Transmission type impact on fuel economy

Maria A. Flores

24th October 2015

Executive summary

The main goal of this work is to explore the relationship between a set of variables and the fuel economy, measured in miles per gallon (mpg). Particulary, we want to know the effect, if any, that an automatic or manual transmission have for the mpg value and quantify the difference. In order to do this, we use the *mtcars* dataset. Firstly, we perform a brief exploratory analysis. Secondly, we try to find out some data relationships that could helps us in the selection of variables and fit different models. The resulting models tend to demonstrate that the type of transmission doesn't have an important effect on fuel economy.

Exploratory analysis

Size of the data, summary for the continous variables and hit table for the discrete ones:

```
## [1] "11 cols x 32 rows"
```

```
##        Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## mpg  10.400  15.420  19.200  20.090   22.80  33.900
## disp 71.100 120.800 196.300 230.700  326.00 472.000
## hp   52.000  96.500 123.000 146.700  180.00 335.000
## drat  2.760   3.080   3.695   3.597    3.92   4.930
## wt    1.513   2.581   3.325   3.217    3.61   5.424
## qsec 14.500  16.890  17.710  17.850   18.90  22.900
```

```
##   Values cyl vs am gear carb
## 1      0  NA 18 19   NA   NA
## 2      1  NA 14 13   NA    7
## 3      2  NA NA NA   NA   10
## 4      3  NA NA NA   15    3
## 5      4  11 NA NA   12   10
## 6      5  NA NA NA    5   NA
## 7      6   7 NA NA   NA    1
## 8      8  14 NA NA   NA    1
```

Model fitting

The strategy for selecting the variables to be included in the model was:

1. Fit mpg vs all the rest. In this model none of the variables was relevant (all the p-values > 0.05).

2. Create a correlation matrix and find those highly correlated with mpg, resulting on *wt, cyl, disp, hp* (negative relationship). This seems to confirm the common sense, but of course, this conclusion could be biased without considering efficiency variables.

3. Use the *step()* function to try and select the best model automatically.

4. Try to fit mpg vs each variable individually. In this model all the variables are potentially relevant (p-values < 0.05). Fit mpg with *am* as a factor with each of the other variables. Add an interaction term with *am*. Compare with the model offered by *step()*. Preselect the best models and compare them, taking in account the adjusted r-squared and residual standard error. We chose the model with the interaction term because it have more balanced values: *mpg ~ wt * factor(am)*

The outliers (high leverage + high influence) deletion process shows that the model doesn't significantly change. But, as the data give strong evidence about the weight being the most important variable in the mpg, and other specialised studies (see *References*) agree on this, is interesting to try to disentangle this effect. In the figures in the *Appendix*, adjusting with *am* reveals that the mean weight for the automatic vehicles are higher than the same value for the manual ones, impacting in the mpg. Reducing the dataset to those vehicles closer to the average weight balance the effect. If we repeat the process of fitting, in this case the best model is not as clear as before. We can choose the second one to be able to evaluate the *am* impact: *mpg ~ hp + factor(am)*

Model	p-values	adj.r.sq	sigma
mpg ~ wt	all highly relevant	0.7446	3.046
mpg ~ hp + factor(am)	all highly relevant	0.767	2.909
mpg ~ wt + qsec + as.factor(am)	>0.05 on itc, low rel. on am	0.8336	2.459
mpg ~ wt * factor(am)	all highly and medium relevant	0.8151	2.591
————————	——————————-	———	——-
mpg ~ cyl	all highly relevant	0.7467	1.57
mpg ~ hp + factor(am)	all highly and medium relevant	0.7091	1.682
mpg ~ cyl + disp + hp + vs + gear	>0.05 on vs, some low rel.	0.8415	1.242

