Exploring the relationship of MPG for Automobiles

Kev Scarr

Saturday, August 09, 2014

1 Introduction

The motor trend magazine are interested in exploring the relationship between miles per gallon and other measurements for a set of cars. They particularly want to understand:-

- Is an automatic or manual transmission better for MPG (miles per gallon)
- Quantify the MPG difference between automatic and manual transmissions

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) see references for further information.

2 Executive Summary

Transmission type in isolation with MPG have a strong correlation together with automatics averaging 24mpg versus manual transmission at 17mpg (indicating a gap). Other attributes are stronger predictors in isolation such as weight of vehicle as our model suggests for each short ton increase in weight a loss of almost 13mpg is observed.

3 Exploratory Data Analysis

Investigating the data highlights a small sample set with the noted characteristics below. I now take a copy of the dataset as I will tweak a few class types.

3.1 Data Characteristics

The dataset consists of 19 automatic transmission cars and 13 manual cars; it's noted this dataset is small 32 rows and thus represents the more expensive types of cars rather than average family vehicle and thus the results should be kept within this context. There are 11 attributes including transmission ('am') which is stored as a numeric.

To simplify the model statements we will convert a number of the attributes to factors to remove the 'as.factors(x)' statements that quickly amass during the investigation.

```
mtc$cyl = factor(mtc$cyl); mtc$vs = factor(mtc$vs); mtc$am = factor(mtc$am)
mtc$gear = factor(mtc$gear); mtc$carb = factor(mtc$carb)
#str(cars) # 5 factor variables and 6 numerics
```

3.2 Best transmission type for MPG?

Figure '1' highlights the MPG range for Automatic transmissions being lower than the range of that of Manual transmissions (max mpg for an automatic is 24.4 and the min mpg for a manual is 15 showing there is an overlap) with a quantifiable difference of 7mpg between the mean average for both types. This indicates on average Manual Transmissions are more fuel efficient than Automatics.

4 Model Strategy

The pairs in figure 2 suggest cylinders, horsepower and weight have a strong correlation with mpg (suggesting they would be good predictors and possibly have a stronger influence on mpg). I will utilise cylinders and weight to see the affects these have in combination on mpg.

```
fit1 <- lm(mpg ~ cyl + wt, mtc)
newdata <- data.frame( cyl=factor(c(4,6,8)), wt=-3.2056)
res <- predict(fit1, newdata)
res[3] - res[1] # -6.071
```

I observe a reduction of 6mpg for an increase from 4 cylinders to 8 cylinders in a car. But what impact does weight have, the illustration below shows two models, with and without in order to discover how influential cylinders are (by holding weight constant). The coefficients are visible in figure 4 of the appendix (intercept as mpg) showing as weight (per ton) increases yields a loss of 3mpg.

```
fit1 <- lm(mpg ~ cyl + wt, data=mtc)
fit2 <- lm(mpg ~ cyl, data=mtc)
newdata1 <- data.frame( cyl=factor(c(4,6,8)), wt=-3.2056)
newdata2 <- data.frame( cyl=factor(c(4,6,8)), wt=0)
res1 <- predict(fit1,newdata1)
res2 <- predict(fit2,newdata2)
res1[3] - res2[3] # 23.1
```

This indicates that by holding weight constant, cylinders appear to have less of an impact on mpg than if weight is disregarded. Now conducting a P-value/Shapiro-Wilk diagnostic test on the model.

```
fit3 <- lm(mpg ~ cyl * wt, mtc)
shapiro.test(fit3$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit3$residuals
## W = 0.9575, p-value = 0.234
```

```
#anova(fit1,fit2,fit3)
#exp(confint(fit3, 'wt')) #95% confidence interval
```

