Regression Model Project
Exploring Data for Motor Trend

by Bruno Berrehuel*

December 19, 2016

1 Overview

The data was extracted from the 1974 Motor Trend US magazine, and comparises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles.[1] This study explores the relationship between a set of variables and miles per gallon (MPG), and answers the two following questions :

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

After statistical analysis of the *mtcars* dataset[2], *manual transmission is better for gas mileage*, and a gain between *7.031 and 21.1 mpg*, and a mean of *14.1 mpg*, with 95 % confidence, can be expected with a manual transmission in comparaison of an automatic transmission. BUT it depends on the weight of the vehicle too. The *weight* of the vehicle, the *qsec* (quarter mile time) and the interaction between *transmission* and *weight* parameters are to consider too. They define a linear regression model for MPG :

$$mpg = 9.72 + 14.08 * am - 2.94 * wt + 1.12 * qsec - 4.14 * am : wt$$

2 Data exploratory

2.1 Basic Exploratory

Graphic view of the gas mileage grouped by transmission type is presented in figure 1. We can see that the manual transmission appears to be better for mpg, as shown by the means of the two groups. The influence of the other parameters, in order to find the best regression model, has to be quantified.

3 Further analysis

3.1 comparing the means

The difference of the means, and most of it, the better performance of the manual transmission, is confirmed by the t.test function : the null hypothesis is rejected ($p < 0.05$) in favor of the alternative hypothesis that the mean of MPG for automatic transmission is less than the mean of MPG for manual transmission.

*Heavily inspired by the book R in action from Robert I. Kabacoff

[1] Source : https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html

[2] data exploratory, t.test confirmation and research of the best linear regression model with the *regsubsets* function of the *leaps* package

```
#compare auto and manual transmission MPG with a t.test :
t.test(mpg~am,mtcars,alt="l")$p.value

[1] 0.0006868192
```

3.2 influence of other parameters

The linear model with all the possible predictors can't lead to any conclusion because the p.values can't allow to reject that a coefficient may be null, as shown in table 3. However, with the use of the *regsubsets* function of the *leaps* package, the linear model can be limited to the following predictors : weight (wt), transmission (am) and quarter mile time (qsec), as shown in the figure 2.

3.3 fitting a model

With the information below, the linear model coefficients are shown in the table 4. The predictors are significant different from zero ($p < 0.05$), and the model seems good as shown with the plots in the figure 3. The model can be improved by considering interactions between transmission (am) and weight (wt), as shown in the figure 4, with the better coefficients shown in table 1, as their p-values are significants. This model describe 0.896 % of the variance.

```
bestFitInter <- lm(mpg~am*wt+qsec,mtcars)
xtable(summary(bestFitInter),
       caption="best fit model - interaction am:wt",
       label="noInter")
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.7231	5.8990	1.65	0.1109
am	14.0794	3.4353	4.10	0.0003
wt	-2.9365	0.6660	-4.41	0.0001
qsec	1.0170	0.2520	4.04	0.0004
am:wt	-4.1414	1.1968	-3.46	0.0018

Table 1: best fit model - interaction am:wt

Finally, the 95 % confidence intervals are the following for each predictors are shown in table 2.

```
xtable(confint(bestFitInter),
       caption="best fit model - confidence interval of the coefficents",
       label="bestFitConf")
```

	2.5 %	97.5 %
(Intercept)	-2.38	21.83
am	7.03	21.13
wt	-4.30	-1.57
qsec	0.50	1.53
am:wt	-6.60	-1.69

Table 2: best fit model - confidence interval of the coefficents

A Tables

```
library(xtable)
xtable(summary(lm(mpg~.,mtcars)),
       caption="linear regression with all predictors",
       label="all")
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	12.3034	18.7179	0.66	0.5181
cyl	-0.1114	1.0450	-0.11	0.9161
disp	0.0133	0.0179	0.75	0.4635
hp	-0.0215	0.0218	-0.99	0.3350
drat	0.7871	1.6354	0.48	0.6353
wt	-3.7153	1.8944	-1.96	0.0633
qsec	0.8210	0.7308	1.12	0.2739
vs	0.3178	2.1045	0.15	0.8814
am	2.5202	2.0567	1.23	0.2340
gear	0.6554	1.4933	0.44	0.6652
carb	-0.1994	0.8288	-0.24	0.8122