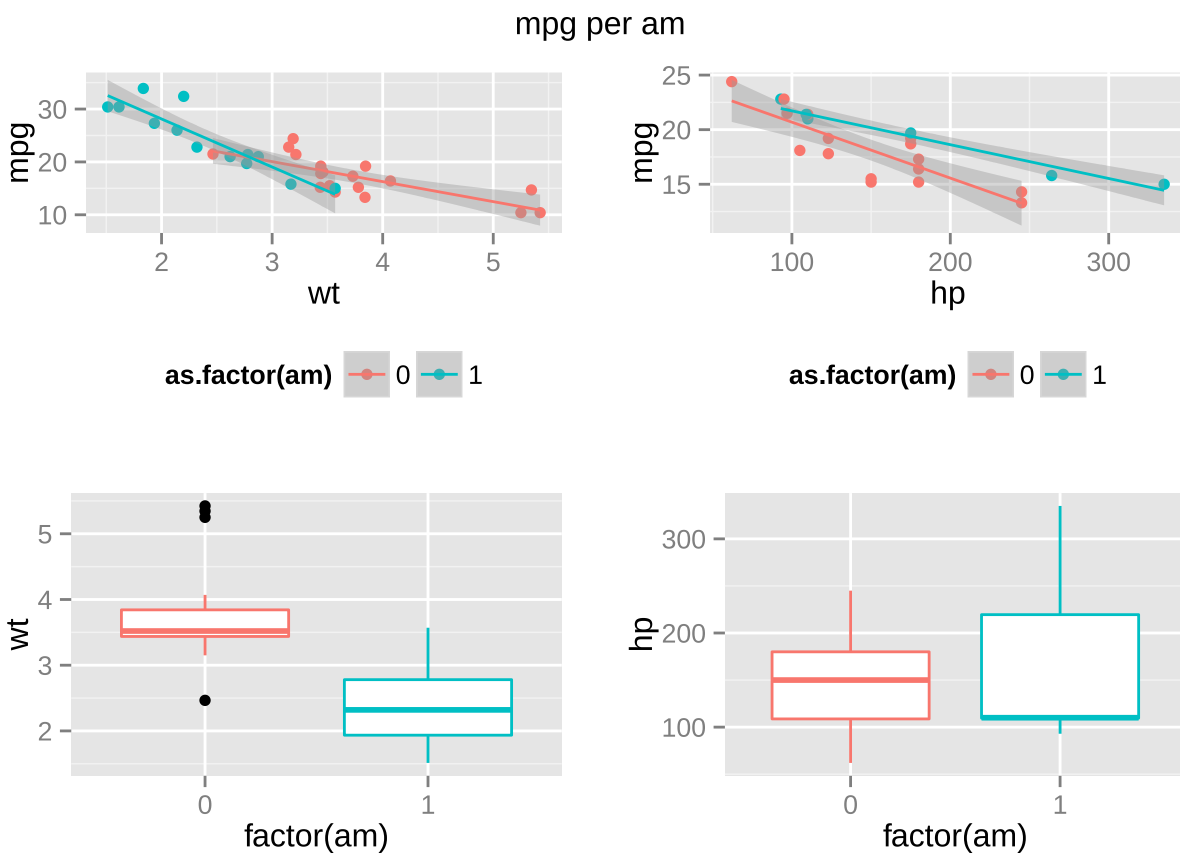
Results

Surely, those regressions aren't optimal, but allow to fit a reasonable model for the analysed vehicles and extract some conclusions. The intuition was that the transmission type could be irrelevant. The key for mpg is to make the engine work in the optimal consumption zone for each speed and rpm and, theoretically, this can be achieved with any transmission type. The data agrees: the manual transmission have a slightly negative impact on the mpg of -5.3 miles per gallon. If we disregard the weight effect, the effect is the opposite, improving +2.2 miles per gallon. In both cases the values are not specially relevant.

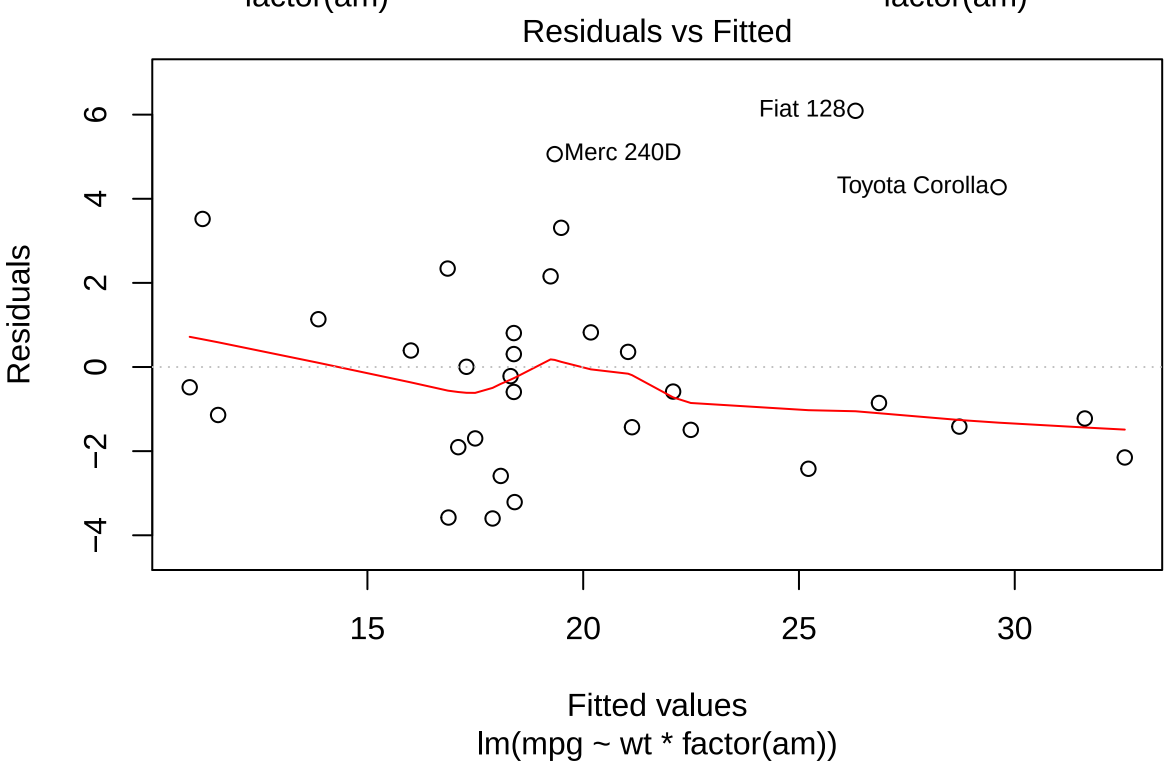
```
##                  Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)     31.416055  3.0201093 10.402291 4.001043e-11
## wt              -3.785908  0.7856478 -4.818836 4.551182e-05
## factor(am)1     14.878423  4.2640422  3.489277 1.621034e-03
## wt:factor(am)1 -5.298360  1.4446993 -3.667449 1.017148e-03
```

```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 24.09740831 0.924152023 26.075156 6.479920e-17
## hp          -0.03977392 0.005497814 -7.234496 5.316725e-07
## factor(am)1  2.22682106 0.771019980  2.888150 9.091235e-03
```

