Motor Trend Analysis

Andrey Alferov

Sunday, May 24, 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:

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

Data Processing

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

```
library(datasets)
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
##               from 'package:stats':
##
##     filter
##
##                from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data(mtcars) #Load mtcars dataset
```

Let's have a brief look at mtcars dataset.

```
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```r
dim(mtcars)
```

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

```r
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

The dataset consists of 32 observations on 11 variables.

Analysis

Is an automatic or manual transmission better for MPG?

To answer that question lets compare average mpg for automatic and manual transmissions.

```r
mean(filter(mtcars, am == 1)$mpg) #Calculate mean mpg for manual transmission
```

```
## [1] 24.39231
```

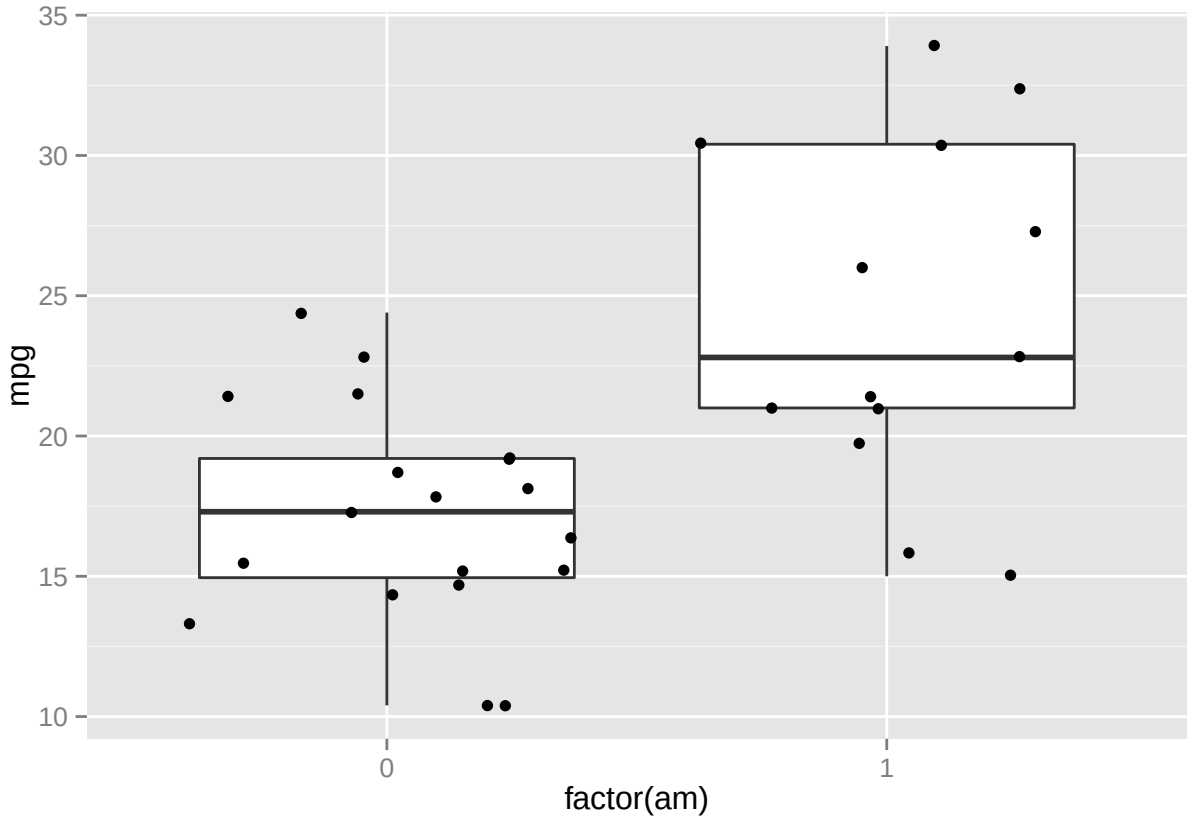
```r
mean(filter(mtcars, am == 0)$mpg) #Calculate mean mpg for automatic transmission
```

```
## [1] 17.14737
```

As we can see cars with manual transmission have greater mpg than cars with manual transmission.

Lets build a boxplot displaying mpg per transmission type.

```
p <- ggplot(mtcars, aes(factor(am), mpg))
p + geom_boxplot() + geom_jitter()
```



The plot also prooves that cars with manual transmission type (represented by 1) has a higher mean for mpg than automatic (represented by 0).

Lets perform a t-test to confirm the null hypothesis that transmission type affects mpg.

```
t.test(mtcars$mpg ~ mtcars$am, conf.level=0.95)
```

```
##
##  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 0 mean in group 1
##        17.14737        24.39231
```

p-value = 0.001374 is less then 0.05, so we reject the null hypothesis that there is no difference in MPG per transmission type.

Quantify the MPG difference between automatic and manual transmissions.

Lets check how all variables affect mpg difference by fitting corresponding linear model.

