**Linear Regression Model**

To build the linear regression model, we remove the unnecessary variables such as ids. Then we transform qualitative data such date into quantitative ones. We also transform the variables that have a lot of 0 values into dummy variables. Then we calculate the age of the houses by subtracting the year of built from year 2016. Afterwards, we regress against the logarithm of price with the new variables.

During our first attempt to build the linear regression model, we find out some variables are linearly dependent on the others. For example, the square foot of living = the square foot of basement + the square foot of upper areas. We delete the items that can be interpreted as linear combinations of the others, and regress again.

During the second attempts, we found some variables are not significant based on t-test and partial F-test, such as the square foot of lot area of the 15 nearest neighbors. We remove those insignificant variables, and regress on the remaining ones.

We also use cook’s distance to check if there is any significant outlier for the model we built. Fortunately there is none.

The R-squared of the resulting model is 77.44%, which is the best among all trials.

As the summary of linear regression model demonstrates, all of those variables are significant by t-test. Then we run a F-test on the general adequacy of the model, which rejects the null hypothesis that the linear relationships does not exist. Afterwards, we run F-tests on the difference of sum of squares due to regression between the full model and the partial model with one variable deleted. The results also show that all of those variables significantly contribute to the final model.

We also make the normal probability plot, which is a straight. We conclude the model does not have normality problem. We then plot the R-student residuals against the fitted values, which does not show any outlier. We then plot the R-student residuals against each of the non-dummy variables, which do not show any severe outliers problem.