EDS241: Assignment 2

Marie Rivers

02/04/2022

In this assignment, test whether the effects of air quality regulations are the same accorss locations with different racial mix. More specifically, test if the NOx Budget Program, a cap-and-trade market for nitrogen oxides (NOx) emissions from power plants, leads to similar effects in counties that are predominantly white versus counties that are predominantly black. The NBP market was in effect in 212 of the 485 counties in the sample from 2003 to 2008.

NBP = 1 if the county was regulated under the NOx Budget Program NBP = 0 if the county was NOT regulated under the NOx Budget Program

PctBlack = the fraction of the county population that is black Dnox_masstons = the change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons). Dnox_masstons values are negative for counties where emissions of NOx from power plants declined.

1 Read Data

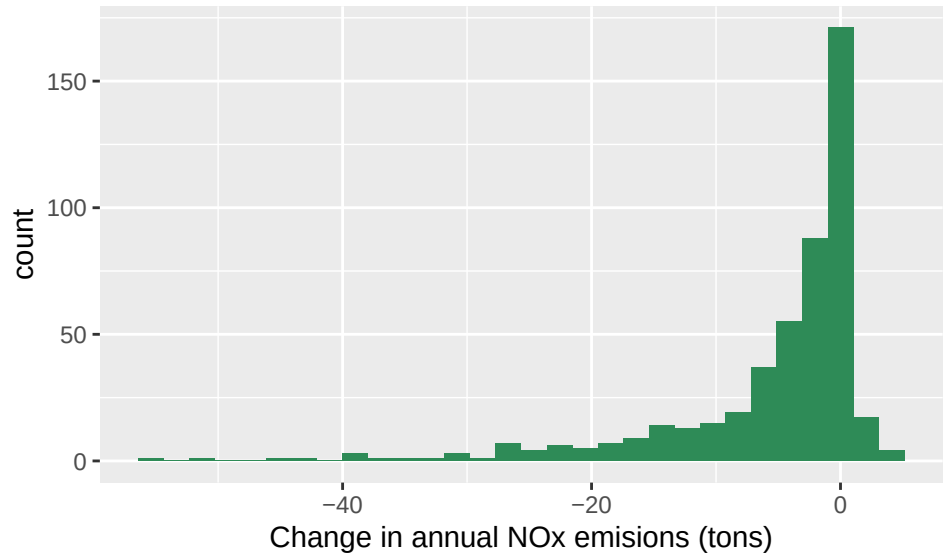
```
data <- read_excel(here("data", "NBP.xls"), sheet = "Sheet1")
```

2 Question a:

Make a histogram depicting the distribution of Dnox_masstons.

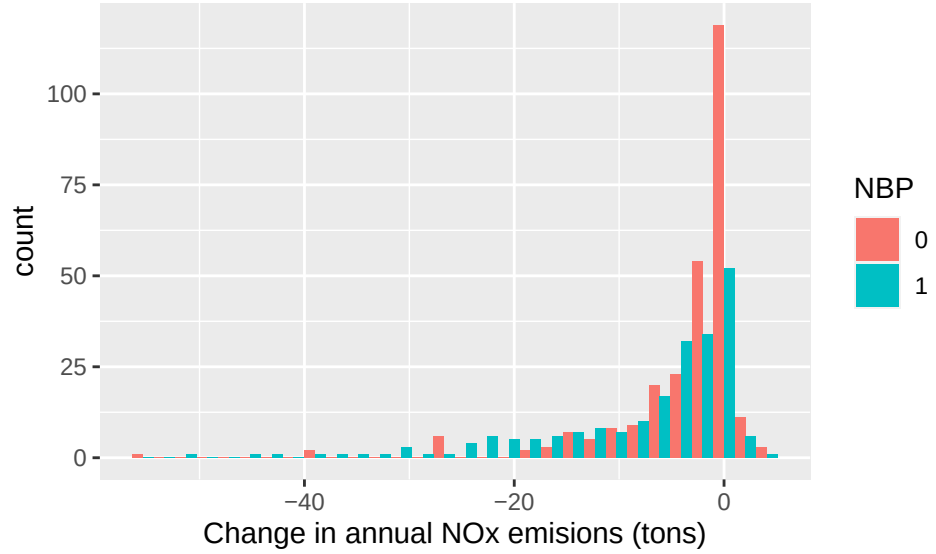
```
Dnox_histogram <- ggplot(data = data, aes(x = Dnox_masstons)) +
  geom_histogram(fill = "seagreen") +
  labs(x = "Change in annual NOx emisions (tons)")
```

