

# EDS241: Assignment 2

Charles Hendrickson

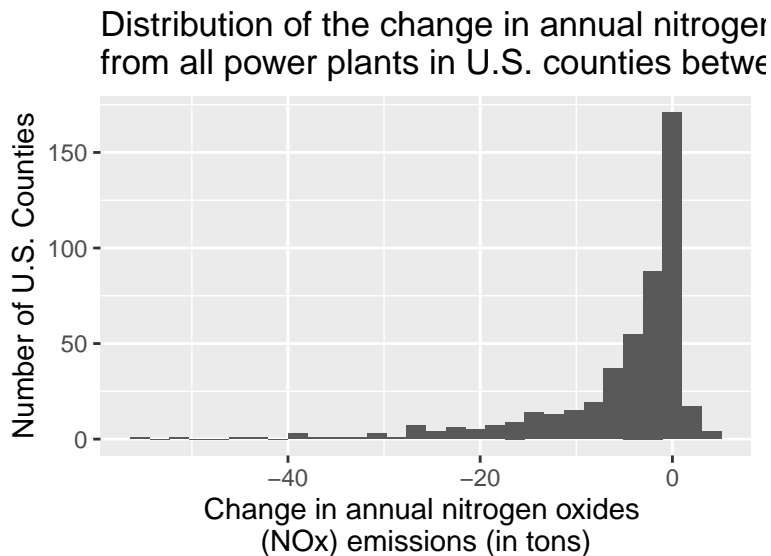
01/30/2022

Load the NBP.xls data

```
# load data
nbp_data <- read_excel("data/NBP.xls")
```

(a) Make a histogram depicting the distribution of Dnox\_masstons.

```
ggplot(data = nbp_data, aes(x = Dnox_masstons)) +
  geom_histogram() +
  labs(title = "Distribution of the change in annual nitrogen oxides (NOx) emissions
from all power plants in U.S. counties between 2000 and 2008",
x = "Change in annual nitrogen oxides
(NOx) emissions (in tons)",
y = "Number of U.S. Counties")
```



(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 average of PctBlack for counties above the median (counties for which D=1) is 19.31375%.**

```
# Create new column called "D" which is an indicator that equals 1 if the county has PctBlack above the
# First, find the sample median for PctBlack
PctBlack_median <- median(nbp_data$PctBlack)

# Second, create a new variable called "D" by using "ifelse", which takes a condition, followed by what
```

```

nbp_data <- nbp_data %>%
  mutate(D = ifelse(PctBlack > PctBlack_median, 1, 0))

# Third, find the average of PctBlack for counties above the median (counties for which D=1)

avg_PctBlack_D1 <- nbp_data %>%
  filter(D == 1)%>%
  summarise(mean_PctBlack = mean(PctBlack))

print(avg_PctBlack_D1)

## # A tibble: 1 x 1
##   mean_PctBlack
##           <dbl>
## 1           19.3

```

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

The estimated intercept means that for U.S. counties that were not regulated under the NOx Budget Program, the average change in annual NOx emissions between 2000 and 2008 from all power plants in that county was -3.622 tons on average.

The estimated coefficient on NBP means that an unregulated county being regulated under the NOx Budget Program will increase the average change in annual NOx emissions by -3.920 tons from all power plants in that county between 2000 and 2008.

```

# Regress Dnox_masstons on NBP
lm_robust(Dnox_masstons ~ NBP, data = nbp_data)

```

```

##           Estimate Std. Error   t value    Pr(>|t|)  CI Lower  CI Upper  DF
## (Intercept) -3.622031  0.4203230 -8.617257 9.830089e-17 -4.447918 -2.796144 483
## NBP         -3.920467  0.7959108 -4.925761 1.156072e-06 -5.484342 -2.356591 483

```

(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.