Regression Models Project

Lucian Pop

Sunday, September 27, 2015

Executive Summary

In this paper we will analyze mtcars data set and determine the relation between mpg and other significant variables. This data is extracted from the 1974 Motor Trend magazine, and denotes fuel consumption and other aspects of automobile performance for 32 73–74 car models.

Upon analysis was found that Manual transmission better Automatic transmission by about 2 mpg given the same weight and power.

Exploratory Data Analysis

At first we plotted mpg vs am (see Annex:Figure 1) and observed that overall automatic transmission achieve less mpg than manual transmission. However there might be other factors involved. We further analyze mean mpg for "automatic"" vs "manual""

```
round(tapply(mtcars$mpg, mtcars$am, mean), 2)
```

```
## Automatic    Manual
##     17.15     24.39
```

We than proceed to pair plot variables and observed that other factors influence mnps: hp, wt, disp, cyl. Upon further analysis "cyl" can be discarded since "disp" is more important. going further we will analyze models based on mpg, mpg+wt, mpg+wt+disp and mps+wt+hp.

Regression Models

```
model_full        <- lm(mpg ~ ., mtcars)
model_am          <- lm(mpg ~ am , mtcars)
model_am_wt       <- lm(mpg ~ am + wt, mtcars)
model_am_wt_disp <- lm(mpg ~ am + wt + disp, mtcars)
model_am_wt_hp    <- lm(mpg ~ am + wt + hp, mtcars)
```

We study all models coeficients and summary values.

```
summary(model_am)$coef
summary(model_am_wt)$coef
summary(model_am_wt_disp)$coef
summary(model_am_wt_hp)$coef
```

Then we perform an Anova Test.

```r
anova(model_am, model_am_wt, model_am_wt_disp,model_am_wt_hp)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + disp
## Model 4: mpg ~ am + wt + hp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 50.2610 1.032e-07 ***
## 3     28 246.56  1     31.76  3.6072   0.06788 .
## 4     28 180.29  0     66.27
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
confint(model_am_wt_hp)
```

```
##                   2.5 %      97.5 %
## (Intercept) 28.58963286 39.41611738
## amManual    -0.73575874  4.90317900
## wt          -4.73232353 -1.02482730
## hp          -0.05715454 -0.01780291
```

Ultimately we choose the model "mpg ~ am + wt + hp" which has the highest Adjusted R-squared: 0.8227 and Residual standard error: 2.538 on 28 degrees of freedom This means that the model can explain about 82% of the variance of the MPG. We believe this is the best option from studied models.

```r
summary(model_am_wt_hp)$coef
```

```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13
## amManual     2.08371013 1.376420152  1.513862 1.412682e-01
## wt          -2.87857541 0.904970538 -3.180850 3.574031e-03
## hp          -0.03747873 0.009605422 -3.901830 5.464023e-04
```

Based on our model given the same "wt" (weight lb/1000) and "hp" (power) a car with manual transmission will have an increase "mpg" by 2.083.

Residual Analysis and Diagnostics

To verify how much an observation has effected the estimate of a regression coefficient we calculate dfbetas sums and found it to be "0".

```r
sum((abs(dfbetas(model_am_wt_hp)))>1)
```

