Automatic transmission increases fuel consuption!

Sunday, November 23, 2014

Such result was obtained while answering these questions:

1. Is an automatic or manual transmission better for miles per gallon?
2. How much is quantitative difference if any?

Synopsis

The analysis is based on the `mtcars` data which was extracted from the 1974 *Motor Trend* US magazine. Data was "factorized" by numers of cylinders, automatic/manual transmission etc. Then was applied algorithm of finding the best model, based on Akaike information criterion. Then there were provided some test to verify that obtained model well to sample data. Was made a conclusion that automatic transmission increases fuel consumtion.

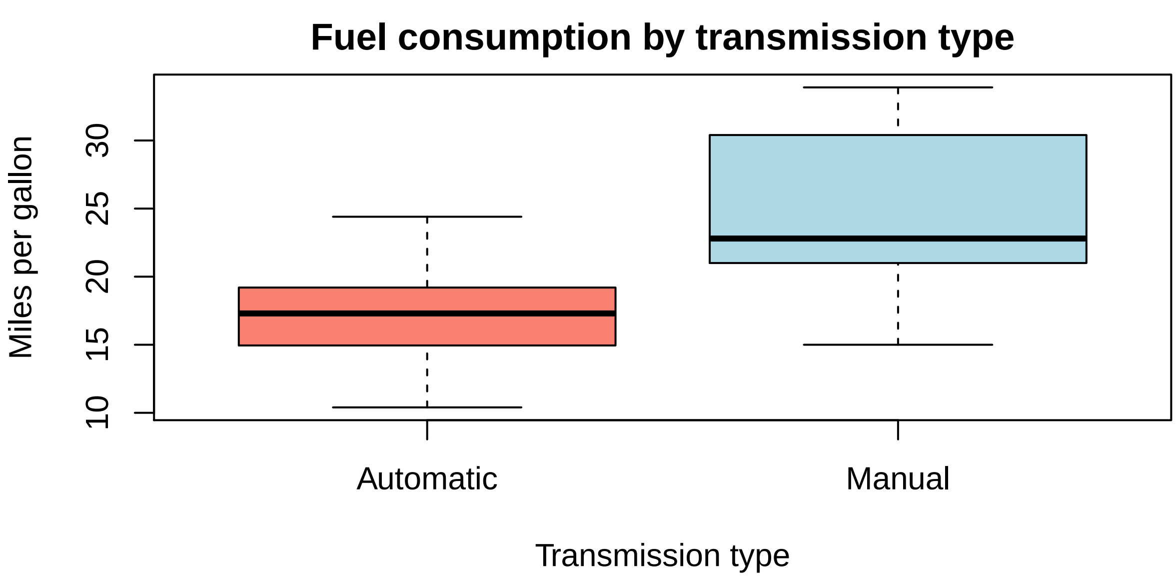
Exploratory analysis

Exploring `mtcars` dataset we can notice, that some variables can be considered as factors. So, first of all, perform such kind of factorization:

```
library(datasets)

mtcars$cyl  <- factor(mtcars$cyl)
mtcars$vs   <- factor(mtcars$vs)
mtcars$am   <- factor(mtcars$am, labels = c("Automatic", "Manual"))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Now, let's find out how fuel consumption (further - mpg) differs for different types of transmission.



