

# Motor Trend MPG Analysis

*MiRo*

*26 Jul 2015*

## Executive summary

In this report, we will analyze `mtcars` data set and explore the relationship between a set of available variables and miles per gallon (MPG). The data was extracted from the 1974 *Motor Trend* US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models). There are 2 main questions to consider: what type of transmission is better for mpg, and how it can be described in numeric terms. The t-test shows that mpg performance difference between cars with automatic and manual transmission. And it is about 7 MPG more for cars with manual transmission than those with automatic transmission. We fit several linear regression models and select the one with highest Adjusted R-squared value. Given model shows that cars that are lighter in weight with a manual transmission and cars that are heavier in weight with an automatic transmission will have higher MPG values.

## Context

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

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

Since p-Value is 0.0013736 the automatic and manual transmissions are from different populations. And the mean for MPG of manual transmitted cars is about 7.2449393 more than that of automatic transmitted cars.

## Regression Model

Testing simple model.

```
simpleMpgModel <- lm(mpg~am, data=mtcars)
summary(simpleMpgModel) # results="hide"
```

Simple model can explain 34% of variance of MPG. Adjusted R-squared value indicates that we need to add other variables to the model to make better predictions.

## Finding proper set of explanation variables

Let's test full model.

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

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  23.87913    20.06582   1.190  0.2525
## cyl6         -2.64870     3.04089  -0.871  0.3975
## cyl8         -0.33616     7.15954  -0.047  0.9632
## disp          0.03555     0.03190   1.114  0.2827
## hp           -0.07051     0.03943  -1.788  0.0939 .
## drat          1.18283     2.48348   0.476  0.6407
## wt           -4.52978     2.53875  -1.784  0.0946 .
## qsec          0.36784     0.93540   0.393  0.6997
## vs1           1.93085     2.87126   0.672  0.5115
## am1           1.21212     3.21355   0.377  0.7113
## gear4         1.11435     3.79952   0.293  0.7733
## gear5         2.52840     3.73636   0.677  0.5089
## carb2        -0.97935     2.31797  -0.423  0.6787
## carb3         2.99964     4.29355   0.699  0.4955
## carb4         1.09142     4.44962   0.245  0.8096
## carb6         4.47757     6.38406   0.701  0.4938
## carb8         7.25041     8.36057   0.867  0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF, p-value: 0.000124
```

Full regression model can explain 78% of MPG variance, but none of the coefficients are significant. We could find more effective set of variables to explain mpg.

```
stepModel <- step(fullModel, k=log(nrow(mtcars)))
summary(stepModel) # results="hide"
mpgModel <- lm(mpg~wt+qsec+am, data=mtcars)
```

Linear model  $\text{mpg} \sim \text{wt} + \text{qsec} + \text{am} - 1$  explains 83% of MPG variance, but consists only significant at 5% s.l. coefficients.

```
anova(mpgModel, simpleMpgModel, fullModel)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + qsec + am
## Model 2: mpg ~ am
## Model 3: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df  RSS Df Sum of Sq    F    Pr(>F)
## 1      28 169.29
```

```
## 2      30 720.90 -2    -551.61 34.3604 2.509e-06 ***
## 3      15 120.40 15     600.49  4.9874  0.001759 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

