MPG between AT and MT

AM

Tuesday, March 17, 2015

1. Executive Summary

This report explores if **automatic** or **manual** transmission (AT/MT) cars gives a better **miles per gallon (MPG)** rating. To derive the answer, the dataset 'mtcars' which is a collection of 32 cars is analysed taking into consideration variables such as cylinders, weight, displacement, etc. This report also quantify the MPG difference between AT and MT.

The analysis is first performed by applying a linear model on the data between variables mpg and am to have a rough idea of the answer. To explore the answer further as a whole, we also picked the other variables which have the most influential values and added them into the linear model.

In conclusion, manual transmission cars does have a better mpg compared to automatic transmission cars.

2. Exploratory Data Analysis

The data set contains 32 observations of different cars with 11 variables. All of the variables are quantifiable in numeric values except for 'vs' and 'am'.
- '**vs**' which stands for "V engine shape" or "Straight engine shape" is represented by 0 and 1 respectively.
- '**am**' which stands for "Auto transmission" or "Manual transmission" is represented by 0 and 1 respectively.

Since the question is interested in **am**'s relation with mpg, let's understand this variable further. Throughout this analysis, 'am' will be presented as a factor.

```
##      am1      am0
## 24.39231 17.14737
```

Looking at the above **coefficients** of 'am' as a factor to mpg, removing the Intercept will give us the mean of mpg grouped by am0 (Auto) and am1 (Manual). A biased assumption of manual giving better mpg over auto is formed.

3. Relationship between the variables and MPG

Other variables should also be considered to support our assumption of manual over auto. The strategy is to use analysis of variance (aov) model to determine the variables that have strong relationship with mpg so that an appropriate model can be built. The pairs plot in **Appendix 1** provides an insight of the data.

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          1  817.7   817.7 116.425 5.03e-10 ***
## disp         1   37.6    37.6   5.353  0.03091 *
## hp           1    9.4     9.4   1.334  0.26103
## drat         1   16.5    16.5   2.345  0.14064
## wt           1   77.5    77.5  11.031  0.00324 **
## qsec         1    3.9     3.9   0.562  0.46166
## vs           1    0.1     0.1   0.018  0.89317
## am           1   14.5    14.5   2.061  0.16586
## gear         1    1.0     1.0   0.138  0.71365
## carb         1    0.4     0.4   0.058  0.81218
## Residuals   21  147.5     7.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The variables with lowest significant codes will have the highest impact on mpg. From the results above, we pick the top 3 which are cyl, wt and disp to fit in our nested model testing.

4. Selecting the best model

Nested model testing is performed by first running a linear model of mpg (as outcome) with only am and cyl (as predictor). Then the next linear model is fitted with predictor 'wt' added and repeated by adding 'disp'. The 3 linear model passed into the anova function to determine which is the best by looking at the p values.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + cyl
## Model 2: mpg ~ am + cyl + wt
## Model 3: mpg ~ am + cyl + wt + disp
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     29 271.36
## 2     28 191.05  1    80.315 11.5085 0.002152 **
## 3     27 188.43  1     2.621  0.3756 0.545093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the results, **model 2** will be choosen the best fitted model.

5. Quantified MPG difference between Auto vs Manual transmissions

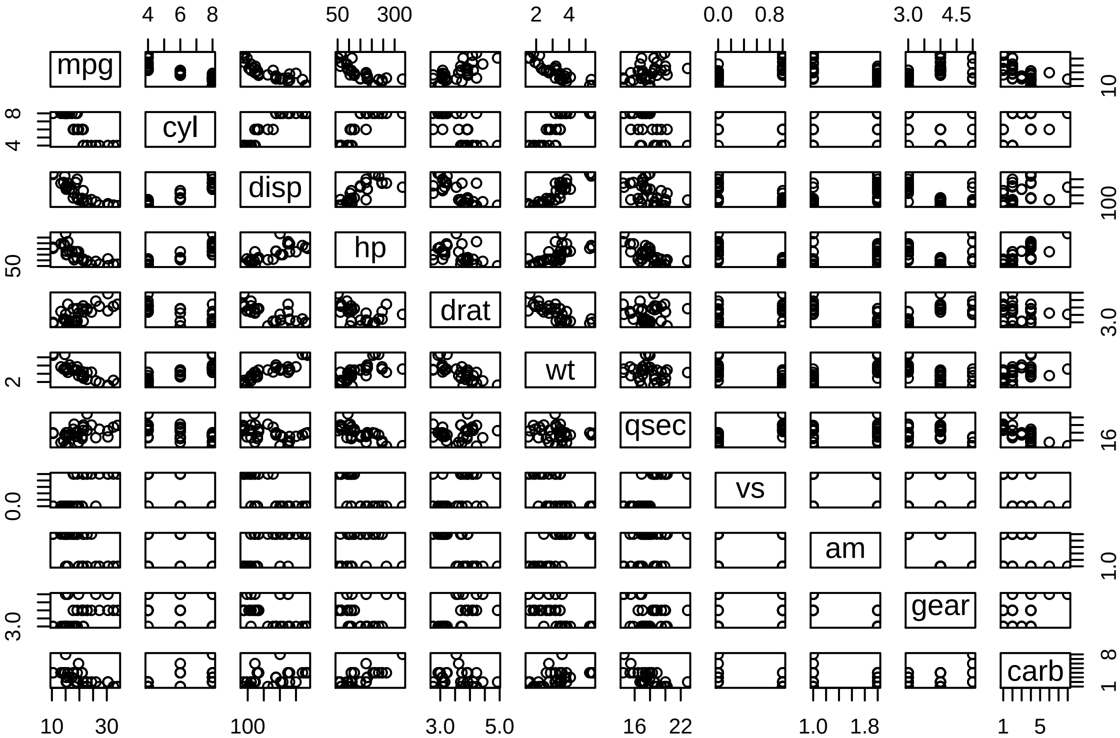
To obtain the mpg difference, a multivariable regression analysis is performed using 'am' as a factor with two levels, *auto (am0)* and *manual (am1)*. By default, the first sorted variable/factor will be the Intercept, which is *manual (am1)* in this case. Our selected model 2 is used for this analysis.

```
##
## Call:
## lm(formula = mpg ~ am + cyl + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.5944     1.8721  21.149  < 2e-16 ***
## am0          -0.1765     1.3045  -0.135  0.89334
## cyl          -1.5102     0.4223  -3.576  0.00129 **
## wt           -3.1251     0.9109  -3.431  0.00189 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 6.51e-11
```

Therefore, the quantified answer is manual transmission cars have better MPG than auto transmission cars by an estimated **0.1765** MPG. Looking at the multiple R squared value, 83% of variance is explained by this model which means this is a reliable model. The residual plot in **Appendix 2** justifies that the distribution of residuals are normal.

References
A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression - R. R. Hocking, Biometrics, Vol. 32, No. 1. (Mar., 1976), pp. 1-49.

Appendix 1 - Pair plot



Appendix 2 - Residual plots