Motor Trend

A. Kramer

Tuesday, March 19, 2015

Executive Summary

In this paper, the mtcars dataset is examined with the goal of determining the impact of transmission type (*am*) on the mile per gallon (*mpg*). We will explore and quantify the differences in fuel consumption. Our analysis indicates that type of transmission is not statistically significant factor in affecting the *mpg* if considered in conjunction with other variables in the data set.

Question: Is an automatic or manual transmission better for MPG?

Building a model for MPG

At first, the model containing all variables in the dataset is created.

```
library(datasets)
library(car)

fit1 <- lm(mpg ~ as.factor(cyl) + as.factor(vs) + factor(am) + as.factor(gear)
           + as.factor(carb) + qsec + drat + hp + wt + disp, data = mtcars)

am1 <- summary(fit1)$coefficient['factor(am)1', ]
```

Now the model of *mpg* versus *am* only is created and the *am* entry is extracted.

```
fit2 <- lm(mpg ~ factor(am), data = mtcars)

am2 <- summary(fit2)$coefficient['factor(am)1', ]
```

Visually presenting displaying comparison between *am1* and *am2* in tabular format.

```
rbind(am1, am2)
```

```
##     Estimate Std. Error   t value     Pr(>|t|)
## am1 1.212116   3.213545 0.3771896 0.7113157275
## am2 7.244939   1.764422 4.1061270 0.0002850207
```

It is observed that (Pr(>|t|)) in *am2* indicates that null hypothesis needs to be rejected, suggesting that there is a relationship between type of transmission and *mpg*. However, when taking in account other variables in the data set, as shows by *am1*, the impact of the transmission type on *mpg* is negated suggesting that null hypothesis cannot be rejected.

Regression Variance

It is also noted that there may be more variables affecting the *mpg*. We select weight, horse power, num. of cylinders, and transmission style to build and run the model.

```
fit3 <- lm(mpg ~ wt + hp + as.factor(cyl) + factor(am), data = mtcars)
```

Now we run the anova test to choose between fit2 and fit2 obtaining p-value for choosing between fit2 and fit2 models. *fit3* is selected.

```
anova(fit2, fit3)$'Pr(>F)'[2]
```

```
## [1] 1.688435e-08
```

The *am* is excluded from the *fit4* below and is compared with *adj.r.squared* from *fit3* and *fit4*.

```
fit4 <- lm(mpg ~ wt + hp + as.factor(cyl), data = mtcars)

am3 = summary(fit3)$adj.r.squared

am4 = summary(fit4)$adj.r.squared

cbind(am3, am4)
```

```
##            am3       am4
## [1,] 0.8400875 0.8360668
```

It appears that the loss of regression variance between *am3* and *am4* is negligible, which strongly suggests that transmission may not be a sole factor affecting *mpg*.

Variance Inflation Factor

The Variance Inflation Factor (VIF) needs to be investigated to check if more variables are required to be included in the analysis. Assessing GVIFˆ(1/(2*Df)) and Df values, it is concluded that all the variables included in the analysis of variance are not redundant and are good to use for the analysis.

```
vif(fit4)
```