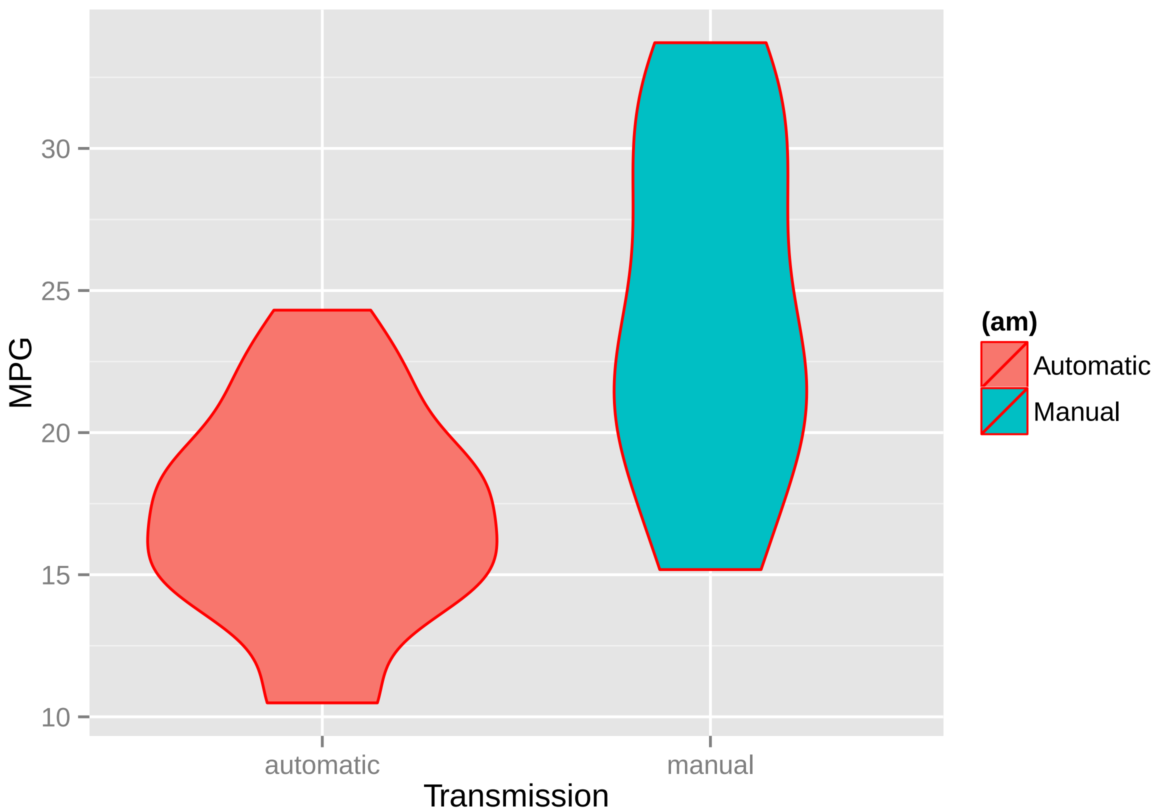
```
## [1] 0
```

From the residual plot (see Annex: Fig3) we observe that "Residuals vs. Fitted"" plot shows no consistent pattern, "Normal Q-Q"" plot indicates that the residuals are normally distributed, "Scale-Location"" plot confirms the constant variance assumption and "Residuals vs. Leverage" argues that no outliers are present.

Appendix: Figures

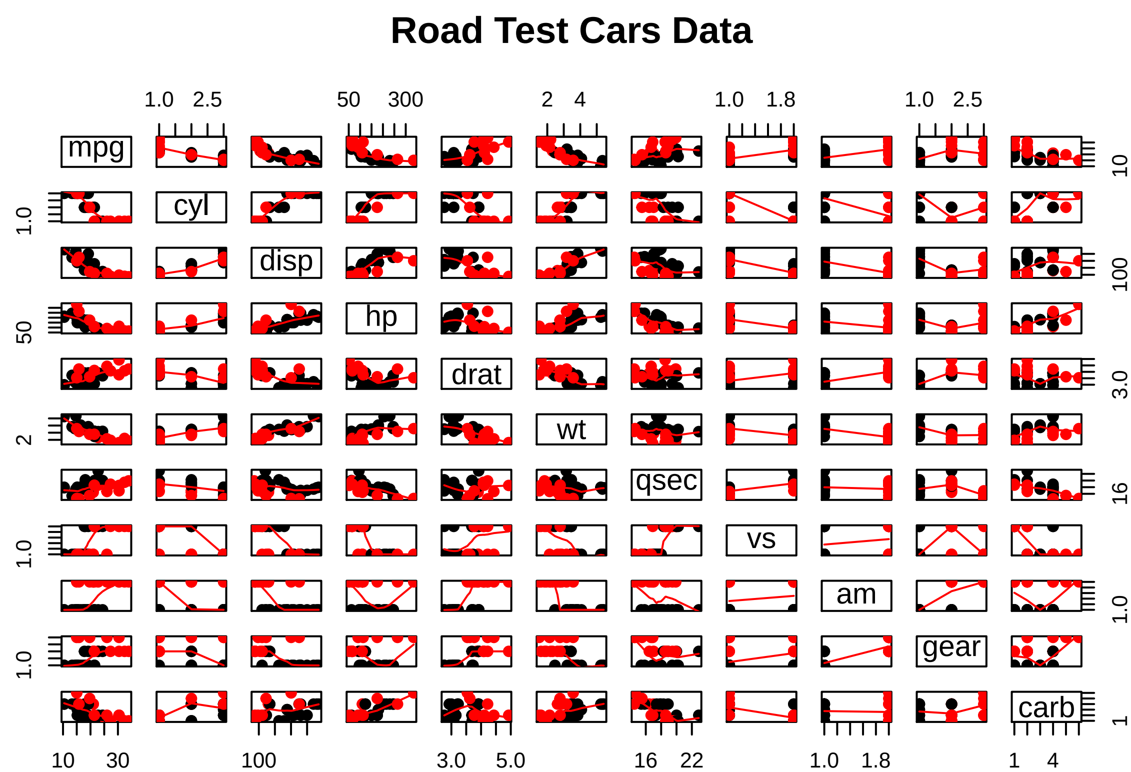
1. Plot of MPG vs. Transmission

```
ggplot(mtcars, aes(y=mpg, x=factor(am, labels = c("automatic", "manual")), fill=(am)))+
  geom_violin(colour="red", size=0.5)+
  xlab("Transmission") + ylab("MPG")
```



