EDS241: Assignment 2

Jake Eisaguirre

02/03/2022

0.0.1 Clean and plot data

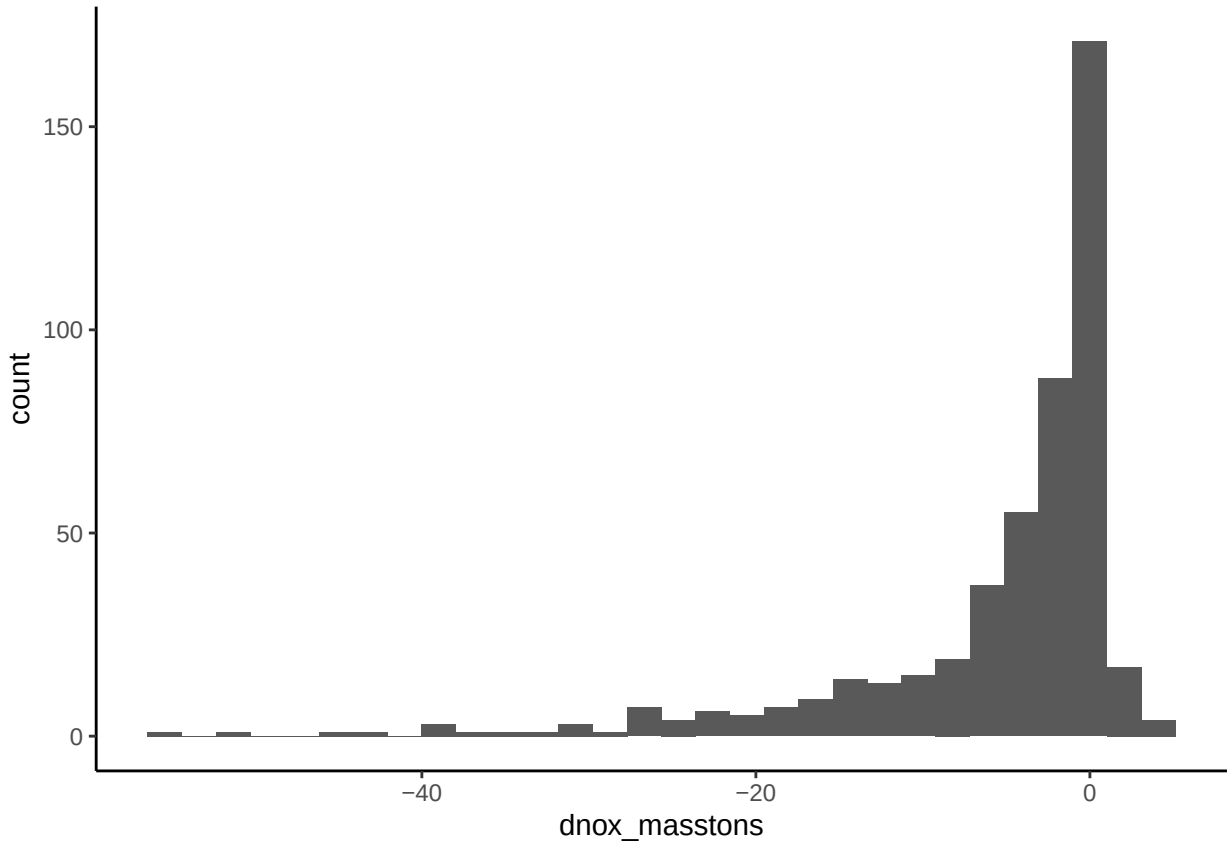
The following code loads and cleans the data.

```
data <- read_excel(here("NBP.xls")) %>%
  clean_names()
```

0.0.2 (a) Make a histogram depicting the distribution of Dnox_masstons

The following code plots the data

```
ggplot(data = data, aes(x = dnox_masstons)) +
  geom_histogram() +
  theme_classic()
```



0.0.3 (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?

