EDS241: Assignment 2

Joe DeCesaro

01/31/2022

1   Assignment 2

The data included in the file NBP.xls, which is available on Gauchospace, are: **fips** (fips code identifying each county), NBP (indicator =1 if the county was regulated under the NOx Budget Program), PctBlack (fraction of the county population that is African American), and Dnox_masstons (change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons)). Note that the NBP market was in effect in 212 of the 485 counties in the sample from 2003 to 2008, so the 2008-2000 change give us a sense of the program's effect on emissions. If emissions of NOx from power plants declined in a county, then Dnox_masstons should be negative.
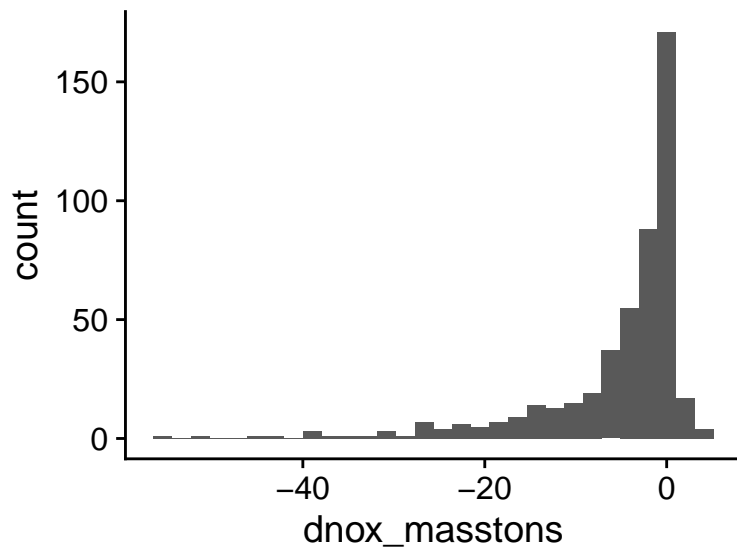
1.1   Clean data

The following code loads and cleans the data.

```
# Read in the first sheet and clean up
NOX_data <- read_xls(here("NBP.xls"), sheet = 1) %>%
  clean_names()
```

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

```
ggplot(data = NOX_data, aes(x = dnox_masstons)) +
  geom_histogram() +
  theme_cowplot(14)
```

1.3 (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?

```r
NOX_data <- NOX_data %>%
  mutate(D = if_else(pct_black > median(pct_black),
                     true = 1,
                     false = 0))

head(NOX_data)
```

```
## # A tibble: 6 x 5
##     fips   nbp pct_black dnox_masstons     D
##    <dbl> <dbl>     <dbl>         <dbl> <dbl>
## 1   1033     1      16.6         -5.89     1
## 2   1039     1      12.4          0.0200   1
## 3   1063     1      80.3         -7.11     1
## 4   1071     1       3.70       -21.2      0
## 5   1073     1      39.4        -25.3      1
## 6   1085     1      73.4          0.00256  1
```

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

```
model1 <- lm_robust(formula = dnox_masstons ~ nbp , data = NOX_data)

tidy_model1 = broom::tidy(model1)

tidy_model1 %>%
  dplyr::select(term, estimate, std.error, p.value) %>%
  knitr::kable() %>%
  kableExtra::kable_styling(full_width = FALSE,
                 latex_options = "HOLD_position")
```

term	estimate	std.error	p.value
(Intercept)	-3.622031	0.4203230	0.0000000
nbp	-3.920467	0.7959108	0.0000012

- Intercept: The change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) for a county not regulated under the NOx Budget Program.

- Coefficient: The difference in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) for a county that was regulated under the NOx Budget Program.

1.4.1 (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.

```
model2 <- lm_robust(formula = dnox_masstons ~ nbp + D + nbp*D , data = NOX_data)

tidy_model2 = broom::tidy(model2)

tidy_model2 %>%
  dplyr::select(term, estimate, std.error, p.value) %>%
  knitr::kable() %>%
  kableExtra::kable_styling(full_width = FALSE,
                 latex_options = "HOLD_position")
```

term	estimate	std.error	p.value
(Intercept)	-2.418075	0.4423052	0.0000001
nbp	-7.141242	1.2572938	0.0000000
D	-2.588031	0.8533574	0.0025542
nbp:D	6.371798	1.6144274	0.0000910

- Intercept: The change in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) for a county not regulated under the NOx Budget Program and that is below the median fraction of the county population that is African American.

- nbp: The difference in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) for a county that was regulated under the NOx Budget Program and that is below the median fraction of the county population that is African American.

- D: The difference in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) for a county that was not regulated under the NOx Budget Program and that is above the median fraction of the county population that is African American.

- nbp:D: The difference in annual NOx emissions from all power plants in a county between 2000 and 2008 (in tons) for a county that was regulated under the NOx Budget Program and that is above the median fraction of the county population that is African American.

1.4.2 (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.

```
new_dat <- data.frame(nbp = 0, D = 1)
ci <- predict(object = model2,
              newdata = new_dat,
              se.fit=TRUE,
              interval = "confidence")

ci$fit
```

```
##           fit       lwr       upr
## [1,] -5.006106 -6.440065 -3.572147
```

The predicted Dnox_masstons in a county that was not regulated under NBP and where PctBlack is above the sample median is -5.006106. We are 95% confident that the true predicted value for a county that was not regulated under NBP and where PctBlack is above the sample median is between -6.4400648 and -3.5721472.