Fuel efficiency and transmission systems

LFR

Sunday, September 21, 2014

Executive summary

This paper addresses the determinants of the miles per gallon (*mpg*) of passenger cars, using the 'mtcars' dataset. We show that the transmission system is highly significant when considered in isolation, but that the move from an automatic to a manual transmission system is not significant when the vehicle's weight is taken into account. The highest fuel economy is obtained with a 4 gear automatic system. The full code of this document can be found at https://github.com/LaurentFranckx/CourseraRegressionProject.

Introduction

This paper addresses the determinants of the miles per gallon (*mpg*) of passenger cars, using the 'mtcars' dataset. This datasets consists of 32 observations of passenger cars - see *help(mtcars)* for a definition of the variables.

Model choice

The fundamental question here is how the transmission type *am* affects *mpg*. As the figure "mpg versus gear and am" in Appendix suggests, the impact of *am* on *mpg* is likely to be affected by *gear* as well. The transmission type also imposes constraints on the number of forward gears, *gears*: cars with an automatic (manual) transmission system have at the most (resp. at the minimum) 4 gears. We therefore create a new factor variable, *trans*, which represents the interaction between *am* and *gears*. Starting with the simplest specification that expresses *mpg* as a function of the transmission system yields the following estimates (coefficients can only be estimated for the *am-gear* combinations that are actually observed):

```
## (Intercept)    trans4.0    trans4.1    trans5.1
##      16.107       4.943      10.168       5.273
```

This first model yields an R squared value of 0.49 and an adjusted R squared of 0.44. While encouraging, this suggests that adding other explanatory variables is worthwhile exploring.

The graph "Plotting the continuous variables" in Appendix suggests that several continuous variables are good candidates for explaining *mpg*, but also that there is a high level of multicollinearity between these continuous variables. It can be verified that, using a linear model, 0.8 % of the variation in *hp* can be explained by the variation in *wt* and *qsec*. Similarly, 0.87 % of the variation in *disp* can be explained by the variation in *wt* and *qsec*.

The graph "Plotting the continuous variables" suggests that *mpg* is a convex quadratic function of weight: *mpg* decreases when weight increases, but at a decreasing rate. We have therefore added a quadratic function of *wt* (centered around its mean value) to the previous specification:

```
## lm(formula = mpg ~ trans + I(wt - mean(wt)) + I((wt - mean(wt))^2),
##     data = mtcars)
```

This second model yields an R squared value of 0.87 and an adjusted R squared of 0.84. The ANOVA tables confirms that adding these two terms is statistically significant. It can be verified that adding *qsec* and *vs* as explanatory variables does not lead to a statistically significant improvement.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ trans + I(wt - mean(wt)) + I((wt - mean(wt))^2)
## Model 2: mpg ~ trans
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     26 152
## 2     28 570 -2      -418 35.8 3.4e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Let us now have a closer look at the preferred model:

```
##                      Estimate Std. Error t value  Pr(>|t|)
## (Intercept)           18.6943     0.7841 23.8425 3.391e-19
## trans4.0               2.8535     1.4272  1.9994 5.612e-02
## trans4.1               0.3463     1.5718  0.2203 8.273e-01
## trans5.1              -2.0821     1.5243 -1.3660 1.837e-01
## I(wt - mean(wt))      -6.0842     0.7196 -8.4547 6.178e-09
## I((wt - mean(wt))^2)   1.3783     0.3596  3.8328 7.220e-04
```

In this model specification, the expected fuel efficiency for a vehicle with average weight and an automatic transmission and three forward gears is 18.69 *mpg*. For a car of any given weight with an automatic transmission system, when moving from 3 to 4 forward gears, the expected gain in fuel efficiency is 2.85 *mpg*. Moving from a 3 gear automatic system to a 4 gear manual system also yields a gain (of 0.35 *mpg*), but a move from a 4 gear automatic system to a 4 gear manual system results in a reduction in fuel efficiency of 0.35 - 2.85 = -2.51 *mpg*. Finally, a move from a 4 gear manual system to a 5 gear manual system results in a decrease in fuel efficiency of -2.08 -0.35 = -2.43 *mpg*. Thus, the highest fuel economy is obtained with a 4 gear automatic system. Note however that the coefficients for *trans4.1* and *trans5.1* are not significant, while the coefficient for *trans4.0* is only significant at the 0.1 level.