Table 3: linear regression with all predictors

```
bestFit <- lm(mpg~am+wt+qsec,mtcars)
xtable(summary(bestFit),
       caption="fit model - linear regression coefficients",
       label="bestFitCoef")
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.6178	6.9596	1.38	0.1779
am	2.9358	1.4109	2.08	0.0467
wt	-3.9165	0.7112	-5.51	0.0000
qsec	1.2259	0.2887	4.25	0.0002

Table 4: fit model - linear regression coefficients

B Figures

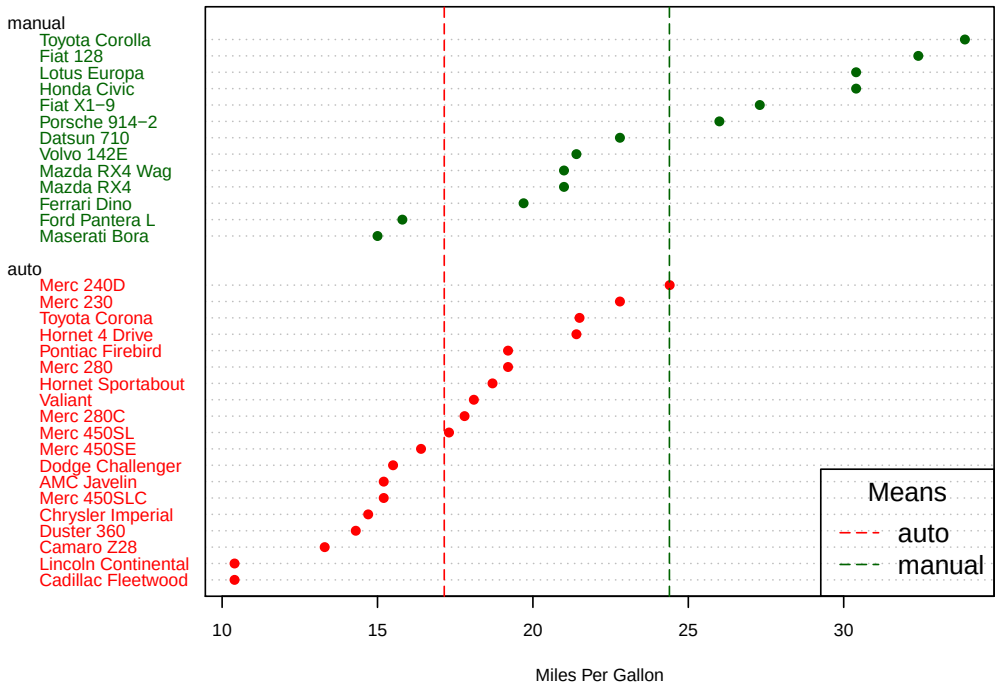

Figure 1: Gas mileage grouped by transmission type

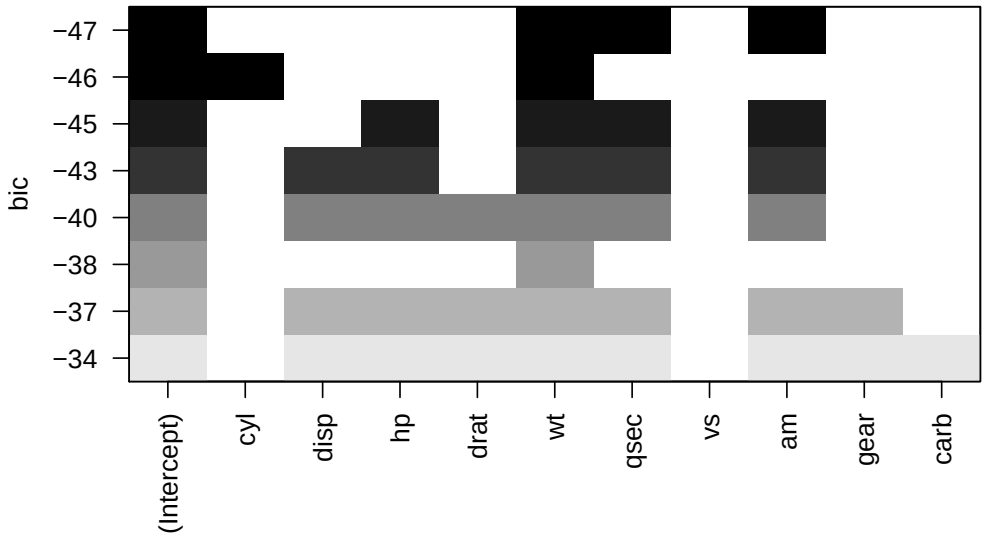

Figure 2: Determination of the best predictors

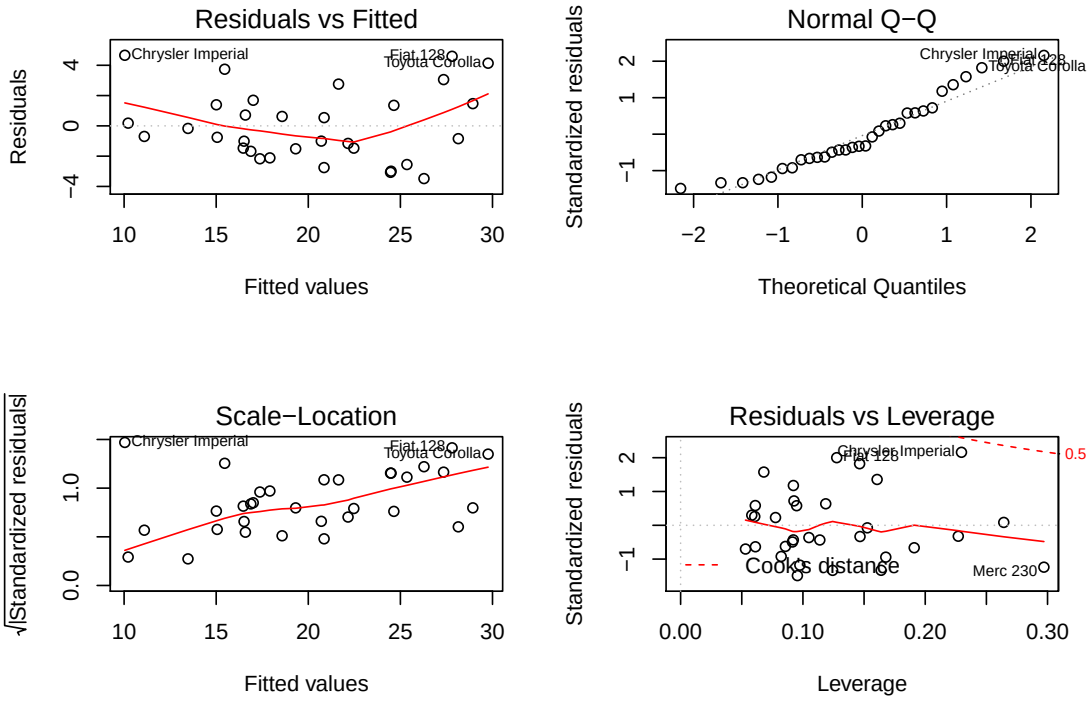

Figure 3: Diagnostic plots for the regression of mpg

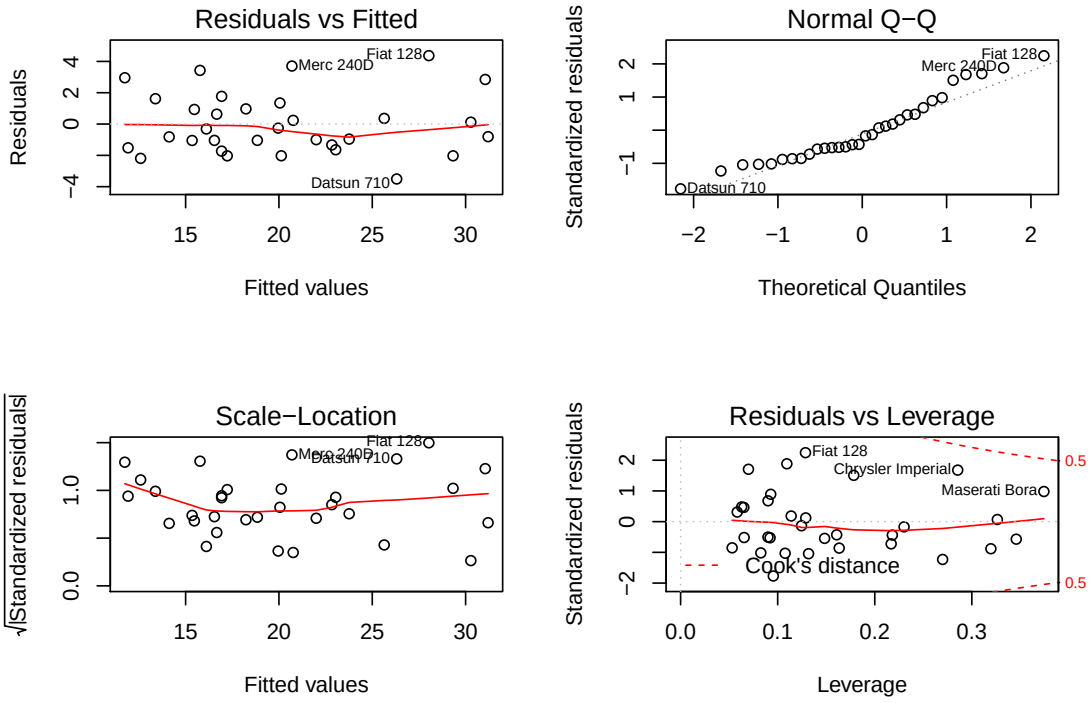

Figure 4: Improved fit model