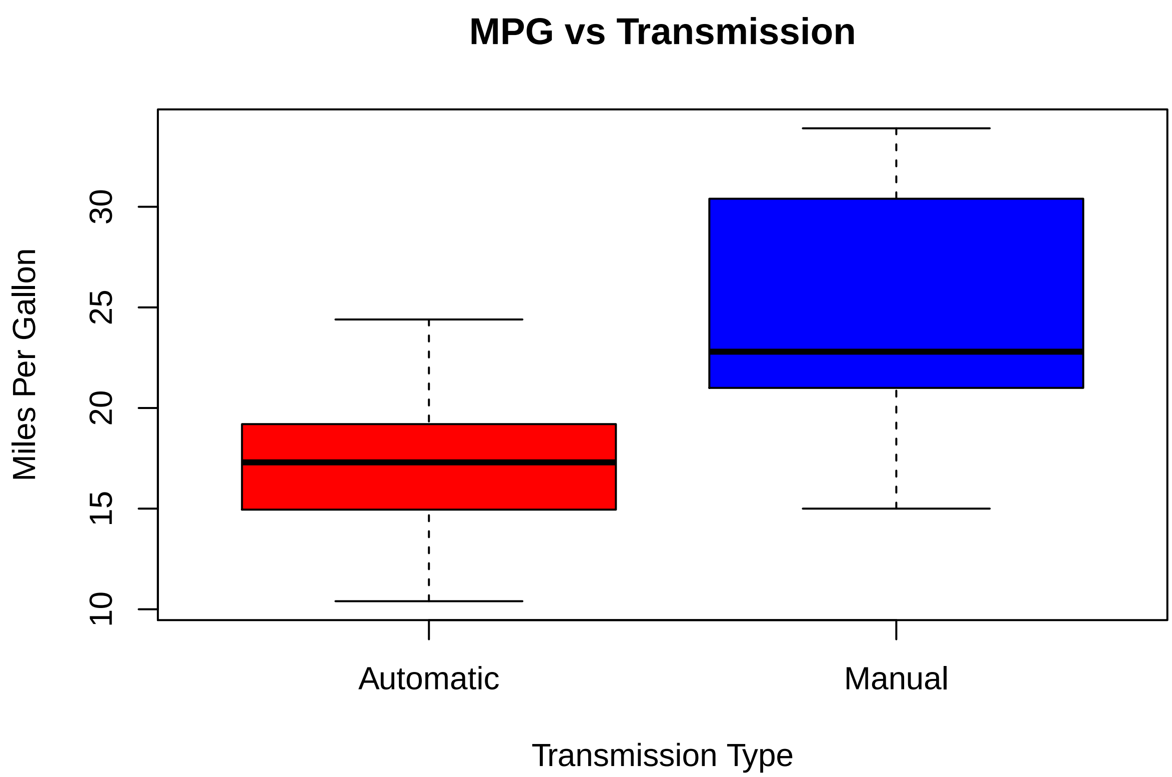
A p-value of 0.3325 fails to reject the null hypothesis (normality being the null hypothesis). Omitting a regressor can bias estimation of the coefficient of certain other regressors. If we were to add additional regressors to the model at this stage we would need to test the residual sums of squares for significance above and beyond that of reducing residual degrees of freedom. So far we've not been shown an automatic way of doing this, so a manual repeating exercise would be the next step.

5 Conclusion

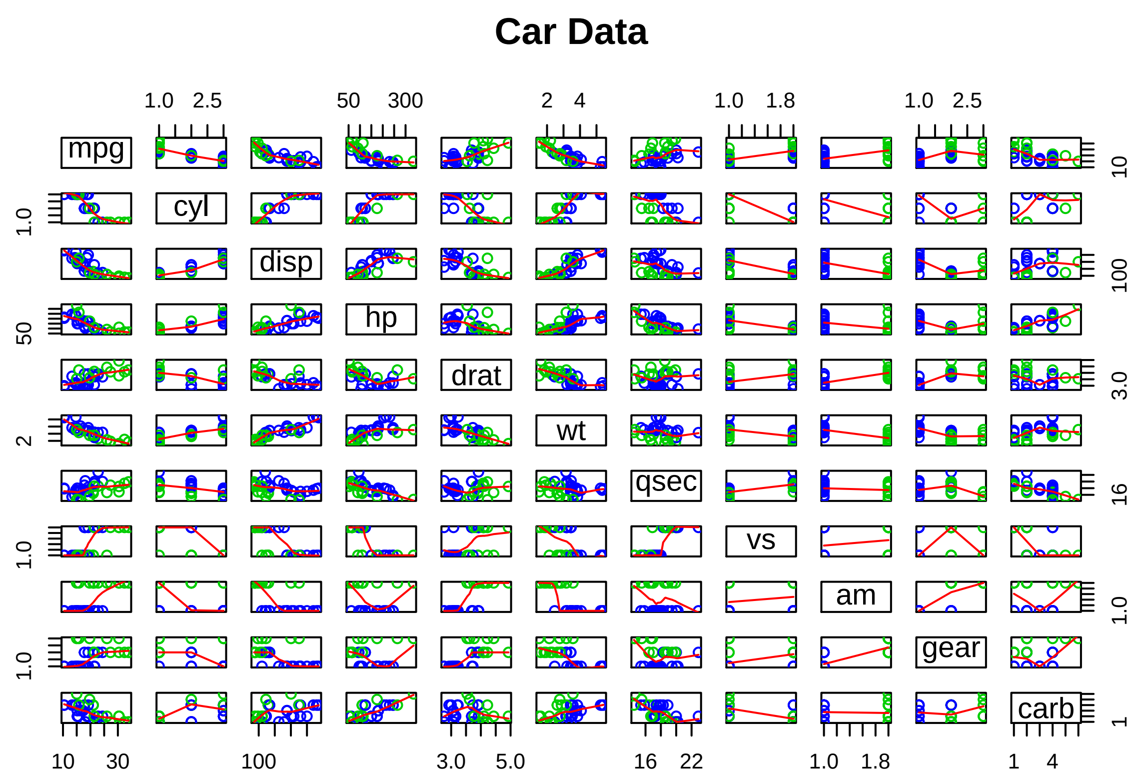
In conclusion we've demonstrated there are singular predictors that are more influential than transmission type (e.g, weight) but also weight and cylinders together perform well as a model for predicting mpg (with weight being one of the most influential factors to consider). There are some outliers in the data but these do not hold influence over the model as demonstrated by figure 3.

