HSE and University of London Double Degree Programme in Data Science and Business Analytics

Elements of Econometrics, 2023-2024

Classteacher: Ksenia Kasianova.

Quiz 2.

Problem 1

We estimate two linear regression to learn $\log(Price)$ determinants for Moscow flats based on CIAN data. The regressors used in the analysis are: $Dist$ – distance from Red Square in km, $Brick$ – dummy variable, 1 – if house if from concrete, 0 – otherwise, $Totsp$ – total space in sq. m., $HouseAge$ – house age in years. $Price$ was measured in mln rub.

Model 1:

```
Call:
lm(formula = LogPrice ~ log(Dist) + Brick + Totsp + HouseAge,
    data = df1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.30669 -0.14870 -0.01638  0.11912  1.59136

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 16.5111476  0.0177128 932.160  < 2e-16 ***
log(Dist)   -0.4609611  0.0047232 -97.594  < 2e-16 ***
Brick        0.0563552  0.0068825   8.188 3.06e-16 ***
Totsp        0.0160821  0.0001155 139.281  < 2e-16 ***
HouseAge    -0.0031191  0.0001027 -30.379  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2409 on 8042 degrees of freedom
  (714 observations deleted due to missingness)
Multiple R-squared:  0.882,     Adjusted R-squared:  0.882
F-statistic: 1.503e+04 on 4 and 8042 DF,  p-value: < 2.2e-16
```

Model 2:

```
Call:
lm(formula = LogPrice ~ log(Dist) + Brick + Totsp + Totsp2 +
    HouseAge + HouseAge2, data = df1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.21542 -0.15016 -0.01509  0.11991  1.58668

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.647e+01  2.400e-02 686.074  < 2e-16 ***
log(Dist)   -4.608e-01  4.749e-03 -97.038  < 2e-16 ***
Brick        4.352e-02  7.426e-03   5.861 4.79e-09 ***
Totsp        1.784e-02  4.692e-04  38.022  < 2e-16 ***
Totsp2      -1.055e-05  2.663e-06  -3.962 7.49e-05 ***
HouseAge    -4.014e-03  2.871e-04 -13.985  < 2e-16 ***
HouseAge2    9.852e-06  3.031e-06   3.250  0.00116 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2405 on 8040 degrees of freedom
  (714 observations deleted due to missingness)
Multiple R-squared:  0.8824,    Adjusted R-squared:  0.8824
F-statistic: 1.006e+04 on 6 and 8040 DF,  p-value: < 2.2e-16
```

(5 points) Interpret each coefficient in Model 2.

(2 points) Which model is better. Provide appropriate statistical test to support your decision.

Problem 2

Consider the following model:

$y_i = \beta_0 + \beta_1 \frac{1}{x_i} + \beta_2 z_i^2 + \beta_3 \frac{z_i}{x_i} + \epsilon_i$

Derive how

(1 point) $1/x_i$ affects y_i? Interpret the effect.

(1 point) x_i affects y_i? Interpret the effect.

(1 point) z_i affects y_i? Interpret the effect.