Motor Trend

Kenrick Setiobudi

22 November 2015

Executive Summary

This paper is an analysis on the relationship between a set of variables and miles per gallon (MPG) for Motor Trend, a magazine about the automobile industry. There are two burning platforms that are investigated in this analysis:

1. "Is an automatic or manual transmission better for MPG"
2. "Quantify the MPG difference between automatic and manual transmissions"

The results have shown that:

1. Cars with automatic transmission provides **worse mileage per gallon** compared to cars with manual transmission;
2. By analysing other characteristics of cars that may contribute to the improvement / deterioration of mileage per gallon into the analysis, it is also noted that:

- heavier cars have less mileage per gallon; and
- increasing 1/4 mile time will improve mileage per gallon.

Automatic vs Manual Transmission on Mileage Per Gallon

To commence the analysis, we first load the data into the system.

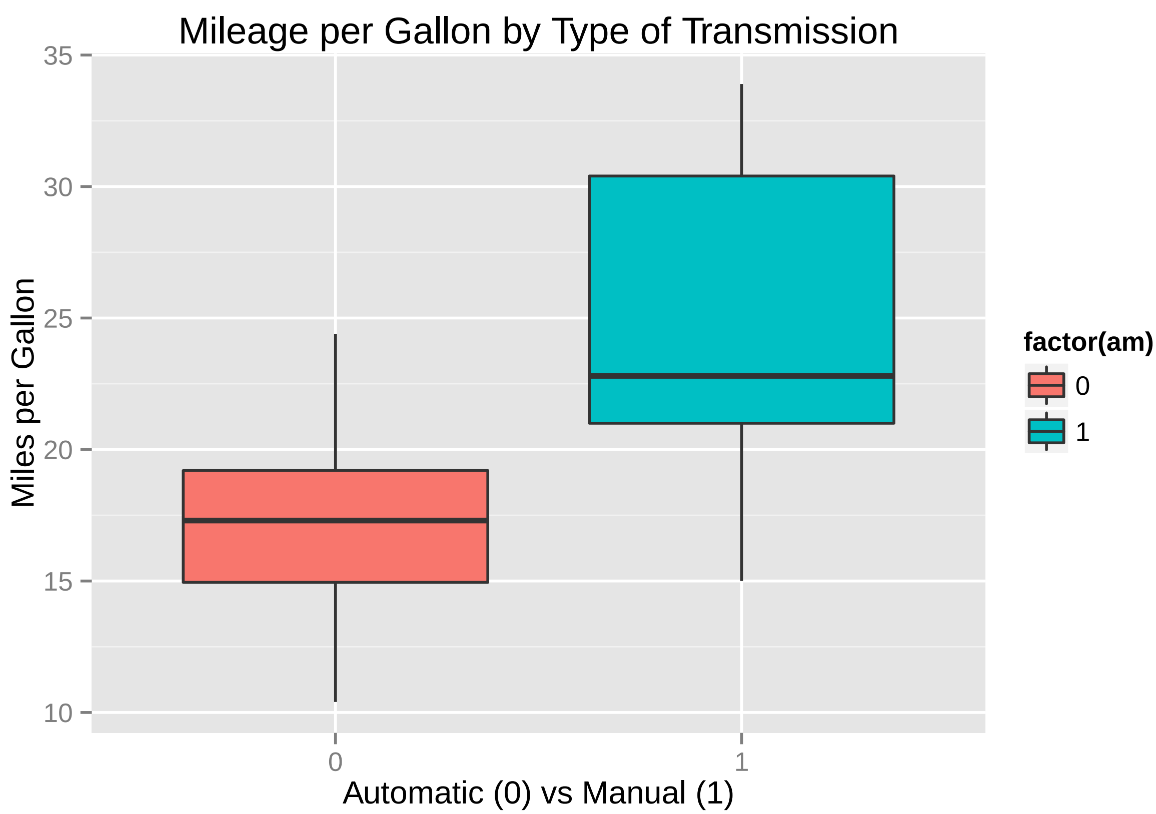
```r
data(mtcars)
str(mtcars)
```

```
## '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 ...
```

Boxplot

When we compare the boxplot of miles per gallon `mpg` based on the type of transmission `am` (shown below), it is clearly evident that automatic cars has less mileage per gallon compared to manual.

```r
library(ggplot2)
g <- ggplot(data = mtcars, aes(x = factor(am), y = mpg))
g <- g + geom_boxplot(aes(fill = factor(am)))
g <- g + ggtitle("Mileage per Gallon by Type of Transmission")
g <- g + xlab("Automatic (0) vs Manual (1)")
g <- g + ylab("Miles per Gallon")
print(g)
```



Univariate Linear Model

We then use a univariate linear model regression to compare the car's mileage per gallon `mpg` against the type of transmission `am`, where:

- `am = 0` for automatic; and
- `am = 1` for manual.

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

```
##
## 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 ***
## am             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
```

The above results have shown that cars with manual transmission has 7.24 miles per gallon advantage over cars with automatic transmission. In addition, because the p-value is very small and less than 0.05, we reject the null hypothesis.

