Regession

Yuani

28 September 2015

Executive Summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

i. “Is an automatic or manual transmission better for MPG”
ii. “Quantify the MPG difference between automatic and manual transmissions”

To investigate the above questions,we look to build a linear regression model for greater understanding.

Exploring the Data Set

```
data("mtcars")
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

```
library(ggplot2)
```

We first start by exploring the data sets using mpg & am vairable to identify any top levels trends.

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
plot1<-ggplot(mtcars, aes(x = am, y = mpg)) +geom_boxplot()
```

Based on the box plot in Appendix, we see that Manual transmission cars are more likely to better mpg perfornace compared to automatic cars.Hoever, there seems to be greate variability among the performance of Manual transmission cars.

Simple Linear Regression

We start with a simple linear regression in an attempt to quantify the effect of transmission type on mpg.

```
model1<-lm(mpg~am, data=mtcars)
summary(model1)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## amManual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

On average, manunal transmission cars run with 7.245 mpgs more than automatic transmission.

The R^2 of model1 however is 0.3598 indicating that the simple linear model only explains for the 35.98% of the variances.

We look to expand further on the linear model to better explain the different factors that affect the mpg for cars.

Multilinear Regression Analysis

```
stepmodel = step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
summary(stepmodel)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Using the stepmodel, we found that wt and qsec are key variables in explaining the variance in mpg. The adjusted R^2 shows that this model explains for 84% of the variance in mpg, a much improved model from the previous model1.

Final Model

```
finalmodel<-lm(mpg~am+wt+qsec,data=mtcars)
anova(model1,finalmodel)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The new finalmodel created captures 84% of the overall variation in mpg. We reject the null hypothesis as p-value is 3.745e-09. The multivariate model presented in finalmodel is significantly different from our simple linear regression model presented in model1.

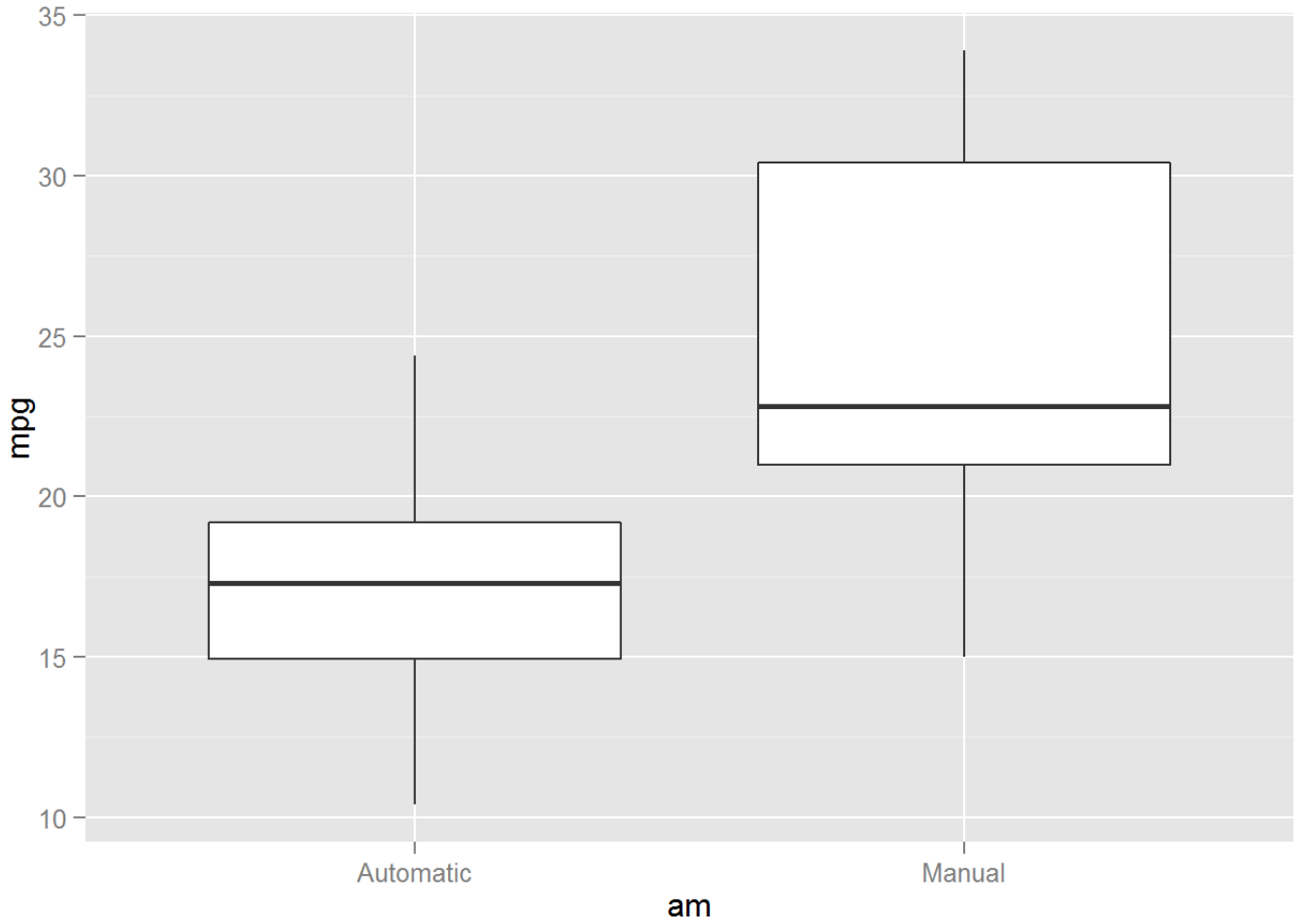
```
summary(finalmodel)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## amManual      2.9358     1.4109   2.081 0.046716 *
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