The following code finds the median of PctBlack and creates an indicator of 1 if above the medium and finds the average of PctBlack for counties above the median

```
med <- print(median(data$pct_black))
```

```
## [1] 4.8
```

```
data <- data %>%
  mutate(D = ifelse(pct_black > med, c(1), c(0)))

ave_pctblack <- data %>%
  filter(D == 1) %>%
  summarise(ave_pctBlack = mean(pct_black))

print(ave_pctblack)
```

```
## # A tibble: 1 x 1
##   ave_pctBlack
##          <dbl>
## 1         19.3
```

```
kable(ave_pctblack)
```

ave_pctBlack
19.31375

The average PctBlack for counties above the median is 19.313.

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

The following code regresses Dnox_masstons on NBP

```
mod1 <- lm_robust(dnox_masstons ~ nbp, data = data)

tidy(mod1) %>%
  kable()
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
(Intercept)	-3.622031	0.4203230	-8.617257	0.0e+00	-4.447918	-2.796144	483	dnox_masstons
nbp	-3.920467	0.7959108	-4.925761	1.2e-06	-5.484342	-2.356591	483	dnox_masstons

Intercept: On average when the county has not been regulated under the NOx Budget program (nbp = 0) the change in NOx emmisions from all power plants in the counties between 2000-2008 will be -3.62 tons.

Slope: When the county has been regulated under the NOx Budget program (nbp = 1) the change in NOx emmissions from all power plants in the counties between 2000-2008 will be -3.92 tons.

0.0.5 (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.

The following code regresses dnox_masstons on NBP, D (median indicator of pct_black), and NBP as an interaction with D.

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

tidy(mod2) %>%
  kable()
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
(Intercept)	-2.418075	0.4423052	-5.466984	0.0000001	-3.287164	-1.5489862	481	dnox_masstons
nbp	-7.141242	1.2572938	-5.679851	0.0000000	-9.611709	-4.6707748	481	dnox_masstons
D	-2.588031	0.8533574	-3.032763	0.0025542	-4.264800	-0.9112619	481	dnox_masstons
nbp:D	6.371798	1.6144274	3.946785	0.0000910	3.199597	9.5439997	481	dnox_masstons

NBP: When all else is held constant, on average counties that have been regulated under the NOx Budget program (nbp = 1) will have a change in NOx emmissions from all power plants in the counties between 2000-2008 of -7.14 tons when above the median compared to below the median.

D: When all else is held constant, on average counties that have an average PctBlack above the median (19.31)(D=1) will have a change in NOx emmissions from all power plants in the counties between 2000-2008 of -2.59 tons when above the median compared to below the median.

NBP:D : The impact of NOx Budget program on NOx emmissions from all power plants in the counties between 2000-2008 is 6.37 tons greater in communities that have an average PctBlack above the median (19.31)(D=1) then communities that have an average PctBlack below the median (19.909)(D=0).

Intercept: On average when the county has not been regulated under the NOx Budget program (nbp = 0) and the PctBlack is below the median (19.31)(D=0) the change in NOx emmisions from all power plants in the counties between 2000-2008 will be -2.42 tons

0.0.6 (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.

The following code regresses dnox_masstons on counties not regulated by NBP and where the PctBlack is above the sample median (D=1)

```
mod3 <- lm_robust(dnox_masstons ~ nbp + D , data = data)
pred_data <- data.frame(nbp=c(0), D=c(1))

tidy(mod3) %>%
  kable()
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df	outcome
(Intercept)	-3.714107	0.4838207	-7.6766193	0.0000000	-4.664766	-2.763449	482	dnox_masstons
nbp	-3.933890	0.8190882	-4.8027670	0.0000021	-5.543314	-2.324465	482	dnox_masstons
D	0.197928	0.7806207	0.2535521	0.7999497	-1.335912	1.731768	482	dnox_masstons

```
predict(mod3, newdata = pred_data, se.fit = T, interval = 'confidence')
```

```
## $fit
##            fit       lwr       upr
## [1,] -3.516179 -4.818788 -2.213571
##
## $se.fit
##         1
## 0.6629394
```

The predicted change in NOx emmissions from all power plants in the counties between 2000-2008 under no regulation from NBP (NBP=0) and when the PctBacl is below the sample median (D=0) is -3.52 tons. The 95% confidence under heterskedasticity-robust standard errors is -4.82 tons - -2.21 tons.