APPLIED STATISTICS EXAM

DATE: 12/07/2022
STUDENT: FILIPPO CIPRIANI
PERSONA CODE: 10956877

EXERCISE NUMBER 3

POINT A)

Let's build the linear model requested. We estimate from the model the coefficients and the standard deviation of the error, sigma:

(Intercept)	2950.82578582
footage	23.86056223
age	-10.04728859
renovation	-7.99891771
transport	0.70526444
center	-0.11015755
supermarket	-0.25634206
park	-0.17230360
two.bathroomsTRUE	76.45448500 (+2950.82578582)
footage:two.bathroomsTRUE	1.64253808 (+23.86056223)
age:two.bathroomsTRUE	15.38915966 (-10.04728859)
renovation:two.bathroomsTRUE	10.69972492 (-7.99891771)
transport:two.bathroomsTRUE	-0.88497751 (+0.70526444)
center:two.bathroomsTRUE	0.03612037 (-0.11015755)
supermarket:two.bathroomsTRUE	0.52644766 (-0.25634206)
park:two.bathroomsTRUE	-0.18265410 (-0.17230360)
sigma:	462.9

Each coefficient for the case where two.bathrooms is TRUE can be obtained by the sum of the coefficient without :two.bathroomsTRUE with the respective coefficent:two.bathroomsTRUE, as shown in the calculation between brackets.

POINT B)

Assumption of the model: Eps ~ N(0, sigma^2)

We check if residuals are normal through a shapiro test (pvalue 0.6531) and we conclude that they are.

the VIFs are: footage

2.495567

age

2.315918

renovation

2.341154

transport

6.302478

center

1.998018

supermarket

6.576233

park

2.059706

two.bathrooms

83.785607

footage:two.bathrooms

8.036454

age:two.bathrooms

6.926501

renovation:two.bathrooms

5.645448

transport:two.bathrooms

107.357046

center:two.bathrooms

6.587780

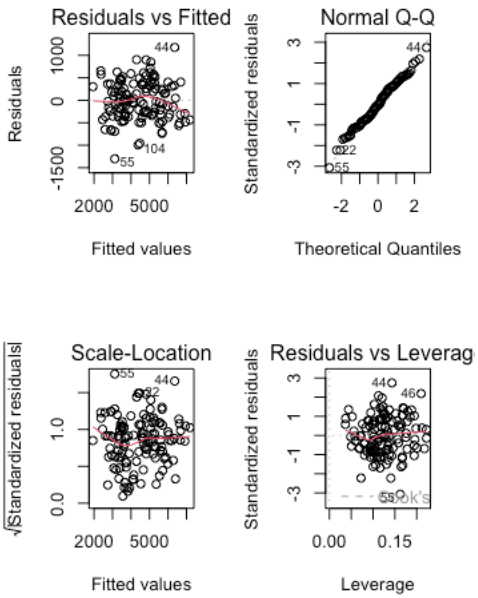
supermarket:two.bathrooms

154.244320

park:two.bathrooms

5.278984

and they’re not good, there’s some collinearity.



The diagnostic is good, we can see that the residuals are more or less homoschedastic.

POINT C)

By performing a LASSO selection with lambda = 45, we get the significant coefficients:

	s1
(Intercept)	2824.34957726
footage	22.88641948
age	-2.06471440
center	-0.08459243
park	-0.15025926
two.bathroomsTRUE	125.78295577
footage:two.bathroomsTRUE	2.58589198
age:two.bathroomsTRUE	6.35257297
supermarket:two.bathroomsTRUE	0.10959016

POINT D)

Using CV to select lambda for the LASSO over the interval [1:100] we get an optimal lamba of: 17.8865 and the selected relevant coefficients are:

(Intercept)	2901.54124846
footage	23.75002774
age	-6.79687171

transport 0.13782762

center -0.09750120

park -0.20322332

two.bathroomsTRUE 28.98388437

footage:two.bathroomsTRUE 1.90217820

age:two.bathroomsTRUE 11.75564954

center:two.bathroomsTRUE 0.01541927

supermarket:two.bathroomsTRUE 0.07068436

POINT E)

The pointwise estimate of the price of an apartment with square footage = 30m2, age = 5, renovation = 5, transport = 300m, center = 1000m, supermarket = 500m, park = 100m, and one bathroom using the LASSO model selected with CV at point D is:

price = 3503.582