However, this model is assuming that all of the other cars' characteristics are not significant. This means that we should compare the univariate regression model with a multivariate regression model through introducing the other cars' characteristics as the regressor.

Multivariate Regression Model

Here we are comparing the univariate model above with a multivariate regression model which starts with using all of the available predictors, then conduct a backward selection algorithm of the predictors that are not significant to the response `mpg`:

```r
full <- lm(mpg~., data = mtcars)
model2 <- step(full, direction = "backward")
```

```r
summary(model2)
```

```
##
## 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 ***
## am            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
```

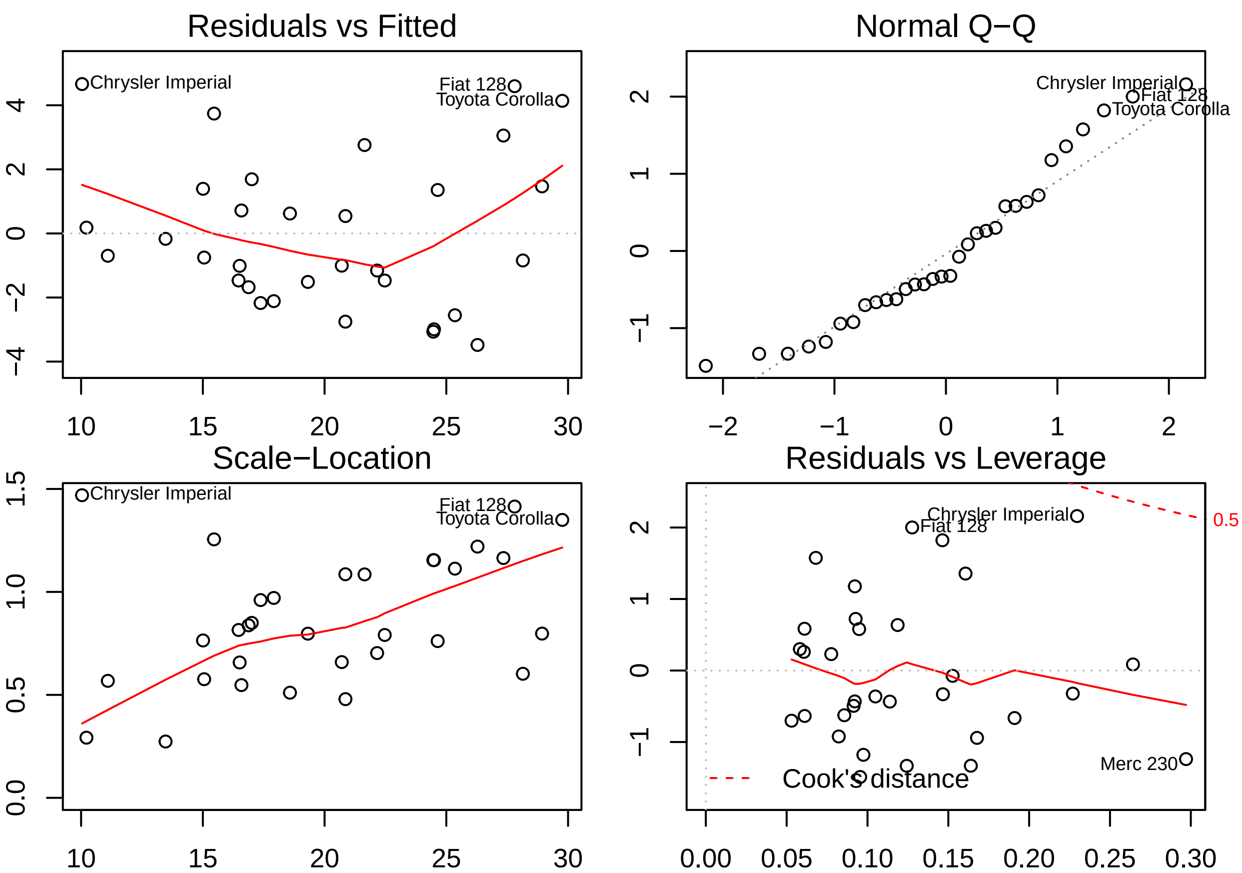
The results have shown that:

1. there are 2 extra variables `wt` (weight of car, lb/1000) and `qsec` (1/4 mile time);
2. the resultant model captures 85% of total variance (compared to 36% in the univariate model), given that R^2 is `0.85`;
3. there is an inverse relationship between mileage per gallon and the weight of the car. For every 1,000 lb weight increase, the mileage per gallon will reduce by 3.92;
4. there is a direct relationship between mileage per gallon and the 1/4 mile time. Mileage per gallon improves by 1.23 for every 1 unit of increase in 1/4 mile time; and
5. manual transmission cars still provide a better mileage per gallon by 2.93.

Appendix

Residual analysis of the multivariate model

```
par(mfrow=c(2,2), mar = c(1.8,1.8,1.5,1.5))
plot(model2)
```



Main insights from the charts above:

- Residual vs Fitted and Scale-Location does not show any pattern
- Normal QQ plot shows that the residuals are normally distributed
- Residual vs Leverage plot doesn't show any particular outliers