# Coursera - Regression Course Project

*Jonathan Stone*

*January 25, 2015*

## Executive Summary

This is a project submitted sdfsaxfbdfor Coursera's Regression Modeling Class. This analyzes the impact of transmission on the miles per gallon of several cars, obtained from the mtcars dataset located in the datasets package. This study uses multivariate regression to determine an increase of 2.9 mpg when the transmission of a car is manual, controlled for both weight and 1/4 mile time with >95% certainty.

## Loading the Data

```
## Loading required package: ggplot2
## Loading required package: MASS
## Loading required package: car
## Loading required package: caret
## Loading required package: lattice
```

```
data(mtcars)
```

## Exploratory Analysis

```
ggplot(mtcars, aes(x=factor(am), y=mpg, fill=factor(am))) +
  geom_boxplot(size=2, colour='black')

featurePlot(x=mtcars, y=mtcars$mpg, plot='pairs')
```

The above code produces two graphs (see Appendix 1.1) that examine both the general impact of the transmission on `mpg` (miles per gallon) and the impact of all variables on mpg. The boxplot confirms that transmission has an impact on mpg, while the multiplot implies that several other variables also have an impact on `mpg`, so multivariate regression is a reasonable approach for this topic.

## Model Selection

```
fit2 <- lm(mpg ~., data=mtcars)
summary(fit2)
```

After examining a simple multivariate regression including all models, it is clear the when all predictors are considered, the results are not statistically significant. Several predictors will have to be removed.

```
step <- step(fit2, direction="backward")
```

The step function (found in the MASS package) indicates that the likely variables to consider for impact are transmission, 1/4 mile time, and weight. Including any more variables than these removes the statistically significant results for `am`.

```
mtcars.lm <- lm(mpg ~ qsec + am + wt, data=mtcars)
summary(mtcars.lm)
mtcars.baselm <- lm(mpg ~ am, data=mtcars)
anova(mtcars.baselm, mtcars.lm)
```

These models indicate a statistically significant outcome when the regression model only considers the variables `qsec` `am` and `wt`. The anova indicates a highly significant results, and we reject the null hypothesis that `qsec` `wt` and `am` are not related to `mpg`.

```
mtcars.basers <- resid(mtcars.baselm)
plot(mtcars$mpg, mtcars.basers, ylab="Residuals", xlab="MPG", main="Residual Plot of MPG in base model")
abline(0, 0)
```

This residual plot (Appendix 1.2) further suggests the multivariate model shown in `mtcars.lm`. There is a clear pattern demonstrated in the residuals of the `mtcars.baselm` model, suggesting more variables are confounding the data.

## Residuals

```
mtcars.rs <- resid(mtcars.lm)
mtcars.basers <- resid(mtcars.baselm)

plot(mtcars$mpg, mtcars.rs, ylab="Residuals", xlab="MPG", main="Residual Plot of MPG in multivariate mo
abline(0, 0)
```

```
require(car)
residualPlots(mtcars.lm)
```

There is a slight pattern in the residuals plot (see Appendix 1.3) of this linear model, indicating another variable may have an effect on `mpg` not considered.

## Inference

```
t.test(mpg ~ am, data=mtcars)
```

A t-test is run to reinforce if `am` does have an impact on `mpg`. This clearly demonstrates that automatic and manual transmissions have an impact on `mpg` with a >99% certainty.

