

EDS241: Take Home Final

Alex Clippinger

03/16/2022

1 Data

The following code loads and cleans the data.

```
# Load data
km_data <- read_csv("KM_EDS241.csv") %>%
  mutate(nearinc = factor(nearinc))

# Create data frames for each year
km_81 <- km_data %>% filter(year==1981)
km_78 <- km_data %>% filter(year==1978)
```

2 Question 1

(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.

term	estimate	std.error	p.value
(Intercept)	101307.51	2944.810	0.0000000
nearinc1	-30688.27	6243.167	0.0000024

The house value “penalty” for houses located near the incinerator (nearinc=1) is \$-30688.27. This means that, based on this simple OLS regression, houses near the incinerator are, on average, worth \$30,688 less than houses away from the incinerator. The estimated coefficient does not correspond to the causal effect of the incinerator because other confounding variables, such as age of the home, square footage, and number of rooms, are not taken into account.

(b) Using the data for 1978, provide some evidence the location choice 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?]

```
price_difference = mean(km_78[km_78$nearinc==0,]$rprice) - mean(km_78[km_78$nearinc==1,]$rprice)
area_difference = mean(km_78[km_78$nearinc==0,]$area) - mean(km_78[km_78$nearinc==1,]$area)
rooms_difference = mean(km_78[km_78$nearinc==0,]$rooms) - mean(km_78[km_78$nearinc==1,]$rooms)
```

Prior to “treatment” (i.e., the construction of the incinerator), the mean average value of a home was \$18824.37 higher for the houses that would be away from the incinerator than for the houses that would be close to the incinerator in 1981. This positive difference indicates that homes further from the incinerator were valued higher (on average) prior to construction, which could mean that the location of construction was not random, but instead selected based on existing home value. Additionally, homes away from construction had 240.11 greater square footage and 0.79 more rooms, on average, supporting the claim that the location of construction was based on home characteristics. These relationships can be examined using simple OLS regression.

term	estimate	std.error	p.value
(Intercept)	82517.23	1878.277	0.0000000
nearinc1	-18824.37	6010.014	0.0020309

The first regression shows that there the average home value is statistically significantly lower for homes near the incinerator prior to construction.

term	estimate	std.error	p.value
(Intercept)	2074.7561	45.82799	0.0000000
nearinc1	-240.1132	120.21379	0.0473153

The second regression shows that the average home square footage is statistically significantly ($p < 0.05$) lower for home near the incinerator prior to construction.

term	estimate	std.error	p.value
(Intercept)	6.829268	0.0718256	0.0000000
nearinc1	-0.793554	0.1589515	0.0000014

The third regression shows that the average number of rooms was statistically significantly lower for homes near the incinerator prior to construction.

(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).

(d) Use a difference-in-differences (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 on the incinerator in (d).

(f) How does your answer in (d) changes when you control for house and lot characteristics? Test 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 key assumption underlying the causal interpretation of the DD estimator in the context of the incinerator construction in North Andover.