Motor Trend Report

Saeed Nusri

11/6/2016

Assessing Effects of Various Parameters on Fuel Efficiency

Executive Summary

This report focuses on exploring how various variables can affect fuel effiecient in different cars and tries to explore if there's a significant difference between automatic and manual transmission. Statistical analysis and regression modeling was used to quantify this.

To summarise the findings, it is found that manual transmissions on average does a little better than automatic transmission but it statistically difficult to draw this conclusion. But weight and cylinders variables were seen to be significantly affecting the fuel effieciency (mpg) the mostout of all the variable and were considered as counfounding variable while finding the effects of transmission on fuel effieciency.

Exploratory Analysis

This section expolres the mtcars cars dataset in R Studio.

```
library (datasets)
data(mtcars)
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

The data set was extracted from the 1974 edition of Motor Trend US Magazine and it deals with 1973 - 1974 models. It consists of 32 observations on 11 variables.

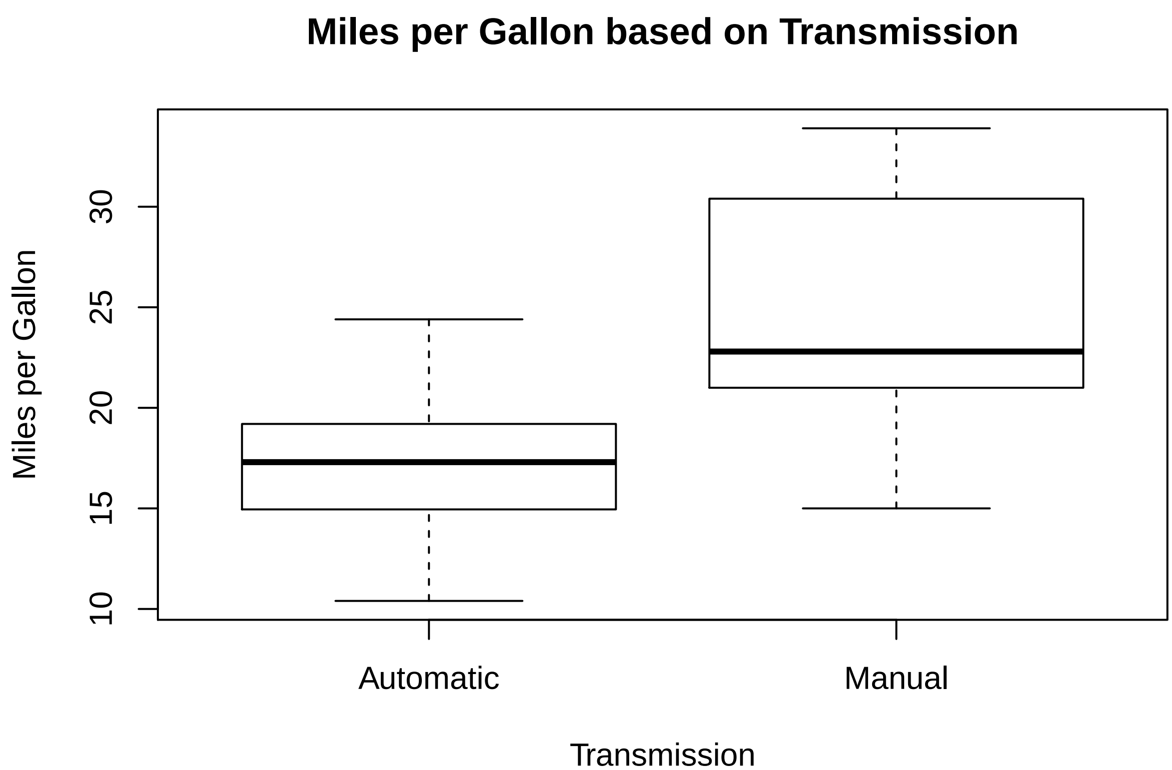
The variable of interest is "am" since it denotes hwether the car is manual or automatic.

After loading the dataset the am variable is simplified to factor variable with properly named levels.

