EDS_241_Final

Marie Rivers

03/15/2022

This assignment examines the impact of the opening of a garbage incinerator on housing values in North Andover, MA. The data for the exercise are a subset of the data in the paper: K.A. Kiel and K.T. McClain (1995): "House Prices During Siting Decision Stages: The Case of an Incinerator from Rumor Through Operation," Journal of Environmental Economics and Management 28, 241-255.

Background

Construction of a new garbage incinerator in North Andover in the early 1980s was controversial due to the increases in ambient pollution that it would create. Rumors of the incinerator began after 1978. Construction started in 1981, and the incinerator began operating in 1985. In Economics, land market theory suggests that local amenities are capitalized in housing values, and predicts that the prices of houses located near the incinerator would fall compared to the price of houses located further away from the incinerator. By 1981, you can assume that all market participants had full information on the upcoming garbage incinerator, so that housing values had capitalized the upcoming arrival of the incinerator.

Read Data

Data Variables

- year = year house was sold (1978 is before rumors of the incinerator; 1981 = during construction of the incinerator)
- age = age of house in years
- rooms = number of rooms
- area = living areas in square feet
- land = lot size in square feet
- nearinc = dummy variable (0 = not near incinerator; 1 = near incinerator)
- rprice = real house values

```r
data_summary <- data %>%
  group_by(year, nearinc) %>%
  summarise(num_houses = n(),
            mean_age = round(mean(age), 1),
            mean_rooms = round(mean(rooms), 1),
            mean_area = round(mean(area), 0),
            mean_land = round(mean(land), 0),
            mean_price = round(mean(rprice), 0))
```

year	nearinc	num_houses	mean_age	mean_rooms	mean_area	mean_land	mean_price
1978	0	123	12.7	6.8	2075	52569	82517
1978	1	56	39.8	6.0	1835	21840	63693
1981	0	102	8.5	6.8	2351	40251	101308
1981	1	40	28.0	6.2	1962	23164	70619

```
data_summary_table <- data_summary %>%
  kable(col.names = c("year", "nearinc", "num_houses", "mean_age", "mean_rooms", "mean_area", "mean_land
  kable_paper(full_width = FALSE) %>%
  row_spec(c(0), background = "lightgray")
data_summary_table
```

Question a:

Using the data for 1981, estiamte 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 amenitites that come with it) on housing values? Explain why or why not?

```
data_1981 <- data %>%
  filter(year == 1981)
```

```
ggplot(data = data_1981, aes(x = nearinc, y = rprice)) +
  geom_jitter(aes(color = nearinc)) +
  geom_boxplot() +
  labs(title = "Real house values in 1981 based on incinerator proximity",
       x = "incinerator proximity", y = "real house value ($)") +
  theme(plot.title.position = "plot")
```
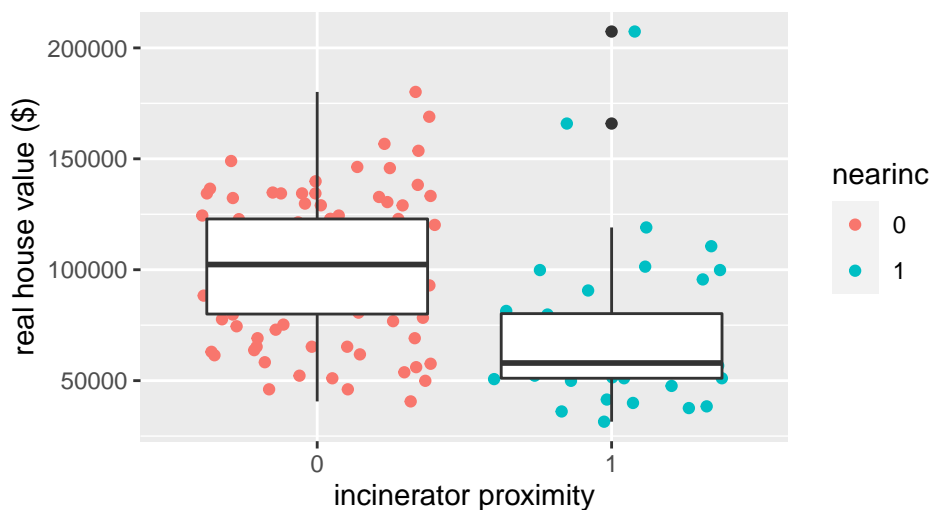
```
model_a <- lm_robust(formula = rprice ~ nearinc, data = data_1981)
```

```
huxreg("real house value" = model_a)
```

	real house value
(Intercept)	101307.515 ***
	(2944.810)
nearinc1	-30688.274 ***
	(6243.167)
N	142
R2	0.165

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

```
model_a_coef <- round(model_a$coefficients[2], 2)
model_a_coef
```

```
##  nearinc1
## -30688.27
```

```
model_a_se <- round(model_a[[2]][2], 2)
model_a_se
```

```
## nearinc1
##  6243.17
```

Based on a simple OLS regression of real house values on the indicator for being located near the incinerator in 1981, the house value "penalty" for houses located near the incinerator was -30688.27 dollars. xxx

Question b:

Using the data for 1978, provide some evidence that 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?

xxx

Question c:

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

xxx # Question 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 estiamted DD coefficient.

xxx

Question e:

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

xxx

Question f:

How does your answer in (d) change 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.

xxx

Question g:

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

xxx

Question h:

Explain in words what is the key assumption underlying the causal interpretation of the DD estimator in the context of th eincinerator construction in North Andover.

xxx