Do Automatic Transmission Systems Reduce the Miles per Gallon (MPG) Rating of Your Car?

Nicholas Ng

Sunday, February 22, 2015

Executive Summary

This report explores the MPG rating and the transmission systems of a number of cars in order to determine if a causal relationship exists, and if so, the effects of this relationship. In doing so, this report will answer the following questions:

1. Is an automatic or manual transmission better for MPG, and
2. What is the MPG difference between automatic and manual transmissions?

By suitably adjusting for confounding factors, the author of this report is 95% confident that cars with a manual transmission system are more efficient than those with an automatic transmission. Additionally, having a manual transmission will increase the MPG rating of a car by approximately 2.94.

Data Exploration

The data set used in this study was provided by the 1974 Motor Trend US magazine, comprising of fuel consumption measures as well as 10 aspects of automobile design and performance for 32 automobiles.

The following packages are used in the analysis of the given dataset:

```
suppressMessages(require("ggplot2"))
suppressMessages(require("reshape2"))
```

In order to understand the data better and pick regressors for the ideal model, pair-wise relationships between the available variables are explored via scatterplots created using the `pairs` command. In addition, the correlation coefficient matrix is calculated via the `cor` command and visually represented in a coloured heatmap after it has been melted into a easily processible data frame.

```
pairsplot <- pairs(mtcars)
```

```
cormatrix <- cor(mtcars)
cormatrix.m <- melt(cormatrix)
ggplot(cormatrix.m, aes(Var1, Var2, fill=value)) +
    geom_tile() +
    scale_fill_gradient(low = "blue", high = "yellow") +
    labs(title = "Correlation Matrix of the mtcars Dataset",
         x = "", y = "")
```

As observed in Appendix A, in addition to the `am` dummy variable (manual cars are represented as 1s and vice versa), there are a number of other variables that appear to influence MPG with expected trends (Fig A). However, it should also be noted that many of these variables are highly correlated amongst each other, and hence would confound the any analysis if included together in the same regression(Fig B).

Regressions & Analysis

To resolve the potential issue of over-specification, a step-wise selection of the appropriate model is carried out. This is achieved via the `step` command, imposing a Bayesian-Schwarz criteria for model selection, while specifying the `am` variable as compulsory.

```
model.full <- lm(mpg ~ ., data = mtcars)
n <- nrow(mtcars)
model.result <- step(model.full, scope = list(upper = model.full, lower = mpg ~
    am), direction = "both", k = log(n))
```

```
summary(model.result)
```

```
##
## 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 ***
## am            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 resulting model keeps the `wt`, `qsec` and `am` variables, which is a logical combination of variables that reduces inter-regressor correlation. Additionally, given that there would be other factors affecting weight and quarter-mile timings, it would be logical to then use the aforementioned variables to mitigate omitted variable bias.

