Regression Models Project - MPG by Transmission Type Analysis

Sherman Wood

August 9, 2016

Executive Summary

This is my submission for the Coursera Regression Models Course Project - Motor Trends MPG Analysis.

The scenarios is that I work for Motor Trends, a magazine about the automobile industry. We are interested in exploring the relationship between a set of variables and miles per gallon (MPG), based on the mtcars data set from R. Of particular interest is the following two questions:

"Is an automatic or manual transmission better for MPG?"

Based on a t-test of MPG versus automatic or manual transmissions in the data set, manual vehicles have a higher (better) MPG compared to automatic vehicles. However the transmission type alone is not a good and sufficient predictor of MPG, so you cannot say that MPG will always be better with manual transmissions.

"What is the quantitative difference for MPG between automatic and manual transmissions?"

With the data set, manual vehicles have MPG that is significantly better (with 95% confidence under a t-test) than automatic vehicles. A linear model that only uses transmission type cannot be used to predict MPG.

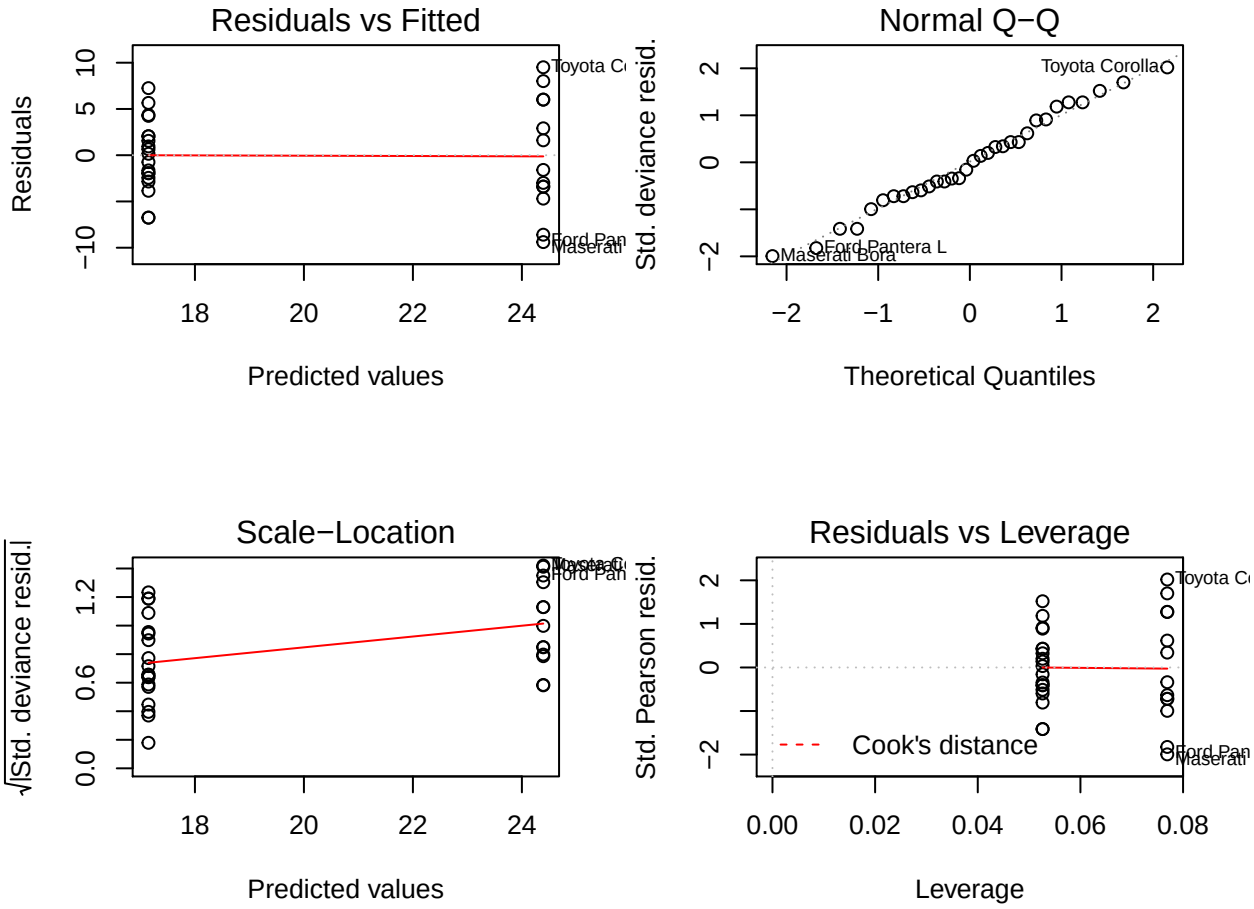
Analysis

To start, let's look at predicting MPG from transmission type only - without other factors.

```
##                Estimate Std. Error  t value     Pr(>|t|)
## transManual    24.39231   1.359578 17.94109 1.376283e-17
## transAutomatic 17.14737   1.124603 15.24749 1.133983e-15
```

Miles per Gallon predicted by Transmission Type

glm(mpg ~ trans − 1)



The 95% confidence interval indicates, based on the transmission type alone, manual cars have significantly better (higher) MPG than automatic cars.

However, the plots of the linear model of MPG predicted by transmission type shows that transmission is a poor predictor of MPG. The residuals vs fitted plot shows that the transmission-only predictive model will only give one value as an outcome for either automatic (17.14737 MPG) or manual (24.39231 MPG) transmissions.

Appendicies

The data: Motor Trend Car Road Tests

The analysis will investigate the mtcars data set, which is available in the R "datasets"" package. 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).

mtcars is a data frame with 32 observations on 11 variables. I made some of the variables factors to ease manipulation.

