RegModProject

Nixon Patel

Sunday, November 23, 2014

Executive Summary

In this report for Motor Trend Magazine, the regression analysis shows changes of Miles per Gallon (MPG) with different variables of cars. In the simplest model, mpg over transmission types, shows that the manual transmission is 7.25 mpg better than automatic transmission. Taking cyliner, displacement, weight and horsepower into account, the multivariate regression model indicates that the manual transmission is 1.81 mpg better than the automatic transmission while the goodness of fit has been reached 86%.

Exploring Dataset

Using density plot and pair() function, those characteristics and definition of variables are examined. For more details, see appendix section.

Finding Nesessary Variables

First, convert numeric values to factor values, and then compare mpg v.s. all other variables The p-values shows cyl, disp and wt are significant predictors for mpg as outcome.

```
setwd("C:/Users/Nixon/Documents/GitHub/nixonpatel/RegressionModels/Regression
Models")
mtcars$am<-as.factor(mtcars$am)     #transmission type
mtcars$cyl<-as.factor(mtcars$cyl)
mtcars$gear<-as.factor(mtcars$gear)
mtcars$carb<-as.factor(mtcars$carb)
mtcars$vs<-as.factor(mtcars$vs)
fit_all<-lm(mpg~.,data=mtcars) #build model mpg over others
summary(aov(fit_all))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          2  824.8   412.4  51.377 1.94e-07 ***
## disp         1   57.6    57.6   7.181   0.0171 *
## hp           1   18.5    18.5   2.305   0.1497
## drat         1   11.9    11.9   1.484   0.2419
## wt           1   55.8    55.8   6.950   0.0187 *
## qsec         1    1.5     1.5   0.190   0.6692
## vs           1    0.3     0.3   0.038   0.8488
## am           1   16.6    16.6   2.064   0.1714
## gear         2    5.0     2.5   0.313   0.7361
## carb         5   13.6     2.7   0.339   0.8814
## Residuals   15  120.4     8.0
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## cyl          2    825     412   51.38 1.9e-07 ***
## disp         1     58      58    7.18   0.017 *
## hp           1     19      19    2.31   0.150
## drat         1     12      12    1.48   0.242
## wt           1     56      56    6.95   0.019 *
## qsec         1      2       2    0.19   0.669
## vs           1      0       0    0.04   0.849
## am           1     17      17    2.06   0.171
## gear         2      5       3    0.31   0.736
## carb         5     14       3    0.34   0.881
## Residuals   15    120       8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.1.2
```

```
cv<-vif(fit_all) #calculate variance inflation
head(cv[order(cv[,3],decreasing=T),],4) #sort the result in descending order
```

```
##           GVIF Df GVIF^(1/(2*Df))
## disp  60.36569  1        7.769536
## hp    28.21958  1        5.312210
## wt    23.83083  1        4.881683
## cyl  128.12096  2        3.364380
```

```
##        GVIF Df GVIF^(1/(2*Df))
## disp  60.37  1           7.770
## hp    28.22  1           5.312
## wt    23.83  1           4.882
## cyl  128.12  2           3.364
```

The result of vif() shows the cylinder, displacement, horsepower and weight are highly corelated with each other.

Multivariate Models

The next step is to build multivariate models by adding above variables on the single variable model. From the p-values and variability inflation, cyl,disp, wt and hp columns are selected to be a part of multivariate regression model. Below R codes show adding each variable one by one. The anova() function shows the degree of freedom and p-values of each model.

```
fit1<-lm(mpg~am,data=mtcars)
fit2<-lm(mpg~am+cyl,data=mtcars)
fit3<-lm(mpg~am+cyl+disp,data=mtcars)
fit4<-lm(mpg~am+cyl+disp+wt,data=mtcars)
```

```r
fit5<-lm(mpg~am+cyl+disp+wt+hp,data=mtcars)
anova(fit1,fit2,fit3,fit4,fit5)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl + disp
## Model 4: mpg ~ am + cyl + disp + wt
## Model 5: mpg ~ am + cyl + disp + wt + hp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     28 264.50  2    456.40 37.9300 2.678e-08 ***
## 3     27 230.46  1     34.04  5.6572  0.025339 *
## 4     26 182.87  1     47.59  7.9102  0.009429 **
## 5     25 150.41  1     32.46  5.3954  0.028621 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl + disp
## Model 4: mpg ~ am + cyl + disp + wt
## Model 5: mpg ~ am + cyl + disp + wt + hp
##   Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1     30 721
## 2     28 264  2       456 37.93 2.7e-08 ***
## 3     27 230  1        34  5.66  0.0253 *
## 4     26 183  1        48  7.91  0.0094 **
## 5     25 150  1        32  5.40  0.0286 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit1)$coefficients[1:2,] #single variable model
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am1          7.244939   1.764422  4.106127 2.850207e-04
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.134e-15
## am1            7.245      1.764   4.106 2.850e-04
```

```r
summary(fit5)$coefficients[1:2,] #multivariate model
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.864276   2.695416 12.563656 2.668321e-12
## am1          1.806099   1.421079  1.270935 2.154510e-01
```

```
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   33.864      2.695  12.564 2.668e-12
## am1            1.806      1.421   1.271 2.155e-01
```