The difference between mpg for automatic/manual transmission is noticalbe even with naked eye. While exploring pairwise `mtcars` plot we can also notice dependencies between mpg and other variables (plot in appendix). So, it's necessary to build a model.

Building a model

First of all, assume that our model is

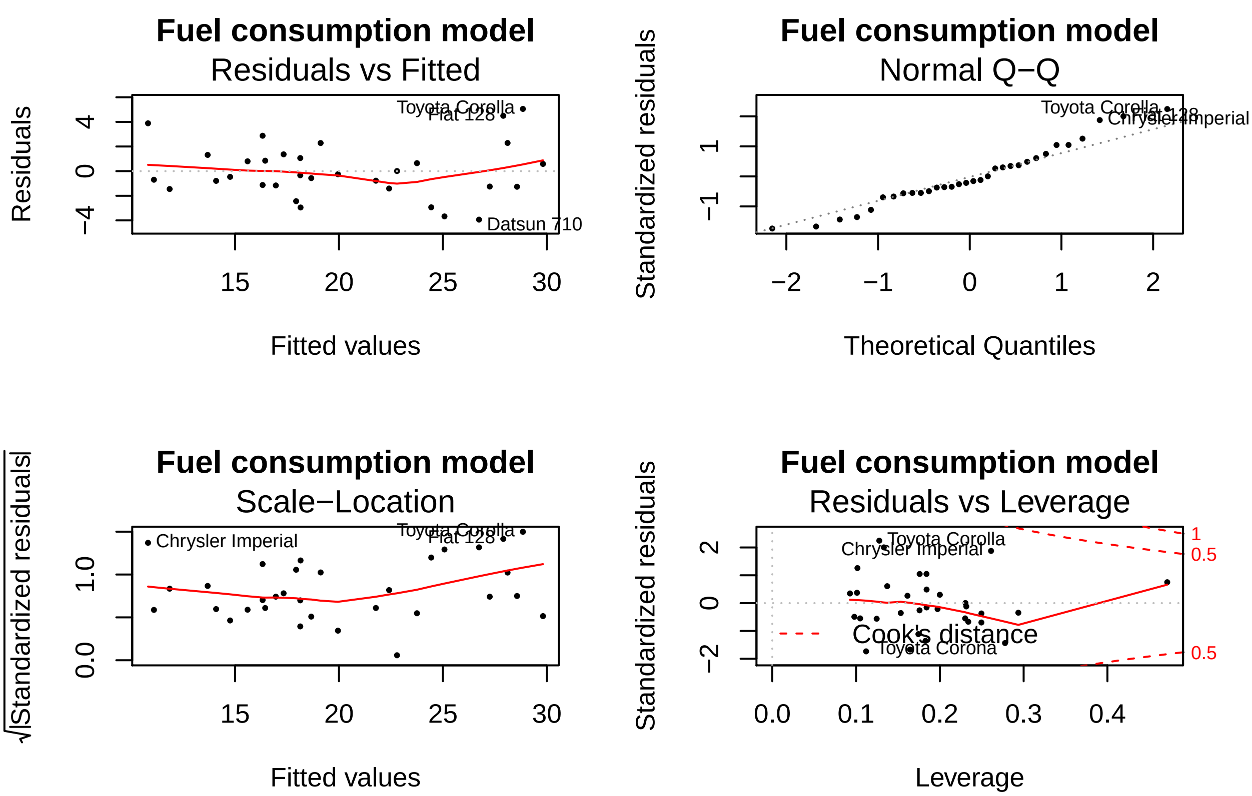
$$mpg = \beta_0 + \beta_1 cyl + \beta_2 disp + \beta_3 hp + \beta_4 drat + \beta_5 wt + \beta_6 qsec + \beta_7 vs + \beta_8 am + \beta_9 gear + \beta_{10} carb + \epsilon$$

Now we have to find out significant predictors i.e. throw away that $\beta_i \approx 0, i = \overline{1, n}$. For this we can apply model selection algorithm, which compares models generated from base and select the best by Akaike information criterion.

```
fit <- lm(mpg ~ ., data = mtcars)
fit <- step(fit, trace = FALSE) ## Applying algo
fit
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Coefficients:
## (Intercept)         cyl6         cyl8           hp           wt
##     33.7083      -3.0313      -2.1637      -0.0321      -2.4968
##    amManual
##      1.8092
```

And appropriate plot



Remarks:

1. Residuals vs. Fitted seems has random pattern.
2. Observations on Q-Q plot are close to line, so residuals are approximately normally distributed.
3. There are few outliers.

Statistical inference

Assume following hypothesis $H_0 : MPG_{Auto} = MPG_{Manual}$ and $H_a : MPG_{Auto} \neq MPG_{Manual}$. Let's provide test.