6 Appendix

6.1 Figure 1

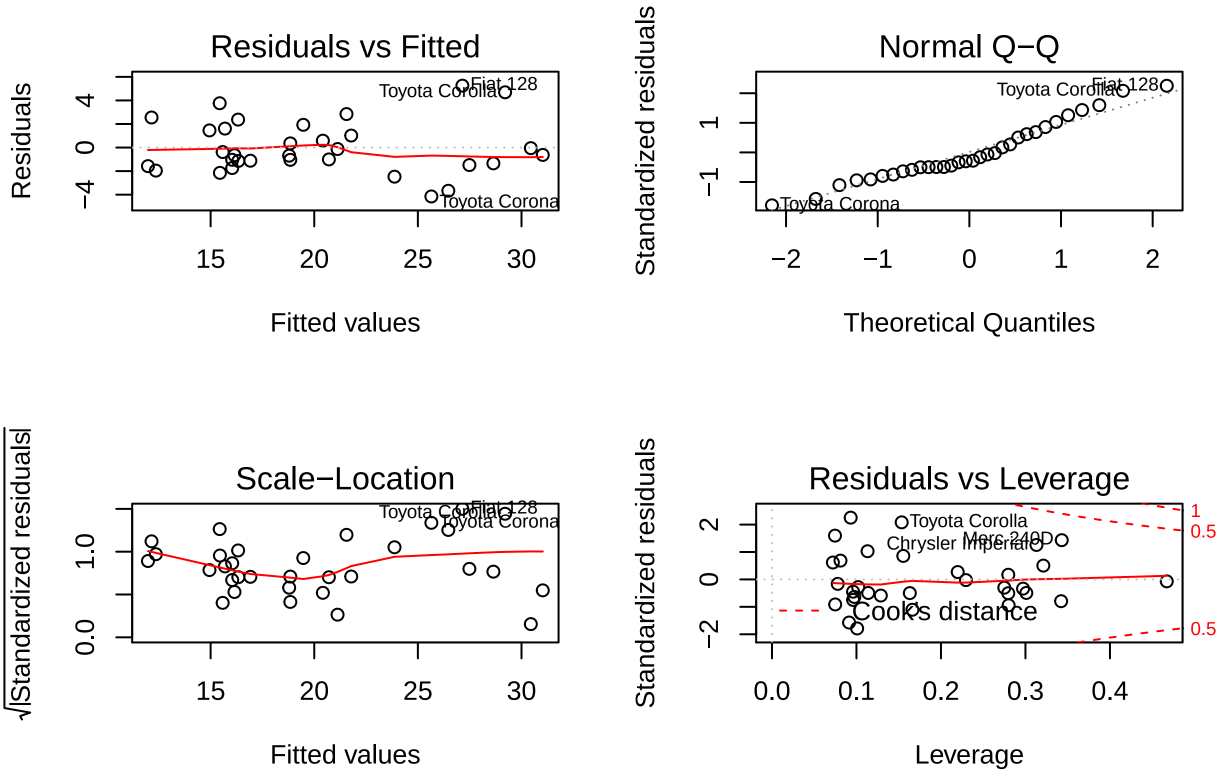


6.2 Figure 2



6.3 Figure 3

```
par(mfrow=c(2,2))
plot(fit3)
```



6.4 Figure 4

```r
summary(fit1)$coef
```

```
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    33.991     1.8878  18.006 6.257e-17
## cyl6           -4.256     1.3861  -3.070 4.718e-03
## cyl8           -6.071     1.6523  -3.674 9.992e-04
## wt             -3.206     0.7539  -4.252 2.130e-04
```

7 References

The following references we're used:-

- Coursera Peer Assignment
- MTCars codebook
- Source .Rmd GitHub Repo