# Regression Models Course Project

*Egor Ignatenkov*

*23.11.2014*

## Relationship between type of transmission and fuel consumption

### Executive summary

In this report we're going to look on the relationship between type of transmission and fuel consumption. Although at first it may seem that using manual transmission significantly increases miles per gallon index, we will show that in reality this change is introduced by other parameters, such as horsepower and weight.

### Research

First, let's plot a simple boxplot to compare miles per gallon (mpg) in relation to type of transmission (0 - automatic, 1 - manual):

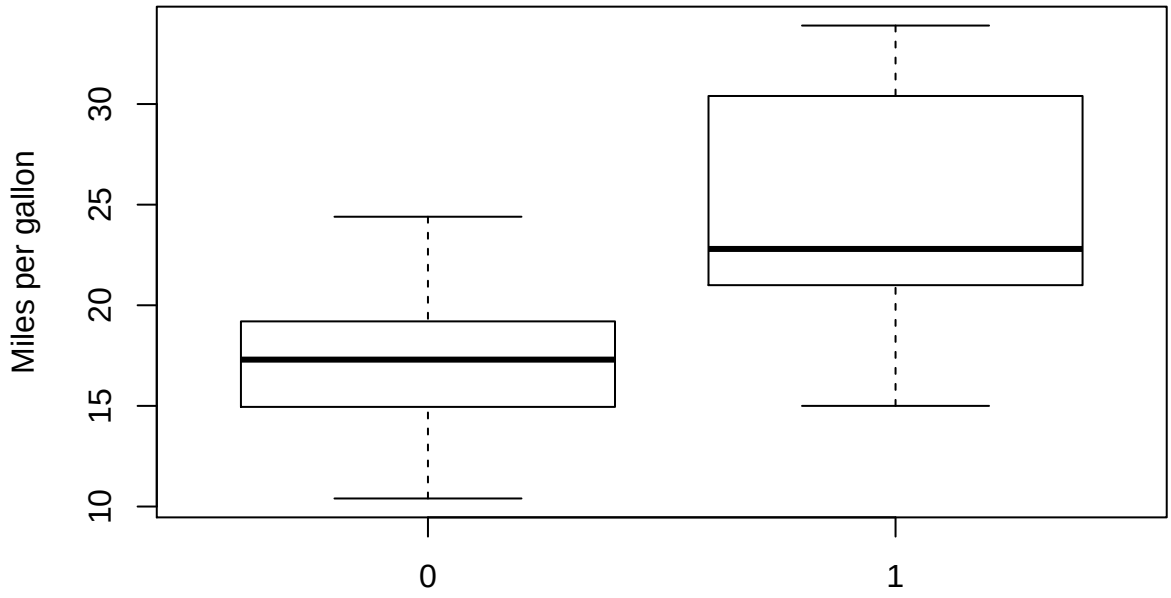



Figure 1. Miles per gallon
for automatic and manual transmission

It seems that manual transmission implies much higher level of mgp. Let's check it with linear regression:

```
fit<-lm(mpg~factor(am),data=mtcars)
summary(fit)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## factor(am)1  7.244939   1.764422  4.106127 2.850207e-04
```

Now let's check some more complicated linear models with more parameters. E.g., it seems logical that mpg can higly depend on weight and horsepower of the car. Let's include these parameters into model and make a comparison:

```
fit2<-update(fit,mpg~factor(am)+wt)
fit3<-update(fit,mpg~factor(am)+wt+hp)
anova(fit,fit2,fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt
## Model 3: mpg ~ factor(am) + wt + hp
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 68.734 5.071e-09 ***
## 3     28 180.29  1     98.03 15.224 0.0005464 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Seems that model with 3 parameters is significantly better. Let's look closer:

```
summary(fit3)$coef
```

```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 34.00287512 2.642659337 12.866916 2.824030e-13
## factor(am)1  2.08371013 1.376420152  1.513862 1.412682e-01
## wt          -2.87857541 0.904970538 -3.180850 3.574031e-03
## hp          -0.03747873 0.009605422 -3.901830 5.464023e-04
```

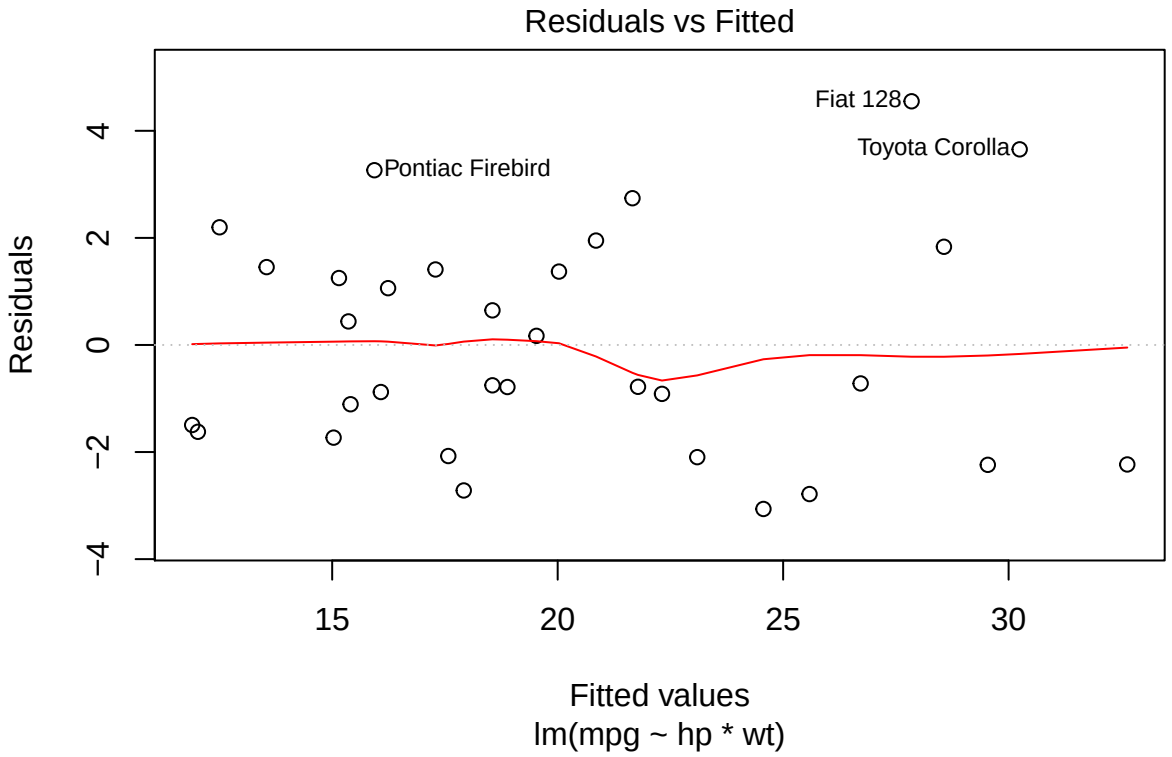
Now it seems that transmission type is quite insignificant (p-value for it is big. Let's look on the model without transmission:

```
fit4<-lm(mpg~hp*wt,mtcars)
summary(fit4)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 49.80842343 3.60515580 13.815887 5.005761e-14
## hp          -0.12010209 0.02469835 -4.862758 4.036243e-05
## wt          -8.21662430 1.26970814 -6.471270 5.199287e-07
## hp:wt        0.02784815 0.00741958  3.753332 8.108307e-04
```

Let's check our residual plot:

```
plot(fit4,which=1)
```




Residuals appear uncorrelated with the fit.

## Conclusion

Main influence on miles per gallon index is done by weight of the car and its horsepower. Having these two parameters in considerations makes type of transmission insignificant.