```
fit_all = lm(data = mtcars, mpg ~ .)
summary(fit_all)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

As we can see there are 2 predictor variables: wt (Weight) and am (Transmission type) which affects mpg variance at most.

Lets try to choose the best set of predictors automatically.

```
fit_best = step(lm(data = mtcars, mpg ~ .), trace=0)
summary(fit_best)
```

```
##
## 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
```

It looks like the best model is the one that includes wt, qsec and am, which means besides transmission types, weight and accelearation also needs to be considered. Weight negatively changes with mpg, and qsec and am positively changes. Every lb/1000 weight increase will cause a decrease of roughly 4 mpg, every increase of 1/4 mile time will cause an increase of 1.2 mpg, and on average, manual transmission is 2.9 mpg better than automatic transmission. The model is able to explain 85% of variance. The residual plots also seems to be randomly scattered (see appendix).

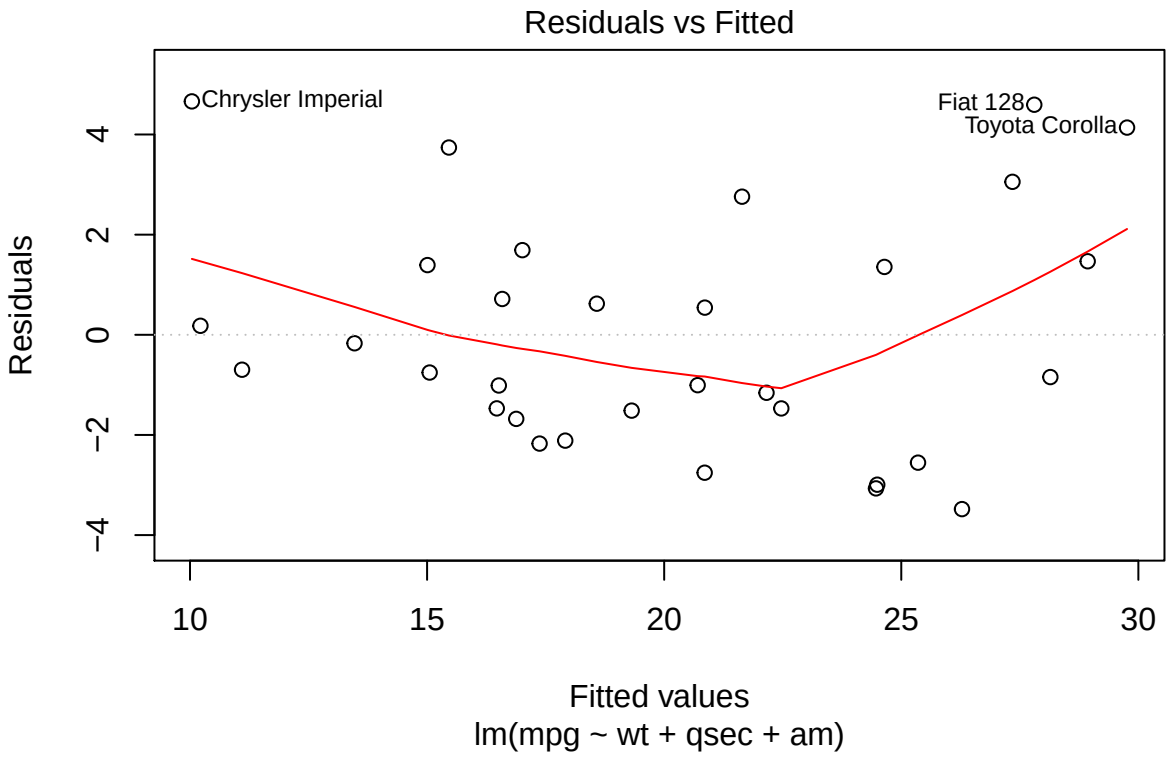
Conclusion

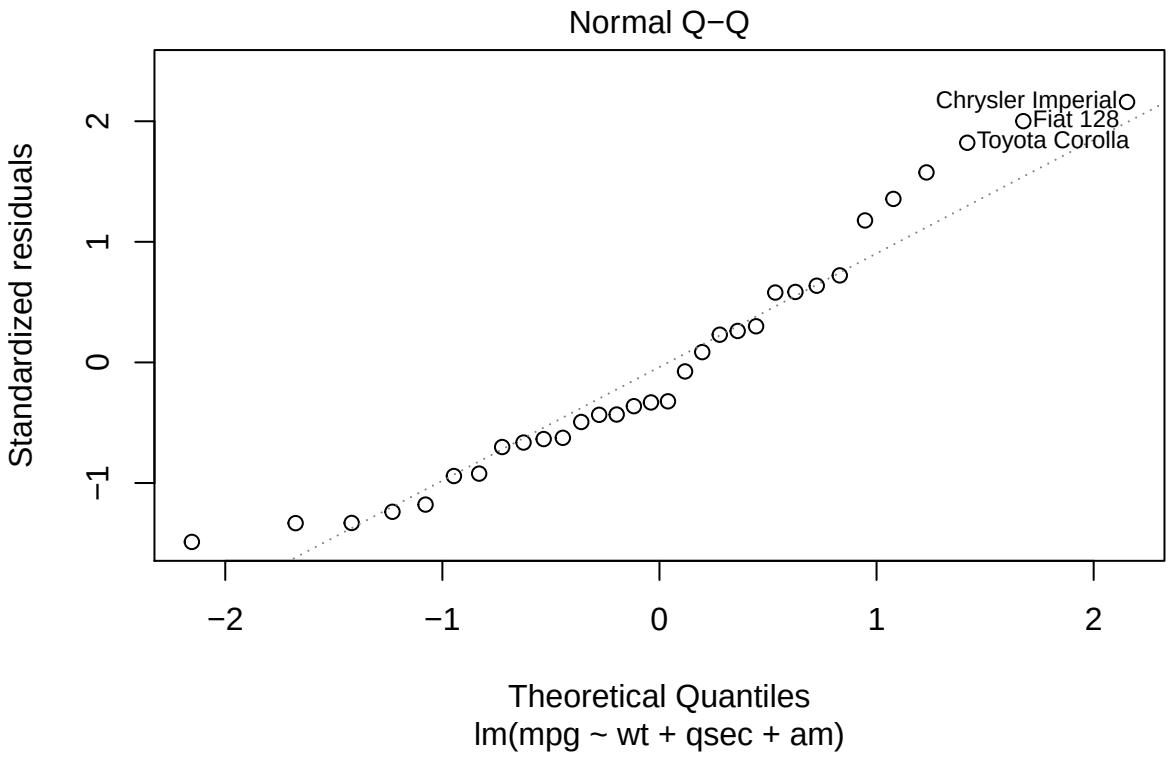
As a result of analysis it can be concluded that manual transmission is better than automatic transmission by 2.9 mpg. However, transmission type is not the only factor accounting for MPG, weight, and acceleration (1/4 mile time) also needs to be considered.