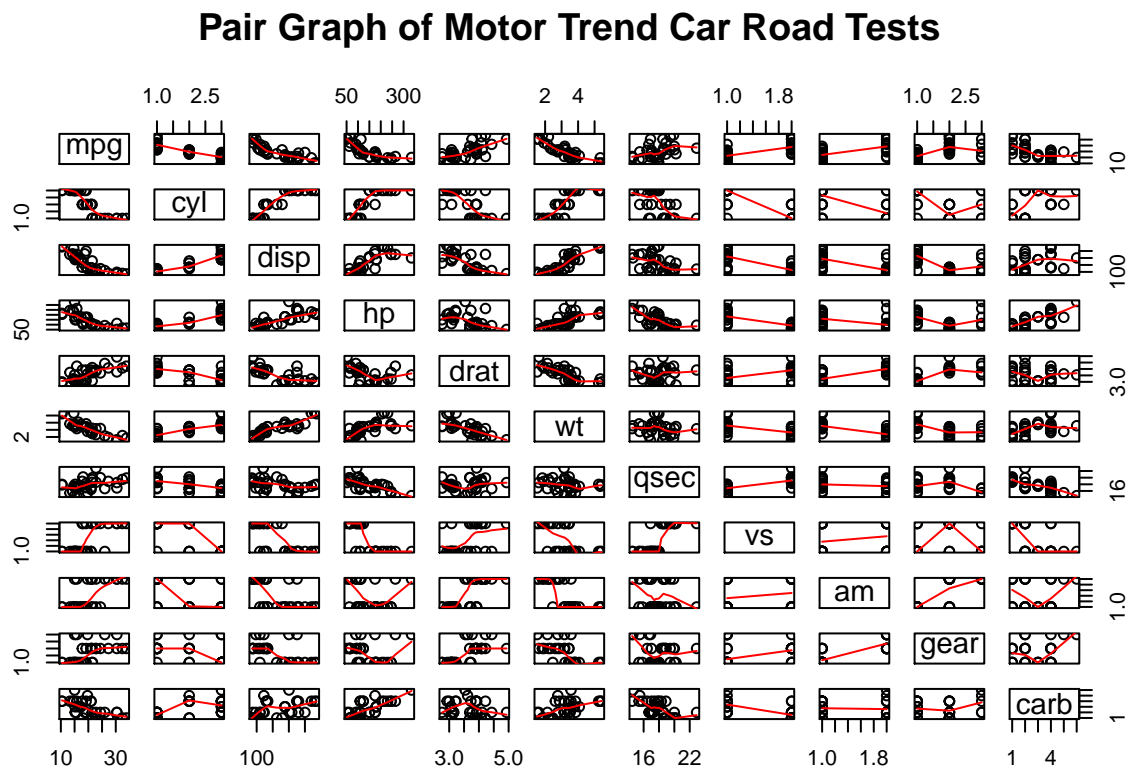
```
confint(mpgModel)
```

```
##              2.5 %    97.5 %
## (Intercept) -4.63829946 23.873860
## wt          -5.37333423 -2.459673
## qsec         0.63457320  1.817199
## am1          0.04573031  5.825944
```

The measure of how much an observation has effected the estimate of a regression coefficient, we get 1  
Therefore, the above analyses meet all basic assumptions of linear regression and well answer the questions.