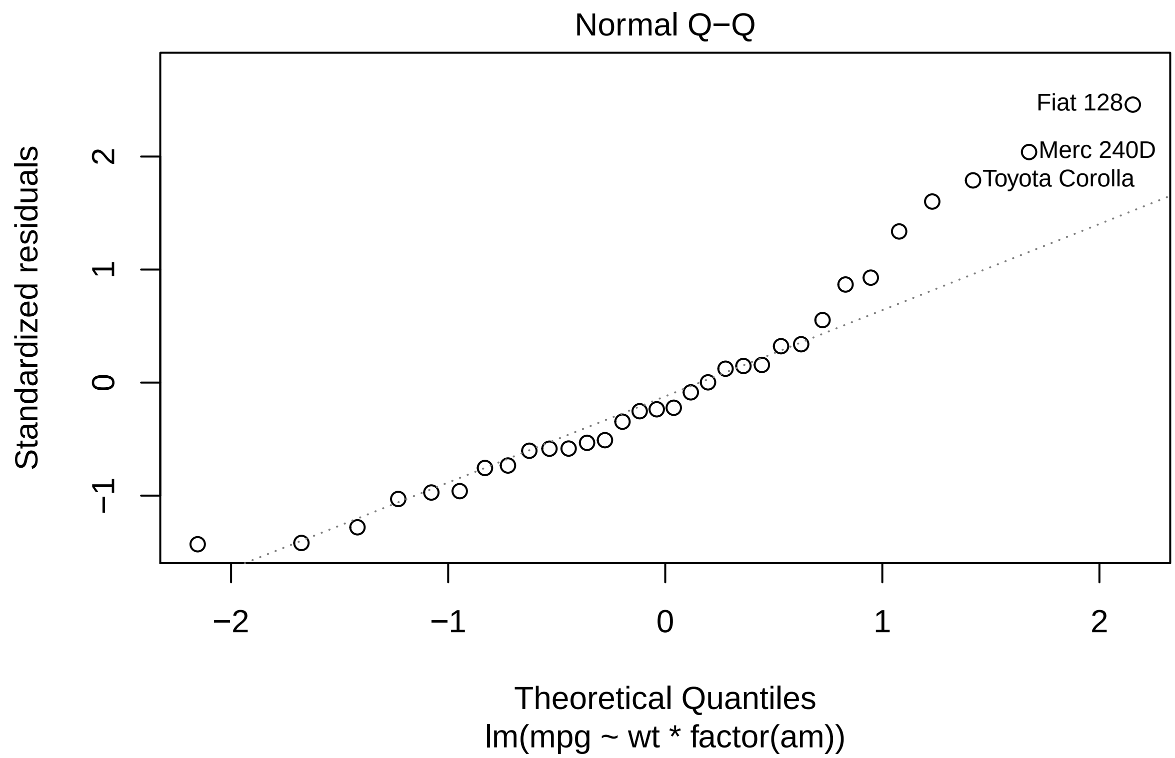
Appendix

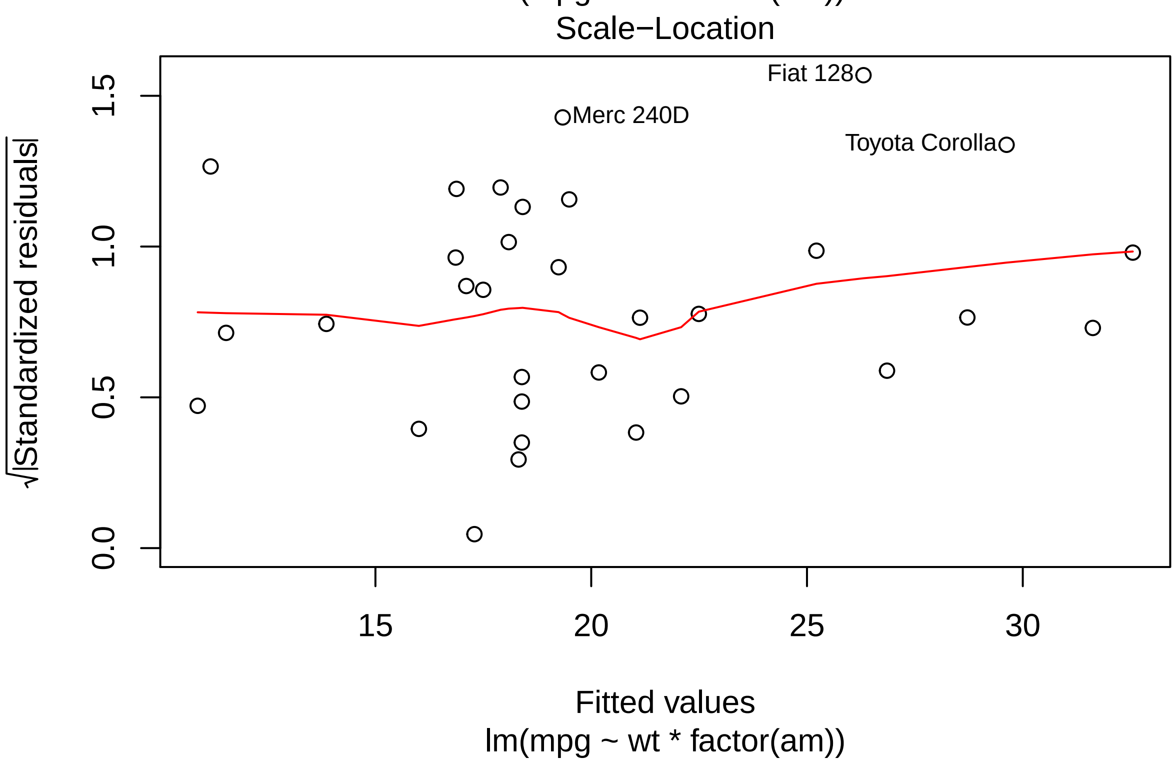