Figure 1: Histogram of county level change in annual NOx emissions



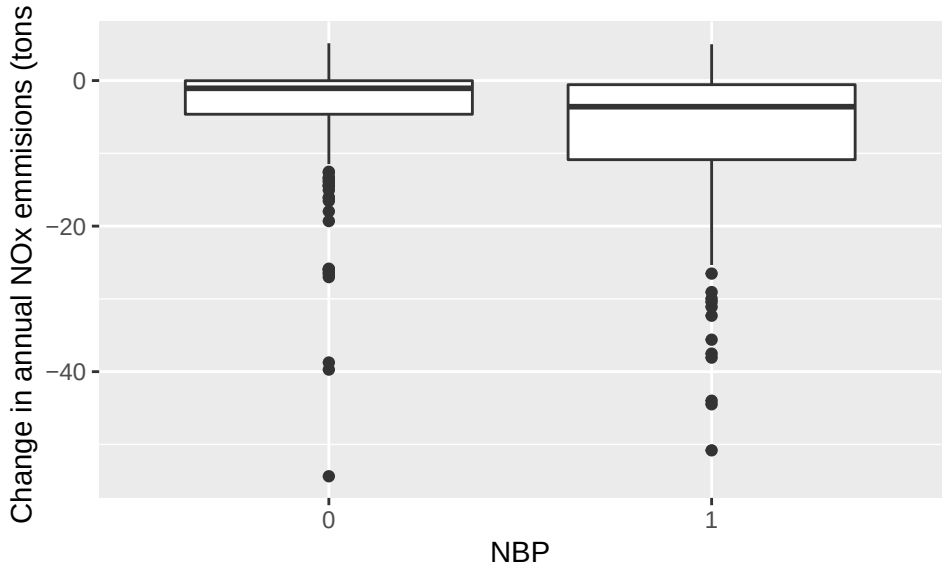

```
Dnox_histogram2 <- ggplot(data = data, aes(x = Dnox_masstons)) +
  geom_histogram(aes(fill = as.factor(NBP)), position = "dodge") +
  labs(x = "Change in annual NOx emisions (tons)", fill = "NBP")
```

Figure 2: Histogram of county level change in annual NOx emissions color coded by NBP status




```
boxplot <- ggplot(data = data, aes(x = as.factor(NBP), y = Dnox_masstons)) +
  geom_boxplot() +
  labs(x = "NBP", y = "Change in annual NOx emmisions (tons)")
```

Figure 3: A box plot to show another way of visualizing the county level change in annual NOx emissions based on NBP status



3 Question b:

Create an indicator = 1 if the county has `PctBlack` above the sample median and 0 otherwise. This is variable D for the rest of the assignment. What is the average of `PctBlack` for counties above the median, ie counties for which D=1?

```
PctBlack_median <- median(data$PctBlack)

data <- data %>%
  mutate(D = if_else(PctBlack > PctBlack_median, true = 1, false = 0))
```

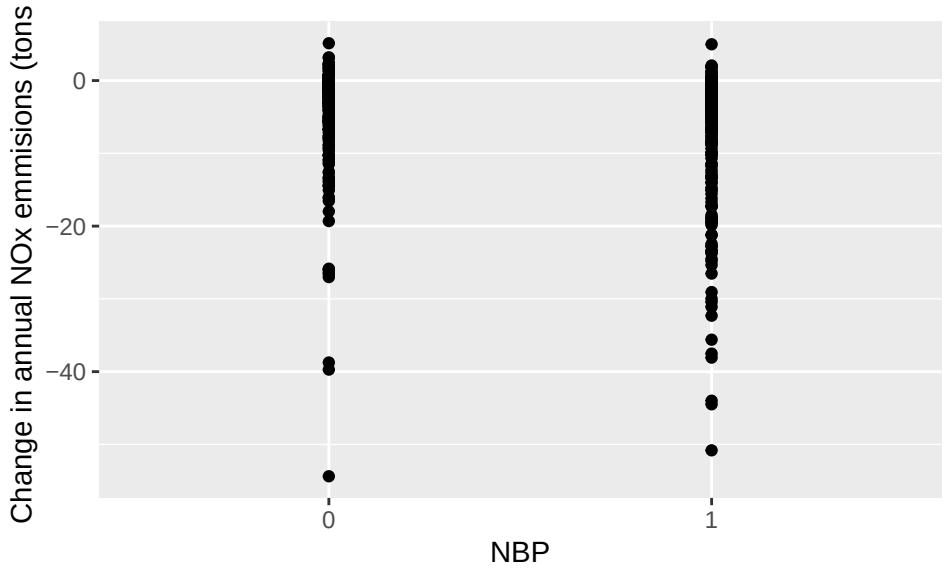
```
mean_PctBlack_D1 <- data %>%
  filter(D == 1) %>%
  summarise(mean(PctBlack)) %>%
  as.numeric()

mean_PctBlack_D0  <- data %>%
  filter(D == 0) %>%
  summarise(mean(PctBlack)) %>%
  as.numeric()
```

The average of `PctBlack` for counties with `PctBlack` above the sample median is 19.314%. The average of `PctBlack` for counties with `PctBlack` below the sample median is 1.557%.

