# Analysis of the Impact of the Transmission Type on Miles Per Gallon

*joergandi*

*17 August 2015*

## Overview

The following analysis of the impact of the transmission type on miles per gallon is based on the R built-in dataset mtcars, extracted from the 1974 Motor Trend US magazine. The objective is to determine if an automatic or manual transmission is better for MPG and to quantify that difference.

## Executive Summary

Although the data indicates that using a manual instead of an automatic transmission would increase MPG by approximately 7.2, this statement only holds if all other parameters which influence MPG are ignored. In fact, fuel consumption can be explained to a large extent by weight and horsepower alone. If one would like to buy a 1973/74 car model with an expected decent consumption, one should rather focus on car weight and horsepower than on the transmission type.

## Exploratory data analysis

Transmission type seems to have an impact on MPG if no other variables are considered, **cf. appendix figure 1**; however, other variables seem to impact MPG as well, **cf. appendix figure 2**.

## Quantification of transmission type impact on mpg

The plots suggests that using a manual transmission increases MPG significantly, which is confirmed by the small p-value of a t-test, allowing to reject the hypothesis of no impact.

```r
data(mtcars)
t.test(mpg~am,data=mtcars,conf.level=0.95)$p.value
```

```
## [1] 0.001373638
```

We fit a linear model to the factor variable of transmission type and mpg.

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

```
##              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
```

The first coefficient in the second row indicates that switching to manual transmission (am==1) will result in an average increase in MPG by 7.2449393 while assuming all other variables fixed.

### Which transmission is better for MPG?

```
fit2<-lm(mpg~wt*hp ,data=mtcars)
fit3<-update(fit2,mpg~wt*hp +factor(am))
anova(fit2,fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt * hp
## Model 2: mpg ~ wt + hp + factor(am) + wt:hp
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     28 129.76
## 2     27 129.72  1  0.042292 0.0088 0.9259
```

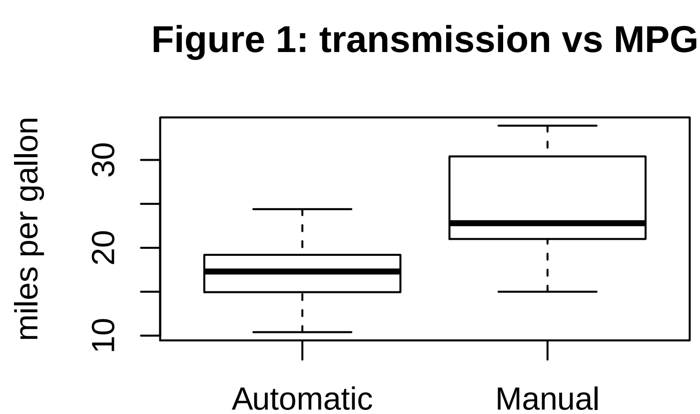
The exploratory data analysis already showed a significant impact of weight and horsepower on MPG. A model selection based on an analysis of the variance of different models (**cf. appendix figure 3**) suggests that a linear model comprising the interaction of weight and horsepower explains 87.2416967% of variance in MPG. The fitting of this model is reasonable when looking at the residuals and their variance which do not exhibit a systematic trend, **cf. appendix figure 4**. Adding the dependency on the transmission type factor variable to this model does not increase the explained variance (86.7734792% ) and the high p-value of a comparative anova analysis suggests that this more complex model is not a better choice (the null hypothesis of “adding the transmission variable has no impact” cannot be rejected). **This suggests that transmission type should not be used as a main indicator to draw conclusions on MPG.**

### Conclusion

Without considering the influence of other variables, using a manual instead of an automatic transmission increases mpg on average by 7.2449393. However, the variance in mpg in this dataset can be explained to a large extent by weight and horsepower and the transmission type seems to have a small impact when considering the impact of other variables.