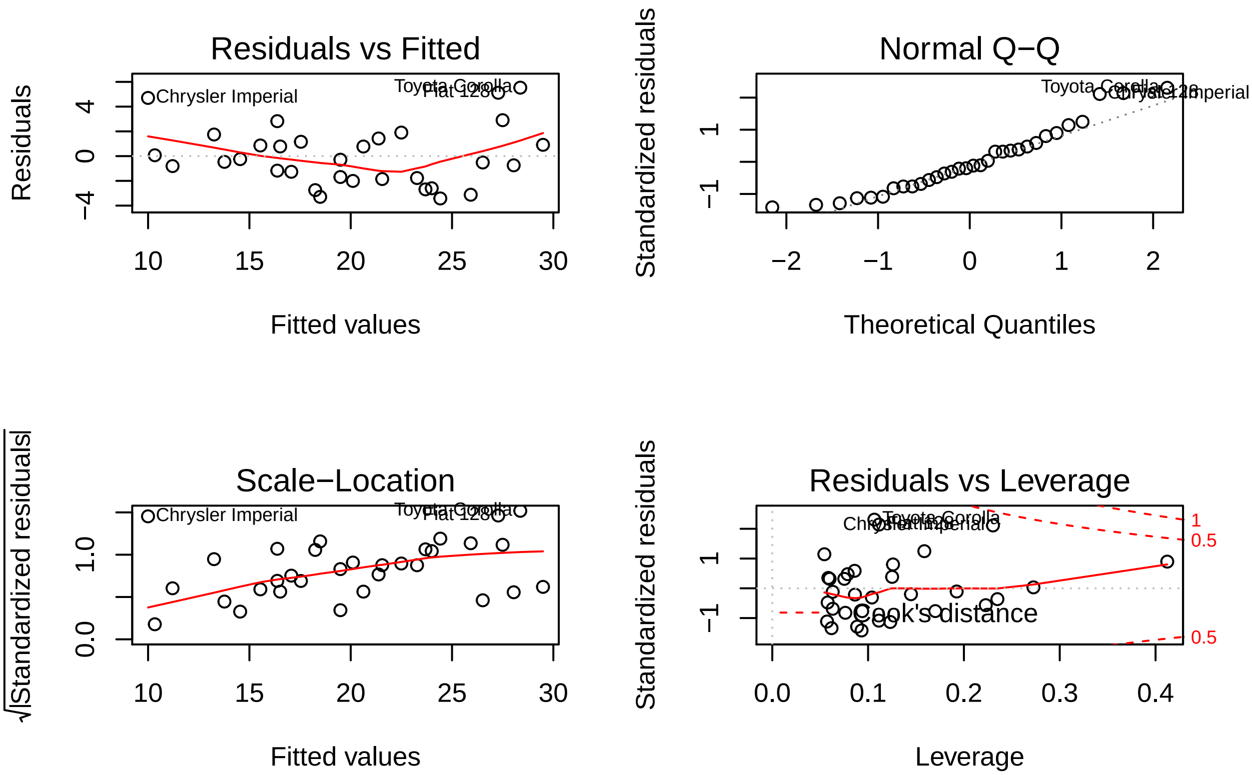
2. Pair Graph of Motor Trend Road Test Cars Data

```
pairs(mtcars, panel=panel.smooth, main='Road Test Cars Data', col=mtcars$am, pch=19)
```



3. Residual Plots

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


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
1
0.5
0.5