.00
##
## over_vlbw_bweight_centralized           -0.0001***                  -0.0001***                 -0.000
##                                          (0.00003)                  (0.00002)                  (0.00
##
## mom_age                                  -0.00000                    -0.00005                   -0.
##                                          (0.0001)                    (0.0001)                  (0.00
##
## as.factor(mom_race)2                     -0.019***                   -0.019***                  -0.01
##                                           (0.001)                     (0.001)                    (0.
##
## as.factor(mom_race)3                      -0.003                      -0.002                     -0.
##                                           (0.003)                     (0.002)                    (0.
```

```
##
## mom_ed2                          0.002*          0.003***        0.00
##                                  (0.001)         (0.001)         (0.
##
## mom_ed3                          -0.001          0.001           -0.
##                                  (0.002)         (0.001)         (0.
##
## mom_ed4                          0.001           0.001           0.
##                                  (0.002)         (0.002)         (0.
##
## as.factor(yob)1984               -0.002          -0.001          -0.
##                                  (0.004)         (0.003)         (0.
##
## as.factor(yob)1985               0.005           0.001           0.
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1986               -0.006*         -0.006**        -0.
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1987               -0.006*         -0.009***       -0.
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1988               -0.007**        -0.009***       -0.
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1989               -0.011***       -0.011***       -0.00
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1990               -0.017***       -0.019***       -0.0
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1991               -0.019***       -0.022***       -0.0
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1995               -0.028***       -0.030***       -0.0
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1996               -0.028***       -0.031***       -0.0
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1997               -0.030***       -0.032***       -0.0
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1998               -0.029***       -0.031***       -0.0
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)1999               -0.028***       -0.030***       -0.0
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)2000               -0.031***       -0.033***       -0.0
##                                  (0.003)         (0.003)         (0.
##
## as.factor(yob)2001               -0.031***       -0.033***       -0.0
##                                  (0.003)         (0.003)         (0.
```

```
##
## as.factor(yob)2002                      -0.035***            -0.036***            -0.03
##                                          (0.003)              (0.003)              (0.0
##
## gest_wks2                                0.023***             0.021***             0.023
##                                          (0.002)              (0.002)              (0.0
##
## gest_wks3                                 0.008                0.006                0.0
##                                          (0.007)              (0.006)              (0.0
##
## nprenatal_2                             -0.005***            -0.007***            -0.00
##                                          (0.001)              (0.001)              (0.0
##
## nprenatal_3                             -0.010***            -0.010***            -0.01
##                                          (0.002)              (0.001)              (0.0
##
## Constant                                 0.066***             0.067***             0.067
##                                          (0.003)              (0.003)              (0.0
##
## ----------------------------------------------------------------------------------------
## Observations                             131,291              194,646              159
## R2                                        0.008                0.009                0.0
## Adjusted R2                               0.008                0.008                0.0
## Residual Std. Error                 0.187 (df = 131261)  0.189 (df = 194616)  0.189 (df
## F Statistic                  38.747*** (df = 29; 131261) 58.179*** (df = 29; 194616) 48.083*** (df
## ========================================================================================
## Note:                                                                      *p<0.1; **p<0
```

```r
#FOr one-year, different calipers
stargazer(OLS_control_65_one_year, OLS_control_105_one_year, OLS_control_85_one_year,
          column.labels=c("caliper=65, one-year", "caliper=105, one-year", "caliper=85, one-year"), typ
```

```
##
## ========================================================================================
##                                                          Dependent variable:
##                                   ------------------------------------------------------
##                                                               agedth5
##                                    caliper=65, one-year  caliper=105, one-year   caliper=85
##                                            (1)                   (2)                  (3
## ----------------------------------------------------------------------------------------
## vlbw_dummy                               -0.005*               -0.003               -0.00
##                                          (0.003)              (0.002)              (0.0
##
## under_vlbw_bweight_centralized          -0.0001**            -0.0001***            -0.000
##                                         (0.0001)             (0.00003)             (0.00
##
## over_vlbw_bweight_centralized          -0.0002***            -0.0001***            -0.000
##                                         (0.00004)            (0.00002)             (0.00
##
## mom_age                                  -0.0002             -0.0002**             -0.00
##                                         (0.0001)              (0.0001)             (0.00
##
## as.factor(mom_race)2                    -0.016***            -0.016***             -0.01
##                                          (0.001)              (0.001)              (0.0
```