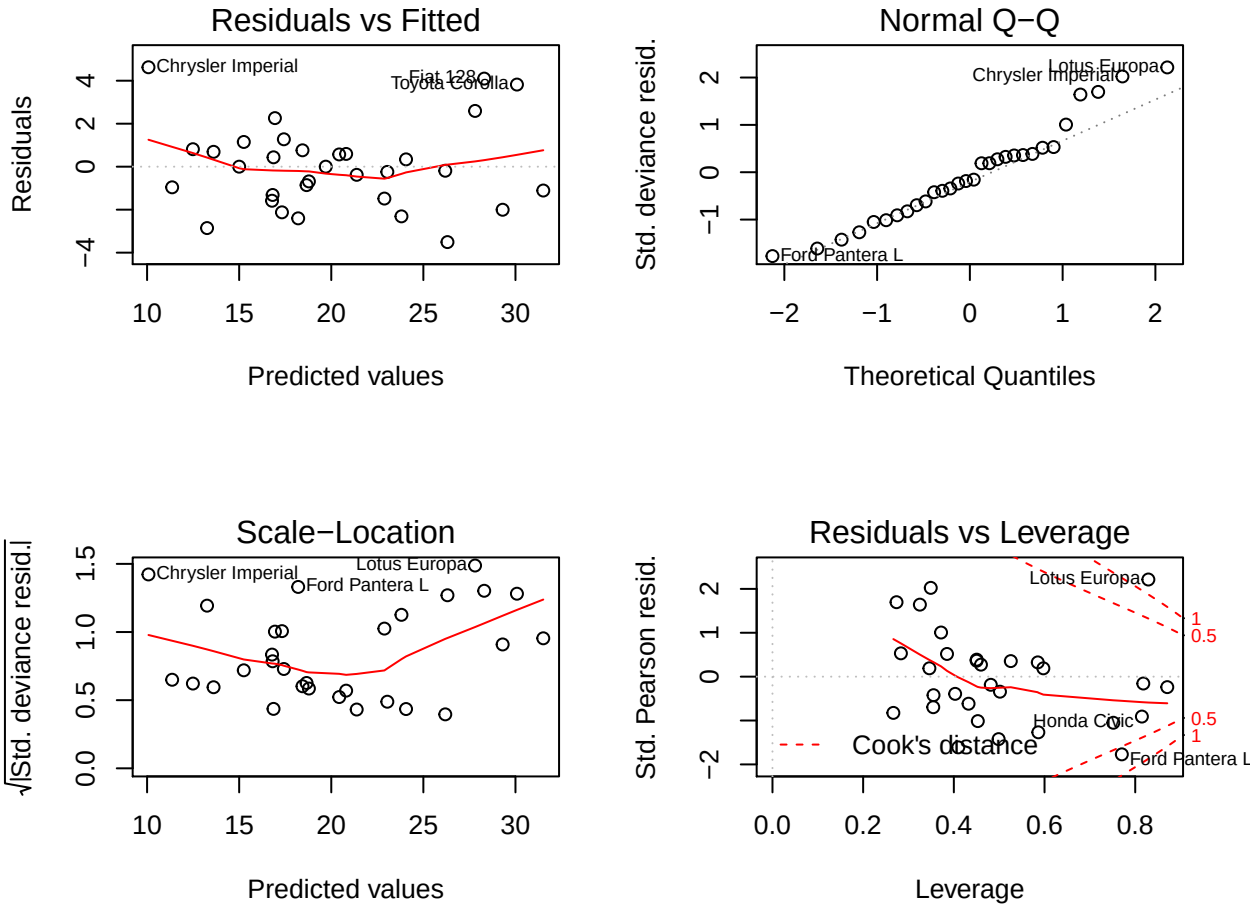
- [, 1] mpg Miles/(US) gallon
- [, 3] disp Displacement (cu.in.)
- [, 4] hp Gross horsepower
- [, 5] drat Rear axle ratio
- [, 6] wt Weight (1000 lbs)
- [, 7] qsec 1/4 mile time
- [,12] trans factor: Automatic or Manual Transmission
- [,13] engType factor V/S: V or straight engine
- [,14] cylf factor No of cylinders
- [,15] gearf factor Number of forward gears
- [,16] carbf factor Number of carburetors

A Deeper Analysis: is transmission type a good predictor for MPG?

Are there other factors influencing this outcome? You would expect that the "more powerful" cars would have lower MPG. Let's model all the factors together, and their variance inflation factors (VIFs), to determine the factors that have the greatest impact on a model.

Miles per Gallon predicted by all factors

glm(mpg ~ . − am − vs − cyl − carb − gear)




```
##              GVIF       Df GVIF^(1/(2*Df))
## disp     7.769536 1.000000        2.787389
## hp       5.312210 1.000000        2.304823
## drat     2.609533 1.000000        1.615405
## wt       4.881683 1.000000        2.209453
## qsec     3.284842 1.000000        1.812413
## trans    3.151269 1.000000        1.775181
## engType  2.843970 1.000000        1.686407
## cylf    11.319053 1.414214        1.834225
## gearf    7.131081 1.414214        1.634138
## carbf   22.432384 2.236068        1.364858
```

From the VIFs, the factors with the most to least effect on the model are:

