Transmission Type Impact on Miles per Gallon

Joy Flowers

January 17, 2016

Executive Summary

This analysis uses linear regression to assess which transmission type (Automatic or Manual) is better for miles per gallon (mpg) using the mtcars dataset. It is first established that there is a significant relationship between transmission type and mpg, and further it is established that while overall, manual transmissions give higher gas mileage, this is only true for lightweight cars. In the case of cars weighing more than a ton, automatic transmission gives better gas mileage.

Introduction

A question is posed, "Is an automatic or manual transmission better for MPG". Secondly, the task is to "Quantify the MPG difference between automatic and manual transmissions" given the mtcars dataset which was extracted from data in the 1974 Motor Trend magazine. The mtcars dataset contains data such as is shown below.

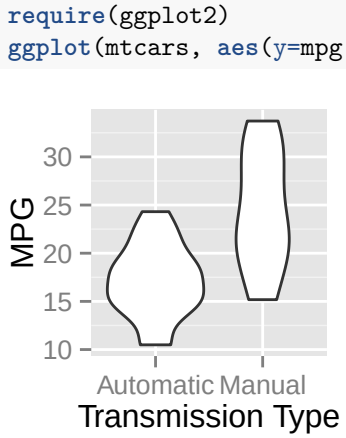
```
head(mtcars,3)
```

```
##                mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
```

Analysis

Having a first glance at the transmission type shows that manual transmission produces higher (better) mpg than automatic transmission.

```
mtcars$am <- as.factor(mtcars$am)
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$vs <- as.factor(mtcars$vs)
require(ggplot2)
ggplot(mtcars, aes(y=mpg, x=factor(am,labels = c("Automatic", "Manual")),am))+geom_violin()+xlab("Transm
```



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

The t-test with a p-value of 0.0013736 shows that this is not by chance. So this is an indication that Transmission Type matters.

To determine which predictors effect miles per gallon, one strategy is to choose predictors based on their relationship with the dependent variable, mpg in this case. Using the formula **all <- lm(mpg~.,mtcars); summary(all)**, starting with all of the potential predictors and then one by one, removing the one with the largest p-value (and each time re-running the lm test) until all remaining predictors are significant will leave only those that are significant. Similarly, using the backward elimination method of stepwise regression, model selection using the Akaike information criterion (AIC) yields the same results:

```
require(knitr)
modsel <- step(lm(mpg~.,data=mtcars),direction = "backward")
```

```
kable(summary(modsel)$coefficients,digits=2)
```

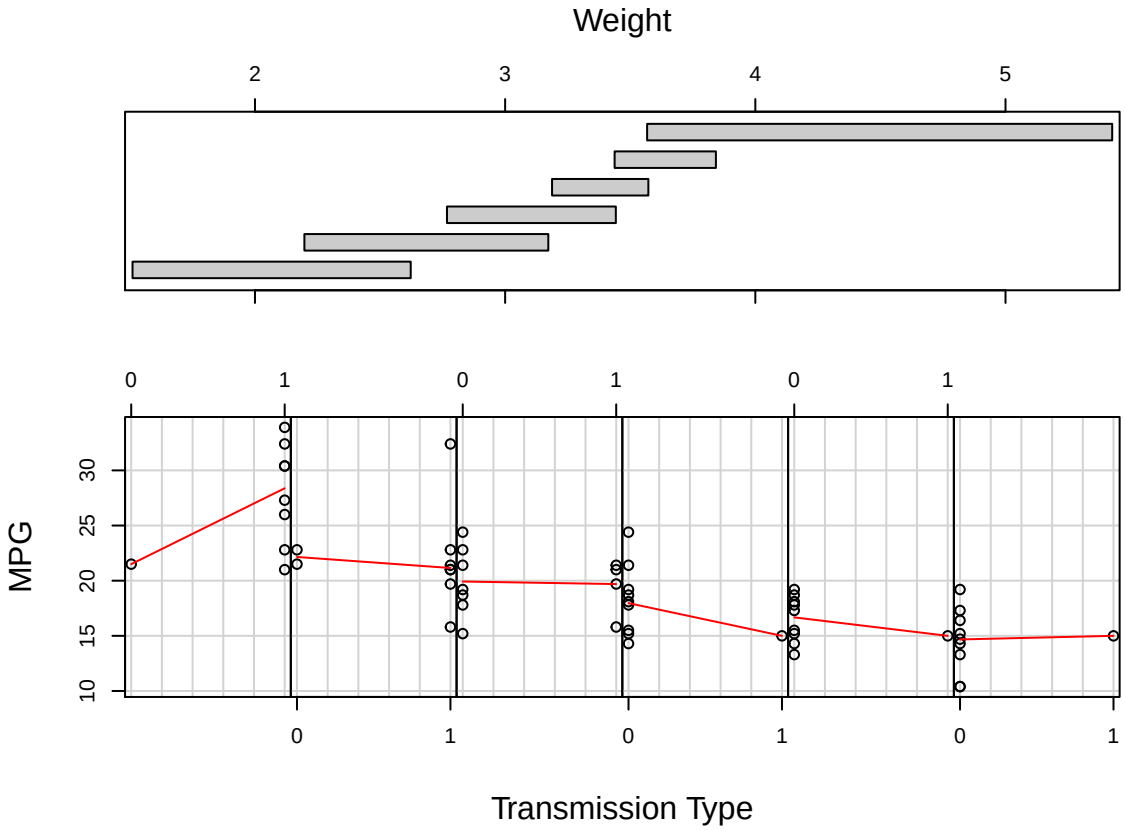
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.62	6.96	1.38	0.18
wt	-3.92	0.71	-5.51	0.00
qsec	1.23	0.29	4.25	0.00
am1	2.94	1.41	2.08	0.05

