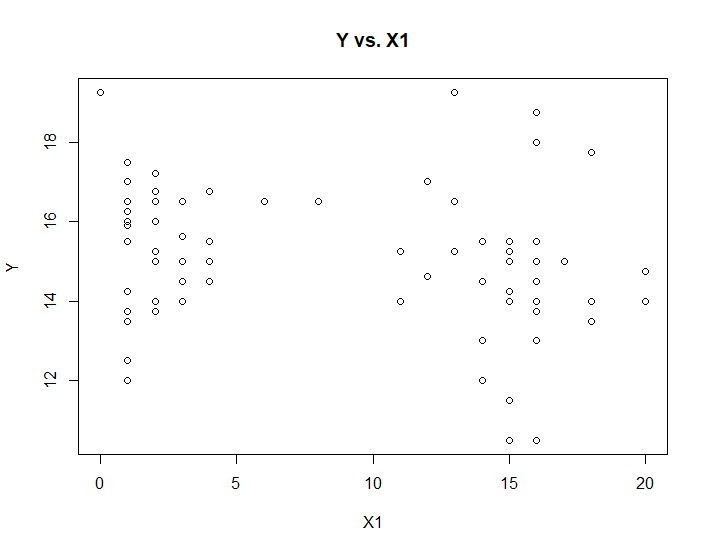
**2、**

(1)

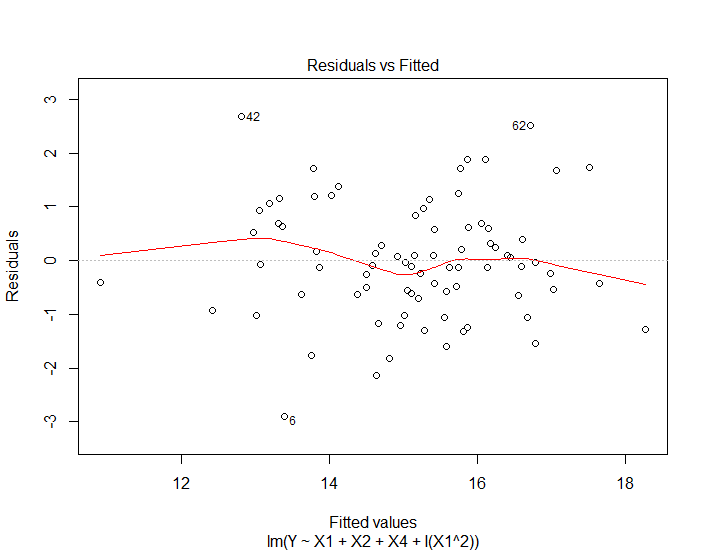


There is no obvious relationship between X1 and Y

(2)



Y = 10.19 - 0.1818\*X1 + 0.314\*X2 + 8.046\*e^-6\*X4 + 0.01415\*X1^2



There is no non-linearity and the variance is the same. So, it is a good fit.

(3)

In model 2 , the R^2 = 0.583, Ra^2 = 0.5667

In this model , the R^2 = 0.6131, Ra^2 = 0.5927

This model fits better than the model 2.

(4)

H0 :  H1 : 



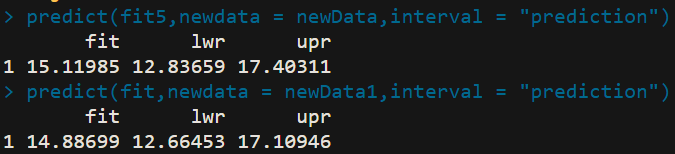
The null distribution is T distribution.

Reject condition : pValue < 0.05

pValue = 0.0174, under the significant of 0.05, we should reject the null hypothesis.

(5)

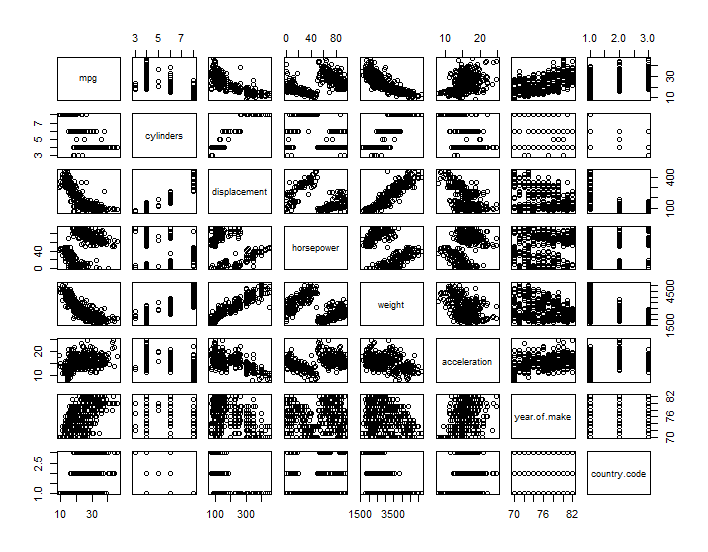
The prediction value is 14.88699



The prediction interval is smaller than the model 2.

**3、**

(1)



The cylinders variable, year of make and country.code variable are qualitative variables.

(2)

The cylinders variable, year of make and country.code variable are qualitative variables.

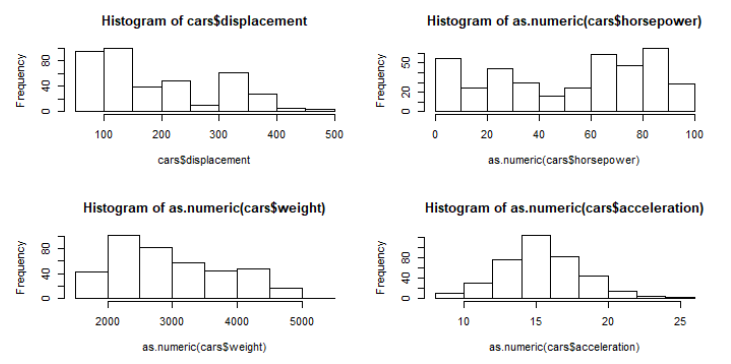
The others are quantitative variables.

(3)

remove the Na value.

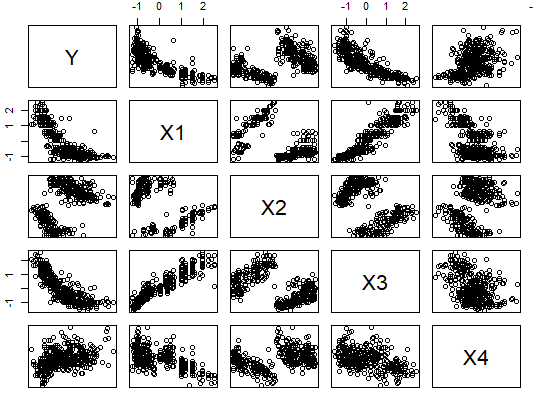
We should also transform those qualitative variables to indicator variables.

(4)



I think we need Z-Score transformation. Because we need to exclude the influence of units.

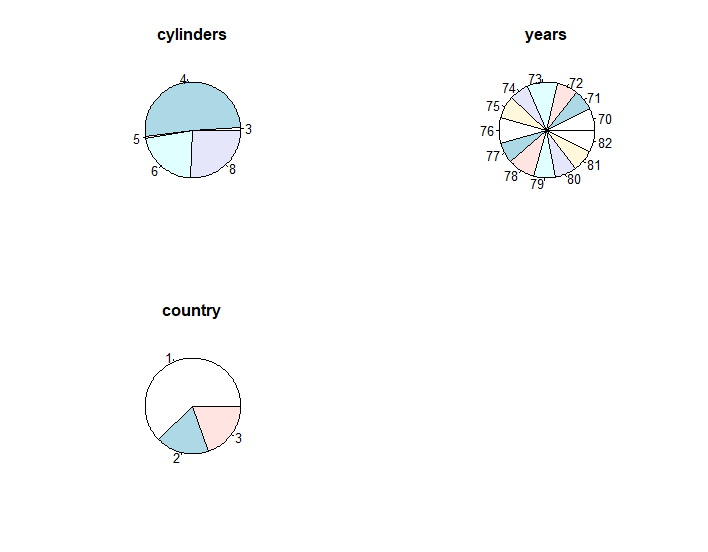
(5)