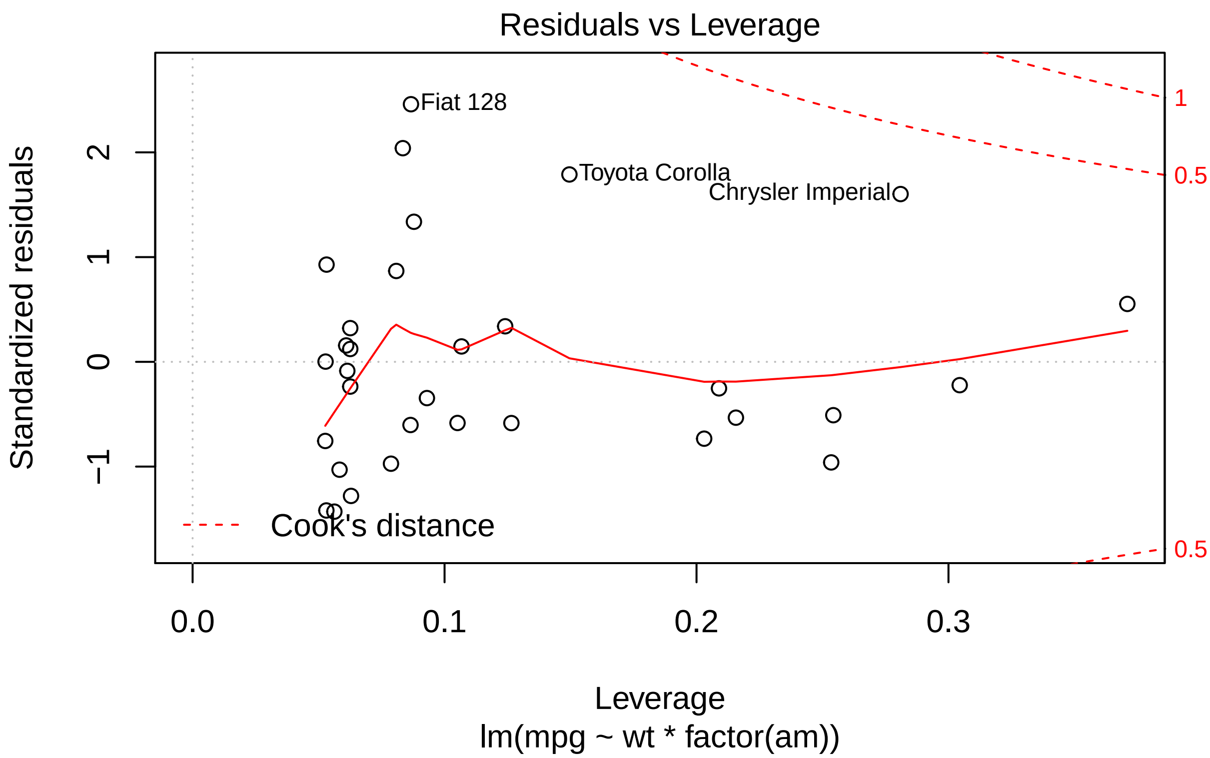
mpg per am
mpg
wt
as.factor(am)
0
1
hp
factor(am)

