# Automatic Vs Maunal Transmission Analysis

*Aanish Singla*

*25 April 2016*

**Aim**: To find out from mtcars dataset - 1. Is an automatic or manual transmission better for MPG? 2. Quantify the MPG difference between automatic and manual transmissions

## mtcars Dataset

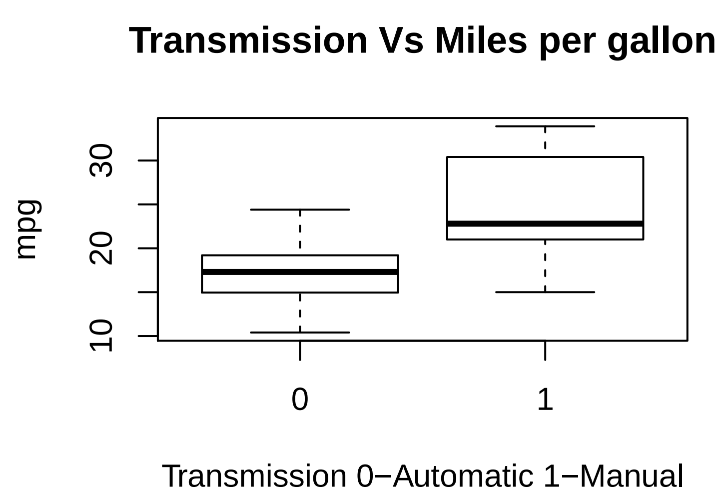
Let see what all does mtcars dataset contains

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Let see how am relates to miles per gallon(mpg).

```
boxplot(mtcars$mpg~factor(mtcars$am), main="Transmission Vs Miles per gallon",
        xlab="Transmission 0-Automatic 1-Manual",
        ylab="mpg")
```




The plot shows manual transmission has appx. 5 mpg higher mileage than automatic. Lets dig deeper with linear models.

Model 1:

```r
fit1 <- lm(mpg~factor(am), mtcars)
fit1$coefficients
```

```
## (Intercept) factor(am)1
##   17.147368    7.244939
```

This shows that mpg increases more rapidly for manual transmission than for automatic.

Model 2: Including Cyl

```r
fit2 <- lm(mpg~factor(am)+factor(cyl), mtcars)
fit2$coefficients
```

```
##  (Intercept)  factor(am)1 factor(cyl)6 factor(cyl)8
##    24.801852     2.559954    -6.156118   -10.067560
```

Model 3: Including Gross Horse Power

```r
fit3 <- lm(mpg~factor(am)+factor(cyl)+hp, mtcars)
fit3$coefficients
```

```
##  (Intercept)  factor(am)1 factor(cyl)6 factor(cyl)8           hp
##  27.29589929   4.15785647  -3.92457850  -3.53341392  -0.04424394
```

**Comparison :** Lets compare the models generated till now:

```r
anova(fit1,fit2,fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + factor(cyl)
## Model 3: mpg ~ factor(am) + factor(cyl) + hp
##   Res.Df   RSS Df Sum of Sq       F   Pr(>F)
## 1     30 720.9
## 2     28 264.5  2     456.4 31.2446 9.43e-08 ***
## 3     27 197.2  1      67.3  9.2141 0.005266 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Result:** Each model seem to improve the F-value and has a small p-value indicating that the included variable is useful. Lets continue to add further variables and analyze.

Model 4: Including disp

```r
fit4<- lm(mpg~factor(am)+factor(cyl)+hp+disp, mtcars)
fit4$coefficients
```

```
##  (Intercept)  factor(am)1 factor(cyl)6 factor(cyl)8           hp
##  29.00423340   3.33382838  -3.22228303  -1.01140831  -0.03856603
##         disp
##  -0.01501139
```

Model 5: Including drat

```r
fit5<- lm(mpg~factor(am)+factor(cyl)+hp+disp+drat, mtcars)
fit5$coefficients
```

```
##  (Intercept)  factor(am)1 factor(cyl)6 factor(cyl)8           hp
##  27.20753247   3.14169676  -3.06220313  -0.76569170  -0.03934070
##         disp         drat
##  -0.01444176   0.47668960
```

**Comparison :** Lets compare the models

