Motor Trend Report

HPSUN

2017年12月3日

## Excutive Summary

For this analysis, We are using the mtcar dataset for analysis.For details of this dataset, your can see this link: <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html>  
We used linear regression model to evaluate the factor affact miles per gallon(mpg). And found that mpg prediction mainly depends on wt,cyl,disp of the car. The manual transmission may increase mpg a little (0.14), but it is not significant.So it’s hard to say that auto transmission is better or not.

## Data Exploratory

library(ggplot2)  
data("mtcars")  
#summary(mtcars)  
m <- mtcars  
m$cyl <- as.factor(m$cyl)  
m$vs <- as.factor(m$vs)  
m$am <- as.factor(m$am)  
m$gear <- as.factor(m$gear)  
m$carb <- as.factor(m$carb)

## Regression Model

From figure 1, It seems all the features are correlated to mpg. So, first, we add all features for regression model.

fit1 <- lm(mpg~.,data=m)  
anova(fit1)

## Analysis of Variance Table  
##   
## Response: mpg  
## Df Sum Sq Mean Sq F value Pr(>F)   
## cyl 2 824.78 412.39 51.3766 1.943e-07 \*\*\*  
## disp 1 57.64 57.64 7.1813 0.01714 \*   
## hp 1 18.50 18.50 2.3050 0.14975   
## drat 1 11.91 11.91 1.4843 0.24191   
## wt 1 55.79 55.79 6.9500 0.01870 \*   
## qsec 1 1.52 1.52 0.1899 0.66918   
## vs 1 0.30 0.30 0.0376 0.84878   
## am 1 16.57 16.57 2.0639 0.17135   
## gear 2 5.02 2.51 0.3128 0.73606   
## carb 5 13.60 2.72 0.3388 0.88144   
## Residuals 15 120.40 8.03   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

From this model, we found cyl, disp, wt are signifcant variables have effect on mpg.So next, we use only disp, wt, cyl and am to predict mpg.

fit2 <- lm(mpg~disp+wt+cyl+am,data=m)  
summary(fit2)

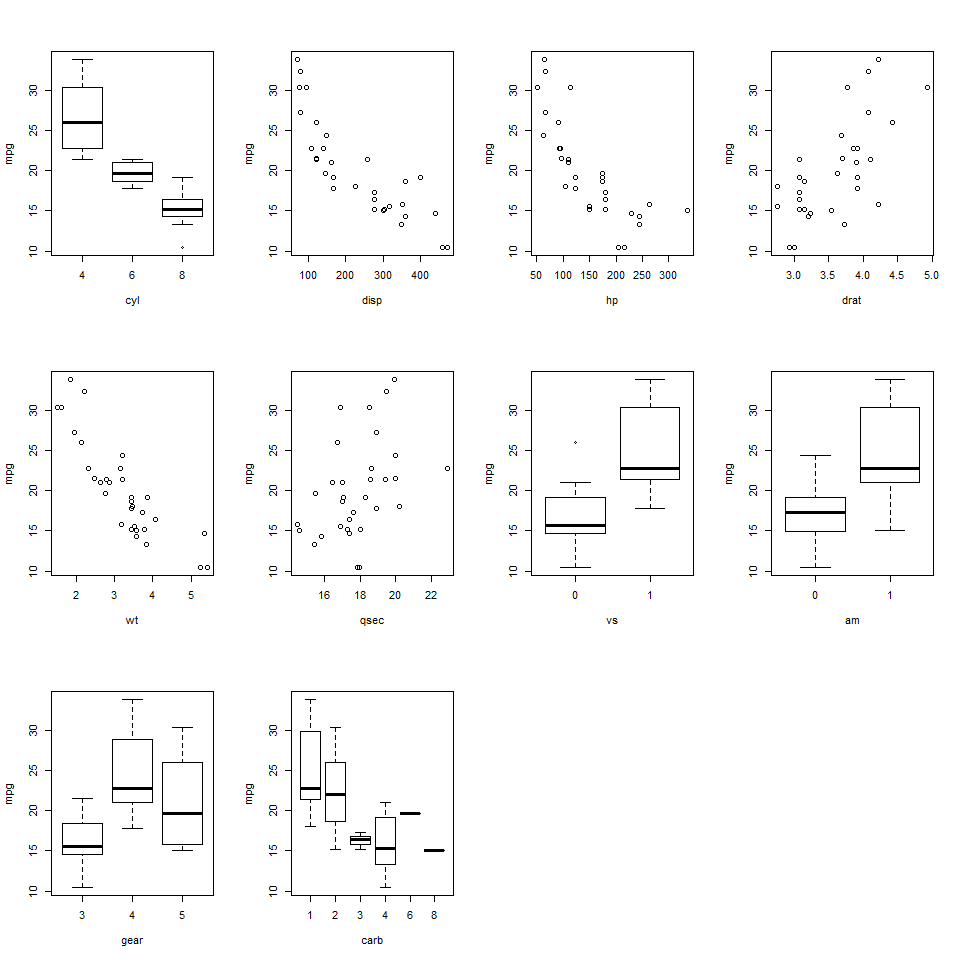
##   
## Call:  
## lm(formula = mpg ~ disp + wt + cyl + am, data = m)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.5029 -1.2829 -0.4825 1.4954 5.7889   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 33.816067 2.914272 11.604 8.79e-12 \*\*\*  
## disp 0.001632 0.013757 0.119 0.90647   
## wt -3.249176 1.249098 -2.601 0.01513 \*   
## cyl6 -4.304782 1.492355 -2.885 0.00777 \*\*   
## cyl8 -6.318406 2.647658 -2.386 0.02458 \*   
## am1 0.141212 1.326751 0.106 0.91605   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.652 on 26 degrees of freedom  
## Multiple R-squared: 0.8376, Adjusted R-squared: 0.8064   
## F-statistic: 26.82 on 5 and 26 DF, p-value: 1.73e-09

This model explains 83.76% of the variance. From this we see the manual transmisson increase 0.141212 mpg,but it’s not significant.The most important factors affect mpg is still cyl and wt.

## Appendix

### Figure 1: Exploratory plot

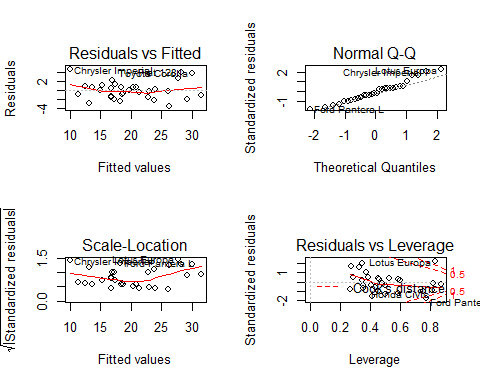
par(mfrow = c(3,4))  
plot(mpg~.,data=m)



### Figure 2 : fit1 Diagnostic

par(mfrow = c(2, 2))  
plot(fit1)

## Warning: not plotting observations with leverage one:  
## 30, 31  
  
## Warning: not plotting observations with leverage one:  
## 30, 31



### Figure 3 : fit2 Diagnostic

par(mfrow = c(2, 2))  
plot(fit2)