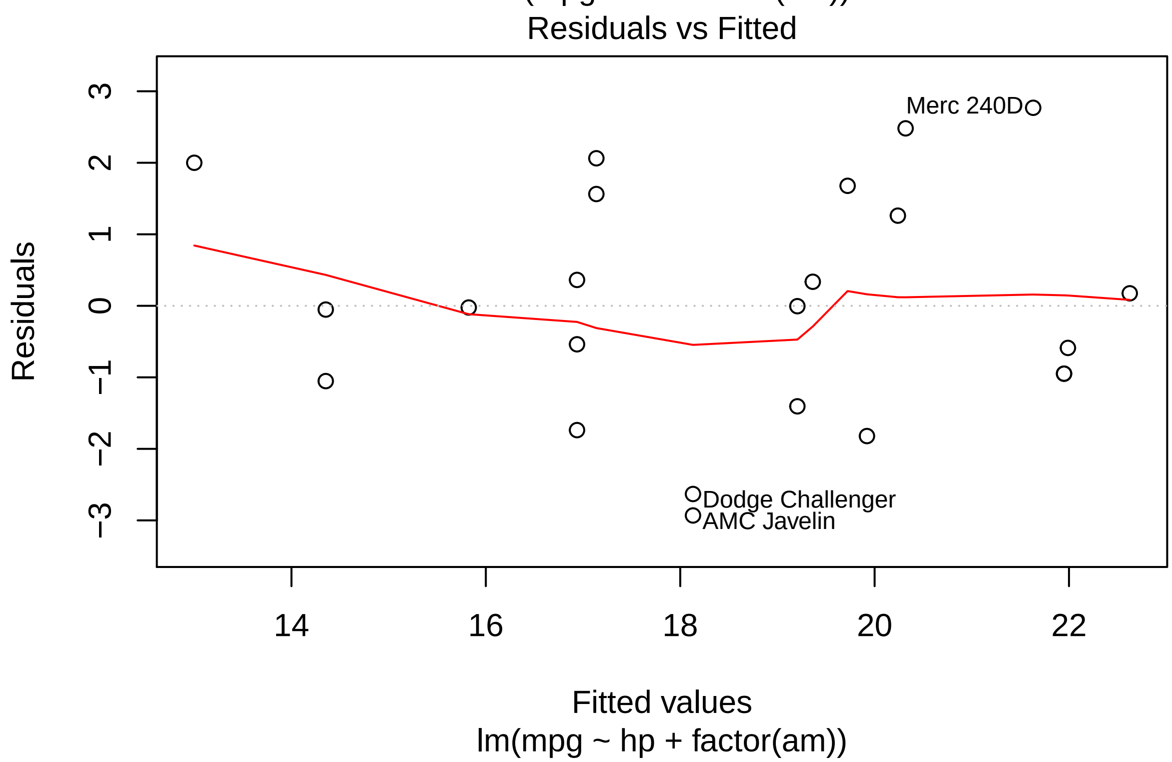


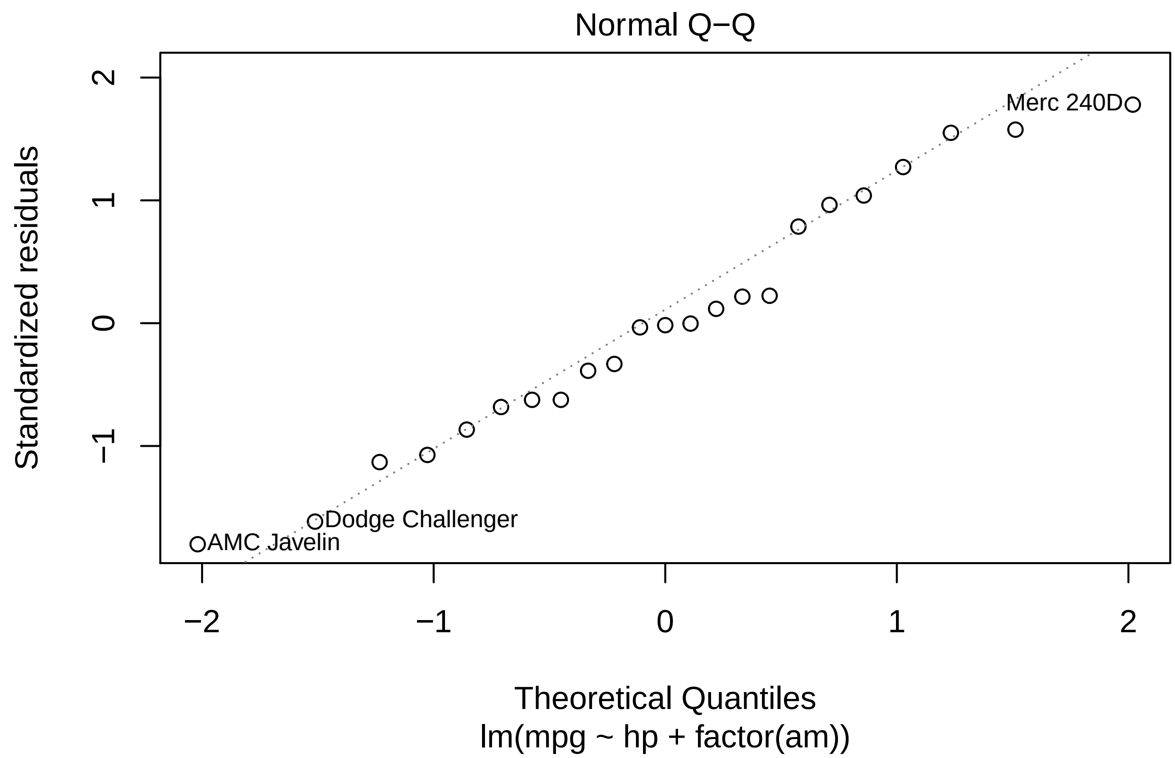
Residuals vs Fitted
Residuals
Fiat 128
Merc 240D
Toyota Corolla
Fitted values
lm(mpg ~ wt * factor(am))