```
t.test(mpg ~ am, data = mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.767, df = 18.33, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.28  -3.21
## sample estimates:
## mean in group Automatic    mean in group Manual
##                   17.15                   24.39
```

Accepting alternative.

Conclusion

Answers to the main questions are based on the best fit model :

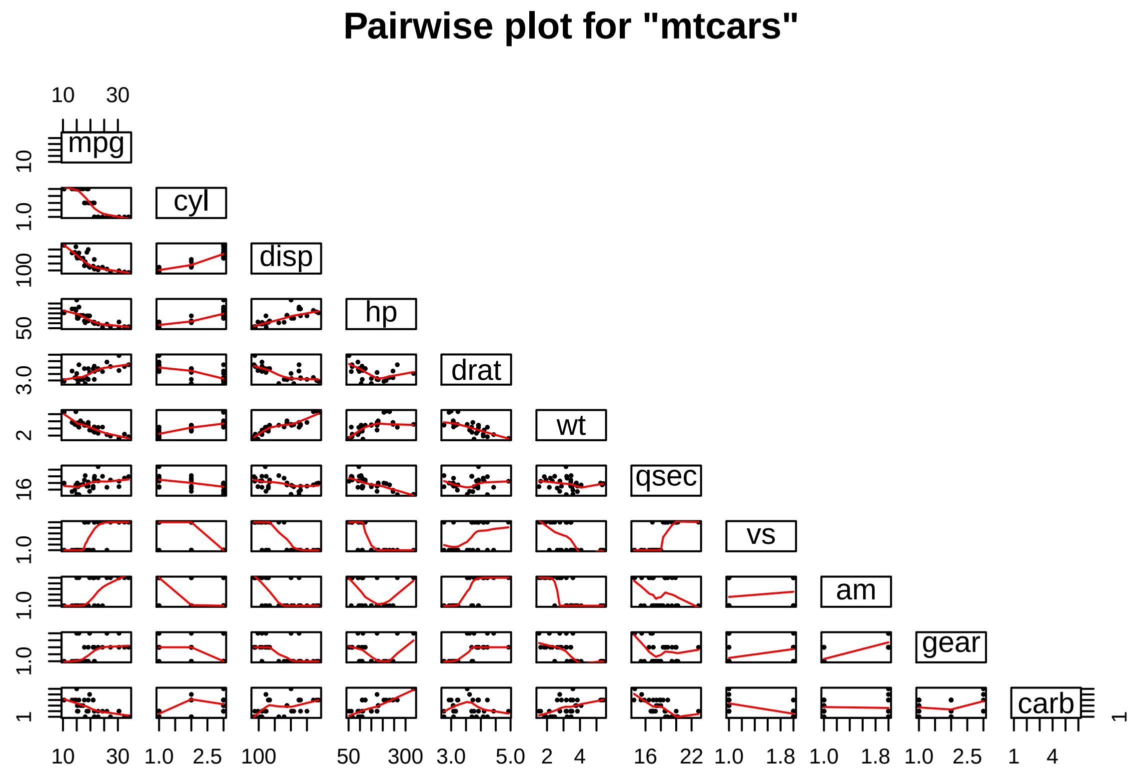
$$mpg = 0 + 33.71cyl_4 + 30.68cyl_6 + 31.54cyl_8 - 0.03hp - 2.5wt + 1.81am_{manual}$$

1. Cars with manual transmission has lower fuel consumption (higher mpg, adjusted by number of cylinders, weigth and horsepower).
2. Cars with manual transmission can drive 1.8 miles further than cars with automatic transmission per gallon of fuel.

Other conclusions:

3. Horspower has negligable impact on fuel consumption (mpg).
4. mpg will decrease by 2.5 for 1000lb increase in weight.

Appendix


Pairwise plot for "mtcars"
mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb
10
30
1.0
2.5
100
50
300
3.0
5.0
2
4
16
22
1.8
1