4 Question c:

Estimate a regression of Dnox_masstons on NBP. Interpret the estimated intercept and the coefficient of NBP.

```
ggplot(data = data, aes(x = as.factor(NBP), y = Dnox_masstons)) +
  geom_point() +
  labs(x = "NBP", y = "Change in annual NOx emmisions (tons)")
```




```
Dnox_NBP_model_robust <- lm_robust(formula = Dnox_masstons ~ NBP, data = data)
```

```
# use lm to estimate coefficients
Dnox_NBP_model_lm <- lm(formula = Dnox_masstons ~ NBP, data = data)

# adjust standard errors using estimatr::starprep() instead of estimatr::lm_robust()
se_robust_model <- starprep(Dnox_NBP_model_lm)
```

Table 1: Change in annaul NOx emissions

	Change in annual NOx emissions
NBP	-3.920^{***}
	(0.796)
Constant	-3.622^{***}
	(0.420)
Observations	485
R^2	0.052
Note:	*p<0.1; **p<0.05; ***p<0.01
	Robust standard errors in parentheses

```
Dnox_NBP_intercept <- Dnox_NBP_model_robust$coefficients[1]
Dnox_NBP_intercept
```

```
## (Intercept)
##   -3.622031
```

```
Dnox_NBP_slope_coef <- Dnox_NBP_model_robust$coefficients[2]
Dnox_NBP_slope_coef
```

```
##       NBP
## -3.920467
```

```
Dnox_NBP_intercept_std_err <- Dnox_NBP_model_robust$std.error[1]
Dnox_NBP_intercept_std_err
```

```
## (Intercept)
##    0.420323
```

```
Dnox_NBP_std_err <- Dnox_NBP_model_robust$std.error[2]
Dnox_NBP_std_err
```

```
##       NBP
## 0.7959108
```

Based on the estimated intercept, the predicted change in annual NOx emissions for a county that was NOT regulated under the NOx Budget Program (NBP = 0) is -3.622 tons. Based on the estimated slope coefficient, change in annual NOx emissions was, on average, -3.92 tons higher for counties that were regulated under the NOx Budget Program than counties that were not regulated. Since negative values of `Dnox_masstons` correspond to a decline in NOx from power plants then the slope coefficient more clearly means that, on average, annual NOx emissions decreased by 3.92 tons more for counties that were regulated under the NOx Budget Program than counties that were not regulated.

```
# included as a check
# For linear regression with one categorical variable that only has two values, the regression estimate
data_NBP1 <- data %>%
  filter(NBP == 1)
mean_Dnox_NBP1 <- mean(data_NBP1$Dnox_masstons)
mean_Dnox_NBP1
```

```
## [1] -7.542498
```

```
data_NBP0 <- data %>%
  filter(NBP == 0)
mean_Dnox_NBP0 <- mean(data_NBP0$Dnox_masstons)
mean_Dnox_NBP0
```

```
## [1] -3.622031
```

```
est <- mean_Dnox_NBP1 - mean_Dnox_NBP0
est
```

```
## [1] -3.920467
```

