# Automatic transmission cars consume more fuel than manual transmission ones

*Ion Scerbatiuc*

*May 21, 2016*

## Summary

By using the `mtcars` dataset from the `Motor Trend` magazine, we're trying to understand what is the relationship between the transmission type and the mileage of cars and how can we quantify it.

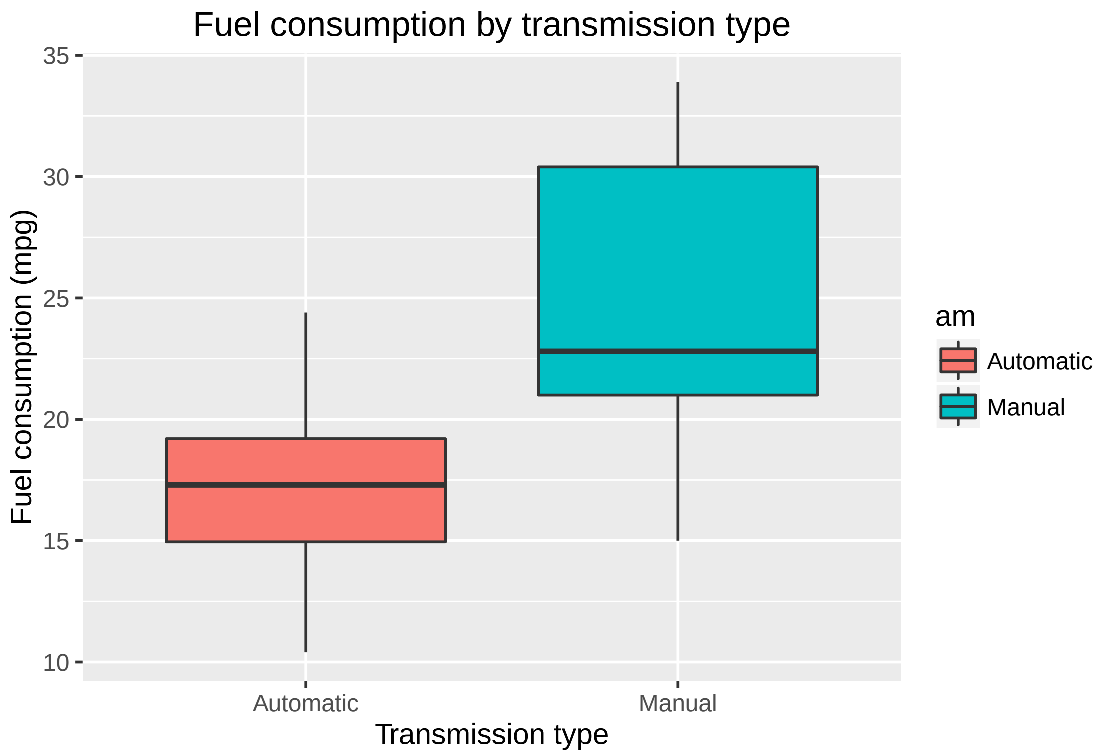
We start with an exploratory analysis and a simple linear model that shows and average increase of fuel consumption for automatic cars of `7.25 mpg` compared to manual cars. After using a stewise model selection algorithm we find a better model that explains about `83.4%` of the mileage variation. Using that model, we conclude that automatic cars consume on average about `2.94 mpg` more than manual cars, given that all the other car caracteristics are kept constant.

## Exploratory Analysis

```
library(datasets)
library(ggplot2)
library(dplyr)
library(car)
data(mtcars)
```

The first step is to identify the variables we have and what values they take (Appendix 1). Next we want to explore what is the measured fuel consumption by transmission type, to see if there is indeed some level of corelation between the two.

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
ggplot(mtcars, aes(x = am, y = mpg, fill = am)) + geom_boxplot() +
    xlab("Transmission type") + ylab("Fuel consumption (mpg)") +
    ggtitle('Fuel consumption by transmission type')
```




From the plot above we can infer that automatic cars are consuming on average more fuel than manual transmission cars. But is that result statistically significant? To answer that question, let's fit a linear model whit `mpg` as the outcome and the transmission type as the predictor.

```
fit <- lm(mpg ~ am, mtcars)
summary(fit)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## amManual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

By inspecting the coefficients table we can conclude that manual cars have an expected mileage `7.25 mpg` higher than the expected mileage of automatic cars. The result is also statistically significant because the p-value for the associated t-test is lower than `0.05`.

