# EDS241: Take Home Finalr

## Allie Cole

## 03/16/2022

# 1 Load and clean data

```
data <- read.csv(here::here("KM_EDS241.csv"))
```

*** (a) Using the data for 1981, estimate a simple OLS regression of real house values on the indicator for being located near the incinerator in 1981. What is the house value "penalty" for houses located near the incinerator? Does this estimated coefficient correspond to the "causal" effect of the incinerator (and the negative amenities that come with it) on housing values? Explain why or why not.***

```
data_1981 <- data %>%
  filter(year == 1981)
model_a <- lm_robust(data = data_1981,
                     formula = rprice ~ nearinc)
huxreg(model_a)
```

|             | (1)               |
|-------------|-------------------|
| (Intercept) | 101307.515 ***    |
|             | (2944.810)        |
| nearinc     | -30688.274 ***    |
|             | (6243.167)        |
| N           | 142               |
| R2          | 0.165             |

*** p < 0.001; ** p < 0.01; * p < 0.05.

The "penalty" for houses located near the incinerator is a decrease of \$30688.27 in the inflation-adjusted sales price. The estimated coefficient corresponds partially to the "causal" effect of the incinerator on housing values. However, housing prices are likely lower in the area where the incinerator was more likely to be sited, meaning that this OLS likely suffers from omitted variable bias.

*** (b) Using the data for 1978, provide some evidence the location of the incinerator was not "random", but rather selected on the basis of house values and characteristics. [Hint: in the 1978 sample, are house values and characteristics balanced by `nearinc` status?] ***

```
data_1978 <- data %>%
  filter(year == 1978) %>%
  group_by(nearinc) %>%
```

```
summarise(mean_rprice = mean(rprice),
          mean_rooms = mean(rooms),
          mean_area = mean(area),
          mean_land = mean(land),
          mean_age = mean(age))
```

*** (c) Based on the observed differences in (b), explain why the estimate in (a) is likely to be biased downward (i.e., overstate the negative effect of the incinerator on housing values).***

The difference in question (b) show that the various housing characteristics are likely omitted variables from the model in question (a) because they are correlated to where a home is near the incinerator or not.

*** (d) Use a difference-in-difference (DD) estimator to estimate the causal effect of the incinerator on housing values without controlling for house and lot characteristics. Interpret the magnitude and sign of the estimated DD coefficient. ***

*** (e) Report the 95% confidence interval for the estimate of the causal effect ont he incinerator in (d).***

*** (f) How does your answer in (d) change when you control for house and lot characteristics? Teset the hypothesis that the coefficients on the house and lot characteristics are all jointly equal to 0.***

*(g) Using the results from the DD regression in (f), calculate by how much did real housing values change on average between 1978 and 1981.*

**(h) Explain (in words) what is the kay assumption underlying the causal interpretation of the DD estimator in the context of the incincerator construction in North Andover.**