Regression Models Project: Motor Trend Data Analysis Report

Anuar Imanbayev

January 30, 2016

Executive Summary

In this report, we analyze 'mtcars' dataset of a collection of cars from Motor Trend, a magazine about the automobile industry. We are interested in exploring the relationship between the outcome variable, miles per gallon **MPG**, and a set of variables. Particularly, we are interested in the transmmission type variable, **automatic** (am = 0) vs. **manual** (am = 1), in order to answer the following questions:

"Is an automatic or manual transmission better for MPG"

"Quantify the MPG difference between automatic and manual transmissions"

Next, we employed both multivariate and stepwise regression models with AIC criteria to derive an optimal model from the starting multivariate model. Results show that MPG performance difference is 1.81 +/- 1.396 MPG higher for cars with manual transmission than those with automatic transmission. In addition to transmission type, the results indicate that cylinders, horsepower, and weight are additional signficantly influencing factors affecting the MPG.

Exploratory Data Analysis

We load the data set `mtcars` and change some variables from `numeric` class to `factor` class.

```
library(ggplot2)
data(mtcars)
mtcars$am <- factor(mtcars$am,labels=c('Automatic','Manual'))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
attach(mtcars)
dim(mtcars)
```

```
## [1] 32 11
```

The pairwise scatter plot between all variables is shown in **Appendix #1**. According to the boxplot shown in **Appendix #2**, we see that manual transmission yields higher values of MPG in general. Thus, it seems that manual transmission is better than automatic transmission for MPG.

Statistical Inference

At this step, we state that the null hypothesis as the MPG of the automatic and manual transmissions are from the same population (assuming the MPG has a normal distribution). We use the two sample T-test to show it.

```
result <- t.test(mtcars$mpg~mtcars$am)
result
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```

Since the p-value is 0.00137, at 5% significance level of $p<0.05$, we reject our null hypothesis that there is no difference in MPG by transmission type. We further conclude that manual transmission is better than automatic transmission for MPG, with assumption that all other conditions remain unchanged.

Regression Analysis

Next, we try to quantify the MPG difference between transmission types. We perform a multivariate linear regression with all variables.

```
multiModel = lm(data = mtcars, mpg~.)
```

We utilize the stepwise regression where predictors are chosen by criterion comparison such as AIC.

```
optModel = step(multiModel, direction = "both") # hidden
```

Summary of stepwise-derived Optimal Model

```
summary(optModel)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

With the model explaning 86.59% of the variance, the results suggest that the best model includes cyl6, cyl8, hp, wt, and amManual variables. Main result is that manual transmission is 1.81 mpg better than automatic transmission. The rest of the variables all changed negatively with mpg, including cylinders (-3.03 miles and -2.16 miles for cyl6 and cyl8 respectively), horsepower (-0.03 miles), weight (-2.5 miles for every 1,000lb). In other words, cylinder, horsepower, and weight performance seem worse with manual transmission and conversely better with automatic transmission.

