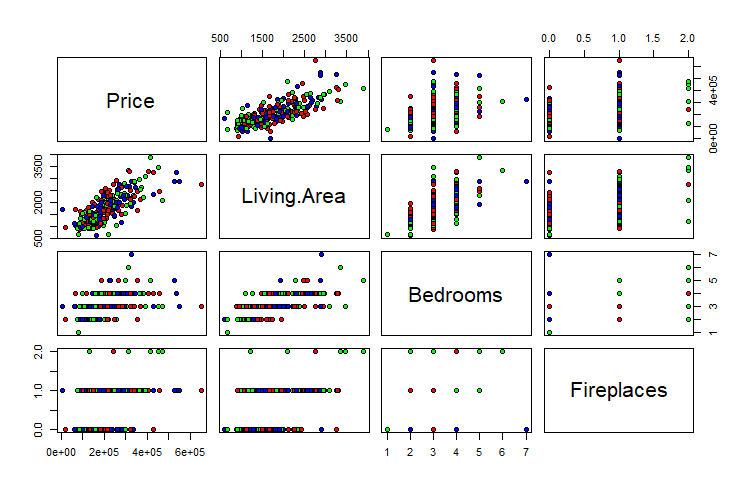
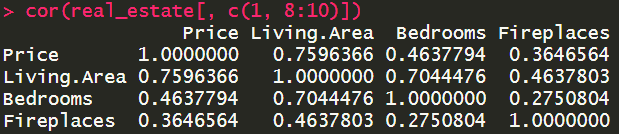
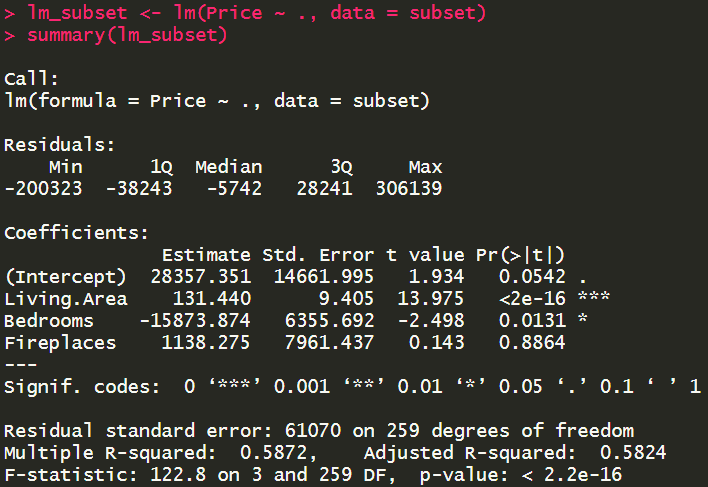
Project Part 2

#Model 1：

1. The scatter plot matrix and the correlation matrix of all the four variables are showed below.



Yes, there is a significant evidence of multicolinearity: in the subplot(2, 3) we notice that there is a strong correlation between Living.Area and Bedrooms, a phenomenon where we can conclude collinearity between these two predictor variables. Same kind of evidence can be provided from correlation matrix of these four variables: we can find that the value of correlation between Living.Area and Bedrooms is 0.7044476, much larger than any other values of different pairs in this correlation matrix.

(2)

LS regression function is:

indicates the selling price of the house in issue.

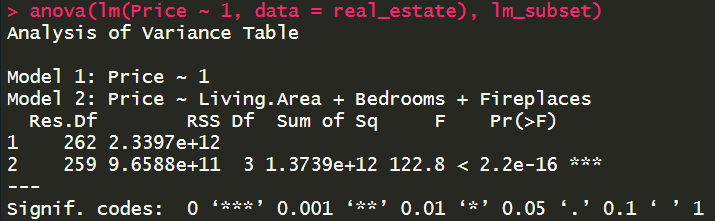
coefficient of X1(Living Area)

coefficient of X2(Bedrooms)

, coefficient of X3(Fireplaces)

(the coefficient of the number of bedrooms) indicates that, given other condition constant, as increases 1 unit, also tends to decrease 15873.874 units.

