## Regression Models - Exploring MPG and other variables

## Executive Summary

In this report we aim to explore if a car with automatic or manual transmission is better for miles per gallon(MPG) and the magnitude of MPG difference between the transmissions.To explore this, we obtained the data from 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models).From the model that best fits,we found that manual transmission have better MPG and 5.3 mpg higher compared to automaic transmissions. Loading the data

data(mtcars)

## Exploratory Data Analysis-explore the variables

tab1<-table(mtcars$am)  
tab2<-summary(mtcars$mpg)  
tab3<-table(mtcars$cyl)  
tab4<-summary(mtcars$disp)  
tab5<-summary(mtcars$hp)  
ttab6<-summary(mtcars$qsec)  
tab7<-table(mtcars$carb)

The variables don't seem to require any preprocessing and there seems to be good distribution between the automatic and manual transmissions.

## Exploratory Data Analysis-correlations and associations

test1<-t.test(mtcars$mpg~factor(mtcars$am))  
test2<-cor(mtcars$am,mtcars$gear)  
test3<-cor(mtcars$hp,mtcars$cyl)  
test4<-cor(mtcars$hp,mtcars$disp)

The test shows a significant difference at p-value < 0.05 between the groups of transmission for MPG. Horsepower is highly corrrelated(>0.75) with number of cylinders and displacement.Thus,We will only use horse power as an adjuster in the multivariate model.Similarly,type of tramsmission and the number of forward gears seem to be correlated.

## Bivariate Models-associations with the outcome

bivar1<-lm(mpg~hp,data=mtcars)  
bivar2<-lm(mpg~carb,data=mtcars)  
bivar3<-lm(mpg~qsec,data=mtcars)

The bivariate models shows a significant associations at p-value<0.05 between our outcome variable MPG and horse power,number of carburetors and 1/4 mile time. Thus, we will build our multivariate models with the above variables as adjusters.

## Multivariate Models

modfit1<-lm(mpg~factor(am),data=mtcars)  
modfit2<-lm(mpg~factor(am)+hp,data=mtcars)  
summary(modfit2)

##   
## Call:  
## lm(formula = mpg ~ factor(am) + hp, data = mtcars)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.384 -2.264 0.137 1.697 5.866   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 26.58491 1.42509 18.65 < 2e-16 \*\*\*  
## factor(am)1 5.27709 1.07954 4.89 3.5e-05 \*\*\*  
## hp -0.05889 0.00786 -7.50 2.9e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.91 on 29 degrees of freedom  
## Multiple R-squared: 0.782, Adjusted R-squared: 0.767   
## F-statistic: 52 on 2 and 29 DF, p-value: 2.55e-10

modfit3<-lm(mpg~factor(am)+hp+carb,data=mtcars)  
modfit4<-lm(mpg~factor(am)+hp+carb+qsec,data=mtcars)  
modfit5<-lm(mpg~factor(am)+hp+carb+qsec+vs,data=mtcars)

The stepwise multivariate models were built with all the variables that were associated with the outcome variable MPG from our exploratory data analysis. All the models show a significant relationship between the type of transmissions used and our outcome variable MPG at p-value<0.05.

## Multivate Model - Final model comparison

anova(modfit1,modfit2,modfit3,modfit4,modfit5)

## Analysis of Variance Table  
##   
## Model 1: mpg ~ factor(am)  
## Model 2: mpg ~ factor(am) + hp  
## Model 3: mpg ~ factor(am) + hp + carb  
## Model 4: mpg ~ factor(am) + hp + carb + qsec  
## Model 5: mpg ~ factor(am) + hp + carb + qsec + vs  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 30 721   
## 2 29 245 1 475 60.41 3e-08 \*\*\*  
## 3 28 229 1 16 2.07 0.16   
## 4 27 224 1 6 0.71 0.41   
## 5 26 205 1 19 2.41 0.13   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

From the multivariate model comparisons above,model 2 seems to be the best fit at p-value<0.05 with the transmission type and horsepower as predictors and significantly associated with MPG at p-value<0.05.The miles per gallon increases by 5.3 mpg for manual transmission compared to automatic transmission cars and the MPG decreases by 0.06 per unit increase in horsepower.

## Residuals and Diagnostics

R-squared value from the model fit is 0.78 and from the Residual and diagnostics plots(see appendix), the model fits very well with constant variance and little to no leverage from the outliers

# Appendix

## Residuals and Diagnostics

plot(modfit2)

   