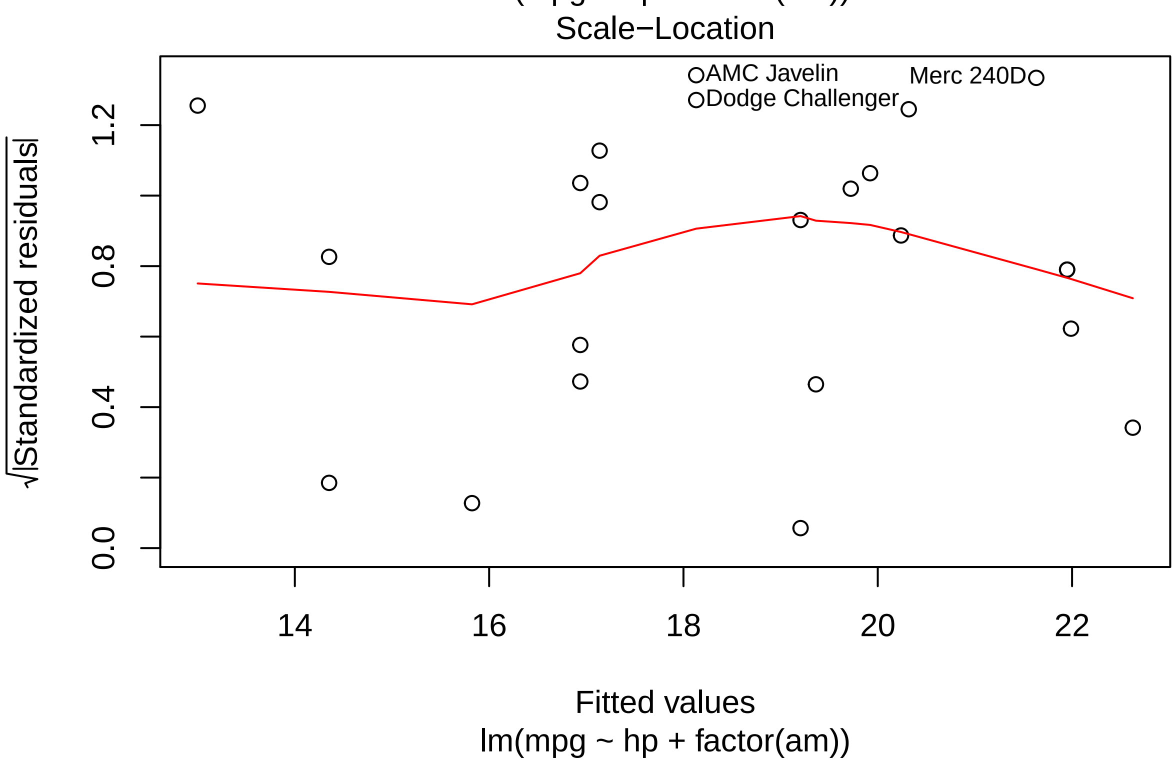

Normal Q−Q
Fiat 128
Merc 240D
Toyota Corolla
Standardized residuals
−1
0
1
2
−2
−1
0
1
2
Theoretical Quantiles
lm(mpg ~ wt * factor(am))