(3)



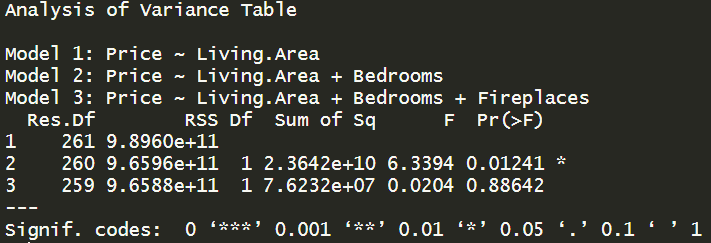
The alternatives of F test in this model are:

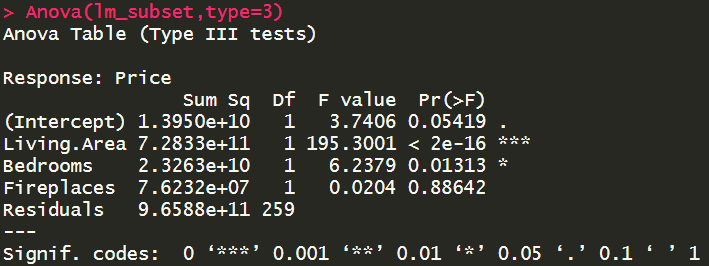
H0: = = = 0

Ha: at least one (i = 1, 2, 3) is not equal to 0.

The value of F test is 122.8, and the P value of F test is less than 2.2e-16, which is close to 0.

The conclusion is that we accept the Ha and reject H0.

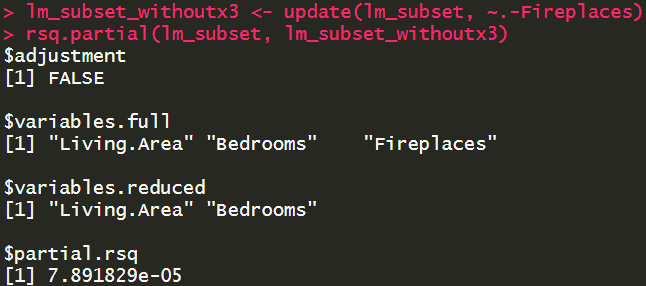
(4) 



the value of extra SSR of X3 is

= 9.6596×1011 – 9.6588×1011

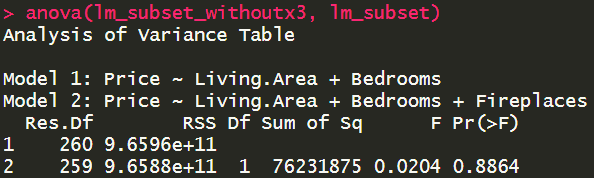
= 7.6232×107.



The coefficient of partial determination :

7.891829×10-5

(5)



The alternatives of F test in this model are:

H0: = 0

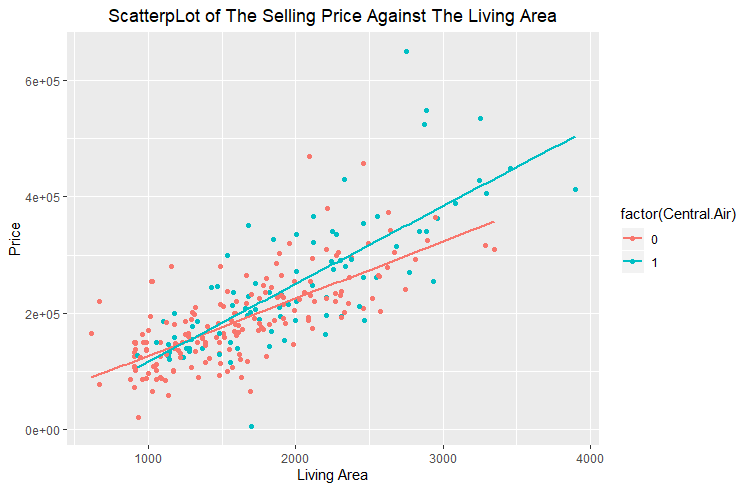
Ha: ≠ 0

The value of F test is 0.0204 and the p-value of F test is 0.8864, which is close to 1.