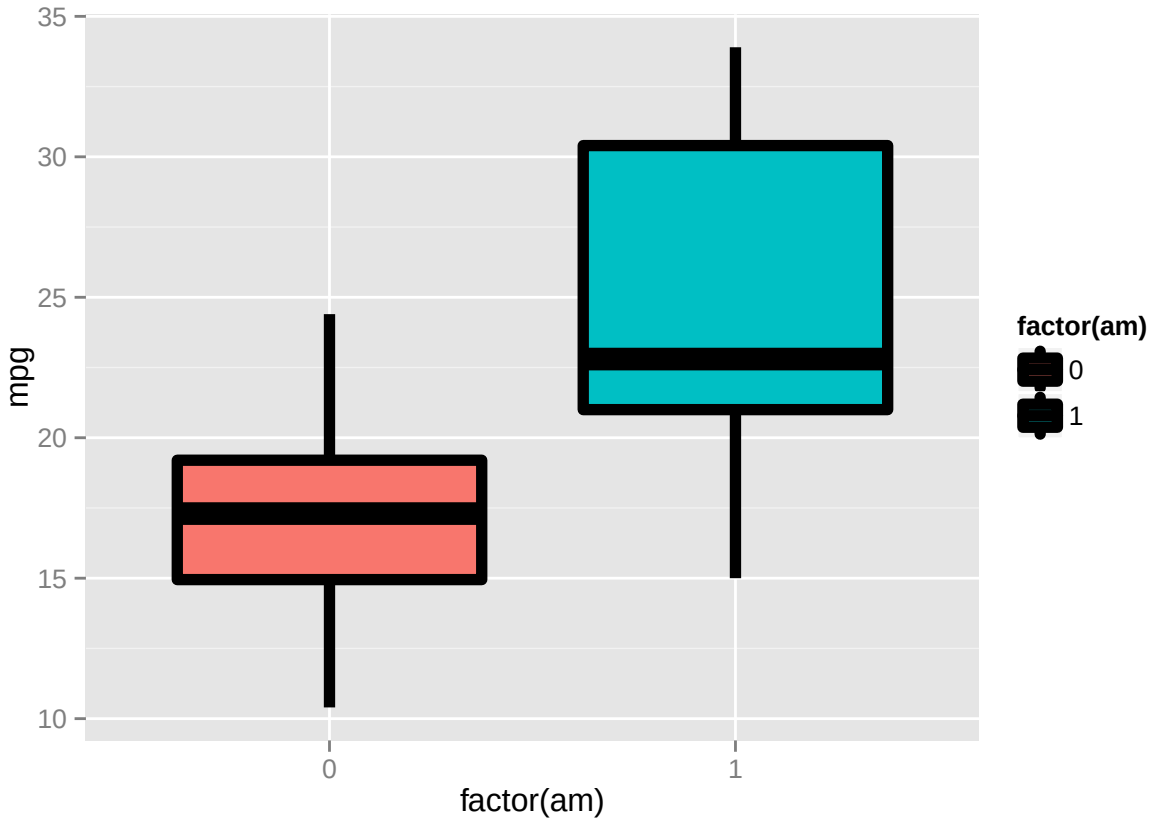
## Conclusions

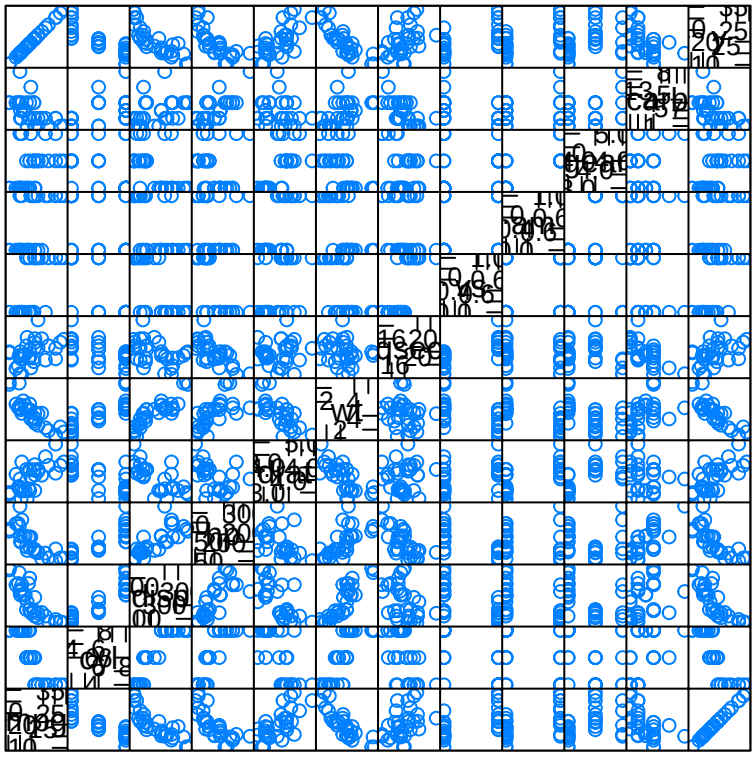
The conclusions reached from this study are:

- When a car has a manual transmission, there is a 2.9 increase in miles per gallon, controlled for both 1/4 mile time and weight. This is determined with a >95% certainty.
- For each 1 second increase in the 1/4 mile time of a car, there is a 1.2 increase in miles per gallon. This is determined with a >99.9% confidence.
- For each 1000 lb interval increase in the weight of a car, there is a 3.9 decrease in miles per gallon. This is determined with >99.9% confidence.