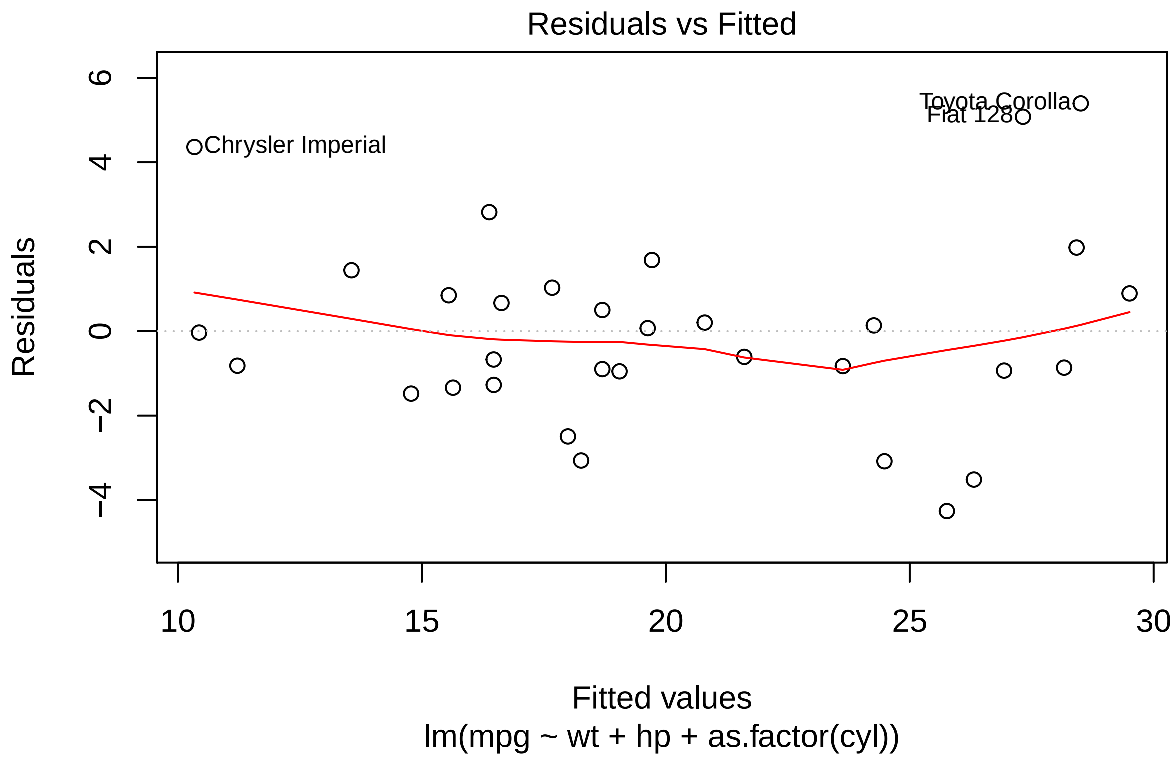
```
##                    GVIF Df GVIF^(1/(2*Df))
## wt             2.580877  1        1.606511
## hp             3.496014  1        1.869763
## as.factor(cyl) 5.105811  2        1.503198
```