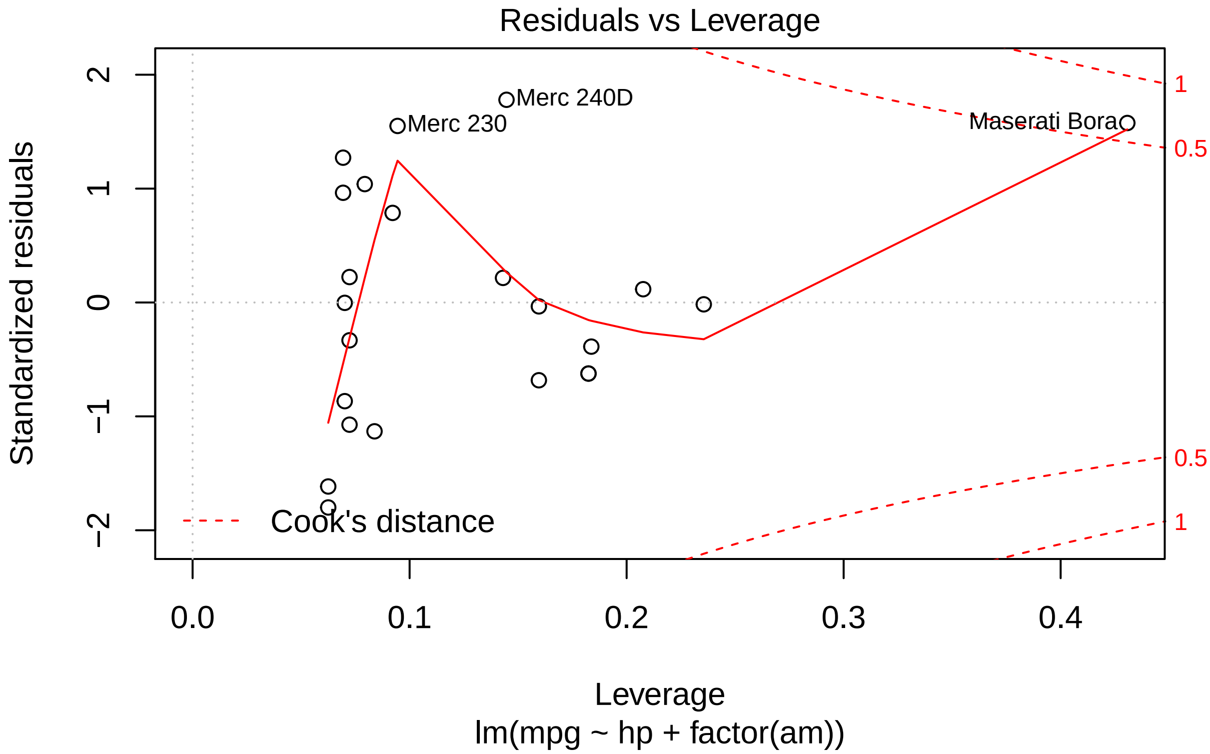

Scale−Location
Fiat 128
Merc 240D
Toyota Corolla
√|Standardized residuals|
0.0
0.5
1.0
1.5
15
20
25
30
Fitted values
lm(mpg ~ wt * factor(am))