```r
anova(fit3,fit4,fit5)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am) + factor(cyl) + hp
## Model 2: mpg ~ factor(am) + factor(cyl) + hp + disp
## Model 3: mpg ~ factor(am) + factor(cyl) + hp + disp + drat
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     27 197.20
## 2     26 183.04  1    14.160 1.9410 0.1758
## 3     25 182.38  1     0.658 0.0902 0.7664
```

**Result:** Both Model 4 and 5 have large p-value and small F-value, indicating that the variables don't add value to our model. Hence we drop disp and drat.

Model 6: Including wt

```r
fit6<- lm(mpg~factor(am)+factor(cyl)+hp+wt, mtcars)
fit6$coefficients
```

```
##  (Intercept)  factor(am)1 factor(cyl)6 factor(cyl)8           hp
##  33.70832390   1.80921138  -3.03134449  -2.16367532  -0.03210943
##           wt
##  -2.49682942
```

Model 7: Including qsec

```r
fit7<- lm(mpg~factor(am)+factor(cyl)+hp+wt+qsec, mtcars)
fit7$coefficients
```

```
##  (Intercept)  factor(am)1 factor(cyl)6 factor(cyl)8           hp
##  21.57616508   2.83269554  -1.90949640  -0.22716240  -0.02480654
##           wt         qsec
##  -2.96274129   0.61917236
```

**Comparison :** Lets compare the models

```r
anova(fit3,fit6,fit7)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am) + factor(cyl) + hp
## Model 2: mpg ~ factor(am) + factor(cyl) + hp + wt
## Model 3: mpg ~ factor(am) + factor(cyl) + hp + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     27 197.20
## 2     26 151.03  1    46.173 8.0172 0.009017 **
## 3     25 143.98  1     7.044 1.2230 0.279293
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

**Result:** wt has improved the model but not qsec.

Continuing with rest of the variable in the same way, we find that Gross horsepower, Weight, No. of cylinders and Transmission are useful variables for the model.

```r
summary(lm(mpg~factor(am)+wt+hp+factor(cyl), mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt + hp + factor(cyl), 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 ***
## factor(am)1   1.80921    1.39630   1.296  0.20646
## 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
## ---
## 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
```

Adjusted R-square is 0.98. Our model seems to be really good!

**Really?** Lets see if Transmission is really adding value.

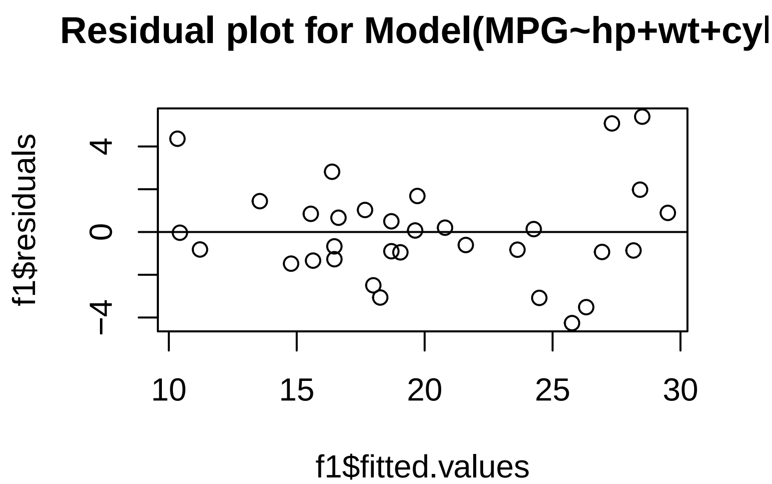
```r
f1 <- lm(mpg~wt+hp+factor(cyl), mtcars)
f2 <- lm(mpg~wt+hp+factor(cyl)+factor(am), mtcars)
anova(f1,f2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + hp + factor(cyl)
## Model 2: mpg ~ wt + hp + factor(cyl) + factor(am)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     27 160.78
## 2     26 151.03  1     9.752 1.6789 0.2065
```

**No its not. Transmission was significant earlier but not with Gross horsepower, Weight, No. of cylinders included.**

Lets see the residual plot for f1.

```r
plot(f1$fitted.values, f1$residuals,
     main="Residual plot for Model(MPG~hp+wt+cyl)")
abline(h=0)
```




Values are random hence the model seems to be good!

**Executive Summary: Given 32 sets of observations for different cars doesn't really show that one type of transmission is better than the other. Though at an initial glance, manual transmission appears better than automated but when analyzed taking into account other important variables, the trend is not conclusive.**