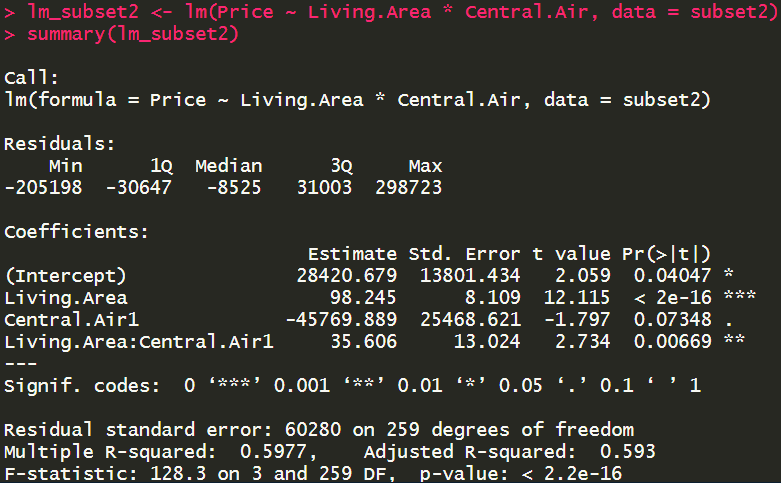
The conclusion is that we accept the H0 and reject Ha.

#Model 2：

(6)



From this scatterplot, we can notice that these two lines are not coincide. The interaction term actually exists because these two lines are not paralleled and indeed intersect within the sample space.

(7) 

indicatesthe selling price of the house in issue.

coefficient of X1(Living Area)

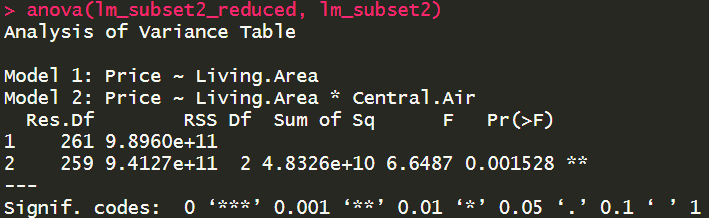
coefficient of X2(Central Air)

, coefficient of interaction term X1X2(Living Area & Central Air)

(The coefficient of Central Air) indicates how much lower for having central air system is than that for not having central air system, holding X1 = 0; However, there is no practical meaning.

(The coefficient of the interaction term between Living Area and Central Air) indicates how much larger/less the magnitude of changes of selling price for a house with a central air system than that of changes of selling price for a house without a central air system.

(8)



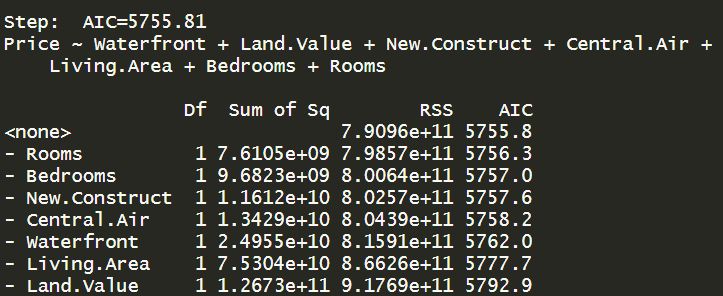
The alternatives of partial F test in this model are:

H0: = = 0

Ha: at least (i = 2, 3) ≠ 0

The value of F-test statistic is 6.6487, the p-value is 0.001528 < 0.05. So, we conclude Ha.

(9)



Therefore, the subset of predicted variables that should be included is: waterfront, Land.Value, New.Construct, Central.Air, Living.Area, Bedrooms and Rooms.