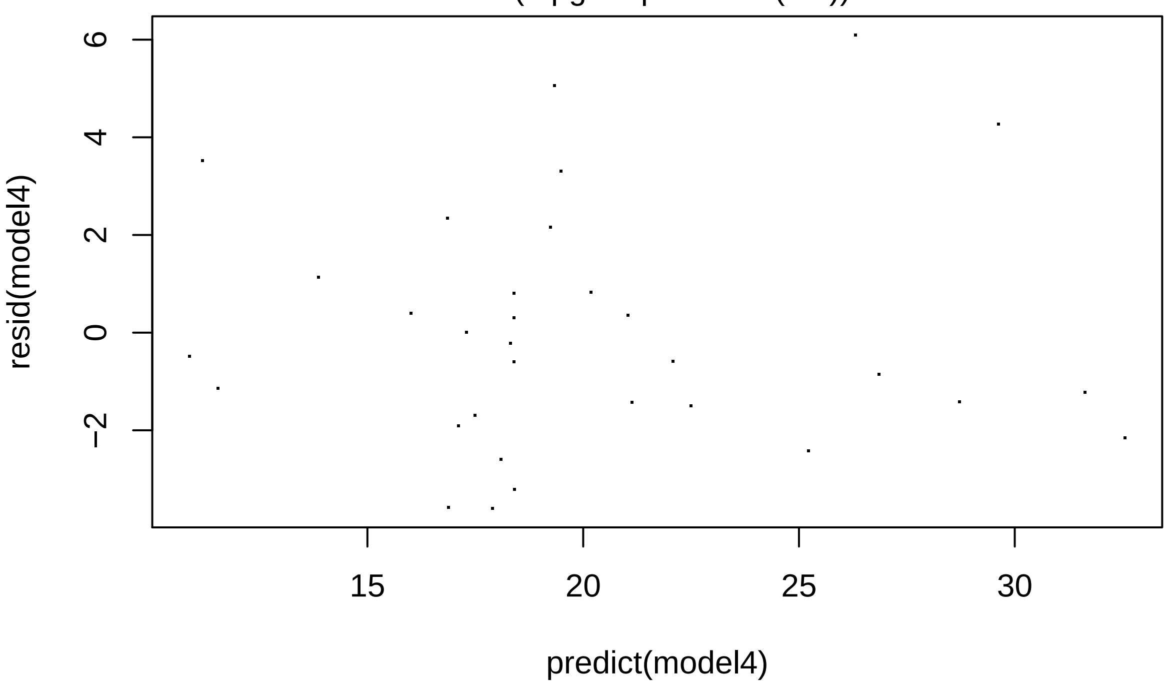

Residuals vs Leverage
Standardized residuals
Fiat 128
Toyota Corolla
Chrysler Imperial
Cook's distance
1
0.5
0.5
Leverage
lm(mpg ~ wt * factor(am))

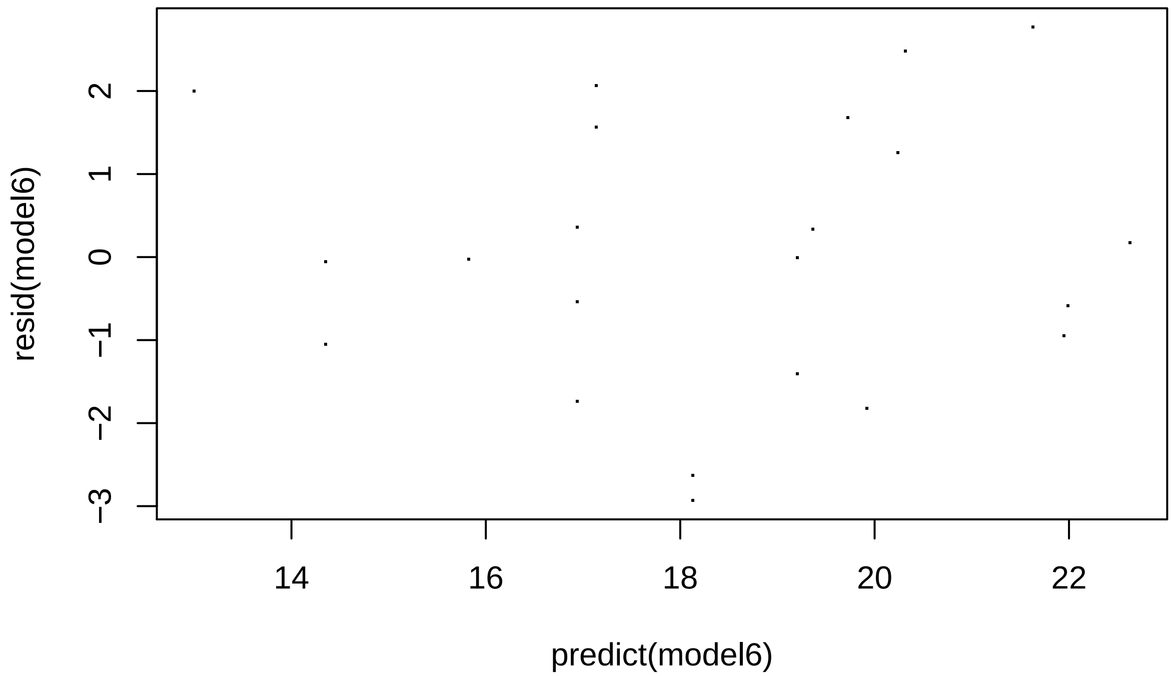

Residuals vs Fitted
Residuals
Merc 240D
Dodge Challenger
AMC Javelin
Fitted values
lm(mpg ~ hp + factor(am))


Normal Q−Q
Standardized residuals
Merc 240D
Dodge Challenger
AMC Javelin
Theoretical Quantiles
lm(mpg ~ hp + factor(am))


Scale−Location
√|Standardized residuals|
AMC Javelin
Dodge Challenger
Merc 240D
Fitted values
lm(mpg ~ hp + factor(am))


Residuals vs Leverage
Standardized residuals
Leverage
lm(mpg ~ hp + factor(am))
Cook's distance
Merc 240D
Merc 230
Maserati Bora
1
0.5
0.5
1
2
1
0
−1
−2
0.0
0.1
0.2
0.3
0.4


resid(model4)
predict(model4)
6
4
2
0
−2
15
20
25
30



References

1. "Factors affecting automotive fuel economy" US environmental protection agency, 1975 link
2. "What factors affect average fuel economy of US passenger vehicles?" Suman Gautman, 2010. Illinois Wesleyan University. link
3. "Prestaciones y consumo: así influyen el peso, el motor y la aerodinámica" Auto10 magazine. Antonio Roncero link