## Appendix

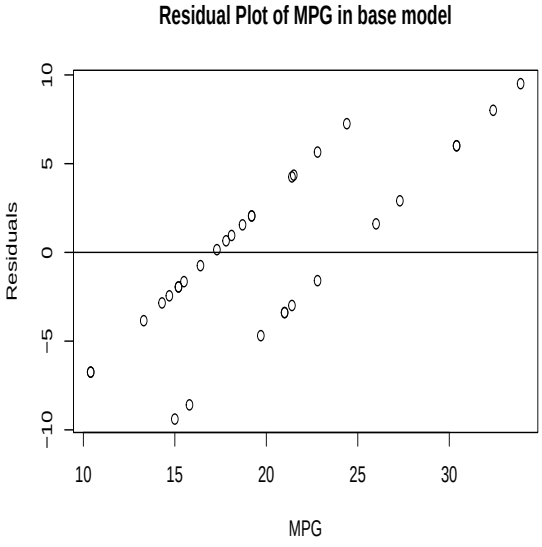
### Appendix 1.1



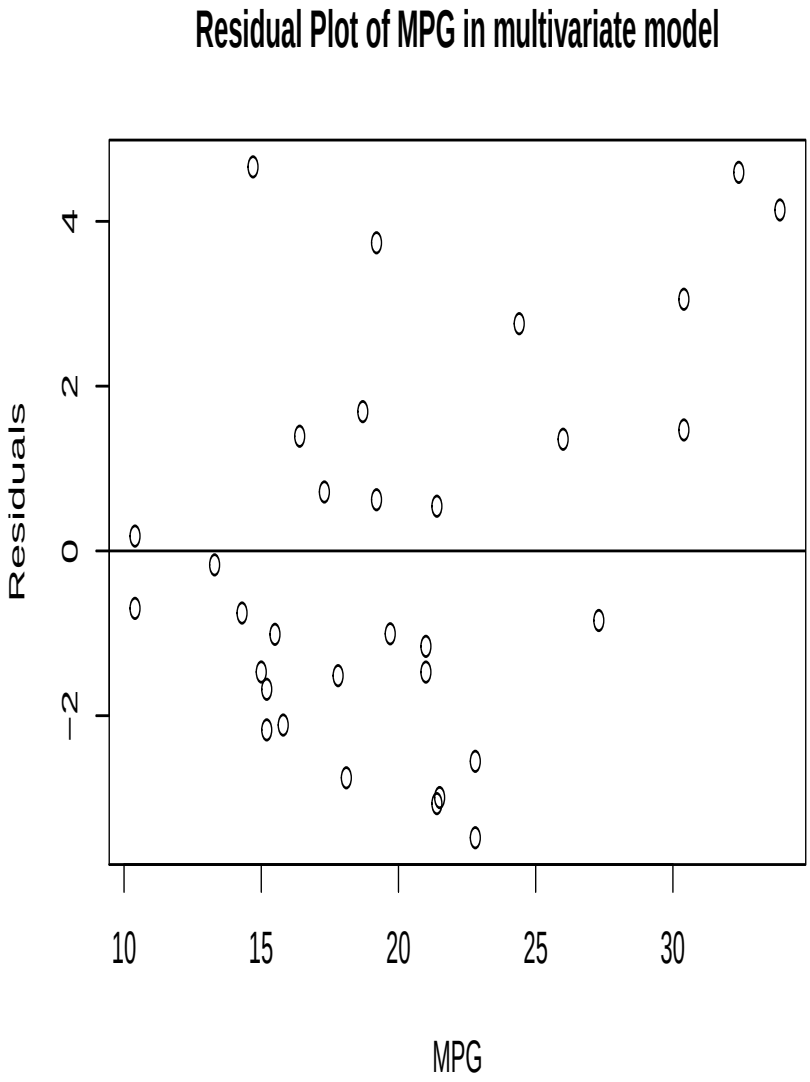


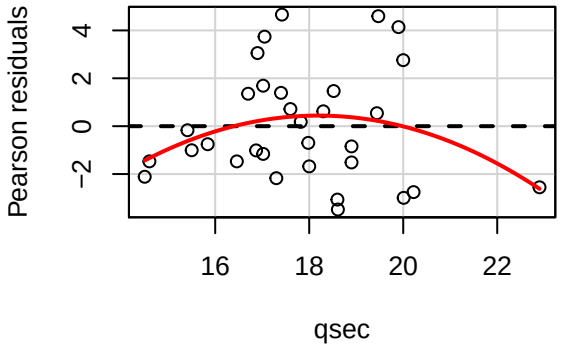
Scatter Plot Matrix

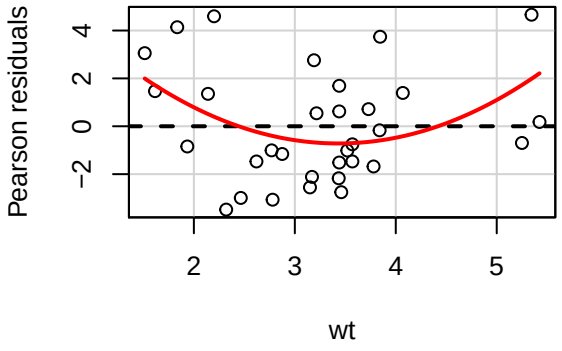
## Appendix 1.2



**Appendix 1.3**




```
##            Test stat Pr(>|t|)
## qsec          -1.565    0.129
## am             1.395    0.174
## wt             2.816    0.009
## Tukey test     3.227    0.001
```