The backward method recommends the following predictors in addition to the automatic transmission: wt + qsec and model coefficients: 9.6177805 -3.9165037 1.225886 2.9358372. While weight (wt) seems intuitive, qsec which is the does not. The adjusted R squared value including qsec is

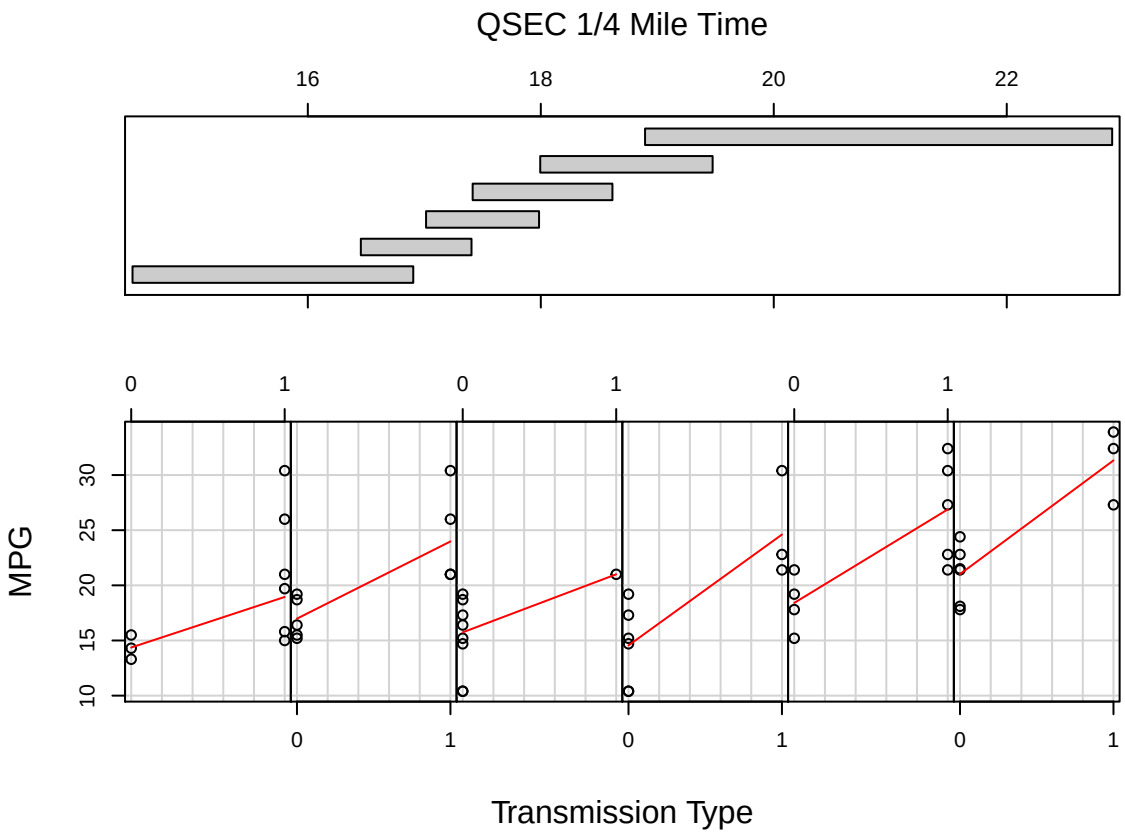
To get an idea of how automatic (0) and manual (1) transmission effect mpg for each weight bucket, see the plot below which shows that with the exception of cars under about one ton, the mpg is a little bit lower for manual transmissions than for automatic transmissions, but relative to qsec, mpg is higher for manual transmissions.

```
require(graphics)
##pairs(mtcars, main = "mtcars data")
```

```
coplot(mpg ~ am | wt, data = mtcars,panel = panel.smooth, rows = 1, xlab = c("Transmission Type","Weigh
```




```
coplot(mpg ~ am | qsec, data = mtcars,panel = panel.smooth, rows = 1, xlab = c("Transmission Type", "QSI
```




```
table(mtcars$am)
```

```
##
##  0  1
## 19 13
```

