

# mtcars

*Conrad Warmbold*

*Saturday, September 20, 2014*

## Executive Summary

Given the data provided in 'mtcars', a linear regression model was calculated using correlation, analysis of predictor variance, and significant linear regression coefficients.

The final resultant model found transmission type to be an insignificant predictor for MPG, therefore no confident conclusions can be made regarding transmission type's correlation with MPG, nor could quantified relationships between the two variables be calculated.

Instead, the only statistically significant parameters – given the provided data – were found to be 'number of cylinders' and 'weight', by which the formula holds true:

```
mpg = 39.6863 + (-1.5078)cyl + (-3.1910)wt
```

## 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:

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions?

## Exploratory Analysis

We're interested in how data variables affect MPG. That said, let's take a quick look at the correlations between the other variables and MPG:

```
##      cyl      disp      hp      drat      wt      qsec      vs      am      gear
## -0.8522 -0.8476 -0.7762  0.6812 -0.8677  0.4187  0.6640  0.5998  0.4803
##      carb
## -0.5509
```

Number of cylinders, displacement, and weight have the strongest correlations with MPG. Let's quantify those relationships further.

## Analysis of the Variance of the Predictors

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## cyl         1     818      818  116.42 5e-10 ***
## disp        1      38       38   5.35 0.0309 *
```

```
## hp          1      9      9      1.33 0.2610
## drat        1     16     16     2.34 0.1406
## wt          1     77     77    11.03 0.0032 **
## qsec        1      4      4     0.56 0.4617
## vs          1      0      0     0.02 0.8932
## am          1     14     14     2.06 0.1659
## gear        1      1      1     0.14 0.7137
## carb        1      0      0     0.06 0.8122
## Residuals   21    147      7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis of the variance of the predictors as they relate to MPG further identifies engine configuration (vs) as the relatively least significant regressor. Let's remove it and reassess our pending model. We'll iteratively continue this until all predictors are significant. Then, we'll calculate coefficients.

## Linear Regression Model

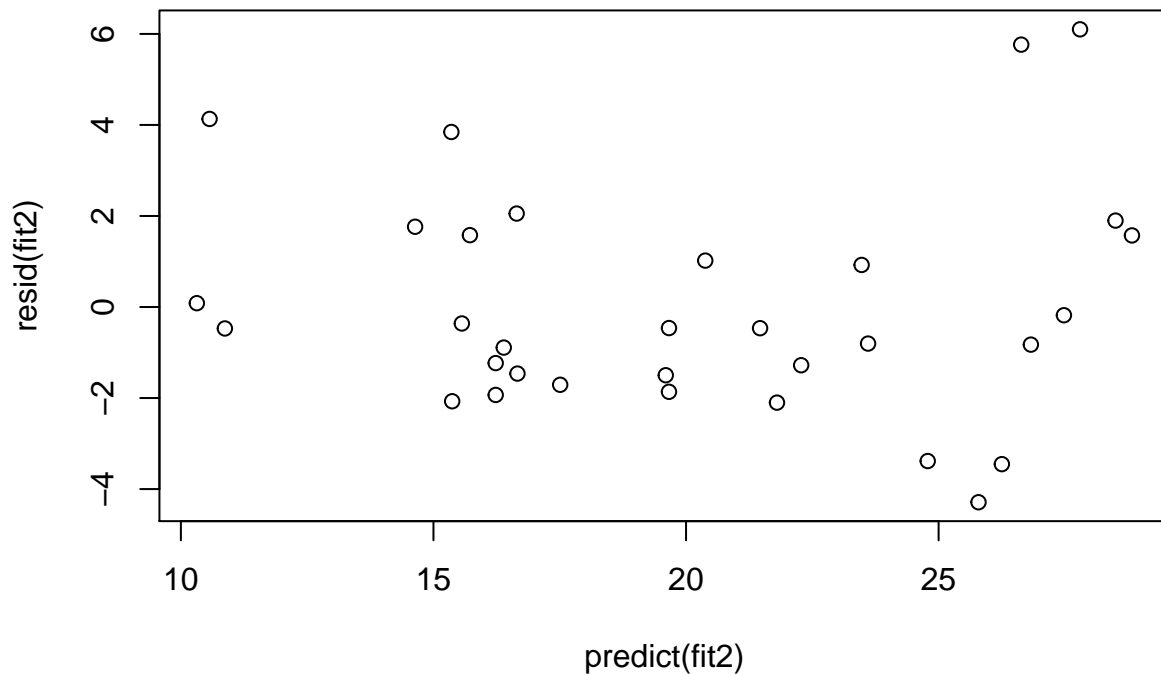
```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 41.107678    2.84243  14.4622 1.620e-14
## cyl         -1.784944    0.60711  -2.9401 6.512e-03
## disp         0.007473    0.01184   0.6309 5.332e-01
## wt          -3.635677    1.04014  -3.4954 1.596e-03
```

With this model, disp becomes insignificant once again, therefore we reduce the data again and reevaluate:

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.686     1.7150   23.141 3.043e-20
## cyl          -1.508     0.4147   -3.636 1.064e-03
## wt           -3.191     0.7569   -4.216 2.220e-04
```

## Residuals

Plotting the residuals produces no obvious patterns:



Therefore, we have a linear model to help answer the pending questions:

1. Is an automatic or manual transmission better for MPG?

*Given the transmission type does not have a significant impact on MPG.*

2. Quantify the MPG difference between automatic and manual transmissions?

*Unable to quantify this regressors impact because it's insignificant by the calculated model*

## Appendix A

All code used in this report:

```
library(datasets)
data(mtcars)
cor(mtcars, mtcars$mpg)[-1,]
summary(aov(mpg ~ ., data=mtcars))
...
fit1 <- lm(mtcars$mpg ~ cyl + disp + wt, data=subset(mtcars, select=c(cyl, disp, wt)))
summary(fit1)$coefficients
...
fit2 <- lm(mtcars$mpg ~ cyl + wt, data=subset(mtcars, select=c(cyl, wt)))
summary(fit2)$coefficients
plot(resid(fit2) ~ predict(fit2))
```