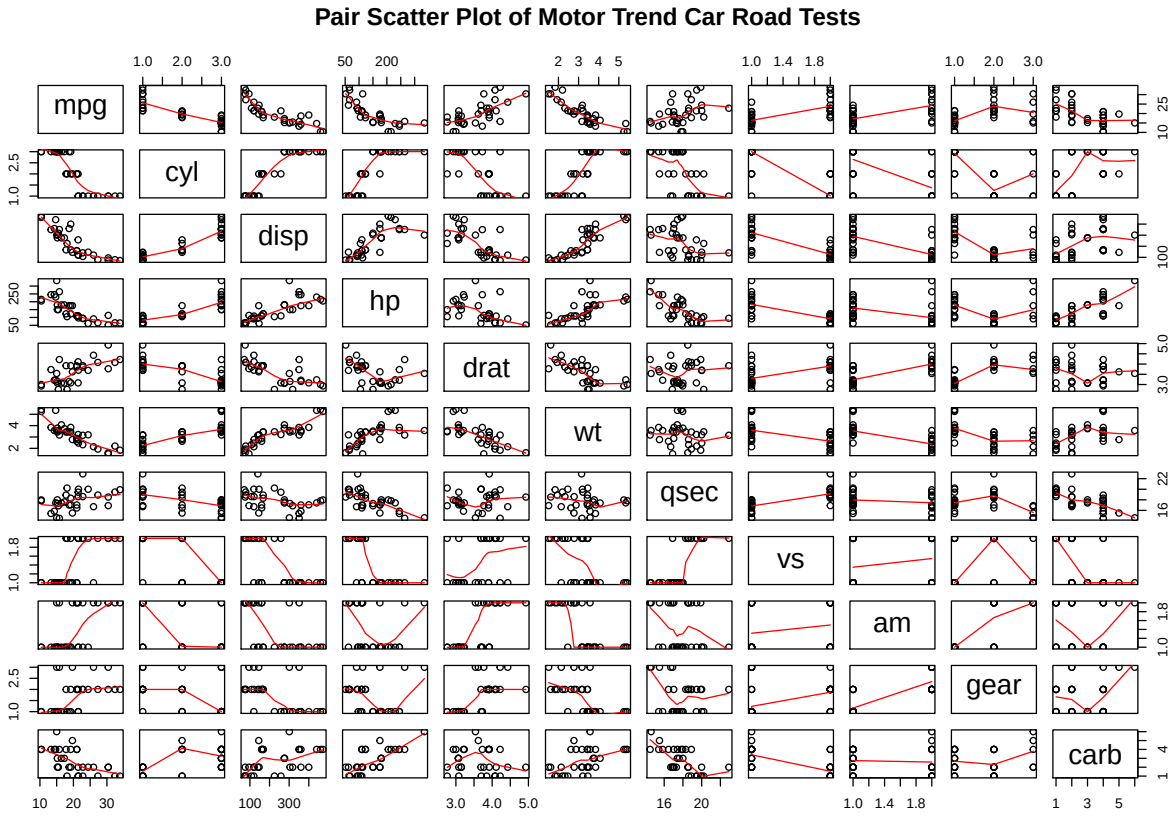
Residual Analysis and Diagnostics

Residual plots seems to be randomly scattered but some transformation may be needed to establish stronger linearity **Appendix #3**. According to the residual plots, we can verify the following underlying assumptions:

Appendix: Figures

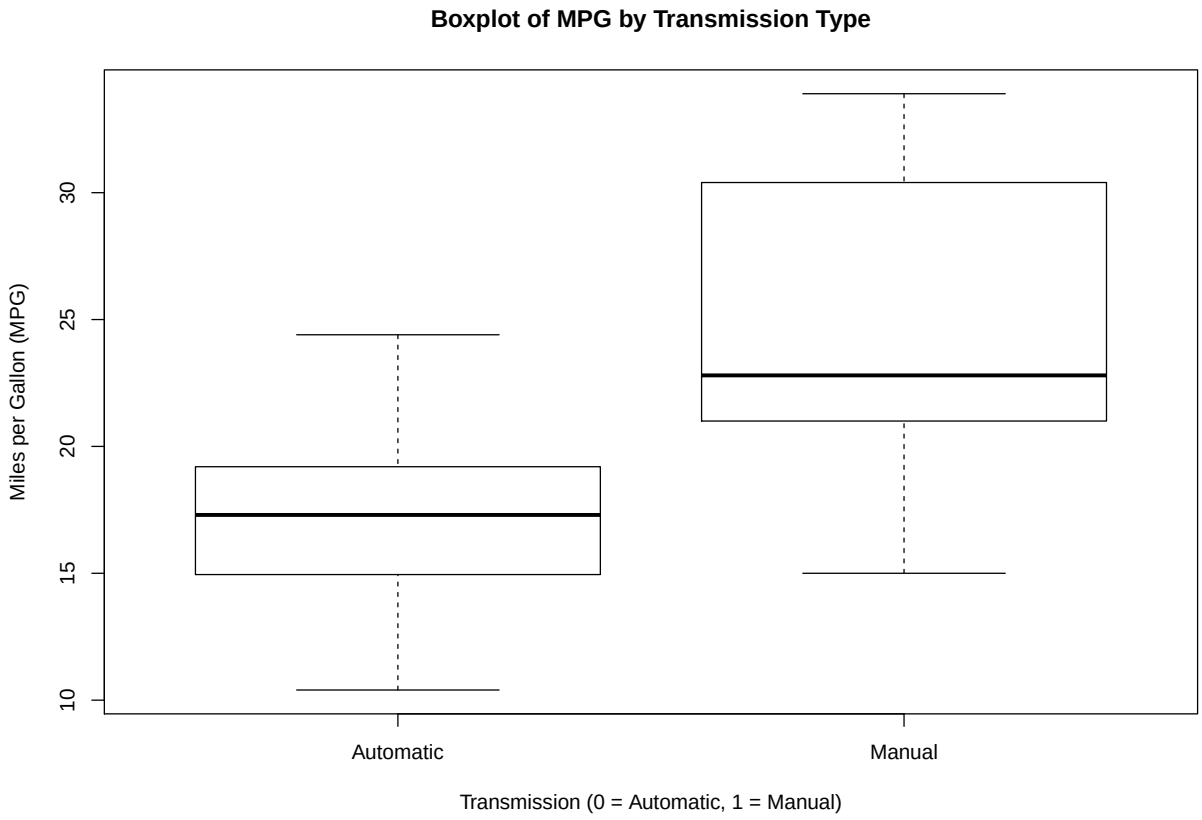
1. Pairwise Scatter Plot ofMotor Trend Car Road Tests

```
pairs(mtcars, panel=panel.smooth, main="Pair Scatter Plot of Motor Trend Car Road Tests")
```