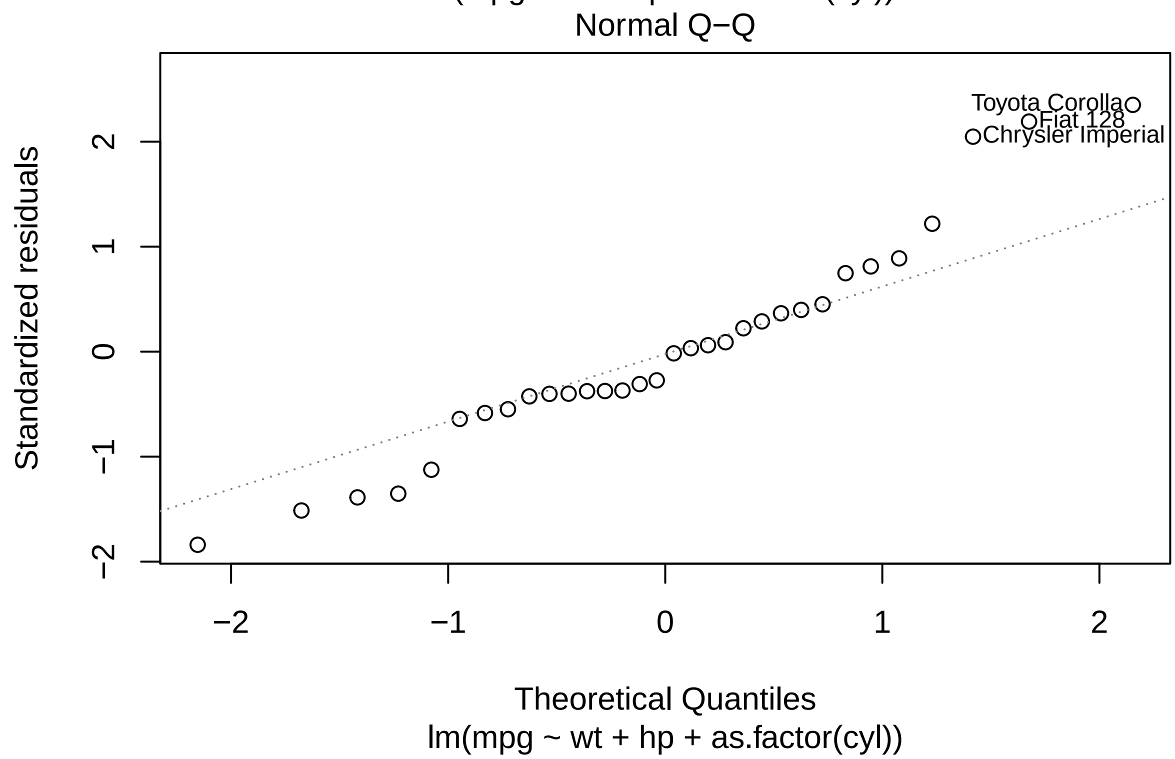
Conclusion

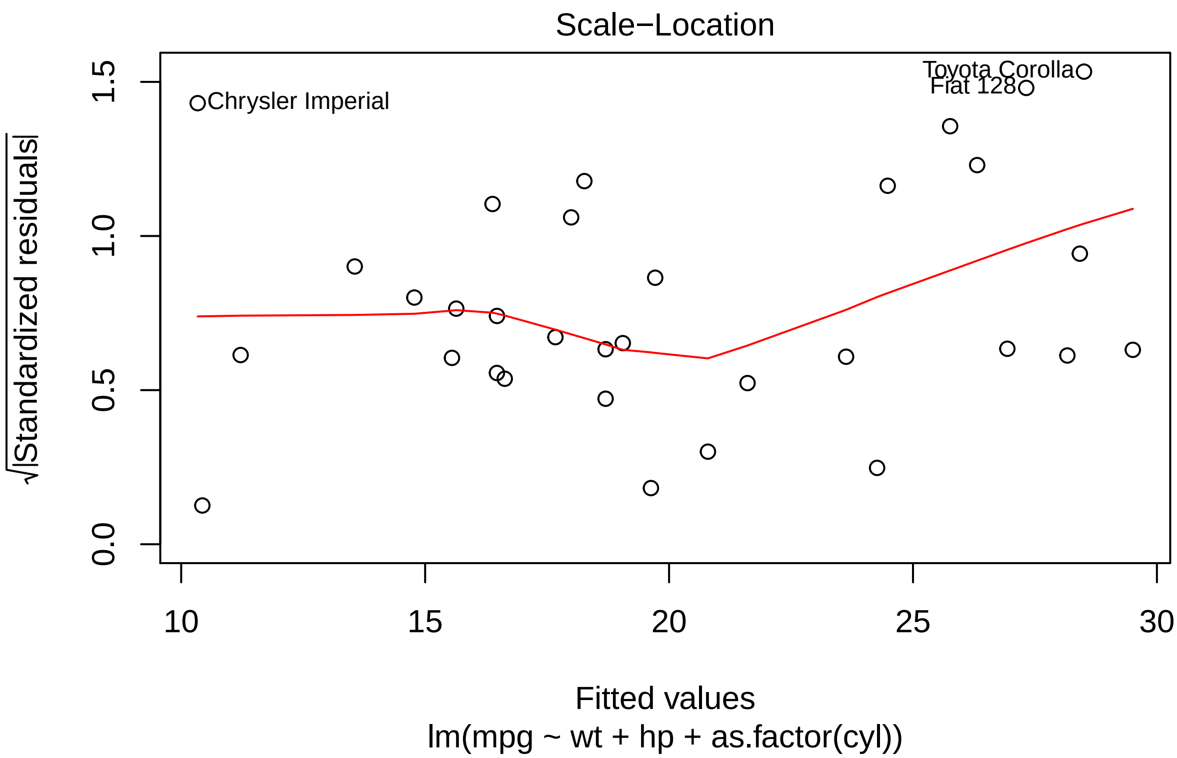
Based on the analysis above we proposed that *fit4* explains mpg. The residual diagnostic and VIF test do not indicate any major flaws with the analysis. As such, the conclusion is that there is no statistically significant impact of transmission type (*am*) on mile per gallon (*mpg*).