```
mtcars$am <- as.factor(mtcars$am)
mtcars$cyl  <- factor(mtcars$cyl)
levels(mtcars$am) <- c("Automatic", "Manual")
```

Through a boxplot, it seems that the type of transmission affects mpg significantly.

```r
boxplot(mpg~am, data=mtcars,
        xlab = "Transmission", ylab = "Miles per Gallon",
        main = "Miles per Gallon based on Transmission")
```



The boxplot shows the relationship between the variables of fuel efficiency (MPG) and the two transmission where distinguisable differences between them is noticeable.

Statistical Analysis & Multi Variate Regression Model

We perform a t-test on the two subsets of mpg data: manual and automatic transmission assuming that the transmission data has a normal distribution and tests the null hypothesis that they come from the same distribution. Based on the t-test results, we reject the null hypothesis that the mpg distributions for manual and automatic transmissions are the same.

```r
t.test(mpg ~ am, data = mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by 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
```

Now we carryout on regression modeling with fuel efficiency as predictor variable. The firstmodel includes all variables as predictors of mpg. Then we perform stepwise model selection in order to select significant predictors for the newmodel. This ensures that we have included useful variables while omitting ones that do not contribute significantly to predicting mpg.

```
firstmodel <- aov(mpg ~ ., data = mtcars)
summary(firstmodel)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          2  824.8   412.4  61.863 2.72e-09 ***
## disp         1   57.6    57.6   8.647  0.00809 **
## hp           1   18.5    18.5   2.776  0.11130
## drat         1   11.9    11.9   1.787  0.19626
## wt           1   55.8    55.8   8.369  0.00900 **
## qsec         1    1.5     1.5   0.229  0.63768
## vs           1    0.3     0.3   0.045  0.83358
## am           1   16.6    16.6   2.485  0.13061
## gear         1    3.8     3.8   0.563  0.46183
## carb         1    1.9     1.9   0.292  0.59485
## Residuals   20  133.3     6.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Variables with p-value below 0.05 are more important. We choose cyl, disp, wt am as predictor variables for first model.

```
newmodel <- lm(mpg ~ cyl + disp + wt + am, data = mtcars)
summary(newmodel)$coefficients
```

```
##                 Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 33.816067258 2.91427152 11.6036090 8.792094e-12
## cyl6        -4.304782473 1.49235501 -2.8845566 7.774612e-03
## cyl8        -6.318405700 2.64765755 -2.3864135 2.458094e-02
## disp         0.001632161 0.01375694  0.1186427 9.064703e-01
## wt          -3.249175911 1.24909839 -2.6012170 1.512685e-02
## amManual     0.141212000 1.32675115  0.1064344 9.160547e-01
```

Coefficient of disp variable has a p-value of 0.906 and is hence removed as one of the confounding variables. The newmodel is corrected to revisedmodel as following.