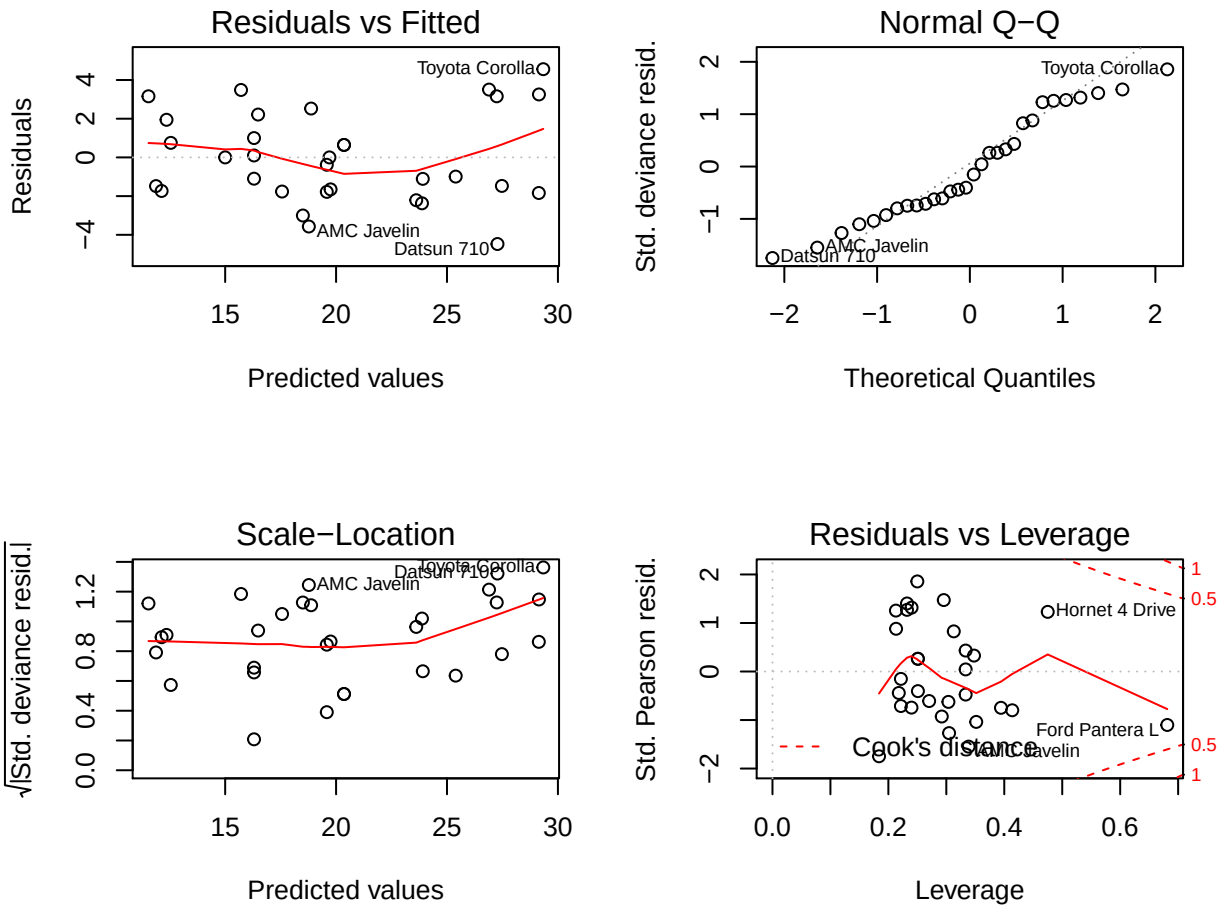
1. carbf - number of carburetors
2. cylf - number of cylinders

3. gear - number of gears
4. disp - Displacement (cu.in.)
5. hp - Gross horsepower

Transmission type is the 7th most effective factor on the model, which is very low, so it is not a strong indicator. Let's look at a model that includes these factors.

Miles per Gallon predicted by highly effective factors

glm(mpg ~ carbf + cylf + gear + disp + hp)



We can compare the strength of a model based on transmission type alone versus a model that includes the most powerful factors suggested by the VIF analysis via ANOVA.

```
## Analysis of Deviance Table
##
## Model 1: mpg ~ (cyl + disp + hp + drat + wt + qsec + vs + am + gear +
##     carb + trans + engType + cylf + gearf + carbf) - am - vs -
##     cyl - carb - gear
## Model 2: mpg ~ carbf + cylf + gear + disp + hp
##   Resid. Df Resid. Dev Df Deviance
## 1        15     120.40
```

```
## 2        21     168.91 -6  -48.503
```