The coefficients for *wt* confirm that *mpg* is a quadratic, convex function of *wt - mean(wt)*, all other things being equal. Because of the quadratic term, the loss in fuel efficiency when weight increases depends on the reference point.

Diagnostics

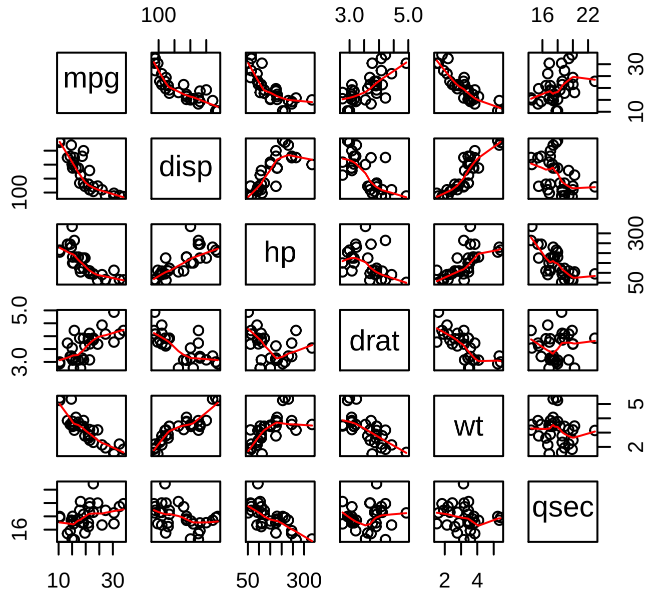
The 'Residuals vs Fitted' and the 'Scale Location' plots in Appendix suggests that the assumptions of linearity and equal variance of the error term are reasonable, but also that there are some outliers (Fiat 128, Pontiac Firebird and Toyota Corolla). From the 'Normal Q-Q' plot in Appendix, we can infer that the observed residuals follow closely the theoretical quantiles of a normal distribution, except for the aforementioned outliers.

The 'Cook's distance' plot in Appendix points to the following influential observations: Fiat 128, Pontiac Firebird and Toyota Corolla, while influence.measures() identifies the following:

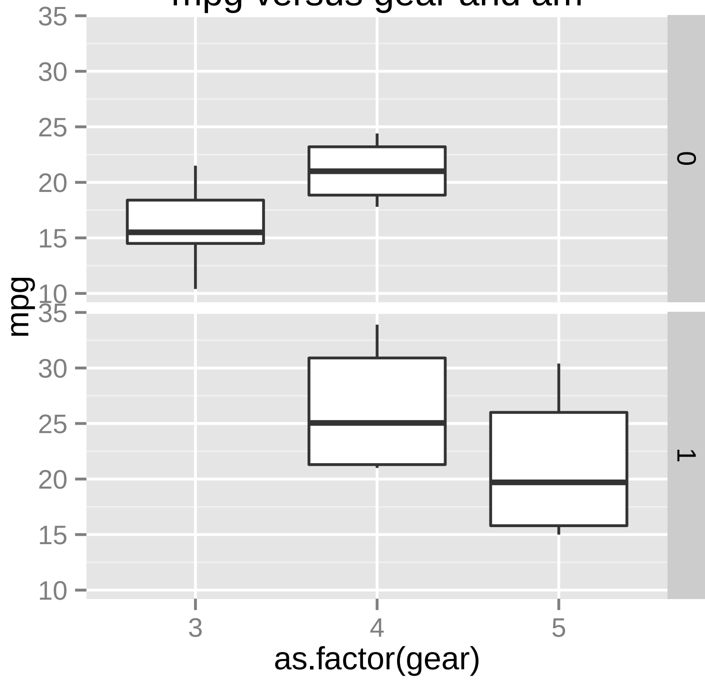
```
## [1] "Lincoln Continental" "Fiat 128"            "Lotus Europa"
## [4] "Maserati Bora"
```

