EDS241: Assignment 2

Jake Eisaguirre

01/28/2022

0.0.1 Clean and plot data
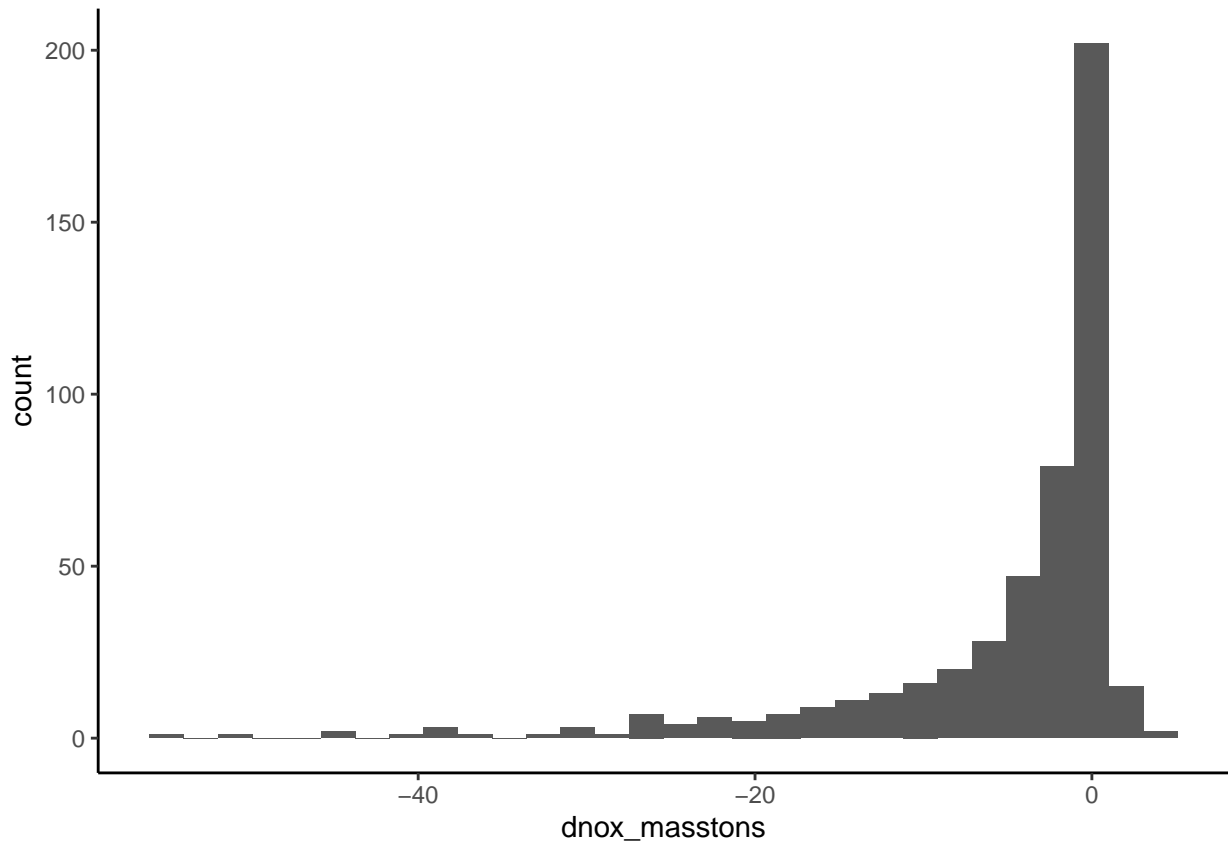
The following code loads and cleans the data.

```
data <- read_csv(here("NBP.csv")) %>%
  clean_names() %>%
  mutate(log_dnox = log(dnox_masstons))
```

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

The following code plots the data

```
#Not normally distributed????
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] 5
```

```
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         19.9
```

```
kable(ave_pctblack)
```

ave__pctBlack
19.90909

The average PctBlack for counties above the median is 19.909.

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)	-3.615385	0.4207345	-8.593030	0.0e+00	-4.442081	-2.788689	483	dnox__masstons
nbp	-3.908200	0.7957798	-4.911158	1.2e-06	-5.471818	-2.344582	483	dnox__masstons

Intercept: On average when the county has not been regulated under the NOx Budget program (nbp = 0) the change in annual 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 annual NOx emmissions from all power plants in the counties between 2000-2008 will be -3.91 tons.

###(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)	-2.601351	0.4690004	-5.546587	0.0000000	-3.522894	-1.6798086	481	dnox__masstons
nbp	-6.332611	1.2158251	-5.208488	0.0000003	-8.721596	-3.9436262	481	dnox__masstons
D	-2.214649	0.8589630	-2.578282	0.0102256	-3.902432	-0.5268653	481	dnox__masstons
nbp:D	5.035403	1.5917320	3.163474	0.0016576	1.907796	8.1630105	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 annual NOx emmissions from all power plants in the counties between 2000-2008 of -6.33 tons.

D: When all else is held constant, on average counties that have an average PctBlack above the median (19.909)(D=1) will have a change in annual NOx emmissions from all power plants in the counties between 2000-2008 of -2.21 tons.

NBP:D : The impact of NOx Budget program on annual NOx emmissions from all power plants in the counties between 2000-2008 is 5.04 tons greater in communities that have an average PctBlack above the median (19.909)(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 (19.909)(D=0) the change in annual NOx emmisions from all power plants in the counties between 2000-2008 will be -2.60 tons

0.0.5 (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)	-3.6131972	0.4897751	-7.3772577	0.0000000	-4.575555	-2.650839	482	dnox_masstons
nbp	-3.9079991	0.8090624	-4.8302817	0.0000018	-5.497724	-2.318274	482	dnox_masstons
D	-0.0047772	0.7681118	-0.0062194	0.9950402	-1.514039	1.504484	482	dnox_masstons

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

```
## $fit
##            fit      lwr       upr
## [1,] -3.617974 -4.89628 -2.339669
##
## $se.fit
##          1
## 0.6505708
```

The predicted annual 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 -3.62 tons. The 95% confidence under heterskedasticity-robust standard errors is -4.87 tons - -2.34 tons.