EDS 241 Assignment 2

Mia Forsline

2/4/2022

1 Set Up

2 Read in data

Variables:

- `fips`: FIPS code identifying each county
- `NPB`: 1 indicates if the county was regulated under the NO_x Budget Program
- `PctBlack`: fraction of the county population that is African American
- `Dnox_masstons`: change in NO_x emissions from all power plants in a county between 2000 - 2008 in tons
  - if `Dnox_masstons` < 0, then the NO_x emissions from power plants declined in that county

```
data <- readxl::read_excel(here("data", "NBP.xls"))

data_clean <- data %>%
  janitor::clean_names()
```

3 (a) Make a histogram depicting the distribution of Dnox_masstons.

```
nox_lab <- expression(paste("Change in NOx"["3"]," emissions from 2000 - 2008 (tons)"))

hist <- ggplot(data = data_clean, aes(x = dnox_masstons)) +
  geom_histogram() +
  theme_classic() +
  labs(x = nox_lab,
       y = "Frequency")
```

Figure 1: Histogram of change in NO_x emissions from all power plants in a county between 2000 and 2008 (in tons)

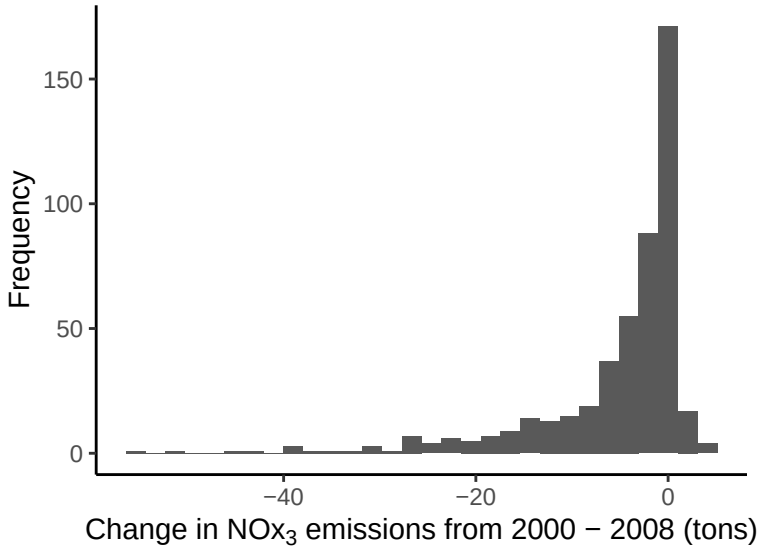

Figure 1 shows the skewed distribution of change of NO_x emissions from 2000 - 2008. There is a long tail to the left, and most of the data centers around zero. Each data point represents one county in California (n = 485).

4 (b) Create an indicator =1 if the county has PctBlack above the sample median, and =0 otherwise (in the rest of the assignment, I refer to this variable as 'D'). What is the average of PctBlack for counties above the median, i.e. counties for which D=1?

```
median <- round(median(data_clean$pct_black), digits = 2)

data_d <- data_clean %>%
  mutate(D = case_when(
    pct_black > median ~ 1,
    pct_black <= median ~ 0),
    D = as.factor(D),
    nbp = as.factor(nbp)
    )

data_above_median <- subset(data_d, D == 1)
avg_pct_black <- round(mean(data_above_median$pct_black), digits = 2)
```

For counties with a percentage of Black residents above the median (n = 240), the average percentage of Black individuals is 19.13%.

5 (c) Estimate a regression of Dnox_masstons on NBP. Interpret the estimated intercept and the coefficient on NBP.

To more formally analyze the relationship between the change in NO_x emissions from all power plants in a county between 2000 - 2008 in tons and if a county was regulated under the NO_x Budget Program or not, we estimate the following regression:

$$Y_i = \beta_0 + \beta_1 X_{1i} + u_i \tag{1}$$

where Y_i is the tons of NO_x emissions from county i, X_{1i} is the dummy variable indicating if a county was regulated under the NO_x Budget Program or not, and u_i the regression error term.

```
mod <- lm_robust(dnox_masstons ~ nbp, data = data_d)
#summary(mod)
abs_value_intercept <- round(mod$coefficients[[1]] * -1, digits = 2)
npb <- round(mod$coefficients[[2]] * -1, digits = 2)
```

The intercept value tells us that, on average, counties that did not participate in the NO_x Budget Program reported a 3.62 tons decrease in the change of NO_x emissions from all power plants in the county from 2000 to 2008.