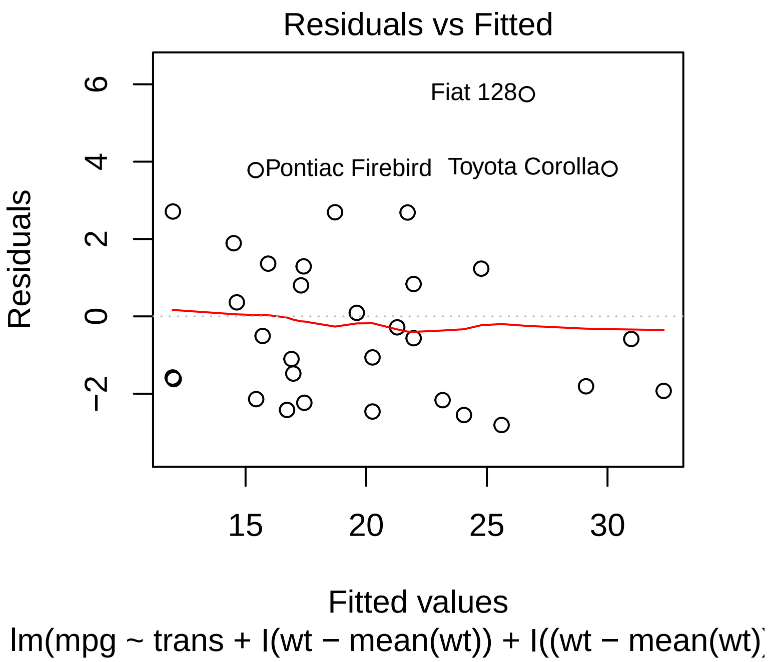
Appendix

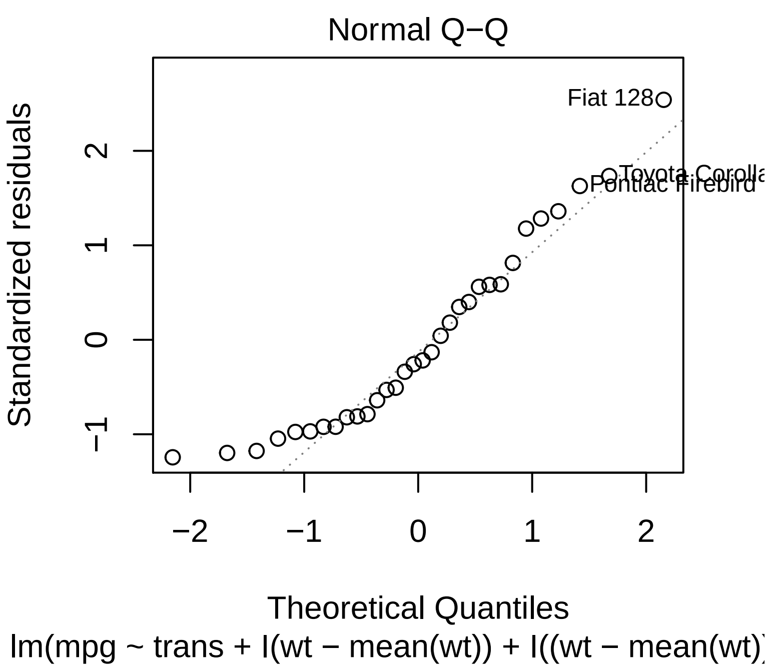
Plotting the continuous variables

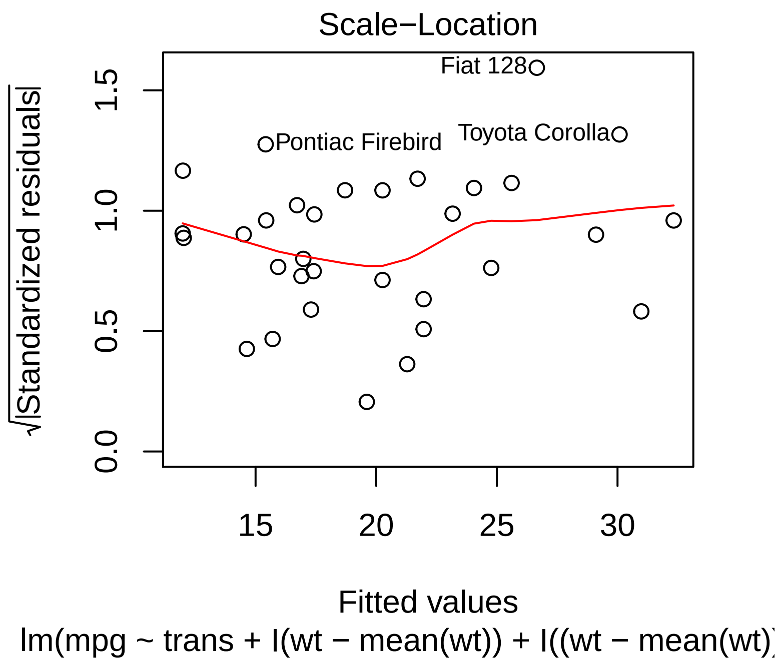


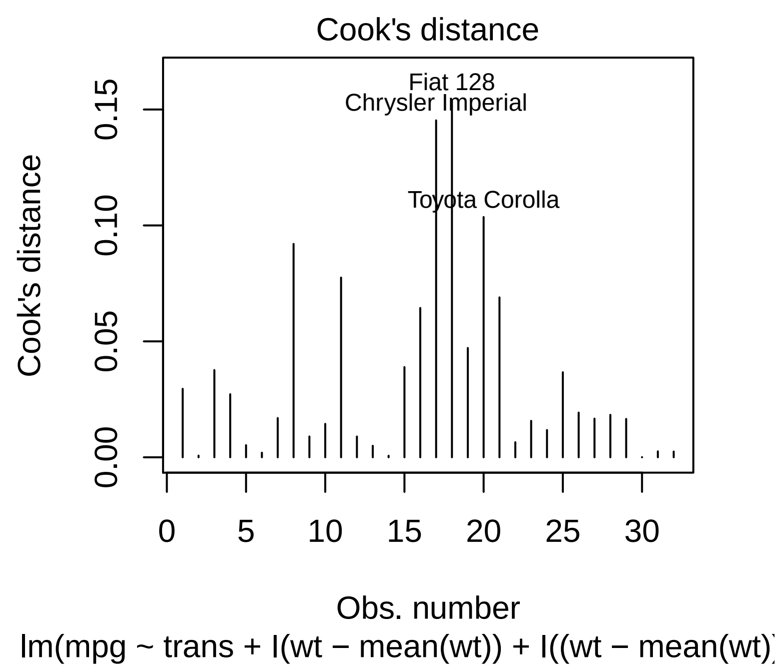
mpg versus gear and am




Residuals vs Fitted
Fiat 128
Pontiac Firebird
Toyota Corolla
Residuals
Fitted values
lm(mpg ~ trans + I(wt − mean(wt)) + I((wt − mean(wt)


Normal Q−Q
Fiat 128
Toyota Corolla
Pontiac Firebird
Standardized residuals
Theoretical Quantiles
lm(mpg ~ trans + I(wt − mean(wt)) + I((wt − mean(wt)


Scale−Location
√|Standardized residuals|
Fitted values
lm(mpg ~ trans + I(wt − mean(wt)) + I((wt − mean(wt)
Fiat 128
Pontiac Firebird
Toyota Corolla
0.0
0.5
1.0
1.5
15
20
25
30


Cook's distance
Cook's distance
Obs. number
lm(mpg ~ trans + I(wt − mean(wt)) + I((wt − mean(wt)
Fiat 128
Chrysler Imperial
Toyota Corolla
0.00
0.05
0.10
0.15
0
5
10
15
20
25
30