Effects of automatic and manual transmission on MPG

Pete Schnettler

February 20, 2016

Executive Summary

Motor Trend magazine has been providing automotive information to their subscribers and interested in answering two questions about fuel efficiency and transmission selection for their readers. Specifically:

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

Utilizing linear regression, this paper will study whether there is a material difference in miles per gallon between cars that have automatic transmissions verses those with manual transmissions.

Data Processing and transformation

For this analysis we will use the mtcars dataset and transform variables into factors

```
data(mtcars)
mtcars$cyl  <- factor(mtcars$cyl)
mtcars$vs   <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$am   <- factor(mtcars$am,labels=c("Automatic","Manual"))
str(mtcars)
```

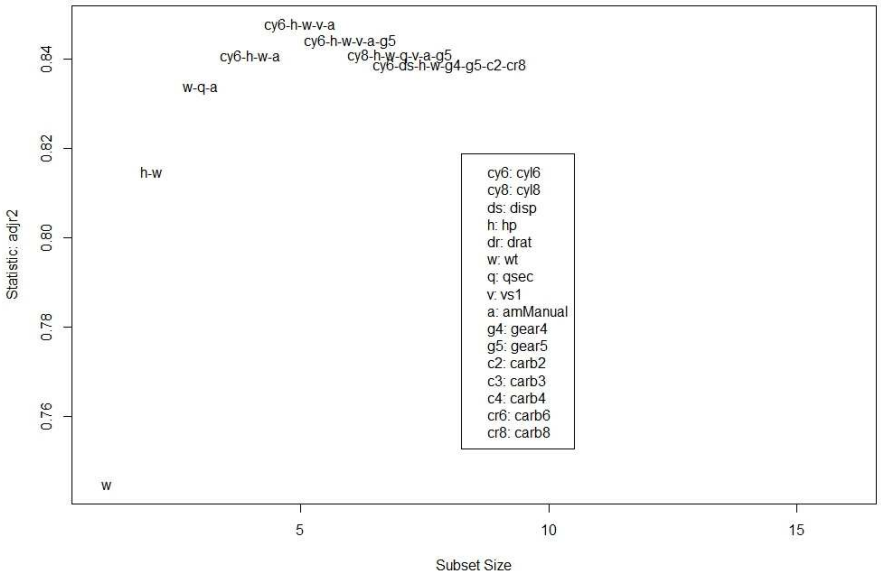
```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : Factor w/ 3 levels "4","6","8": 2 2 1 2 3 2 3 1 1 2 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : Factor w/ 2 levels "0","1": 1 1 2 2 1 2 1 2 2 2 ...
##  $ am  : Factor w/ 2 levels "Automatic","Manual": 2 2 2 1 1 1 1 1 1 1 ...
##  $ gear: Factor w/ 3 levels "3","4","5": 2 2 2 1 1 1 1 2 2 2 ...
##  $ carb: Factor w/ 6 levels "1","2","3","4",..: 4 4 1 1 2 1 4 2 2 4 ...
```

Exploratory Analysis

Model Selection

In order to determine the predictors, we will start with building a model with all of the variables as predictors and perform all-subsets regression using the leaps( ) function from the leaps package. Finnaly, the car library will be used to look at the adjusted r Square values for combinations of variables.

```
library(leaps)
attach(mtcars)
leaps <- regsubsets(mpg ~ ., data = mtcars, nbest=1)
library(car)
subsets(leaps, statistic = "adjr2", legend=c(20, 20))
```

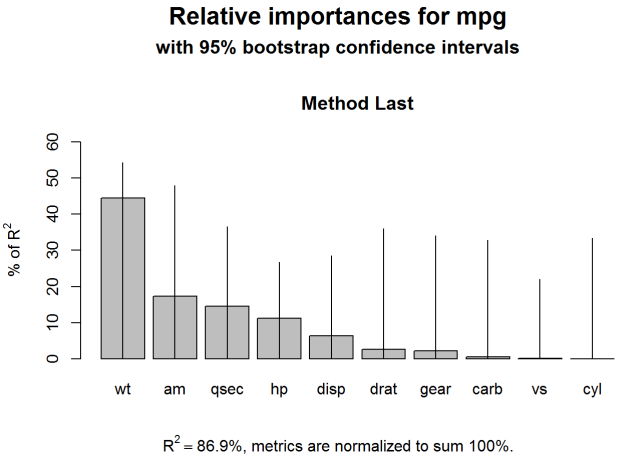


From the above model details,the adjusted R^2 peaks at about 0.85 considering all combinations of variables. Thus, we can conclude that more

than 85% of the variability is explained by the above model.

```
data(mtcars)
fit <- lm(mpg~., data = mtcars)
library(relaimpo)
```

```
relimp <- calc.relimp(fit,type=c("last"),rela=TRUE)
boot <- boot.relimp(fit, b = 100, type = c("last"), rank = TRUE, diff = TRUE, rela = TRUE)
plot(booteval.relimp(boot,sort=TRUE))
```