More than half (7/13) of the cars with manual transmission are lightweight cars most under a ton. Because of the strong interaction between weight and transmission type, a comparison is made using AIC to determine if an interaction model is better. To do this, the recommended model above is compared to using just the transmission type. The interaction model with a lower AIC is better.

```
Int <- lm(mpg ~ am + wt + qsec + am*wt, data=mtcars)
AIC(modsel,Int)
```

```
##        df      AIC
## modsel  5 154.1194
## Int     6 144.3736
```

```
kable(summary(Int)$coefficients,digits=2)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.72	5.90	1.65	0.11
am1	14.08	3.44	4.10	0.00
wt	-2.94	0.67	-4.41	0.00
qsec	1.02	0.25	4.04	0.00
am1:wt	-4.14	1.20	-3.46	0.00

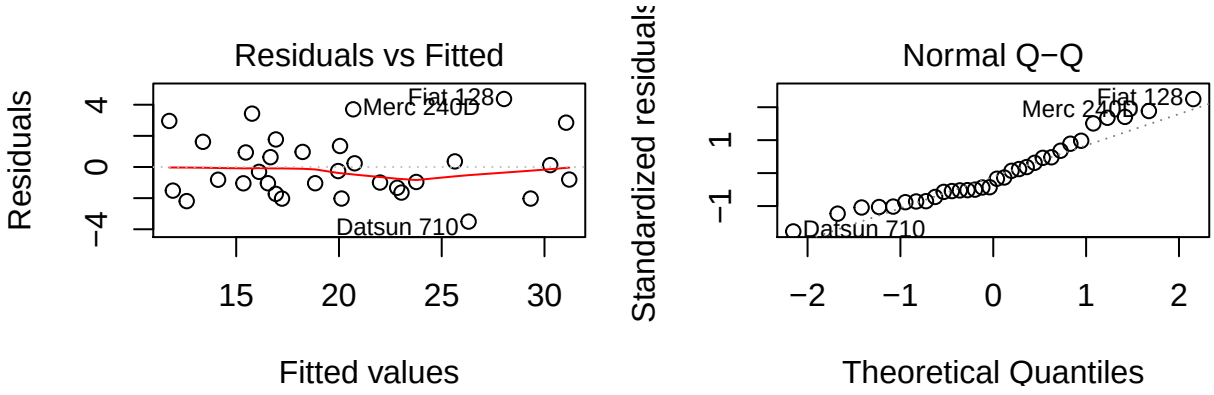
The new adjusted R squared value is 0.8804219. Now an analysis of variance will help determine if there is any confounding variables.

```
amonly <- lm(mpg~am,mtcars)
anova(modsel,amonly)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + qsec + am
## Model 2: mpg ~ am
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     28 169.29
## 2     30 720.90 -2   -551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value of tells us that the variables do contribute to the model.

```
#plot(modsel)
par(mfrow=c(1,2), mar=c(4,4,2,1)); plot(Int,which=1); plot(Int,which=2)
```



Note that the residuals versus fitted values appear random. The residuals appear to be normally distributed

Conclusion

This analysis has shown that manual transmissions give higher gas mileage for lightweight cars. In the case of cars weighing more than a ton, automatic transmission gives better gas mileage. A car having manual transmission, holding other variables fixed, with weight in thousands, will have a difference in mpg from that of an automatic transmission given by the following equation: 14.08 - 2.94wt + 1.02qsec - 4.14wt + 9.72 which is 23.8 - 7.08wt + 1.02 qsec.