EDS241: Take Home Final

Scout Leonard

03/15/2022

Read in and clean data

```r
#read in data
housing <- read.csv(here("data", "KM_EDS241.csv"))
```

Question 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.

The code chunk below estimates a simple OLS regression of real house values on the indicator for being located near the incinerator in 1981.

```
#filter housing for only observations from 1981
housing_1981 <- housing %>%
  filter(year == 1981)

#model simple ols regression
ols_mod_1981 <- lm_robust(nearinc ~ rprice, data = housing_1981)
```

The code chuck below inputs the outputs of the simple OLS regression `ols_mod_1981` into a table using the function `kable()`.

```
#create table with regression results
ols_mod_1981_table <- tidy(ols_mod_1981)

#print table
ols_mod_1981_table %>%
  select(term, estimate, std.error, p.value, conf.low, conf.high) %>%
  kable()
```

term	estimate	std.error	p.value	conf.low	conf.high
(Intercept)	0.7808873	0.1358348	0.0000001	0.5123346	1.0494401
rprice	-0.0000054	0.0000014	0.0001646	-0.0000081	-0.0000026

The house value "penalty" for houses located near the incinerator is 0.0000054

This estimated coefficient does not correspond to the 'causal' effect of the incinerator. It is likely that there are other variables which effect the proximity of houses to the incinerator, causing omitted variables bias in this model.

Question 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?]

```
#filter housing for only observations from 1981
housing_1978 <- housing %>%
  filter(year == 1978)
```

Houses Near the Incinerator: 1978 Summary

```
housing_1978_nearinc <- housing_1978 %>%
  filter(nearinc == 1)

summary(housing_1978_nearinc) %>%
  kable()
```

year	age	rooms	area	land	nearinc	rprice
Min. :1978	Min. : 0.00	Min. :4.000	Min. : 750	Min. : 1710	Min. :1	Min. : 31000
1st Qu.:1978	1st Qu.: 17.00	1st Qu.:5.000	1st Qu.:1336	1st Qu.: 8143	1st Qu.:1	1st Qu.: 44000
Median :1978	Median : 28.00	Median :6.000	Median :1581	Median : 10684	Median :1	Median : 50950
Mean :1978	Mean : 39.79	Mean :6.036	Mean :1835	Mean : 21840	Mean :1	Mean : 63693
3rd Qu.:1978	3rd Qu.: 56.00	3rd Qu.:6.250	3rd Qu.:2093	3rd Qu.: 17724	3rd Qu.:1	3rd Qu.: 62250
Max. :1978	Max. :189.00	Max. :9.000	Max. :5078	Max. :282704	Max. :1	Max. :300000

Houses Far from the Incinerator: 1978 Summary

```
housing_1978_farinc <- housing_1978 %>%
  filter(nearinc == 0)

summary(housing_1978_farinc) %>%
  kable()
```

year	age	rooms	area	land	nearinc	rprice
Min. :1978	Min. : 0.00	Min. : 4.000	Min. : 960	Min. : 7858	Min. :0	Min. : 26000
1st Qu.:1978	1st Qu.: 0.00	1st Qu.: 6.000	1st Qu.:1819	1st Qu.: 43560	1st Qu.:0	1st Qu.: 69000
Median :1978	Median : 2.00	Median : 7.000	Median :2071	Median : 44431	Median :0	Median : 84300
Mean :1978	Mean : 12.75	Mean : 6.829	Mean :2075	Mean : 52569	Mean :0	Mean : 82517
3rd Qu.:1978	3rd Qu.: 9.00	3rd Qu.: 7.000	3rd Qu.:2443	3rd Qu.: 48593	3rd Qu.:0	3rd Qu.: 94000
Max. :1978	Max. :188.00	Max. :10.000	Max. :3792	Max. :544500	Max. :0	Max. :142500

In comparing summary statistics for variables from 1978 Andover houses from near and far from the incinerator, we can see that houses far from the incinerator have higher mean and median area, land, and price, and lower mean and median age. This suggests location choice of the incinerator was *not* random.

In the code chunk below, linear regressions return the average difference in outcomes between houses near and far from the incinerator.

```
mod1_1978 <- lm_robust(area ~ nearinc, data  = housing_1978)

mod2_1978 <- lm_robust(land ~ nearinc, data = housing_1978)

mod3_1978 <- lm_robust(rprice ~ nearinc, data = housing_1978)
```

The results indicate that on average, houses near the incinerator have an area -240.11324 square feet smaller than houses far from the incinerator.

On average, houses near the incinerator have -30729.1283391 fewer square feet of lot compared to houses far from the incinerator.

On average, houses near the incinerator have an inflation-adjusted sales price that is -18824.3704994 less than houses far from the incinerator.

While these are the returned average differences in 1978 between houses near and far from the incinerator, none of the Beta 1 values, which represent the average difference between the two groups, are significant below the 0.1 significance level.

Question 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).

Based on the observed differences in (b), the estimate in (a) is likely to be biased downward (overstate the negative effect of the incinerator on housing values), because there are variables that are correlated with whether or not a home is near or far from the incinerator.

Question 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.

```
#dd <- lm(formula = rprice ~ nearinc + )
```

Question E

Report the 95% confidence interval for the estimate of the causal effect on the incinerator in (d).

Question 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.

Question 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.

Question 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.