```
##
## as.factor(mom_race)3                    0.002                  0.001                  0.0
##                                        (0.003)                (0.003)                (0.0
##
## mom_ed2                               -0.004**               -0.003**               -0.0
##                                        (0.002)                (0.001)                (0.0
##
## mom_ed3                               -0.010***              -0.008***              -0.00
##                                        (0.002)                (0.002)                (0.0
##
## mom_ed4                               -0.010***              -0.009***              -0.00
##                                        (0.002)                (0.002)                (0.0
##
## as.factor(yob)1984                     -0.004                 -0.002                 -0.0
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1985                      0.004                  0.002                  0.0
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1986                    -0.009**               -0.009***               -0.0
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1987                    -0.010**               -0.011***              -0.00
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1988                     -0.006                -0.009***               -0.0
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1989                    -0.016***              -0.015***              -0.01
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1990                    -0.023***              -0.024***              -0.02
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1991                    -0.024***              -0.026***              -0.02
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1995                    -0.036***              -0.037***              -0.03
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1996                    -0.037***              -0.039***              -0.03
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1997                    -0.040***              -0.040***              -0.03
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1998                    -0.040***              -0.040***              -0.03
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)1999                    -0.040***              -0.041***              -0.03
##                                        (0.004)                (0.003)                (0.0
##
## as.factor(yob)2000                    -0.044***              -0.044***              -0.04
##                                        (0.004)                (0.003)                (0.0
```

```
##
## as.factor(yob)2001                          -0.041***                 -0.042***                 -0.0
##                                              (0.004)                   (0.003)                   (0.0
##
## as.factor(yob)2002                          -0.047***                 -0.047***                 -0.0
##                                              (0.004)                   (0.003)                   (0.0
##
## gest_wks2                                   0.030***                  0.029***                  0.03
##                                              (0.002)                   (0.002)                   (0.0
##
## gest_wks3                                     0.009                    0.015**                   0.0
##                                              (0.008)                   (0.007)                   (0.0
##
## nprenatal_2                                 -0.008***                 -0.009***                 -0.0
##                                              (0.001)                   (0.001)                   (0.0
##
## nprenatal_3                                 -0.011***                 -0.012***                 -0.0
##                                              (0.002)                   (0.002)                   (0.0
##
## Constant                                    0.099***                  0.099***                  0.09
##                                              (0.004)                   (0.003)                   (0.0
##
## -------------------------------------------------------------------------------------------------------
## Observations                                 131,291                   194,646                   159
## R2                                            0.010                     0.010                    0.0
## Adjusted R2                                   0.009                     0.009                    0.0
## Residual Std. Error                    0.224 (df = 131261)       0.225 (df = 194616)       0.225 (df
## F Statistic                        44.399*** (df = 29; 131261) 65.153*** (df = 29; 194616) 54.055*** (df
## =======================================================================================================
## Note:                                                                              *p<0.1; **p<0
```

Compared with the caliper of 85, the estimates are similar. The difference that originates from the size of caliper is the difference in variance. The model with the biggest caliper has the smallest variance (noticeable from the second and third row of the above table). So when increasing the caliper, we can decrease the variance of each estimate. However, from the first row of the above table, we could also see that the size of the treatment effect may not be so accurate for the model with bigger caliper. While the models with caliper of 65 and 85 both present the effect size of -0.005, the model with the caliper of 105 solely has the result of -0.004. The differences support the trade off between bias and variance when choosing a caliper size.

The implication also applies to the one-year mortality estimates. While the variance of the estimates (row 1-3) is smaller for the model with a larger caliper, the model's treatment effect estimate is solely insignificant and different from other models.

Q9. Synthesize your findings and what kind of supplementary info need to make a cost-benefit analysis of treatment received by newborns close to the very low birth weight threshold

Through the problem set, the main topic of focus was figuring out the relationship between mortality rate (representative of health outcome) and VLBW diagnosis which often times lead to special treatment with medical expenses. Around the threshold, those newborns classified as VLBW have lower mortality rate and it could be connected with the additional medical treatment that receive when diagnosed as VLBW. In order to do the cost-benefit analysis, I need to know the additional medical expenses incurred for the newborns under 1500 due to the special treatment. The economic value of the newborn's better health outcomes (i.e. 'benefits' from the treatment) also needs to be clarified. Additionally, if there are social costs and benefits from the special treatment to VLBW infants, they should also be considered.

EOD