Linear Regression Study of the Motor Trend Car Road Tests

N MA

17 September 2015

Executive Summary

The mtcars data set extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles in the years 1973-1974. This study will focus mainly on the effect of Transimission on the fuel efficency miles per gallon. The two question of particular interest are as follows:

- Is an automatic or manual transmission better for MPG and result shows manual is better than automatic transmission for fuel efficency
- Quantify the MPG difference between automatic and manual transmissions, and the result shows that for all the quantites estimated 17.1 and 24.4 respectively

```
library(ellipse);library(ggplot2); data(mtcars); attach(mtcars)
```

Exploring the Data

```
## The following object is masked from package:ggplot2:
##
##     mpg
```

```
mtcars1<-mtcars
cor1<-round(cor(mtcars1),2)
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)
```

Summary of the data shows that mpg stands for Miles Per Gallon, Cylinders(cyl), Displacemnet(disp), Gross horsepower(hp), weight(wt, lb/1000), qsec(1/4 mile time), vs(V/S), transmission (am, 1=manual, 0=automatic), Number of forward gears (gears) and Number of carburetors.

The plot (see Appendix) and correaltion matrix shows strong correlation between the variables. The variable 'mpg' is strong negative correlation between number of cyclinder, disp and hp and wt, and a weak negative correlation to carb. However, there is also a strong positive correlation between for example disp and cyl, wt etc. Hence, trying to predict mpg with all the other predictors won't be predictive as shown in the lm function below. Some of the relations are shown in the appendix of the this document. Since the data is highly collinear and hence the approcach used is AIC to fit the model

Analysis through Regression Regression is used to analysis the best model to fit the best outcome for mpg. This method used are build on model by model and to see which model best describes the variablilty the best. #### MPG as a function of other predictors

```r
FirstModel <- lm(mpg ~ ., data = mtcars)
FinalModel<- step(FirstModel, direction = "both")
```

The method used to develop using forwqrd and backward elimination methods by AIC algorithm since many of the column are highly collinear. The Final model include predictors cyl, hp, wt and am. The model describes roughly 83% of all variability. #### Transmission to describe mpg

```r
lm.mpg<-lm(mpg~as.factor(am), data=mtcars)
confint(lm.mpg)
```

```
##                   2.5 %   97.5 %
## (Intercept)    14.85062 19.44411
## as.factor(am)1  3.64151 10.84837
```

```r
lm.mpg2<-lm(mpg~hp+as.factor(am), data=mtcars)
```