5 Question 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.

```
Dnox_NBP_D_model_robust <- lm_robust(formula = Dnox_masstons ~ NBP + D + NBP:D, data = data)
```

```r
Dnox_NBP_D_intercept <- Dnox_NBP_D_model_robust$coefficients[1]
Dnox_NBP_D_intercept
```

```
## (Intercept)
##   -2.418075
```

```r
Dnox_NBP_D_slope_coef_NBP <- Dnox_NBP_D_model_robust$coefficients[2]
Dnox_NBP_D_slope_coef_NBP
```

```
##       NBP
## -7.141242
```

```r
Dnox_NBP_D_slope_coef_D <- Dnox_NBP_D_model_robust$coefficients[3]
Dnox_NBP_D_slope_coef_D
```

```
##         D
## -2.588031
```

```r
Dnox_NBP_D_slope_coef_NBP_D <- Dnox_NBP_D_model_robust$coefficients[4]
Dnox_NBP_D_slope_coef_NBP_D
```

```
##    NBP:D
## 6.371798
```

```r
Dnox_NBP_D_std_err_NBP <- Dnox_NBP_D_model_robust$std.error[2]
Dnox_NBP_D_std_err_NBP
```

```
##      NBP
## 1.257294
```

```r
Dnox_NBP_D_std_err_D <- Dnox_NBP_D_model_robust$std.error[3]
Dnox_NBP_D_std_err_D
```

```
##         D
## 0.8533574
```

```r
ci_lower_NBP <- Dnox_NBP_D_model_robust$conf.low[2]
ci_lower_NBP
```

```
##       NBP
## -9.611709
```

```r
ci_upper_NBP <- Dnox_NBP_D_model_robust$conf.high[2]
ci_upper_NBP
```

```
##       NBP
## -4.670775
```

```r
ci_lower_D <- Dnox_NBP_D_model_robust$conf.low[3]
ci_lower_D
```

```
##       D
## -4.2648
```

```r
ci_upper_D <- Dnox_NBP_D_model_robust$conf.high[3]
ci_upper_D
```

```
##          D
## -0.9112619
```

```r
# this model doesn't have robust standard errors

Dnox_NBP_D_model_lm <- lm(formula = Dnox_masstons ~ NBP + D + NBP:D, data = data)

se_robust_model2 <- starprep(Dnox_NBP_D_model_lm)
```

```r
# constant is the intercept
# not sure why type = "latex" didn't work
stargazer(Dnox_NBP_D_model_lm, se = se_robust_model2,
          type = "text", ci=FALSE, no.space = TRUE,
          header = FALSE, omit.stat = c("adj.rsq","ser", "f"),
          covariate.labels = c("NBP", "D", "NBP:D"), dep.var.labels = c("Change in annual NOx emissions
          dep.var.caption = c(""), notes = c("Robust standard errors in parentheses"),
          title = "Change in annaul NOx emissions based on NBP and PctBlack", table.placement = "H")
```

```
##
## Change in annaul NOx emissions based on NBP and PctBlack
## ================================================
##                 Change in annual NOx emissions
## ------------------------------------------------
## NBP                          -7.141***
##                               (1.257)
## D                            -2.588***
##                               (0.853)
## NBP:D                        6.372***
##                               (1.614)
## Constant                     -2.418***
##                               (0.442)
## ------------------------------------------------
## Observations                    485
## R2                             0.086
## ================================================
## Note:                  *p<0.1; **p<0.05; ***p<0.01
##              Robust standard errors in parentheses
```

Base on the estimated intercept, the model predicts the average change in annual NOx emissions for a county without an NBP market (NBP = 0) and with a percent black population below the median (D = 0) to be -2.418 tons.

Based on the coefficient on NBP, the average change in annual NOx emission changes by -7.141 tons when a county has an NBP market (NBP = 1) for counties with a percent black population below the median (D = 0). Since this coefficient is negative, the model predicts that annual NOx emissions due to the NBP market.

Based on the coefficient on D, on average, the change in annual NOx emissions are 2.588 tons lower for counties with above median black populations (D = 1) relative to counties with black populations below the median (D = 0) when there is no NBP market (NBP = 0).

```
test <- data %>%
  filter(NBP == 0) %>%
  group_by(D) %>%
  summarise(mean_dnox = mean(Dnox_masstons))
test
```

```
## # A tibble: 2 x 2
##       D mean_dnox
##   <dbl>     <dbl>
## 1     0     -2.42
## 2     1     -5.01
```

Based on the coefficient on the interaction term NBP:D, the impact of the NBP market on the change in annual NOX emissions is 6.372 tons lower for counties with black populations above the median (D = 1) than it is in counties with percent black populations below the median (D = 0). This coefficient tells us how the relationship between change in annual NOx emissions varies with percentage black population. This coefficient tells us the slope fo the relationship between change in annual NOx emissions and NBP market for counties with D = 1 minus the same slope for counties with D = 0.

6 Question e:

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

```
NBP <- 0
D <- 1

predicted_Dnox <- Dnox_NBP_D_intercept + (Dnox_NBP_D_slope_coef_NBP * NBP) + (Dnox_NBP_D_slope_coef_D *
predicted_Dnox
```

```
## (Intercept)
##   -5.006106
```

```
Pred <- data.frame(NBP=c(0), D=c(1))
fit <- predict(object = Dnox_NBP_D_model_robust, newdata = Pred, se.fit = TRUE, interval = 'confidence'
```

```
fit_value <- fit$fit[1]
fit_value
```

```
## [1] -5.006106
```

```
conf_lower <- fit$fit[2]
conf_lower
```

```
## [1] -6.440065
```

```
conf_upper <- fit$fit[3]
conf_upper
```

```
## [1] -3.572147
```

The predicted change in annual NOx emissions in a county that is not regulated by NBP (NBP = 0) and where the percent black population is able the sample median (D = 1) is -5.006tons. Since this value is negative, emissions of NOx from power plants in this county declined. Using heteroskedasticity-robust standard errors, the confidence interval for this prediction is from -6.44 tons to -3.572 tons.

Note: if you use lm, the predicted value would be the same but the confidence interval would be smaller because you're under estimating the error in the model.