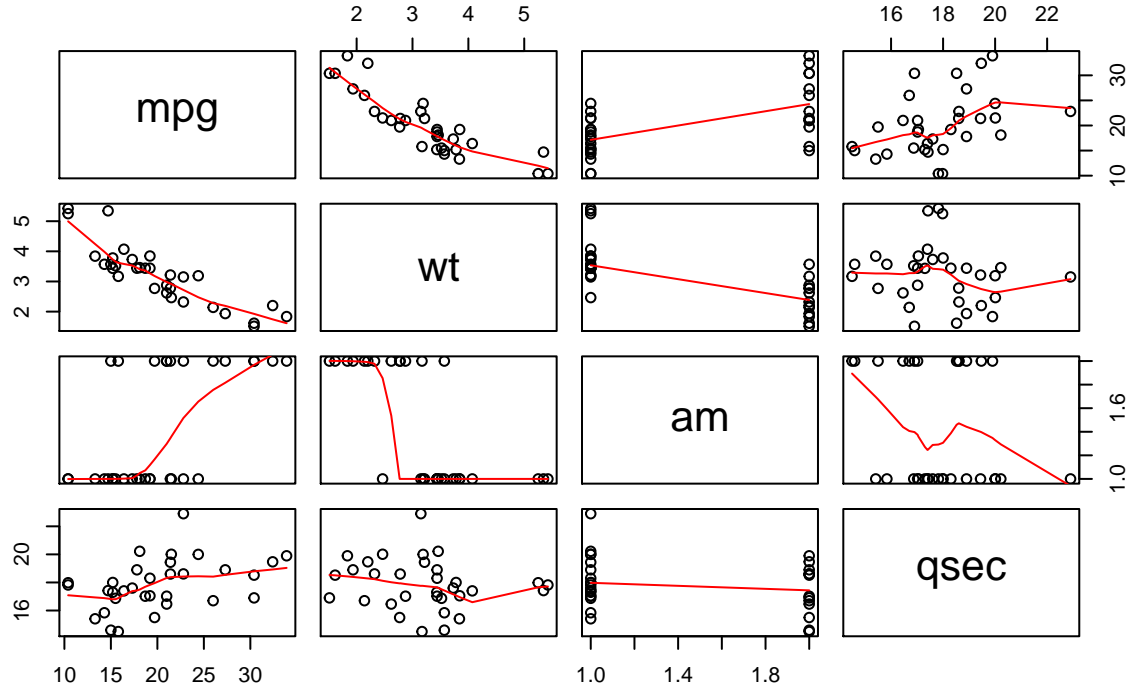
## Appendix: Figures

### 1. Full Pair Graph



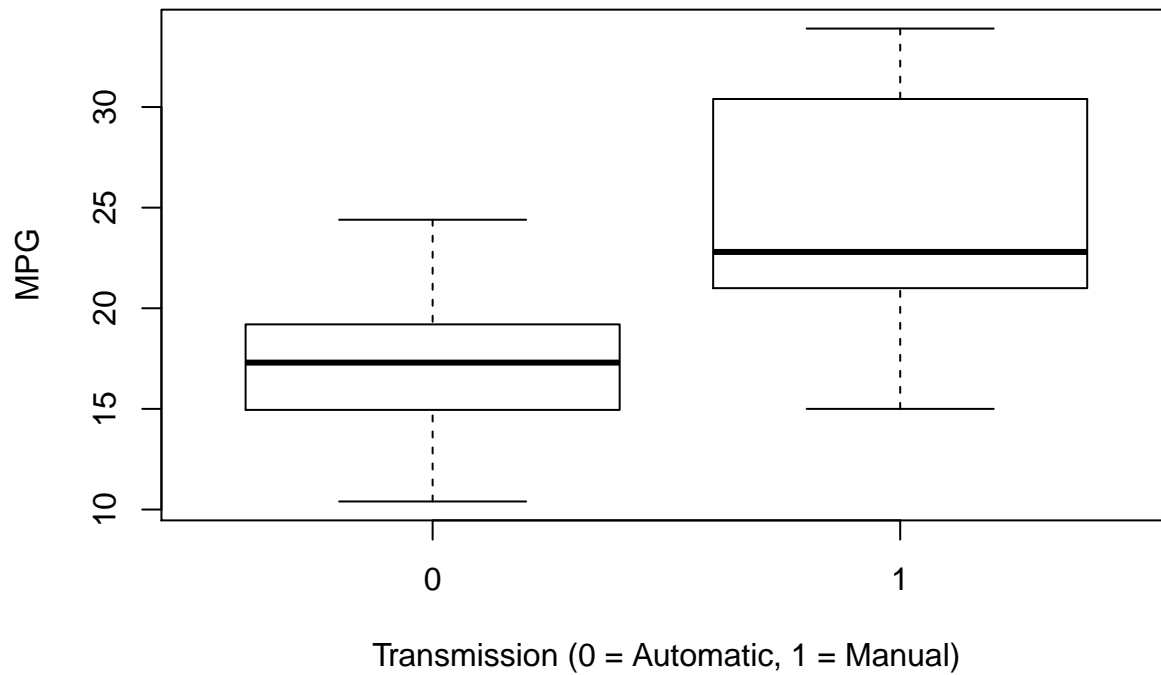
### 2. Selected Dimensions Pair Graph

## Pair Graph of Motor Trend Car Road Tests



3. Box Plot

## Boxplot of MPG vs. Transmission



4. Residual Plots