On average, Manual transmission cars are better as they will experience 2.94 mpg increase in performance than automatic transmision cars. Answering the 2 questions of interest along however is insufficient to completely understanding the relationship among a set of factors and mpg. This was evident when we found a higher value for mpg different in our initial simple regression model where other factors were not accounted for.

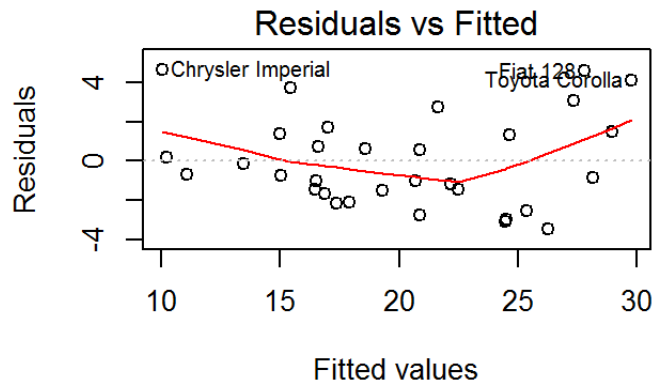
Appendix

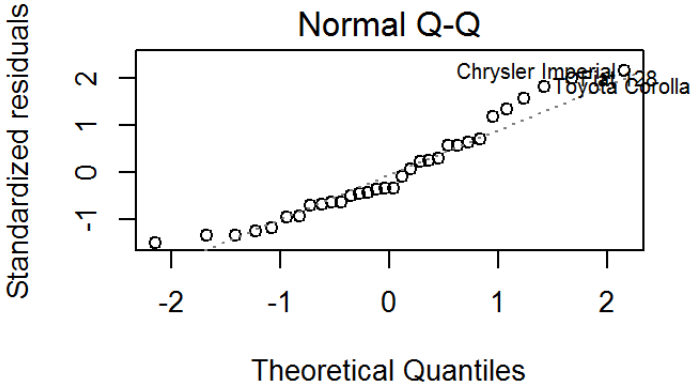
```
plot1
```

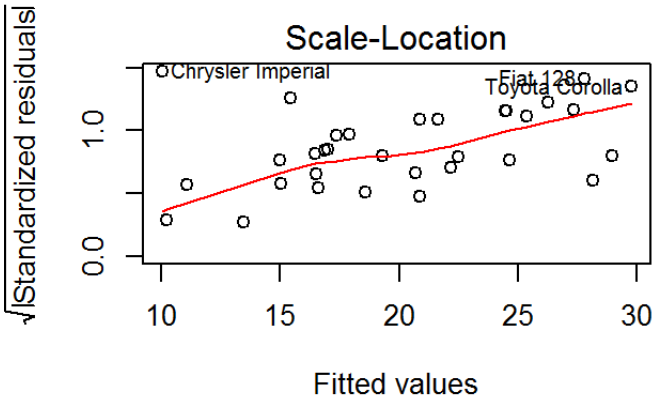


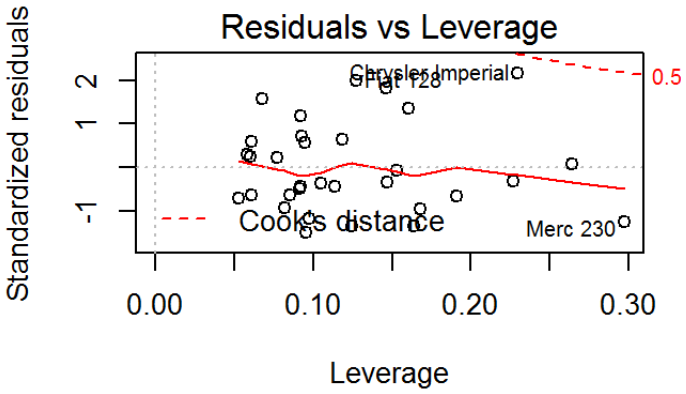
###Diagnostic

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


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla


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


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


Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
0.5
Chrysler Imperial
Fiat 128
Merc 230