```
par(mfrow = c(2, 2))
plot(model.result)
```

To ensure that the data do not distort the results, plots of the residuals are created via the code above. Despite the highly influential points as marked on Fig C (Appendix B), most residuals fall within an acceptable range. Therefore, no transformations of the data would be required.

Conclusion

In conclusion, it can be observed from the regression summary above that the coefficient of `am` is significant at the 5% significance level. Therefore the type of transmission system used in a car does affect the fuel consumption of the car. Additionally, given the listed coefficient, we may conclude that having a manual transmission adds 2.94 to the MPG rating of the car.

Appendix A - Exploratory Graphs

Fig A. Pair-wise Relationship between Variables

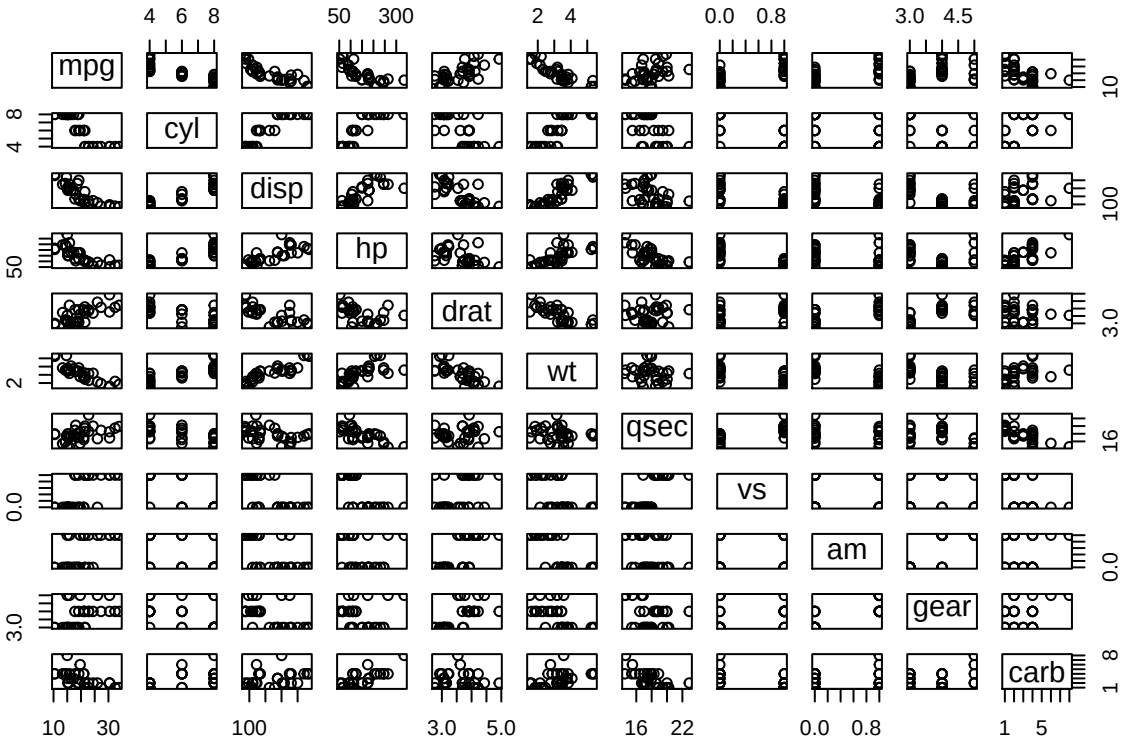
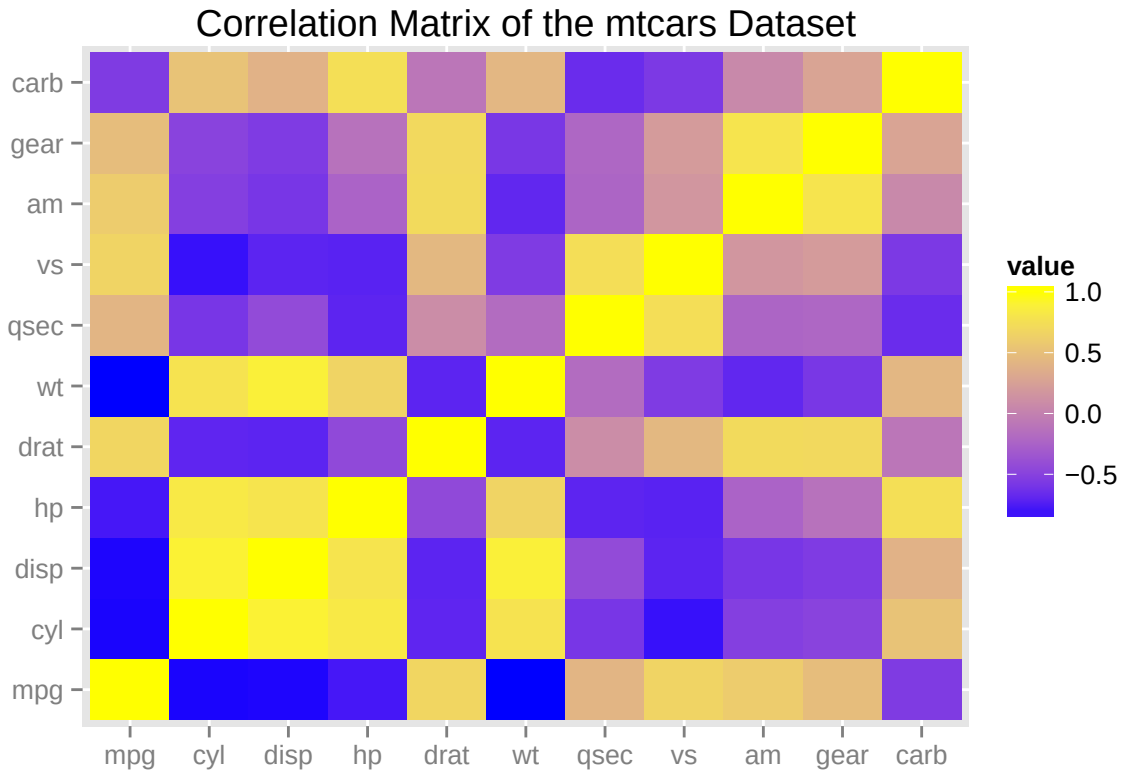


Fig B. Variable Correlation Matrix Heatmap



Appendix B - Residual Plots

Fig C. Residual Diagnostic Plots