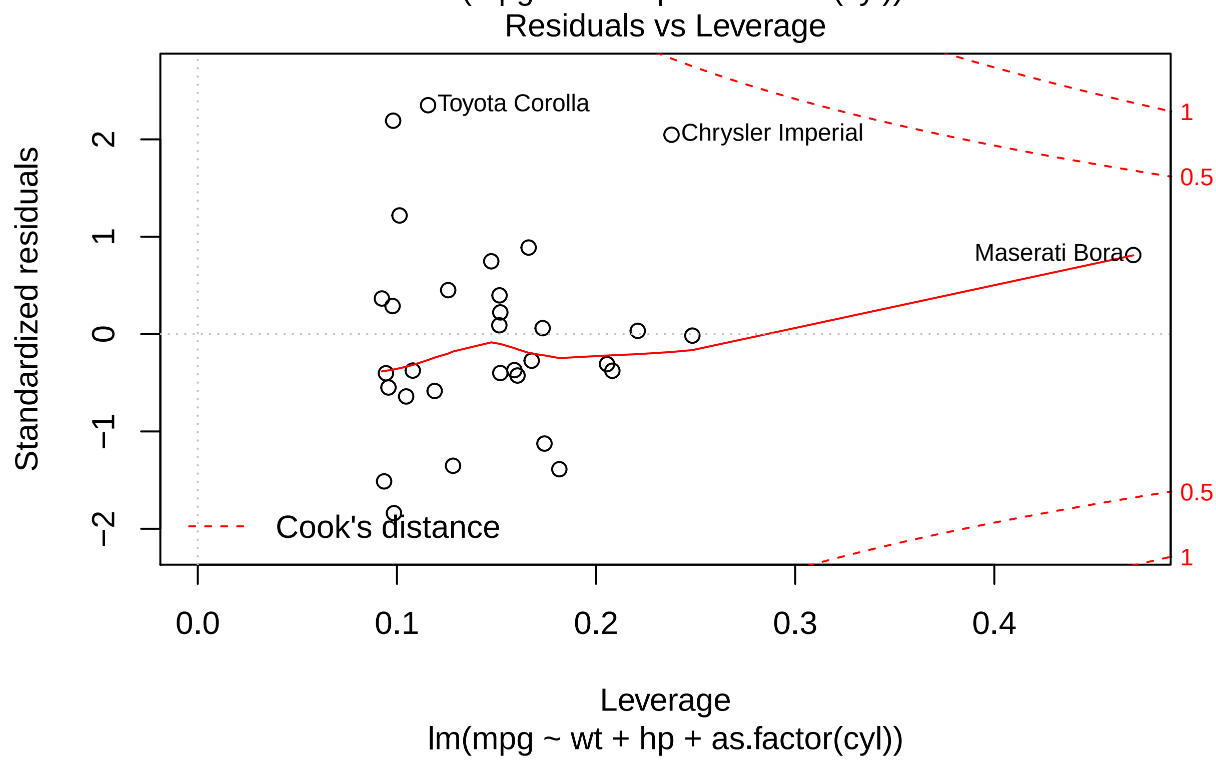
Appendix

```r
plot(fit4)
```






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


Residuals vs Leverage
Standardized residuals
−2
−1
0
1
2
0.0
0.1
0.2
0.3
0.4
Toyota Corolla
Chrysler Imperial
Maserati Bora
1
0.5
0.5
1
Cook's distance
Leverage
lm(mpg ~ wt + hp + as.factor(cyl))