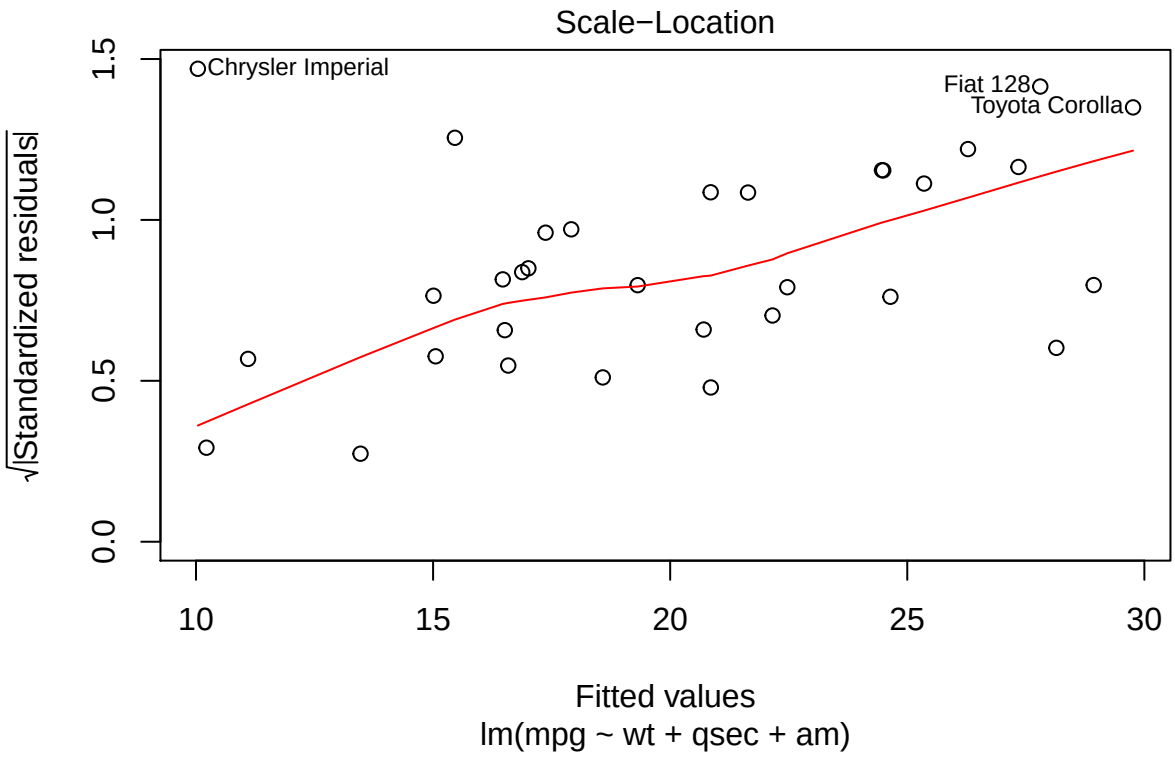
Appendix

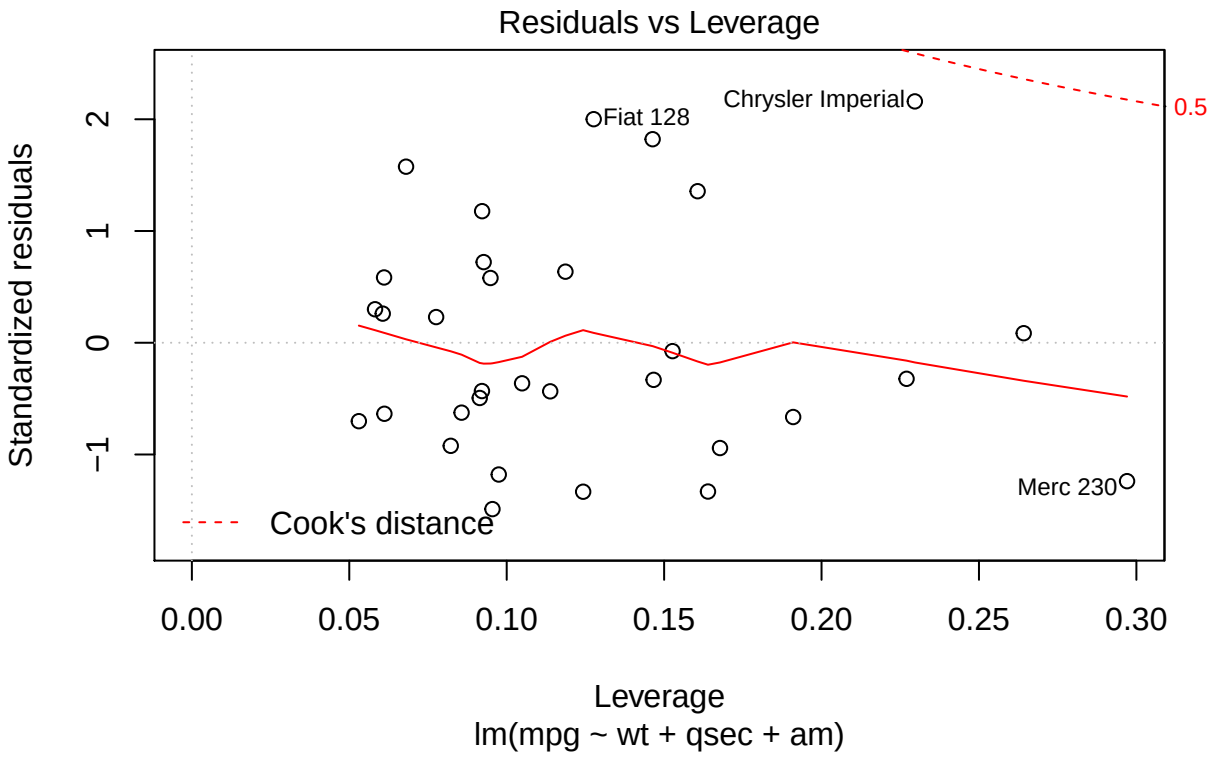
Residual plots

```
plot(fit_best)
```


Residuals vs Fitted
Chrysler Imperial
Fiat 128
Toyota Corolla
Residuals
4
2
0
−2
−4
10
15
20
25
30
Fitted values
lm(mpg ~ wt + qsec + am)


Normal Q−Q
Chrysler Imperial
Fiat 128
Toyota Corolla
Standardized residuals
2
1
0
−1
−2
−1
0
1
2
Theoretical Quantiles
lm(mpg ~ wt + qsec + am)


Scale−Location
Chrysler Imperial
Fiat 128
Toyota Corolla
√|Standardized residuals|
0.0
0.5
1.0
1.5
10
15
20
25
30
Fitted values
lm(mpg ~ wt + qsec + am)


Residuals vs Leverage
Standardized residuals
Leverage
lm(mpg ~ wt + qsec + am)
Fiat 128
Chrysler Imperial
Merc 230
0.5
Cook's distance
2
1
0
−1
0.00
0.05
0.10
0.15
0.20
0.25
0.30