# Regression Models - Week 4 Course Project

*Sanjay Lonkar*

*30 Jun 2018*

## Overview

Analyze data set "mtcars" of collection of cars and answer following two questions:

1. Is an automatic or manual transmission better for MPG
2. Quantify the MPG difference between automatic and manual transmissions

## Result

**Manual transmission is better for MPG than Autaomatic transmission. Also, manual transmission has 1.8 greater MPG than automatic transmission.**

### Load data set and perform exploratory data analysis

See Appendix 1 for brief exploratory data analysis.

```
data (mtcars)
```

### Visual Analysis

Verify impact on MPG by generating a box plot of Transmission Type against MPG. See Appendix 2 for generated box plot. It shows MPG is less for Aumotatic transmission type cars.

### Quantified Regression Analysis

```
aggregate(mpg~am, data = mtcars, mean)
```

```
##   am      mpg
## 1  0 17.14737
## 2  1 24.39231
```

```
mtcars$cyl  <- factor(mtcars$cyl)
mtcars$vs   <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am   <- factor(mtcars$am,labels=c("Automatic","Manual"))
tManual <- mtcars [mtcars$am == "Manual", ]
tAutomatic <- mtcars [mtcars$am == "Automatic", ]
t.test(tAutomatic$mpg, tManual$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  tAutomatic$mpg and tManual$mpg
## t = -3.7671, df = 18.332, p-value = 0.001374
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

It shows that Autmoatic cars have lower 7.25 lower MPG (24.39231 - 17.14737) than Manual cars. Corresponding P-Value is 0.001374 which indicated substantial difference.

```
regressionModel <- lm (mpg ~ am, data = mtcars)
summary (regressionModel)
```

```
##
## 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
```

Manual car average MPG is 7.2 higher than that of Autmoatic cars which has MPG of 17.1 MPG. Constructing new multivariate regression model as above shows only 36% of the variance (R squared).

```
multivariateModel <- lm (mpg ~ am + cyl + disp + hp + wt, data = mtcars)
anova (regressionModel, multivariateModel)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl + disp + hp + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     25 150.41  5    570.49 18.965 8.637e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(multivariateModel) # See Appendix 3 for result
```

We have lesser P-Value of 8.637e-08 indicating better model. Pair plot in Appendix 4 shows correlations between difference data points and MPG.

**Conclusion, new model gives us 86.64% variance. Taking into account disp, hp, cyl and wt affect on correlation between mpg and am, we can say that difference between Manual and Automatic cars MPG is 1.81 MPG.**

**Appendix 1**

Brief Exploratory Analysis

```r
head (mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs       am gear
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0   Manual    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0   Manual    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1   Manual    4
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1 Automatic   3
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0 Automatic   3
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1 Automatic   3
##                   carb
## Mazda RX4            4
## Mazda RX4 Wag        4
## Datsun 710           1
## Hornet 4 Drive       1
## Hornet Sportabout    2
## Valiant              1
```

```r
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 : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
##  $ 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  : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...
```

**Appendix 2**

Box plot visually showing Manual cars give higher MPG than Automatic cars

```r
boxplot(mpg ~ am, data = mtcars, col = (c("red","blue")), ylab = "Miles Per Gallon",
        xlab = "Transmission Type")
```
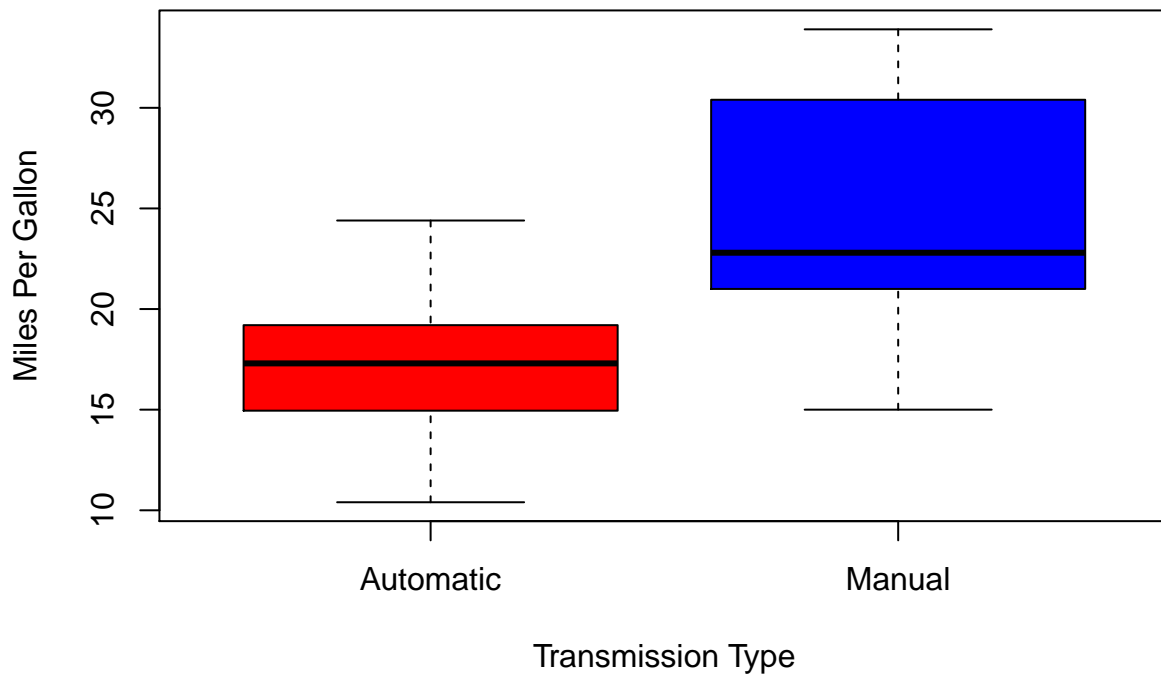
**Appendix 3**

Summary of multivariate model that gave minimum P-Value

```
summary(multivariateModel)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + disp + hp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9374 -1.3347 -0.3903  1.1910  5.0757
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.864276   2.695416  12.564 2.67e-12 ***
## amManual     1.806099   1.421079   1.271   0.2155
## cyl6        -3.136067   1.469090  -2.135   0.0428 *
## cyl8        -2.717781   2.898149  -0.938   0.3573
## disp         0.004088   0.012767   0.320   0.7515
## hp          -0.032480   0.013983  -2.323   0.0286 *
## wt          -2.738695   1.175978  -2.329   0.0282 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 2.453 on 25 degrees of freedom
## Multiple R-squared:  0.8664, Adjusted R-squared:  0.8344
## F-statistic: 27.03 on 6 and 25 DF,  p-value: 8.861e-10
```

**Appendix 4**

Pair plot indicating correlations between difference data points and MPG

```
pairs(mpg ~ ., data = mtcars)
```