Course Project

## Executive Summary

This document presents an analysis of the ‘mtcars’ dataset. This analysis wants to explain if there is some kind of relationship between miles per gallon and other variables presented in the data. The questions we want to answer in particular are the following:

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

From our analysis we can conclude that manual transmission is better for MPG(1.8 units higher)

## Exploratory Analysis

mtcars <- mtcars  
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 ...

names(mtcars)[9] <- paste('transmission')  
mtcars$transmission <- factor(mtcars$transmission,labels = c('Automatic','Manual'))  
mtcars$cyl <- factor(mtcars$cyl)  
mtcars$gear <- factor(mtcars$gear)  
mtcars$carb <- factor(mtcars$carb)  
mtcars$vs <- factor(mtcars$vs,labels = c('V-shaped','straight'))  
summary(mtcars)

## mpg cyl disp hp drat   
## Min. :10.40 4:11 Min. : 71.1 Min. : 52.0 Min. :2.760   
## 1st Qu.:15.43 6: 7 1st Qu.:120.8 1st Qu.: 96.5 1st Qu.:3.080   
## Median :19.20 8:14 Median :196.3 Median :123.0 Median :3.695   
## Mean :20.09 Mean :230.7 Mean :146.7 Mean :3.597   
## 3rd Qu.:22.80 3rd Qu.:326.0 3rd Qu.:180.0 3rd Qu.:3.920   
## Max. :33.90 Max. :472.0 Max. :335.0 Max. :4.930   
## wt qsec vs transmission gear   
## Min. :1.513 Min. :14.50 V-shaped:18 Automatic:19 3:15   
## 1st Qu.:2.581 1st Qu.:16.89 straight:14 Manual :13 4:12   
## Median :3.325 Median :17.71 5: 5   
## Mean :3.217 Mean :17.85   
## 3rd Qu.:3.610 3rd Qu.:18.90   
## Max. :5.424 Max. :22.90   
## carb   
## 1: 7   
## 2:10   
## 3: 3   
## 4:10   
## 6: 1   
## 8: 1

## Regression Analysis

Auto <- mtcars[mtcars$transmission=='Automatic',]  
Manu <- mtcars[mtcars$transmission=='Manual',]  
mean(Manu$mpg)-mean(Auto$mpg)

## [1] 7.244939

We can see that the difference of 7.24 is significant so we conduct a t.test

ttest <- t.test(Auto$mpg,Manu$mpg)  
ttest$conf.int

## [1] -11.280194 -3.209684  
## attr(,"conf.level")  
## [1] 0.95

ttest$p.value < 0.05

## [1] TRUE

We can see that the confidence interval does not contain 0 and the p-value is less than 0.05 so we can reject the null hyphotesis(no difference in means).

Now we can try to see the relationship between MPG and transmission to see if they are related and if the variance in MPG is explain by transmission. And also how much of this variation is explain by the model(R2)

model\_1 <- lm(mpg~transmission,data = mtcars)  
summary(model\_1)

##   
## Call:  
## lm(formula = mpg ~ transmission, 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 \*\*\*  
## transmissionManual 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

The intercept tells us that the median of Automatic is 17.14 and the median of Manual is 7.24 units higher. The p-value for this model is 0.00028, which tells us that the coefficients are statistically significant. Finally, Transmission explains 36% of the variance between means, it might be suitable to explore more variables so the variance can be explained better. Now we compute a second model.

model2 <- lm(mpg~transmission+cyl+disp+hp+wt,data = mtcars)  
summary(model2)

##   
## Call:  
## lm(formula = mpg ~ transmission + 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 \*\*\*  
## transmissionManual 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

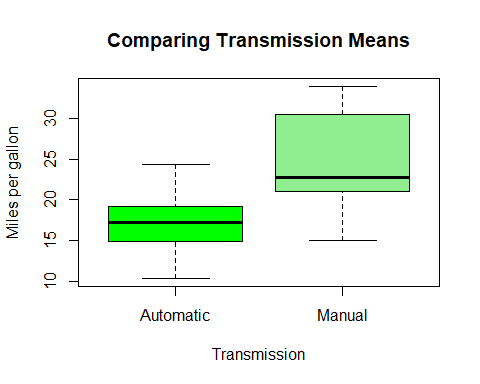
anova(model\_1,model2)

## Analysis of Variance Table  
##   
## Model 1: mpg ~ transmission  
## Model 2: mpg ~ transmission + 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

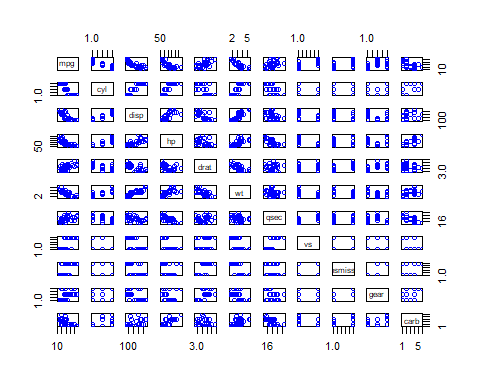
In this second model we can see that there are other variables that are statistically significant and now the variance in MPG is exaplained better with the inclussion of other variables (R-squared=87%). The p-value also is less than 0.05. And we can conclude that having a manual transmission could lead to a difference in means of 1.81 in MPG. Finally, check for homoskedastic in the appendix and according to the behaviour of the residuals this assumption holds.

## Appendix

boxplot(mpg~transmission,data = mtcars,col=c('green','light green'),ylab='Miles per gallon',xlab='Transmission',main='Comparing Transmission Means')



pairs(mpg~.,data = mtcars,col='blue')



par(mfrow = c(2,2))  
plot(model2)

