# Regression Models Course Project : The Effects of Transmission Type on MPG

*Chris Gomes*

*July 22, 2016*

## Executive Summary

We study the `mtcars` data set to examine the relationships between several variables (e.g. transmission type, displacement, weight, and cylinder number) and miles per gallon (MPG).

We answer two questions:

- Is an automatic or manual transmission better for MPG?
- How different is the mpg between automatic and manual transmissions?

We determine that there is a difference in mpg for automatic and manual transmissions with manual transmissions giving an increase of 1.8 miles per gallon over automtic transmissions.

## Exploratory Analysis

We load the data set and and try to determine which variables are statistically significant. Looking at the internal structure of `mtcars`, we see that the variables: cyl, vs, am, gear, and carb are categorical. So we convert them using factor().

See figure 1 in the appendix for exploratory analysis and visualizations.

Since the goal of this analysis is to explore potential relationships between mpg and transmission type, we look at a box plot and note that there is a difference (see figure 2).

## Regression Analysis

### Simple Linear Regression

First we look at a simple liinear model with `am` as the predictor of `mpg`.

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

```
summary(simple)
```

### Generalized Linear Models

Next, we look at a model using all variables as predictors of mpg, then we refine this naive initial model using the `step` method to select the best model.

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

```
summary(best)
```

We see that the best model uses cyl, hp, wt, and am. Since the adjusted R-squared is approximately 0.84, we conclude that over 84% of the variability is explained by the above model.

### Comparing the models

We use analysis of variance (ANOVA) to compare the simple regression model with the generalized linear model.

```
anova(simple, best)
```

Looking at the above results, the *p-value* is highly significant. Hence, we reject the null hypothesis that the confounding variables `cyl`, `hp`, and `wt` do not contribute to the accuracy of the model.

### Residual Plot and Diagnostics

See figure 3 for the residual plots.

The points in the Residuals vs Fitted plot appear to be randomly scattered confirming the independence condition.

Most of the points in the Normal Q-Q plot lie in a straight line confirming that the residuals are normally distributed.

The residuals in the Scale-Location plot appear to be spread equally along the ranges of predictors confirming the assumption of equal variance.

The Residuals vs Leverage plot reveals some potential leverage points.

The points with the most leverage are given by

```
leverage <- hatvalues(best)
tail(sort(leverage),3)
```

```
##       Toyota Corona Lincoln Continental       Maserati Bora
##          0.2777872           0.2936819           0.4713671
```

The points with the strongest influence on model coefficients are given by

```
influence <- dfbetas(best)
tail(sort(influence[,6]),3)
```

```
## Chrysler Imperial          Fiat 128     Toyota Corona
##         0.3507458         0.4292043         0.7305402
```

These are the same cars appearing in the plots above (the Maserati Bora is off the scale). Hence, we conclude that our analysis is correct.

## Statistical Inference

We perform a 2-sample t-test on the two type of transmission: automatic and manual.

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

The *p-value* 0.001374 is significant and so we reject the null hypothesis: The true difference in means equals 0.

## Conclusions

From the summary of `best`, we conclude that:

- Cars with manual transmissions get 1.8 more miles per gallon than cars with automatic transmissions (after adjusting for weight, horsepower, and nuber of cylinders).
- cars lose about 2.5 miles per gallon for every increase in weight of 1000 lbs.
- cars lose about 0.032 miles per gallon for every increase of 1 horsepower.
- If the number of cylinders increases from 4 to 6 and 6 to 8, respectively, then the decreases in miles per gallon will be approximately 3 and 2.2, respectively.