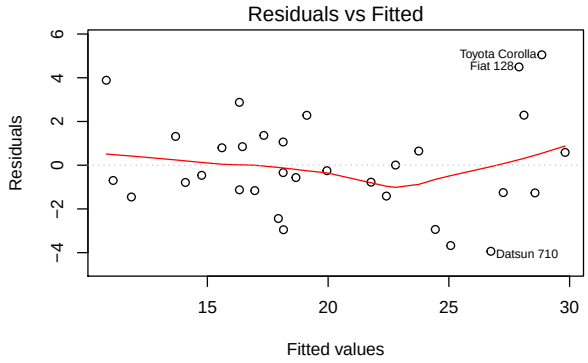
2. Boxplot of MPG by Transmission Type

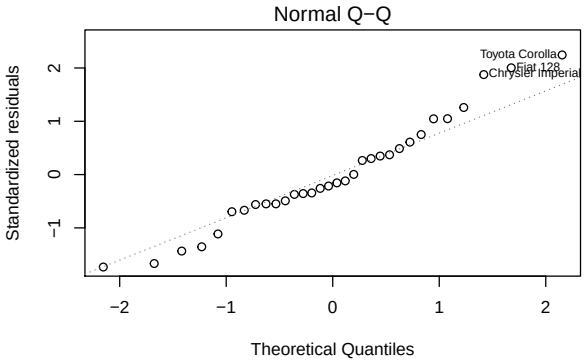
```
boxplot(mpg ~ am, data=mtcars,
        xlab="Transmission (0 = Automatic, 1 = Manual)",
        ylab="Miles per Gallon (MPG)",
        main="Boxplot of MPG by Transmission Type")
```

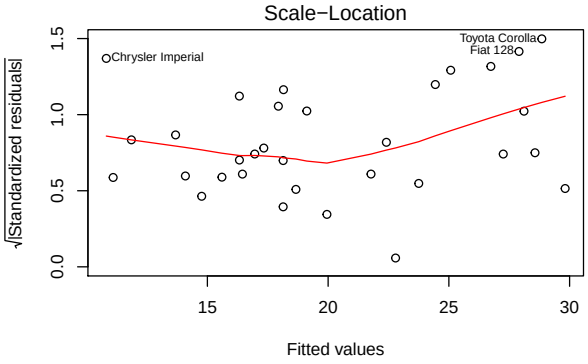


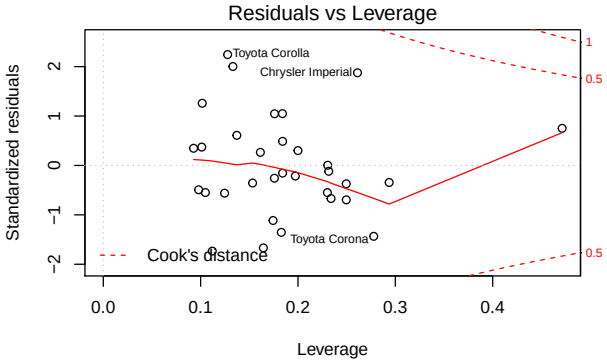
3. Residual Plots

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


Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710


Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Chrysler Imperial


Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128


Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
1
0.5
0.5