

EDS241: Take Home Final

Alexandra Yousefivand

03/16/2022

```
# Load data
data_raw <- read.csv(here("KM_EDS241.csv"))

data <- data_raw %>%
  mutate(year = as.factor(year),
         age = as.numeric(age),
         rooms = as.numeric(rooms),
         area = as.numeric(area),
         land = as.numeric(land),
         nearinc = as.factor(nearinc),
         rprice = as.numeric(rprice))
```

(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.

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

```
model <- lm_robust(formula = rprice ~ nearinc, data = data_1981)
summary(model)
```

```
##
## Call:
## lm_robust(formula = rprice ~ nearinc, data = data_1981)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
## (Intercept)   101308      2945   34.402 3.633e-70   95485   107130 140
## nearinc1      -30688      6243   -4.915 2.442e-06  -43031  -18345 140
##
## Multiple R-squared:  0.1653 ,    Adjusted R-squared:  0.1594
## F-statistic: 24.16 on 1 and 140 DF,  p-value: 2.442e-06
```

```
penalty <- abs(round(model$coefficients[2]))
```

The house value “penalty” for houses located near the incinerator is $\$3.0688 \times 10^4$. The very low p-value indicates that this is a statistically significant result and this estimated coefficient might correspond to the

‘causal’ effect of the incinerator. There are other variables that may contribute to the difference in housing prices as well, which implies the possibility of omitted variables bias.

(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?]

```
# subset data
data_1978 <- data %>% filter(year == 1978)

# unadjusted mean difference
data_nearinc <- data_1978 %>% filter(nearinc == 1)
nearinc_mean_price <- round(mean(data_nearinc$rprice), 2)

data_not_nearinc <- data_1978 %>% filter(nearinc == 0)
not_nearinc_mean_price <- round(mean(data_not_nearinc$rprice), 2)

difference <- not_nearinc_mean_price - nearinc_mean_price
difference
```

```
## [1] 18824.37
```

```
# unadjusted mean difference using linear regression
model_age <- lm_robust(formula = age ~ nearinc, data = data)
age <- round(model_age$coefficients[[2]], 2)
summary(model_age)
```

```
##
## Call:
## lm_robust(formula = age ~ nearinc, data = data)
##
## Standard error type: HC2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|) CI Lower CI Upper DF
## (Intercept)    10.82      1.951   5.548 6.084e-08   6.984   14.66 319
## nearinc1       24.03      3.966   6.060 3.835e-09  16.230   31.83 319
##
## Multiple R-squared:  0.1145 ,    Adjusted R-squared:  0.1117
## F-statistic: 36.72 on 1 and 319 DF,  p-value: 3.835e-09
```

```
model_rooms <- lm_robust(rooms ~ nearinc, data = data)
rooms <- round(model_rooms$coefficients[[2]], 2)

model_area <- lm_robust(area ~ nearinc, data = data)
area <- round(model_area$coefficients[[2]], digits = 2)

model_land <- lm_robust(land ~ nearinc, data = data)
land <- round(model_land$coefficients[[2]], digits = 2)
```