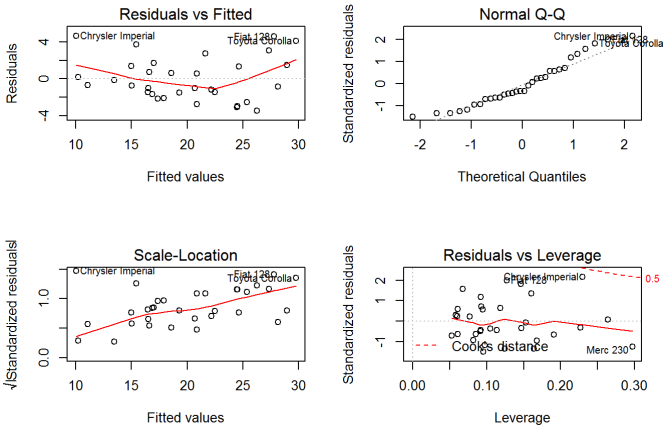
From the plot above, “am” which represents automatic vs manual, ranks second in relative importance for mpg behind weight.

Residuals and Diagnostics

Lets take a look at some residual plots from the regression model along with computation of regression diagnostics for the model. This will examine the residuals and finding leverage points to find any issues with the model.

```
baseModel <- lm(mpg ~ ., data = mtcars)
stepModel <- step(baseModel, direction = "both")
```

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



From the above plots, we can make the following observations,

- The Residuals vs. Fitted plot shows no consistent pattern, supporting the accuracy of the independence assumption.
- The Normal Q-Q plot consists of the points which for tthe most part fall on the line indicating that the residuals are normally distributed.
- The Scale-Location plot confirms the constant variance assumption, as the points are randomly distributed.
- The Residuals vs. Leverage plot indicates that no outliers are present, as all values fall well within the 0.5 bands in the upper right of the plot.

Inference

A t-test will be run to look whether the transmission data has a normal distribution

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

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## 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 in group 0 mean in group 1
##        17.14737        24.39231
```

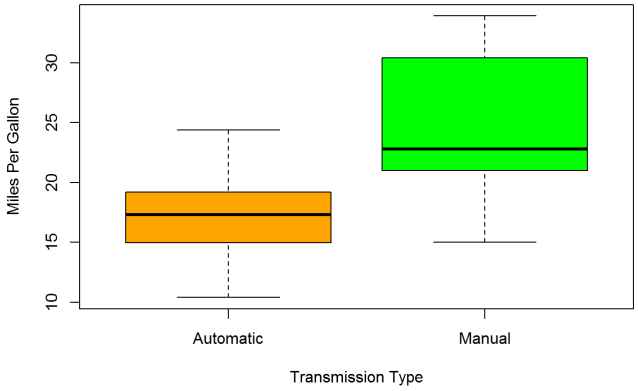
Based on the t-test results and a p-value of .0014, we reject the null hypothesis that the mpg distributions for manual and automatic transmissions are the same.

Conclusion

The first question to answer is:

- Is an automatic or manual transmission better for MPG

```
boxplot(mpg ~ am, data = mtcars, col = (c("orange","green")), names = c("Automatic", "Manual"), ylab = "Miles Per Gallon", xlab = "Transmission Type")
```



From the box plot above we can see that Manual transmissions vehicles have better MPG

The second question to answer:

- Quantify the MPG difference between automatic and manual transmissions

```
fit <- lm(mpg ~ am, data = mtcars)
coef(summary(fit))
```

```
##             Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

From the linear model,

- The average MPG for automatic transmissions is 17.147 MPG (Which is also illustrated in the boxplot above), and
- We can calculate the average MPG for manual transmissions as 24.392 MPG (7.245 + 17.147).
- To answer the question above, the MPG difference between automatic and manual transmission is 7.245 MPG;

```
summary(fit)
```

```
##
## 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 ***
## am             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
```

The model an Adjusted R-squared value is 0.3385, which means that the model can explain about 34% of the variance of the MPG variable. The low Adjusted R-squared value would indicates that we need to add other variables to the model such as weight which we observed in the relative im,portance for mpg chart above..