Chapter 6 -

Thalles Quinaglia Liduares

31/03/2022

Exercise 6.1

Upload packages

```
library(wooldridge)
library(lmreg)
```

Upload database

```
data<-wooldridge::kielmc

attach(data)
```

Use the data in KIELMC.RAW, only for the year 1981, to answer the following questions. The data are for houses that sold during 1981 in North Andover, Massachusetts; 1981 was the year construction began on a local garbage incinerator

(i) To study the effects of the incinerator location on housing price, consider the simple regression model

$$log(price) = \beta_0 + \beta_1 log(dist) + u$$

where price is housing price in dollars and `dist` is distance from the house to the incinerator measured in feet. Interpreting this equation causally, what sign do you expect for β_1 if the presence of the incinerator depresses housing prices? Estimate this equation and interpret the results.

```
summary(lm1<-lm(lprice~ldist))
```

```
##
## Call:
## lm(formula = lprice ~ ldist)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.22356 -0.28076 -0.05527  0.27992  1.29332
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.25750    0.47383  17.427  < 2e-16 ***
## ldist        0.31722    0.04811   6.594 1.78e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4117 on 319 degrees of freedom
## Multiple R-squared:  0.1199, Adjusted R-squared:  0.1172
## F-statistic: 43.48 on 1 and 319 DF,  p-value: 1.779e-10
```

The estimated equation is expressed as follows

$$\widehat{log(price)} = 8.25 + 0.31log(dist)$$

In theory, the expected sign would be negative. However, the estimated coefficient for `dist` is positive. In this case, there's a postive relatioshinp between the distance of incinerator and the houses.

(ii) To the simple regression model in part (i), add the variables log(intst), log(area), log(land), rooms, baths, and age, where `intst` is distance from the home to the interstate, area is square footage of the house, land is the lot size in square feet, rooms is total number of rooms, baths is number of bathrooms, and age is age of the house in years. Now, what do you conclude about the effects of the incinerator? Explain why (i) and (ii) give conflicting results.

```
summary(lm2<-lm(lprice~ldist+lintst+larea+lland+rooms+baths+age))
```

```
##
## Call:
## lm(formula = lprice ~ ldist + lintst + larea + lland + rooms +
##     baths + age)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.35838 -0.18221  0.00117  0.20533  0.82180
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.2996221  0.5960536  10.569  < 2e-16 ***
## ldist        0.0282046  0.0532130   0.530  0.59647
## lintst      -0.0438028  0.0424358  -1.032  0.30277
## larea        0.5124039  0.0698229   7.339 1.87e-12 ***
## lland        0.0782203  0.0337208   2.320  0.02100 *
## rooms        0.0503141  0.0235113   2.140  0.03313 *
## baths        0.1070541  0.0352304   3.039  0.00258 **
## age         -0.0035631  0.0005774  -6.171 2.10e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2828 on 313 degrees of freedom
## Multiple R-squared:  0.5925, Adjusted R-squared:  0.5834
## F-statistic: 65.02 on 7 and 313 DF,  p-value: < 2.2e-16
```

In this case, the magnitude of coefficient associated to the `dist` variable is very small, equal to 0.02, despites it's positive sign. The others explanatory variables has expected signs.

(iii) Add $[log(intst)]^2$ to the model from part (ii). Now what happens? What do you conclude about the importance of functional form?

```
summary(lm3<-lm(lprice ~ ldist + lintst + larea + lland + rooms +
    baths + age + lintstsq))
```

```
##
## Call:
## lm(formula = lprice ~ ldist + lintst + larea + lland + rooms +
##     baths + age + lintstsq)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.41716 -0.17776  0.01025  0.19295  0.72087
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.7916130  2.2956018  -1.652  0.09960 .
## ldist        0.1897791  0.0626879   3.027  0.00267 **
## lintst       1.9026565  0.4304872   4.420 1.37e-05 ***
## larea        0.5137107  0.0677316   7.585 3.86e-13 ***
## lland        0.1068861  0.0333137   3.208  0.00147 **
## rooms        0.0494857  0.0228076   2.170  0.03078 *
## baths        0.0898826  0.0343833   2.614  0.00938 **
## age         -0.0035701  0.0005601  -6.374 6.62e-10 ***
## lintstsq    -0.1128526  0.0248446  -4.542 7.96e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2743 on 312 degrees of freedom
## Multiple R-squared:  0.6178, Adjusted R-squared:  0.608
## F-statistic: 63.04 on 8 and 312 DF,  p-value: < 2.2e-16
```

There's a considerable increase in the R^2 of the estimated equation, equal to 61.78% and almost all variables has statistical significance in this case.

(iv) Is the square of log(dist) significant when you add it to the model from part (iii)?

```
ldist_sqr<-ldist^2

summary(lm4<-lm(lprice ~ ldist + lintst + larea + lland + rooms +
    baths + age + lintstsq+ldist_sqr))
```

```
##
## Call:
## lm(formula = lprice ~ ldist + lintst + larea + lland + rooms +
##     baths + age + lintstsq + ldist_sqr)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42171 -0.17728 -0.00087  0.19645  0.71830
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.110e+01  7.006e+00  -1.585  0.11405
## ldist        2.109e+00  1.739e+00   1.213  0.22613
## lintst       1.521e+00  5.520e-01   2.755  0.00622 **
## larea        5.062e-01  6.805e-02   7.439 9.94e-13 ***
## lland        9.696e-02  3.449e-02   2.811  0.00525 **
## rooms        4.776e-02  2.285e-02   2.090  0.03743 *
## baths        8.939e-02  3.437e-02   2.600  0.00976 **
## age         -3.524e-03  5.615e-04  -6.275 1.17e-09 ***
## lintstsq    -8.891e-02  3.297e-02  -2.697  0.00739 **
## ldist_sqr   -1.026e-01  9.286e-02  -1.104  0.27027
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2742 on 311 degrees of freedom
## Multiple R-squared:  0.6193, Adjusted R-squared:  0.6083
## F-statistic: 56.21 on 9 and 311 DF,  p-value: < 2.2e-16
```

No. The square of `log(dist)` isn't significant at any level.