Regression Models Course Project

Omar Nooreddin

12/22/2018

Executive Summary

This paper will endeavour to answer two questions: (1)Is an automatic or manual transmission better for a car's Miles Per Gallon (MPG); (2) And the difference, in MPG, between automatic and manual. For this we shall use the "mtcars" dataset in R.

Exploring Data

As per the above, we will be using the mtcars data for this analysis:

```
data(mtcars)
dim(mtcars)
```

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

We can see that there are 11 variables and 32 observations provided in this data set.

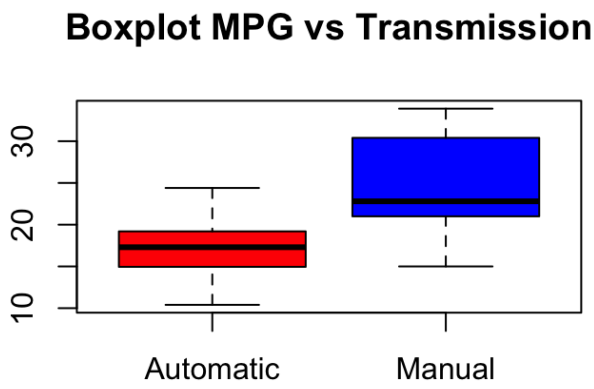
To facilitate our analysis we will be converting the am variable into factor:

```
mtcars$am   <- factor(mtcars$am,labels=c("Automatic","Manual"))
```

Inferential Analysis

As per our summary, we're going to look into the effect of manual vs automatic transmission on the MPG. For this initial analysis we can look at boxplot for MPG split by transmission:

```
boxplot(mpg~am,data=mtcars,col=c('red','blue'),main="Boxplot MPG vs Transmission")
```



We can see from the plot above that the mean MPG of the Manual transmission is higher/better than automatic transmission. But to confirm our observation, we will perform a T Test on mpg split by transmission to attain the P-Value:

```
t.test(mtcars$mpg[mtcars$am=="Automatic"],mtcars$mpg[mtcars$am=="Manual"])$p.value
```

```
## [1] 0.001373638
```

We can see that the P-Value of 0.0013736 strongly rejects the null hypothesis, further confirming the difference of MPG between Manual and Automatic.

Regression Analysis Investigating MPG

Initially we'll investigate the relationship between MPG and transmission by using Least Squares method:

```
fit1<-lm(mpg ~ am , data=mtcars)
summary(fit1)
```

```
##
## 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
```

We can see that the R^2 is 0.3597989 which indicates that the variable "am" accounts for 36% of MPG. As such we're going to use the rest of the variables in the dataset to investigate which variables contribute to MPG. For this we will be using the **step** function, which will go through all variables and selects model based on AIC metric (https://en.wikipedia.org/wiki/Akaike_information_criterion). (Note: for 'step' function output check Appendix):

```
fit2<-lm(mpg ~ . , data=mtcars)
fit3<-step(fit2)
```

```
summary(fit3)
```

```
##
## 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
```

The model selected “fit3” has an R^2 value of 85%, i.e this model (which includes weight, am and qsec) explains 85% of the variation in MPG with each coefficient(values) having a p-value of less than 5% (i.e they’re statistically significant).

Residual plots

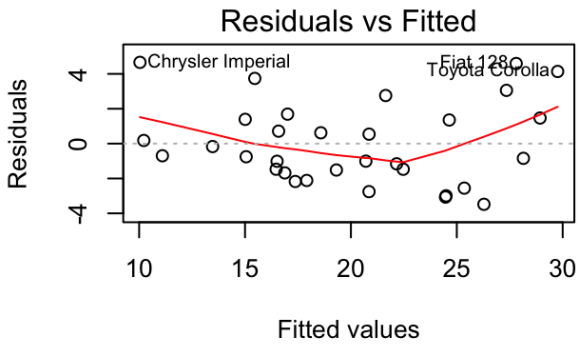
For the plots check the appendix. We can see that the data is close to a Normal Distribution and the Residual plot shows an approximate homoscedastic variance.

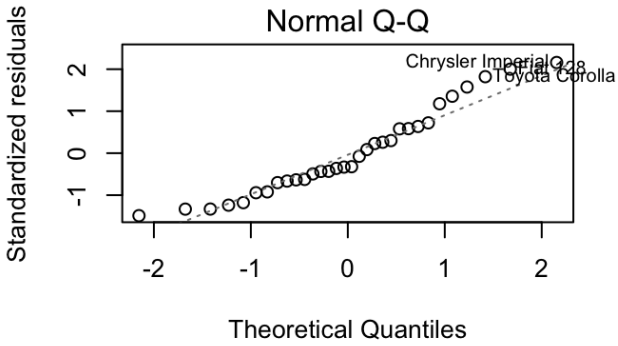
Conclusions

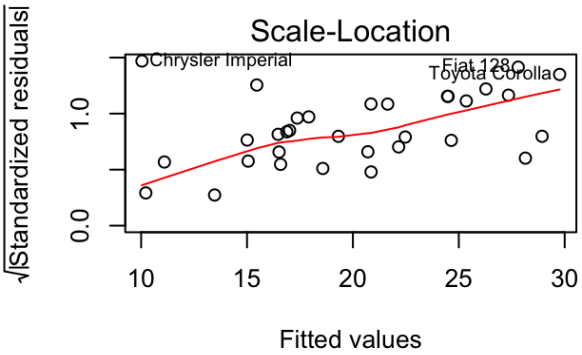
We can conculde that overall Manual transimission is better than Automatic when it comes to MPG. Also by looking at the chosen model fit3 the average MPG for automatic is 9.6, while manual is 3 MPG higher than automatic.

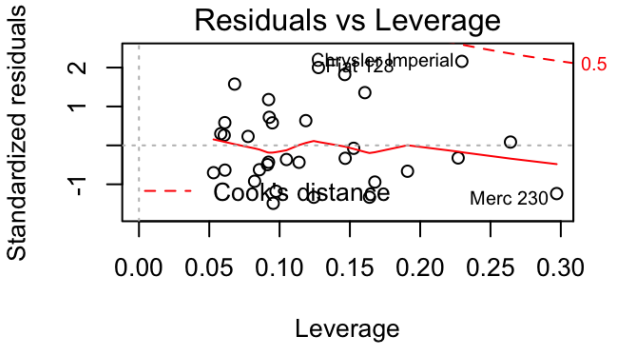
APPENDIX

```
par(mfrow = c(2,2))
plot(fit3)
```


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