Regression

Multi Linear Regression

Assignment 2

BP: Predict sales of the computer.

Do transformations for getting better predictions of profit and

Make a table containing R^2 value for each prepared model.

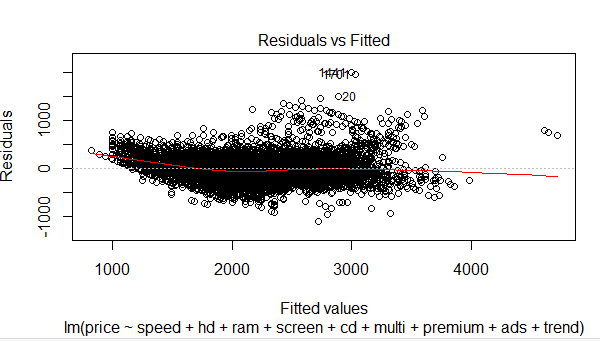
1. Now, from the given data we can observe that there is qualitative data after analysing the data we can eliminate the qualitative data as it doesn’t have any effect on the output .  
   This can be done by using dummies and dummy packages.
2. Now we have to create a multi linear model for this , by the formula :

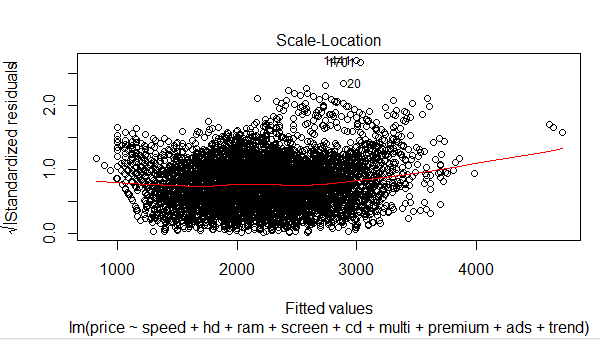
m1 <- lm(price ~ hd + ram + cd + premium + screen+ads+premium + trend,data=cd1)

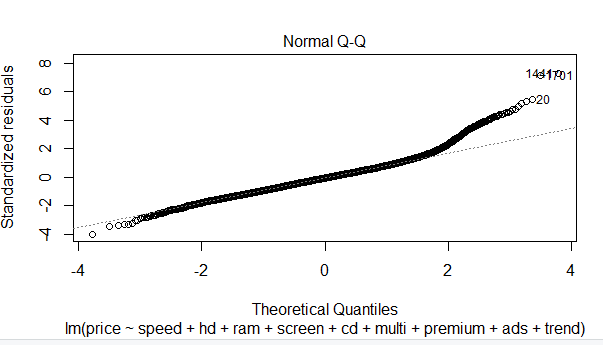
on using summary ( m1) we get,

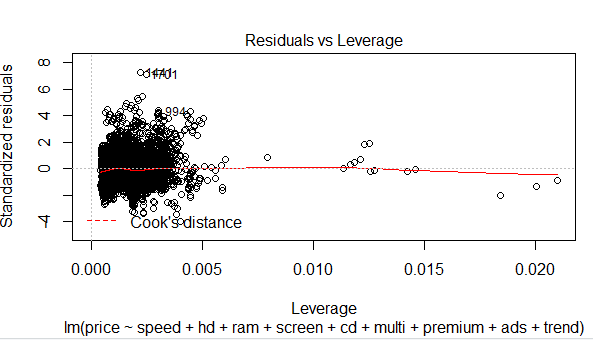
Now and value of R2= 0.7756

EDA of the data is as follows:





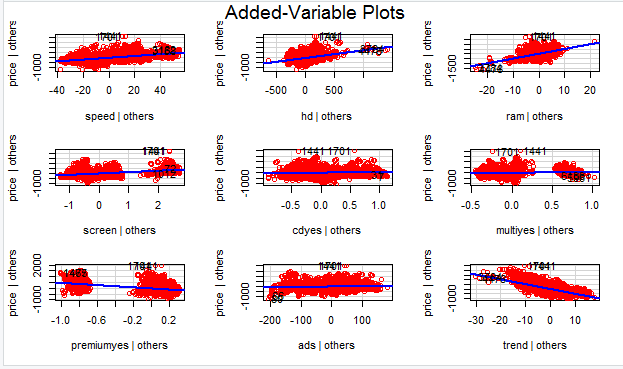




1. Now we have to try and build model , we can use variance inflation models,

If vif > 10 collinearity is good, but from the given values we can see that collinearity is on lower side , we cannot remove all, but can eliminate at least the least one.  
So let’s check which factor is effecting the correlation to bring it down.

Now on plotting added variable plots, we get as follows



From the following plots we can assume that , ads , multi , cd has been pulling down the value of correlation . But we cannot judge directly like that , so we use another method of regression called as Akaike information criterion model .  
  
on doing so we get as follows,

Df Sum of Sq RSS AIC

<none> 469770713 70273

- cd 1 3200767 472971480 70314

- multi 1 6291077 476061789 70354

- ads 1 12326753 482097466 70433

- hd 1 60143054 529913767 71025

- screen 1 70555708 540326421 71147

- premium 1 129290419 599061132 71793

- ram 1 156041652 625812364 72066

- speed 1 191824969 661595682 72414

- trend 1 513654981 983425694 74895

Call:

lm(formula = price ~ speed + hd + ram + screen + cd + multi +

premium + ads + trend, data = cd1[-1441, ])

103.9607 -509.6809 0.6543 -51.7588

According to AIC regression, it is suggesting to remove the data entry at the point 1441 which is trying to reduce the collinearity.

The value of R2 is = 0.7767.

Which is increased by a very small value.

To increase the value of R2 we can try some transformations.

The value before and after transformation are:

   
  
So we can conclude that after transformations, square root on both the axis is working good as we can see increase in R2 value up to 0.8007 which is pity decent.