The MPG difference for an automatic and transmission are given by 17.147 and 24.38 respectively. The 95% confidence interval value is between 14.85 and 19.44 for automatic, while 95% confidence interval is between 18.49 and 30.39. The anova function is used to study whether the model is a significant improvement on the model with just am being used.

```r
anova(lm.mpg, lm.mpg2, FinalModel)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ as.factor(am)
## Model 2: mpg ~ hp + as.factor(am)
## Model 3: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 81.853 1.634e-09 ***
## 3     26 151.03  3     94.41  5.418  0.004961 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

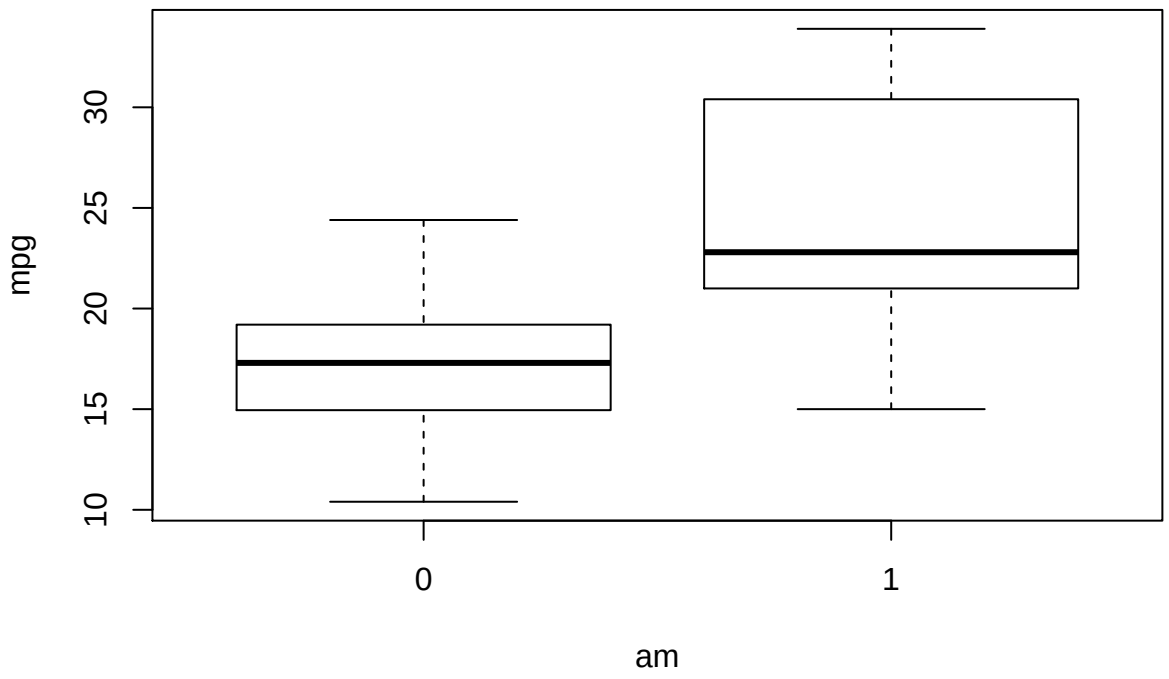
The anova function shows both the models are a better fit than the with automation alone. Residual plot is plotted in the Appendix section. The plots shows that residuals are spread randomly showing no signs of underlying function dependence. Similarly Normal Q-Q plot shows that distribution falls a staight line indicating that the residuals are normally distributed. There are some outliners like Chrysler Imperidla and Fiat and Toyota Corolla.

Conclusion

- Manual has higher Mile per Gallon thn Automatic car.
- Data shows heavy collinearity between various predictors while modelling the outcome MPG
- For an automatic cars, for an 100 hp increase mpg decreases by 5.88 and for manual the decrease by 5.21 mpg.

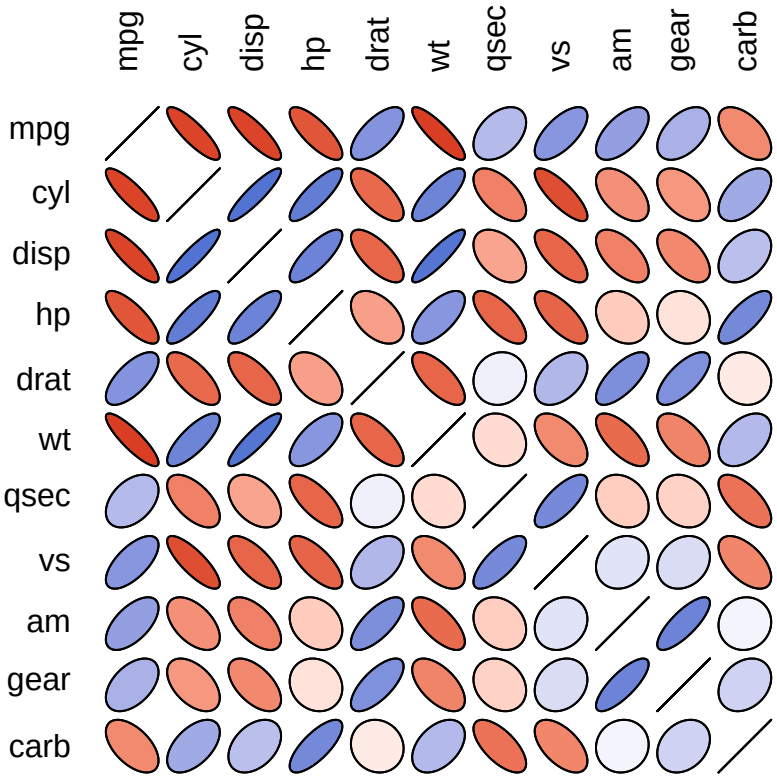
Appendix

```
plot(mpg~am, data=mtcars)
```




```
c <- ggplot(mtcars, aes(x=hp, y=mpg, colour=factor(am)))+stat_smooth(method=lm) + geom_point()
```

```
cor1<-round(cor(mtcars1),2)
colorfun <- colorRamp(c("#CC0000","white","#3366CC"), space="Lab")
plotcorr(cor1, col=rgb(colorfun((cor1+1)/2), maxColorValue=255),mar = c(0.1, 0.1, 0.1, 0.1))
```




```
par(mfrow=c(2,2))
plot(FinalModel)
```