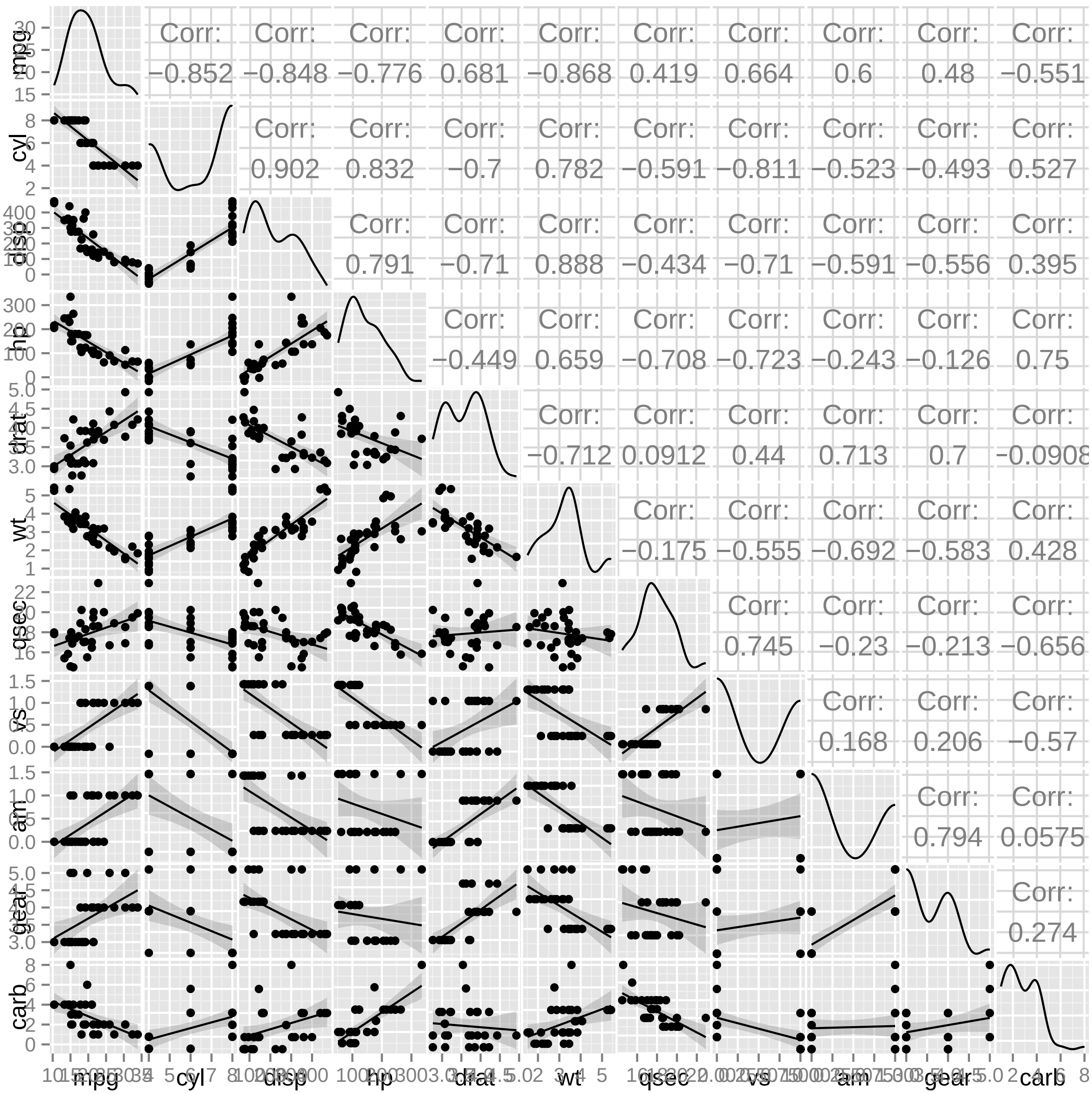
### Appendix

```
boxplot(mpg~factor(am),data=mtcars, #xlab="transmission type",
        names = c("Automatic", "Manual"),
       ylab="miles per gallon", main="Figure 1: transmission vs MPG")
```



```r
require(GGally)
require(ggplot2)
g=ggpairs(mtcars,lower=list(continuous="smooth"),params=c(method="loess"),
         title="Figure 2: Exploratory analysis of variable interactions")
g
```

Figure 2: Exploratory analysis of variable interactions




```r
#Figure 3: Comparing different linear models using analysis of variance
#  with an increasing number of variables
m0<-lm(mpg~wt+hp,data=mtcars)
m1<-update(m0,mpg~wt+hp+factor(cyl))
m2<-update(m0,mpg~wt+hp+factor(am))
m3<-update(m0,mpg~wt+hp+factor(cyl)+factor(am))
```

```r
m4<-lm(mpg~wt*hp,data=mtcars)
anova(m0,m1,m2,m3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + hp
## Model 2: mpg ~ wt + hp + factor(cyl)
## Model 3: mpg ~ wt + hp + factor(am)
## Model 4: mpg ~ wt + hp + factor(cyl) + factor(am)
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1     29 195.05
## 2     27 160.78  2    34.270 2.9499 0.07005 .
## 3     28 180.29 -1   -19.513 3.3594 0.07830 .
## 4     26 151.03  2    29.265 2.5191 0.10000 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(m3)
```

```
##
## Call:
## lm(formula = mpg ~ wt + hp + factor(cyl) + factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.70832    2.60489  12.940 7.73e-13 ***
## wt           -2.49683    0.88559  -2.819  0.00908 **
## hp           -0.03211    0.01369  -2.345  0.02693 *
## factor(cyl)6 -3.03134    1.40728  -2.154  0.04068 *
## factor(cyl)8 -2.16368    2.28425  -0.947  0.35225
## factor(am)1   1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

```r
summary(m4)
```

```
##
## Call:
## lm(formula = mpg ~ wt * hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.0632 -1.6491 -0.7362  1.4211  4.5513
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.80842    3.60516  13.816 5.01e-14 ***
## wt          -8.21662    1.26971  -6.471 5.20e-07 ***
## hp          -0.12010    0.02470  -4.863 4.04e-05 ***
## wt:hp        0.02785    0.00742   3.753 0.000811 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.153 on 28 degrees of freedom
## Multiple R-squared:  0.8848, Adjusted R-squared:  0.8724
## F-statistic: 71.66 on 3 and 28 DF,  p-value: 2.981e-13
```

```r
par(mfrow=c(2,2))
plot(m4)

title("Figure 4: Analysis of residuals for selected model", outer=TRUE)
```