In the multivariate model, the manual transmission is 1.81 mpg better than automatic transmission. The R-squared of the multivariate model, which indicates how good the model fits data, increased from 36% to 86%. Additionally, the residual plot of fit5 is shown in Appendix section.

```
c(summary(fit1)$r.squared, summary(fit5)$r.squared)
```

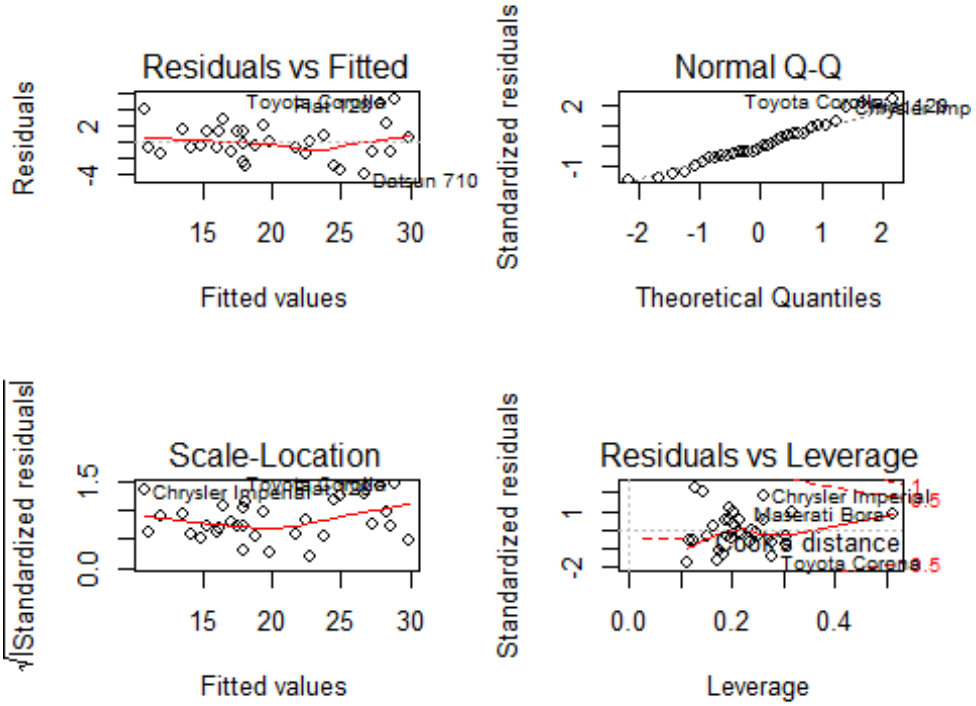
```
## [1] 0.3597989 0.8664276
```

```
## [1] 0.3598 0.8664
```

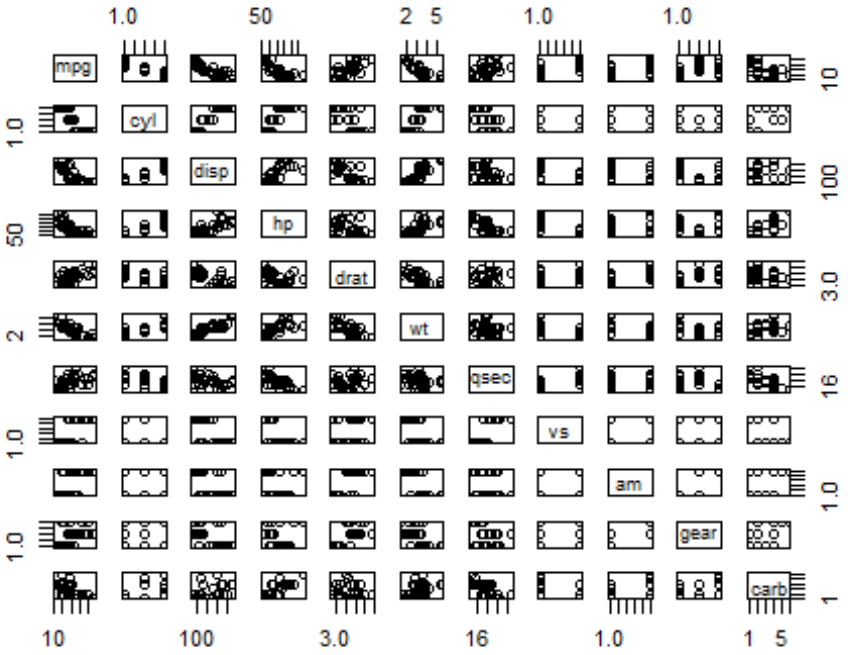
Appendix

Red: Automatic Transmission / Green: Manual Transmission Residual Plots

```
par(mfrow=c(2,2))
plot(fit5)
```




```
##Plotting Pairs
pairs(mtcars)
```


mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb