Module 5 Assignment

Umakanth Sai Balguri - (U01403700)



Screenshot from Canvas

# Preprocessing

### Loading Library files

rm(list=ls())  
library(rmarkdown)  
library(rio)  
library(moments)

### Loading the file into R

cars=import("6304 Module 5 Assignment Data.xlsx",  
 sheet="Sheet 1")  
colnames(cars)=tolower(make.names(colnames(cars)))

### Selecting Random Samples based on Seed number = U-number and taking a sample of 250 observations meeting the specified condition

cars.sub = subset(cars, make == "cadillac" & cylinders %in% c(6,8) & year >2005 & year <2012)  
set.seed(01403700)  
sample.cars=cars.sub[sample(1:nrow(cars.sub),250,replace=FALSE),]  
sample.cars$year=as.factor(sample.cars$year)  
attach(sample.cars)

# Analysis

### Response to Q1,Q2

cars.out=lm(price~odometer+year+cylinders,data=sample.cars)  
summary(cars.out)

##   
## Call:  
## lm(formula = price ~ odometer + year + cylinders, data = sample.cars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10252.7 -1986.3 540.7 2010.6 23595.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.366e+03 2.366e+03 -3.959 9.92e-05 \*\*\*  
## odometer -2.647e-02 7.193e-03 -3.680 0.000287 \*\*\*  
## year2007 3.891e+03 8.987e+02 4.330 2.18e-05 \*\*\*  
## year2008 3.002e+03 9.430e+02 3.184 0.001644 \*\*   
## year2009 4.699e+03 1.289e+03 3.646 0.000327 \*\*\*  
## year2010 6.312e+03 9.568e+02 6.597 2.63e-10 \*\*\*  
## year2011 7.274e+03 9.508e+02 7.650 4.74e-13 \*\*\*  
## cylinders 2.673e+03 2.977e+02 8.977 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4214 on 242 degrees of freedom  
## Multiple R-squared: 0.3613, Adjusted R-squared: 0.3428   
## F-statistic: 19.56 on 7 and 242 DF, p-value: < 2.2e-16

1. Based on the graphical analysis, the independent variables were treated as factors. However, this did not change the R\_squared value. Hence the model is left as is. Although this is not a good fit due to the low R-squared value, it is being considered. (Note: As the condition of car, an important dependent factor is not being considered, I feel the confidence is low.)
2. Coeffients B0,B1,B2,B3 as as per the COffeicent outputs with esitmated values.

B0=INtercept= -9.366e+0.3 P VALUE= 9.92e-05

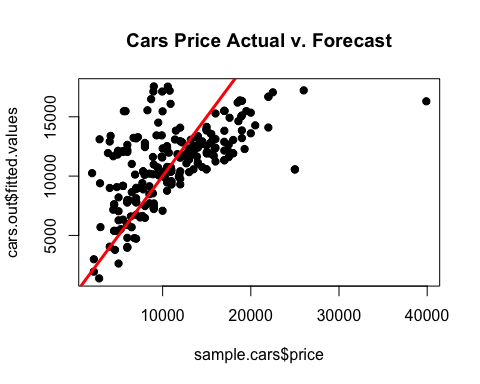
B1=Odometer= -2.647e-0.2 Pvalue= 0.000287

b2=year = multiole values observed P Value= Multiple Values

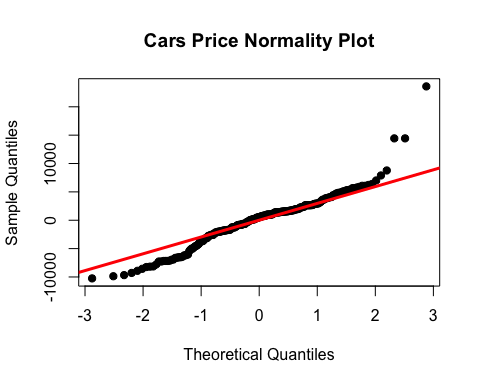
b3-cylinders = 2.673e+0.3 P Value= < 2e-16

### Response to Q3, Q4

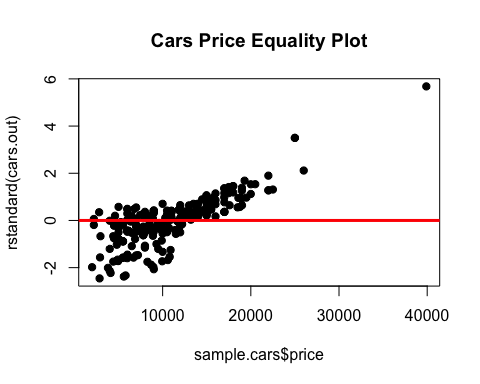
plot(sample.cars$price,cars.out$fitted.values,pch=19,main="Cars Price Actual v. Forecast")  
abline(0,1,lwd=3,col="red")



qqnorm(cars.out$residuals,pch=19,main="Cars Price Normality Plot")  
qqline(cars.out$residuals,lwd=3,col="red")



plot(sample.cars$price,rstandard(cars.out),pch=19, main="Cars Price Equality Plot")  
abline(0,0,col="red",lwd=3)

 3) The model theoretically insnt a good fit for the data because Multiple R-squared is 0.3613. However, this may be practically accepted. This is further analysed from the LINE analysis below.

4)Linearity= It observes Linearity. (However deviation is marginally high) Normality= For the most part the model is normal Equality= except for 2 observations, the model follows normality

Overall, the model is marginally a good fit with certain amount of deivation as observed from the linemar model analysis. ### Response to Q5

pdt=data.frame(odometer=175757, year=2011, cylinders=8)  
pdt$year=as.factor(pdt$year)  
predict(cars.out,pdt,interval="predict")

## fit lwr upr  
## 1 14638.36 6168.631 23108.08

predict(cars.out,pdt,interval="confidence")

## fit lwr upr  
## 1 14638.36 12952.69 16324.03

As per the prediction the ideal price would be 14638.36 with lower and uppper values of 6168.631 and 23108.08 respectively. For the 95% confidence interval, lower and uppper values of 12952.69 and 16324.03 respectively.

No, Pricing advice will not be usable and not accurate because there is less than 50% confidence on our model as it accounts for too many points out of standard.Moreover, he mentions the condition to be excellent, which is not taken into account by the model.