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 annual 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 annual NOx"["3"]," emissions 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 annual NO_x emissions from all power plants in a county between 2000 and 2008 (in tons)

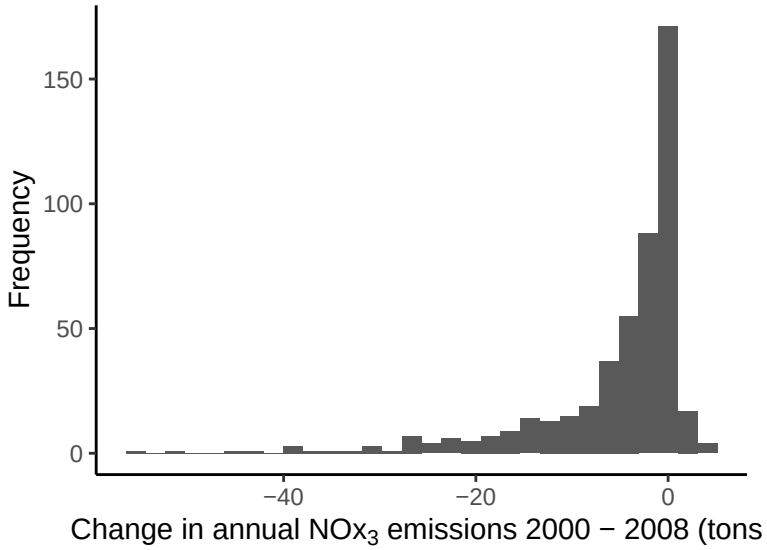

Figure 1 shows the skewed distribution of change of annual NO_x emissions from 2000 - 2008. There is a long tail to the left, and most of the data centers around zero. Each data point represents one county in California (n = 485).

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 <- median(data_clean$pct_black)

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 19.31%.

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

```
mod <- lm_robust(dnox_masstons ~ nbp, data = data_d)
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 total 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 total 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)
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 coeffients from estimating equation (1).

Table 1: Title

	NOX (tons)	
	(1)	(2)
NBP	-3.920^{***}	-7.141^{***}
	(0.763)	(1.068)
D		-2.588^{***}
		(0.995)
Interaction		6.372^{***}
		(1.505)
Constant	-3.622^{***}	-2.418^{***}
	(0.504)	(0.679)
Observations	485	485
R^2	0.052	0.086
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 annual 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 and have a fraction of the county population that is African American below the median value reported 7.14 less tons of annual NO_x emissions compared to 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.

The `D` coefficient 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 2.59 less tons of annual NO_x emissions compared to counties that did not participate in the NO_x Budget Program and have a fraction of the county population that is African American *over* 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 *over* the median reported value of 7.14, 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 6.37 more tons of annual 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.

```
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')
lower <- round(pred$fit[2], digits = 2)
upper <- round(pred$fit[3], digits = 2)
```

The 95% confidence interval is bounded by a lower bount of -6.44 and an upper bound of -3.57.