## Appendix

**Figure 1 - the pair plot between mpg and the other variables**

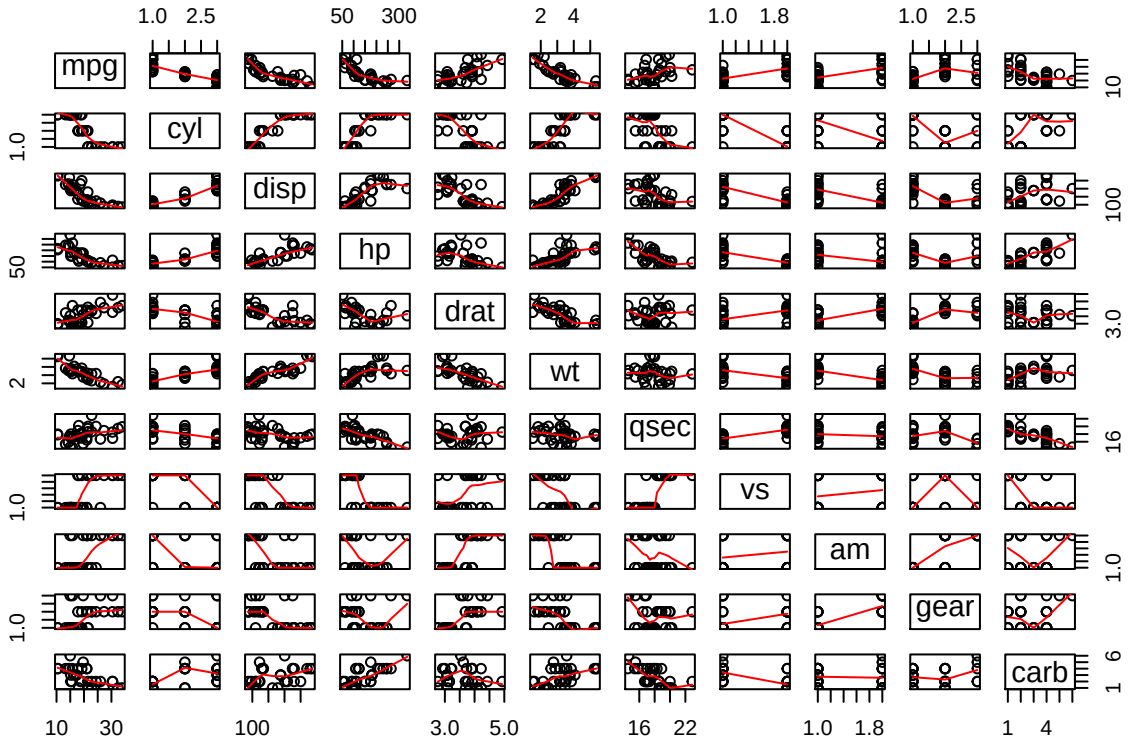

**Figure 2 - Box plot of mpg vs transmission type**

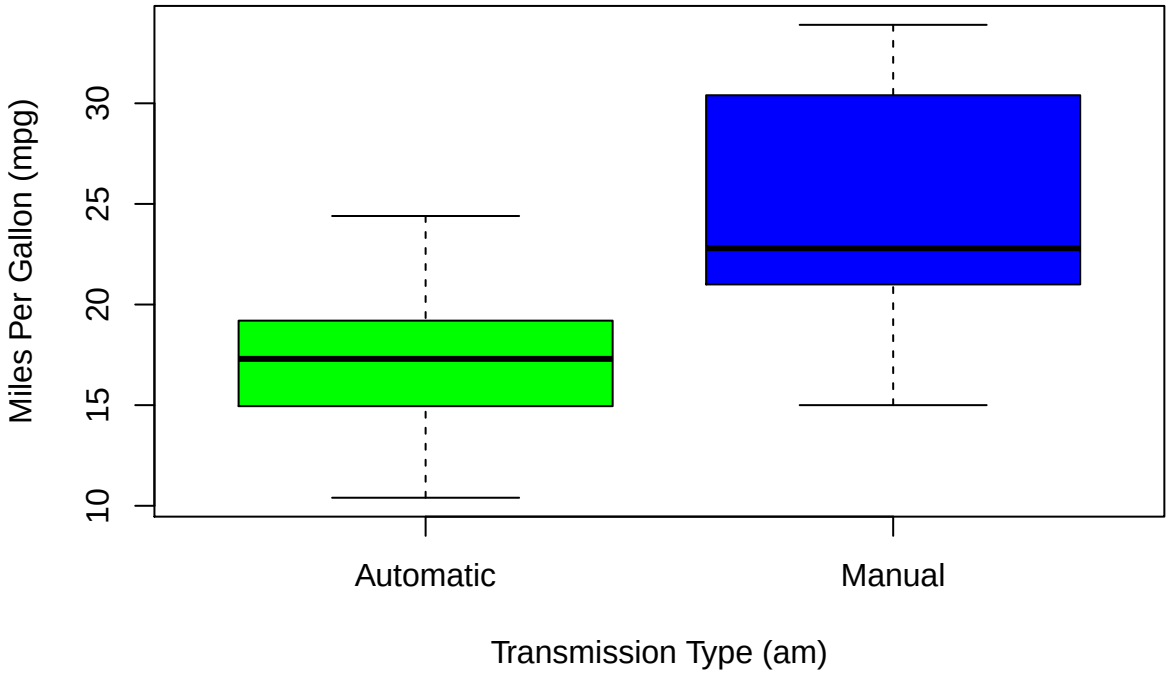

**Figure 3 - Residual Plots**