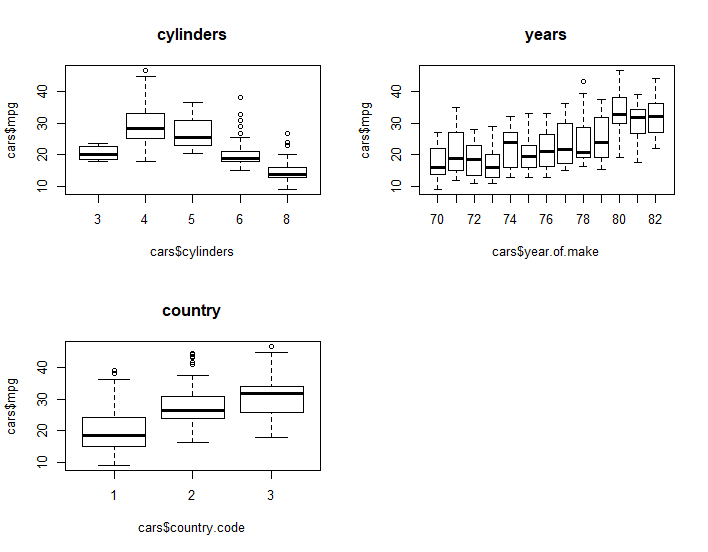


The displacement, horsepower and the weight are nonlinear relationship with response variable.

We can transform response variable .

(6)





The mpg is smaller if cylinders are larger. The mpg is larger if years are larger. The mpg is larger if the country code is larger.

(7)

**4、**

(1)



(2)



5、

(1)

There are no correct model among the models being considered. Because the key X variables are sin(x) and sin(2x). But the polynomial model dose not contain any key X variables.

(2)

Under the error variance of 0.5:

The in-sample variance of 1 model is : 0.5147214

The in-sample variance of 2 model is : 0.7689784

The in-sample variance of 3 model is : 1.001592

The in-sample variance of 5 model is : 1.472747

The in-sample variance of 7 model is : 1.987825

The in-sample variance of 9 model is : 2.51138

The variance will be changed if the error variance change.

(3)

The model with high polynomial will have the lower bias.

The bias will not change if the error variance change.

(4)

The model variance is the dominant component in the (in-sample) mean-squared-estimation-error if the polynomial is high.

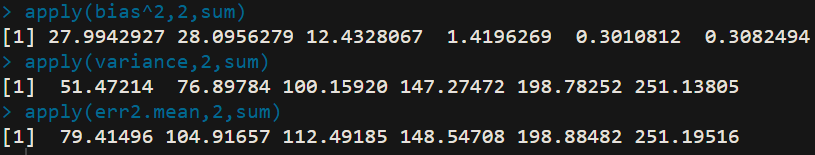
The model bias is the dominant component in the (in-sample) mean-squared-estimation-error if the polynomial is low.

Because there are many nuisance X variables in the model with high polynomial, the variance will be larger and the bias will be smaller.

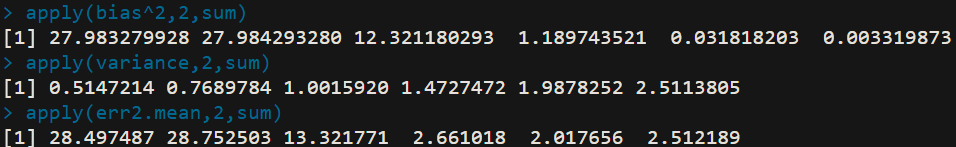
There is no nuisance X variables in the model with low polynomial, the variance will be smaller and the bias will be larger.

(5)

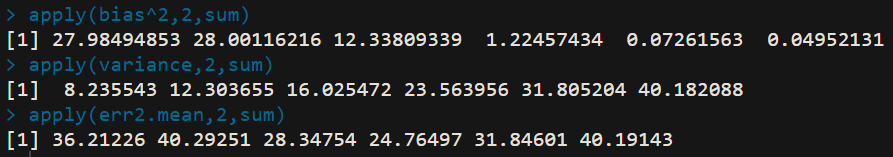
Model under error variance is 5:



Model under error variance is 0.5:



Model under error variance is 2:



It depends on the value of error variance.

If the error variance is large, we should choose the lower polynomial model .

If the error variance is small, we should choose the higher polynomial model.

If the error variance is in the medium value, we should choose the medium polynomial model.

Because the mean-squared-estimation-error is equal with variance plus squared bias. The bias of each model are not change but variance will change hugely.

(6)

Model under error variance is 0.5:



Model under error variance is 2:



Model under error variance is 5:



E(SSE) will be smaller the high polynomial model if the error variance is same.

E(SSE) will be larger if the error variance is larger and the model is the same. Because mean-squared-estimation-error is equal with variance plus squared bias. the variance of this model will larger but the bias will not change significantly.So, the E(SSE) will be larger.