Regression Model Assignment - Analysis of impact of Automobile Transmission on Fuel efficiency

Executive Summary

This assignment focusses on exploring and analysing the impact of Automobile Transmission on Fuel efficiency. Looking at a data set of a collection of cars, we wshall be exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). We will be particularly interested in answering the following two questions:

"Is an automatic or manual transmission better for MPG" "Quantify the MPG difference between automatic and manual transmissions"

For this purpose, will be performing some Exploratory data analysis and then proceed to a fit a regression model that best fits the data set and draw conclusion based on the models.

Data Analysis

In this section we will load mtcars data set and convert transmission type to factor.

```
data(mtcars)

# Convert am from nymeric column to "Automatic/Manual" factor

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"))
```

Explore coorelation for mpg with other variables (Appendix - Fig A)

Fig A suggests that mpg has strong corelation with cyl, disp, hp, drat, wt, vs and am. We will explore their interdependence and influence on the model using step function for selecting the best model.

Understand impact of Transmission Type on MPG (Appendix Fig B)

Fig B suggests that Manual transmissions have better MPG compared to Automatic Transmission.

```
ampg <- mtcars[mtcars$am == "Automatic",]
mmpg <- mtcars[mtcars$am == "Manual",]

t.test(ampg$mpg, mmpg$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  ampg$mpg and mmpg$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
```

The mean difference in MPG between the Automatic and Manual Transimission is about 7.245. Also, the p-value above does not suggest that the mean difference in MPG between Automatic and Manual Transmission is significant ignoring coorelation between any of the other features (wt, hp, cylinder etc. We will however check the potential influence of other factors on the MPG in the next section.

Regression Model

We will now fit a linear model for (uni variate) and best fit (multi variate)

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

```
##
## 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
```

This model only explains about 34% variance (based on R^2 value) suggesting an influence of other factors as well. We will thus need to select the best variables. We will use AIC alogorithm to choose the best variables.

```
basefit <- lm(mpg ~ ., data = mtcars)
bestfit <- step(basefit, direction = "both")
```

```r
summary(bestfit)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + 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 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     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
```

The above model explains about 84% variance (based on R^2 value) which is significantly higher than the previous model. It also suggest that based on the bestfit model, the mean MPG for Manual Transmission Cars is 1.8 MPG more than that for Automatic Transmission.

Appendix

Fig A

```r
pairs(mpg ~ ., data=mtcars, main = "Pair Plot for mtcars")
```

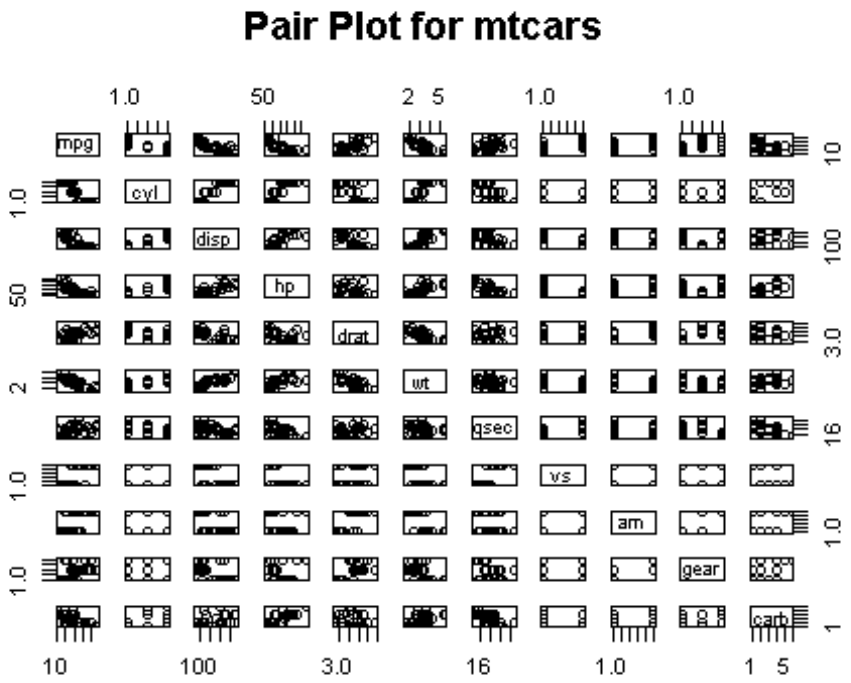
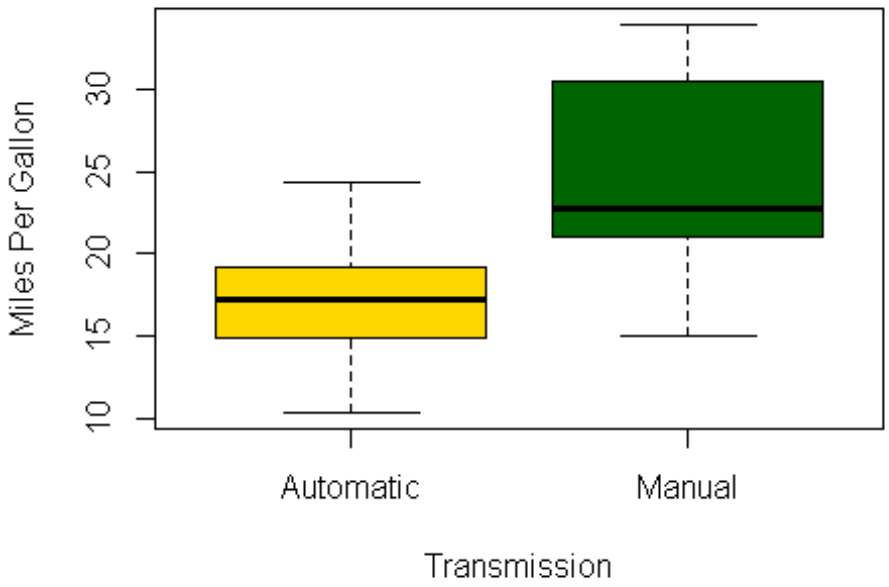


Fig B

```
boxplot(mpg ~ am, data = mtcars, col = (c("gold","darkgreen")), ylab = "Miles
Per Gallon", xlab = "Transmission", notch = FALSE)
```


Miles Per Gallon
10
15
20
25
30
Automatic
Manual
Transmission