The `npb` coefficient tells us that, on average, counties that participated in the NO_x Budget Program reported a 3.92 tons decrease in the change of NO_x emissions from all power plants in the county from 2000 to 2008 compared to counties that did not participate in the NO_x Budget Program.

6 (d) Create an interaction between the variables NBP and D. Estimate a regression of Dnox_masstons on NBP, D, and this interaction. Interpret each estimated regression coefficient, including the intercept.

```
mod2 <- lm_robust(dnox_masstons ~ nbp + D + nbp:D, data = data_d)
#summary(mod2)
abs_value_intercept <- round(mod2$coefficients[[1]] * -1, digits = 2)
npb_coefficient <- round(mod2$coefficients[[2]] * -1, digits = 2)
D_coefficient <- round(mod2$coefficients[[3]] * -1, digits = 2)
interaction <- round(mod2$coefficients[[4]], digits = 2)
```

Table 1 shows the estimated coefficients from estimating Model 1 (`dnox_masstons ~ nbp + D`) and Model 2 (`dnox_masstons ~ nbp + D + nbp:D`).

Table 1: Model Coefficients

	NOx Change (tons)	
	(1)	(2)
Constant	-3.920^{***}	-7.242^{***}
	(0.763)	(1.073)
NBP		-2.540^{**}
		(0.993)
D		6.484^{***}
		(1.504)
NPB:D	-3.622^{***}	-2.422^{***}
	(0.504)	(0.683)
Observations	485	485
R^2	0.052	0.087
Note:	*p<0.1; **p<0.05; ***p<0.01	

The intercept value tells us that, on average, counties that *did not participate* in the NO_x Budget Program and have a fraction of the county population that is African American *below* the median value reported a 2.42 tons decrease in the change of NO_x emissions from all power plants in the county from 2000 to 2008.

The `nbp` coefficient tells us that, for all counties that have a fraction of the county population that is African American *below* the median value, on average, counties that *participated* in the NO_x Budget Program reported a 7.24 tons decrease in NO_x emissions compared to counties that *did not participate* in the NO_x Budget Program.

The `D` coefficient tells us that, for all counties that *did not participate* in the NO_x Budget Program, on average, counties that have a fraction of the county population that is African American *above* the median value reported 2.54 tons decrease in the change NO_x emissions compared to counties that have a fraction of the county population that is African American *below* the median value.

The interaction tells us that, on average that:

- a) counties that *participated* in the NO_x Budget Program, regardless of their fraction of the county population that is African American
- b) counties that have a fraction of the county population that is African American *above* the median reported value of 4.8 %, regardless of whether they participated in the NO_x Budget Program
- c) and counties that *participated* in the NO_x Budget Program and have a fraction of the county population that is African American *above* the median value reported 6.48 more tons of NO_x emissions from all power plants from 2000 to 2008.

7 (e) What is the predicted Dnox_masstons in a county that was not regulated under NBP and where PctBlack is above the sample median (i.e., where D=1)? Report the 95% confidence interval for this prediction. Make sure to use "heteroskedasticity-robust" standard errors.

```r
mod2 <- lm_robust(dnox_masstons ~ nbp + D + nbp:D, data = data_d)

pred_constraints=data.frame(nbp=c("0"), D=c("1"))

pred <- predict(mod2, newdata=pred_constraints, se.fit=TRUE, interval='confidence')
pred_value <- round(pred$fit[[1]] * -1, digits = 2)
lower <- round(pred$fit[[2]], digits = 2)
upper <- round(pred$fit[[3]], digits = 2)
```

A county that did not participate in the NO_x Budget Program and has a fraction of the county population that is African American above the median is predicted to have emitted approximately 4.96 tons less from 2000 to 2008.

The 95% confidence interval is bounded by a lower bount of -6.38 and an upper bound of -3.55.