However, if we look at the value of R-squared, we see that this model only explains about `33.85%` of the variation. There has to be a better model that explains more of the variation in mileage, so let's dig deeper using a multivariate linear model.

## Model Selection

To find the best model from the variables in the dataset, let's use a stepwise model fit algorithm.

```
bestFit <- step(lm(data = mtcars, mpg ~ .), trace=0, steps=1000)
summary(bestFit)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

From the results above it seems that by adding the vehicle weight and acceleration to the initial model, we obtain a better model that explains about `83.4%` of the mileage variation. To validate the result, we performed an analysis of variance and found a statistically significant difference between the two models (Appendix 2). We also performed residual diagnostics and found that the residuals are normally distributed and mostly patternless (Appendix 3).

## Conclusions

The interpretation of the coefficients in the above table is that manual transmission cars consume about `2.94 mpg` less fuel that automatic transmission cars, holding all the other caracteristics constant. That is, if we take two cars, with automatic and manual transmissions respectively, with the same weight and acceleration, we expect the automatic car to consume `2.94 mpg` more fuel than the manual car.

## Appendixes

### Appendix 1 - Dataset definition

```
A data frame with 32 observations on 11 variables.
```

```
[, 1]   mpg Miles/(US) gallon
[, 2]   cyl Number of cylinders
[, 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
[, 8]   vs  V/S
[, 9]   am  Transmission (0 = automatic, 1 = manual)
[,10]   gear    Number of forward gears
[,11]   carb    Number of carburetors
```

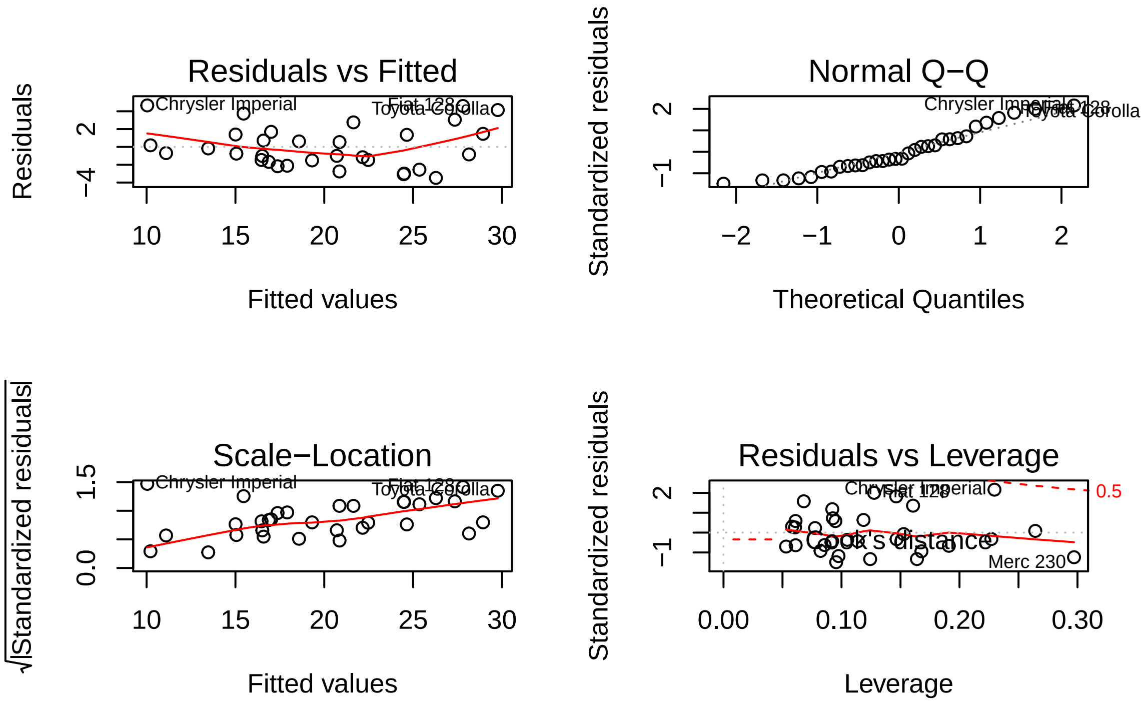
### Appendix 2 - Analysis of variance for the fitted model

```r
anova(fit, bestFit)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Appendix 3 - Residual diagnostics of the fitted model

```r
par(mfrow = c(2,2))
plot(bestFit)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Cook's distance
Merc 230
0.5