Car characteristics affecting MPG in cars

Dominic Tabeta

May 9, 2016

Executive Summary

In this report we try to answer the question "Is an automatic or manual transmission better for MPG?" and we also try to quantify the MPG difference between automatic and manual transmissions. We compare MPG of 32 cars with the following features and try to quantify the MPG impact of some of these features.

Data Processing

```
data(mtcars)
apply(mtcars,2,class)
```

```
##       mpg       cyl      disp        hp      drat        wt      qsec
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##        vs        am      gear      carb
## "numeric" "numeric" "numeric" "numeric"
```

The am predictor is a numeric class so for better interpretability we convert it to a factor class instead, converting 0s to Automatic and 1s to Manual.

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("Automatic", "Manual")
```

Exploratory analysis

First we'll try to address the question "Is an automatic or manual transmission better for MPG?". We see from the below graphs that at manual cars have higher mean and median mpg than automatic transmission cars. I have added to the below boxplots the mean mpg for each group indicated by a red point and text. At first glance the boxplot indicates that manual cars have higher mpg, but there is significant overlap of mpg between the two groups which suggest that determing the mpg of a car depends on more than just whether the transmission is manual or automatic as the am feature does not explain all the variance of mpg.

The means of Automatic and Manual transmission cars are 17.15 and 24.39 respectively. Using t.test, we test if this MPG difference is significant.

```
##
##  Welch Two Sample t-test
##
## data:  mtcars[mtcars$am == "Automatic", "mpg"] and mtcars[mtcars$am == "Manual", "mpg"]
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

Assuming an alpha level of 0.05 and all other car characteristics being equal p-value of 0.001374, we can reject the null hypothesis and conclude that automatic vehicles have lower mpg than manual vehicles. However the response of mpg to other characteristics needs further investigation.

Simple Linear regression model using only am

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## amManual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

A linear regression model using only am as a predictor has an R-squared value of only 0.3598, meaning this simple model only explains 35.98% of mpg's variance, which suggests a more accurate model predicting a car's mpg should use other predictors as well. Under this simplified model, the amManual coefficient suggests that manual cars get 7.245 more miles per gallon than automatic vehicles.

A more complex regression model

We call the step() function to return a more complex regression model using mpg again as the response.

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
```

```
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

The stepwise algorithm returned a linear regression model that used predictors am, qsec and wt. The R-squared value of this model is 84.97% which explains the mpg variance to a much higher degree. While the model uses am as a predictor, the p-value for am is higher than that of the other predictors used which suggests it is the least significant of the three predictors used.

The amManual coefficient in the summary quantifies the effect of transmission type on mpg. For vehicles with the same qsec and wt characteristics, manual vehicles get 2.9358 additional miles per gallon than automatic vehicles.

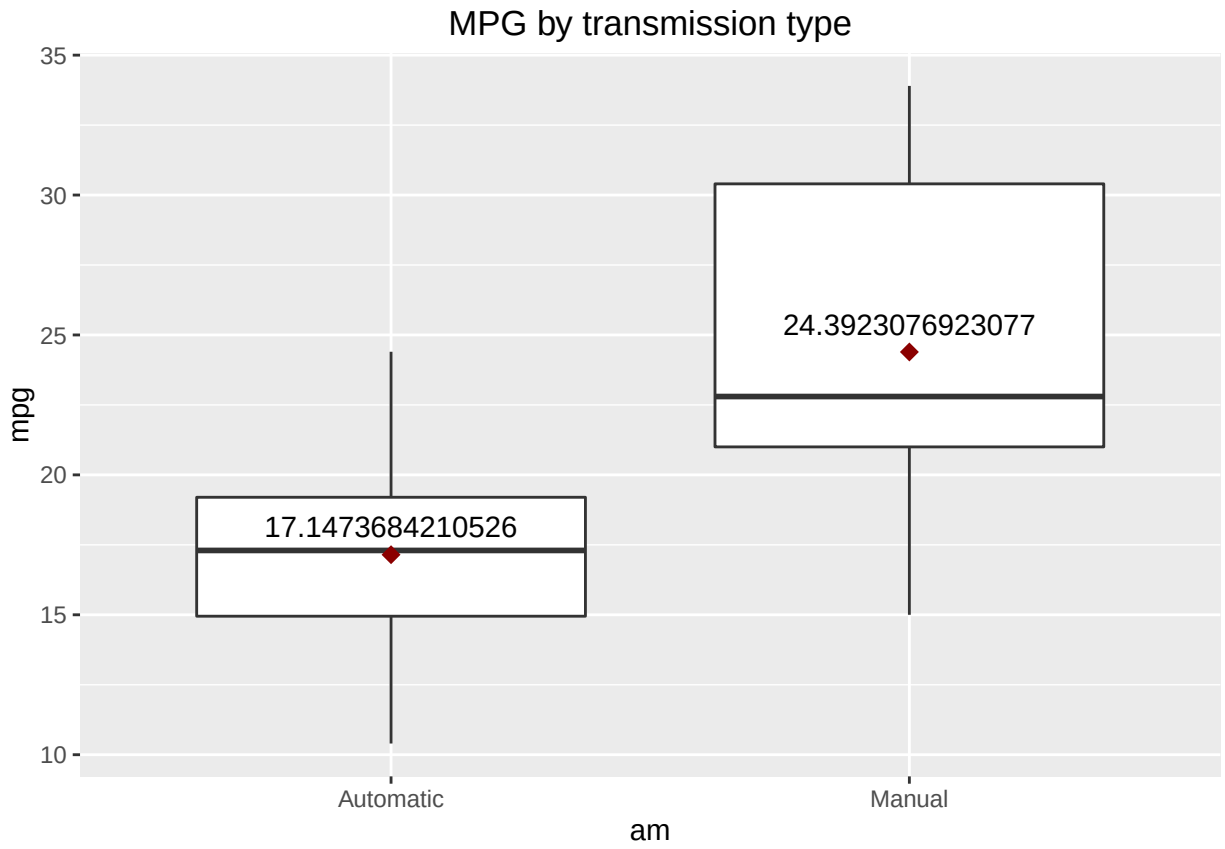
The residual diagnostics indicate normality and a random pattern of variation. Refer to the appendix for the residual diagnostic plots for the model returned by the stepwise function.

Results

In both the simple and final regression models we used we oberve a tendency for manual cars to have greater mpg than automatic cars. The simpler model indicated that mpg variation could be well explained by am alone. By using additional predictors qsec and wt, the second model explained 84% of the variation and also quantified the mpg response to transmission type. For vehicles with the same qsec and wt characteristics, manual vehicles get 2.9358 additional miles per gallon than automatic vehicles.

Appendix 1: Boxplot of am vs mpg

```
means <- aggregate(mpg ~  am, mtcars, mean)
suppressWarnings(g<-ggplot(data=mtcars,aes(y=mpg,x=am))+
        geom_boxplot()+stat_summary(fun.y=mean, colour="darkred", geom="point", shape=18, size=3,show_g
```


MPG by transmission type
35
30
25
20
15
10
mpg
24.3923076923077
17.1473684210526
Automatic
Manual
am

Appendix 2: Residual diagnostics for second regression model