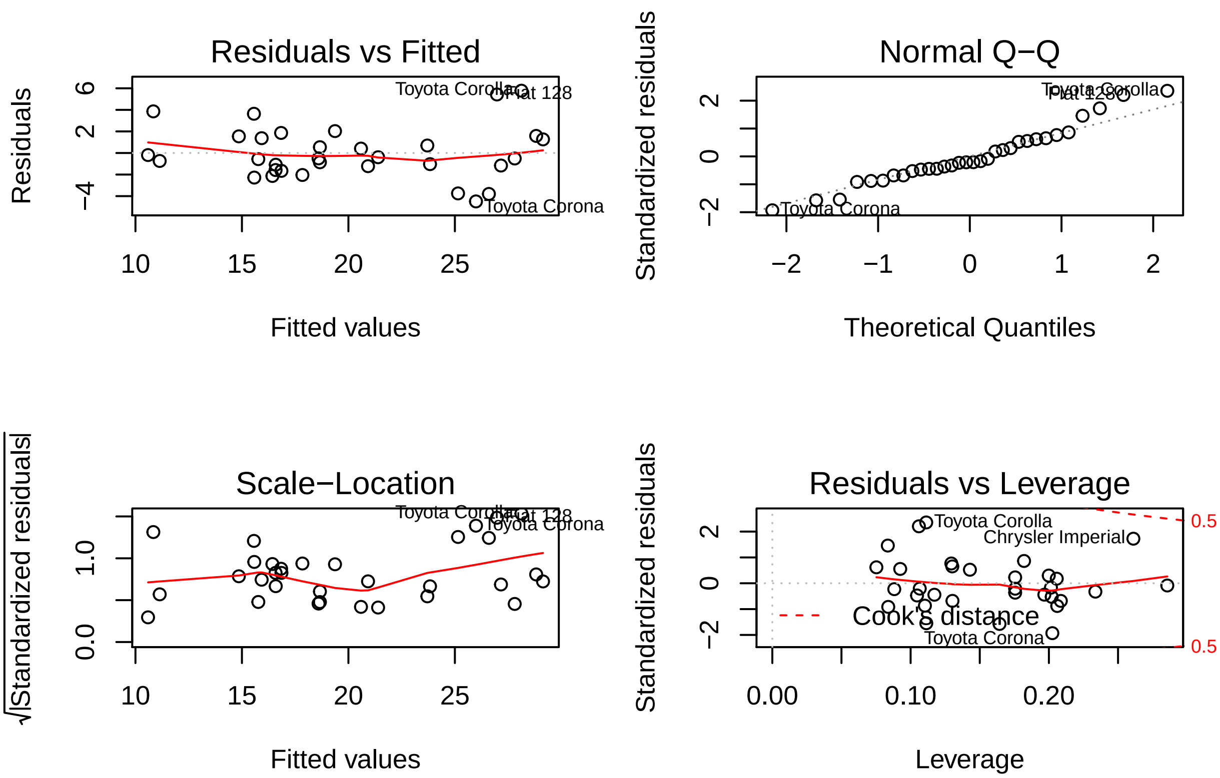
```
revisedmodel<- lm(mpg ~ cyl + wt + am, data = mtcars)
summary(revisedmodel)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4898 -1.3116 -0.5039  1.4162  5.7758
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.7536     2.8135  11.997  2.5e-12 ***
## cyl6         -4.2573     1.4112  -3.017  0.00551 **
## cyl8         -6.0791     1.6837  -3.611  0.00123 **
## wt           -3.1496     0.9080  -3.469  0.00177 **
## amManual      0.1501     1.3002   0.115  0.90895
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.603 on 27 degrees of freedom
## Multiple R-squared:  0.8375, Adjusted R-squared:  0.8134
## F-statistic: 34.79 on 4 and 27 DF,  p-value: 2.73e-10
```

The adjusted r-squared is 0.83 and this is our final model. Clearly, with cylinders and weight as confounding variables, the ceofficient of the am variable is small but has a large p-value.

```r
par(mfrow = c(2, 2))
plot(revisedmodel)
```



Following deduction can be made from the plots

- The points in the Residuals vs. Fitted plot are randomly scattered on the plot that verifies the independence condition.
- The Normal Q-Q plot consists of the points which mostly fall on the line indicating that the residuals are normally distributed.
- The Scale-Location plot consists of points scattered in a constant band pattern, indicating constant variance.

- There are some distinct points of interest (outliers or leverage points) in the top right of the plots that may indicate values of increased leverage of outliers.

Conclusion

Based on the analysis done the following can be concluded:

- Cars with Manual transmission get 0.15 more miles per gallon compared to cars with Automatic transmission. (adjusted for cyl, and wt).
- mpg will decrease by 3.15 for every 1000 lb increase in wt.
- If number of cylinders, cyl increases from 4 to 6 and 8, mpg will decrease by a factor of 4.3 and 6.1 respectively (adjusted by wt, and am).

Appendix

Pairs plot for the mtcars dataset

```r
pairs(mpg ~ ., data = mtcars, main = "Pairs plot for the mtcars dataset")
```