Fuel Economy Dependence on Transmission Type

Matthew Gast

June 2015

Executive Summary

At first glance, manual transmissions seem to offer improved fuel economy. However, after controlling for the engine power and weight of the car, the effect of a manual transmission is not statistically significant. The improved fuel economy of manual transmissions is apparently due to the increased prevalence of manual transmissions in lighter cars with smaller engines. The message for car buyers is clear: if you want better fuel economy, purchase a lighter car with a smaller engine, and pay only slight attention to the transmission type.

Exploratory Analysis

The data set analyzed in this paper, `mtcars`, consists of data on 32 automobiles across 10 dimensions. Begin by taking an exploratory look at the data, as shown by figures in the appendix. The boxplot shows that in general, fuel economy is higher for manual transmission cars, but of course, the purpose of this analysis is to determine if that analysis is due to the transmission type or other confounding factors. For example, in the 1970s, automatic transmissions were expensive extra-cost options on smaller and therefore higher-mileage cars, but standard on larger low-mileage cars.

Regression Analysis

To assess the effect of transmission type, we will consider all the available variables that can explain fuel economy. The purpose of this analysis is to determine whether the type of transmission affects fuel economy. Intuitively, the expectation is that a manual transmission will increase fuel economy, and large cars and their associated large engines will decrease fuel economy.

A pairwise plot of variables showing bivariate correlations is in the appendix. From the pairwise plot and calculated correlation values, it is apparent that the number of cylinders, engine displacement, and power are all strongly correlated. In this analysis, we choose horsepower because that is the most commonly known statistic to buyers.

In addition to power, we expect that the weight of the car matters to fuel economy because heavier cars are harder to move. Finally, it is common for mechanically simpler manual transmissions to have more gears. To control for whether the apparent fuel economy of manual transmissions is due to having more forward gears, we will include that in our modeling.

With potentially four predictors to choose from (transmission type, power, weight, and the number of gears), we first perform an analysis of variance between the four models. The analysis of variance shows that adding the number of cylinders and weight offer significant improvements to the model at the .01 level of significance, but the number of gears does not.

```
fit1 <- lm (mpg ~ am, data=mtcars)
fit2 <- lm (mpg ~ am + hp, data=mtcars)
fit3 <- lm (mpg ~ am + hp + wt, data=mtcars)
fit4 <- lm (mpg ~ am + hp + wt + gear, data=mtcars)
anova(fit1, fit2, fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
## Model 4: mpg ~ am + hp + wt + gear
##   Res.Df    RSS Df Sum of Sq       F   Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 68.9298 8.76e-09 ***
## 3     28 180.29  1     65.15  9.4449 0.004923 **
## 4     26 179.34  2      0.95  0.0689 0.933571
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Consider the results of the third model. It explains 84% of the variation in the data, and all coefficients except the transmission type are significant at the .01 level. The coefficients show that increasing power decreases fuel economy, with each 10 horsepower resulting in a 3 mpg decrease. Also as expected, increases in weight decrease fuel economy, with each thousand pounds decreasing fuel economy by about 2.9 mpg. The coefficient on the transmission is positive, well within a standard deviation of zero, indicating that a manual transmission is expected to increase fuel economy by 2 mpg, but the effect is not statistically signficant.

```r
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ am + hp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## ammanual     2.083710   1.376420   1.514 0.141268
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## wt          -2.878575   0.904971  -3.181 0.003574 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

Finally, we consider the robustness of the model. In the appendix, a residual plot shows that the residuals from the model are normally distributed, with only one outlier on the residual plot. However, there are no strong single outliers that warp the model. Not surprisingly, a high-performance Maserati is the biggest single influencer.

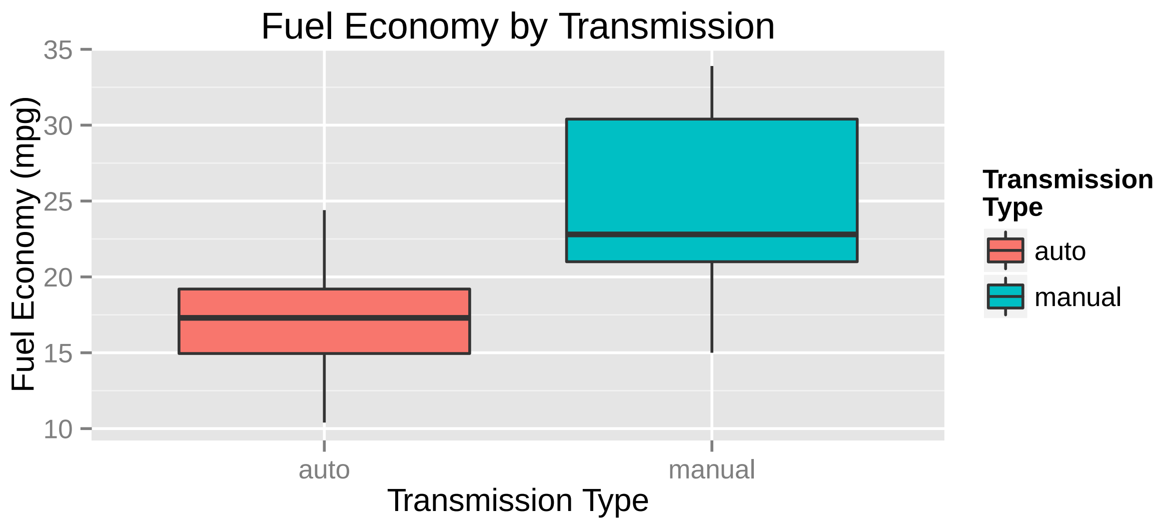
```r
head(sort(hatvalues(fit3),decreasing=TRUE),n=3)
```

```
##       Maserati Bora Lincoln Continental  Cadillac Fleetwood
##           0.4121968           0.2725847           0.2349639
```

Appendix of Figures

1. Exploratory analysis - transmisssion type

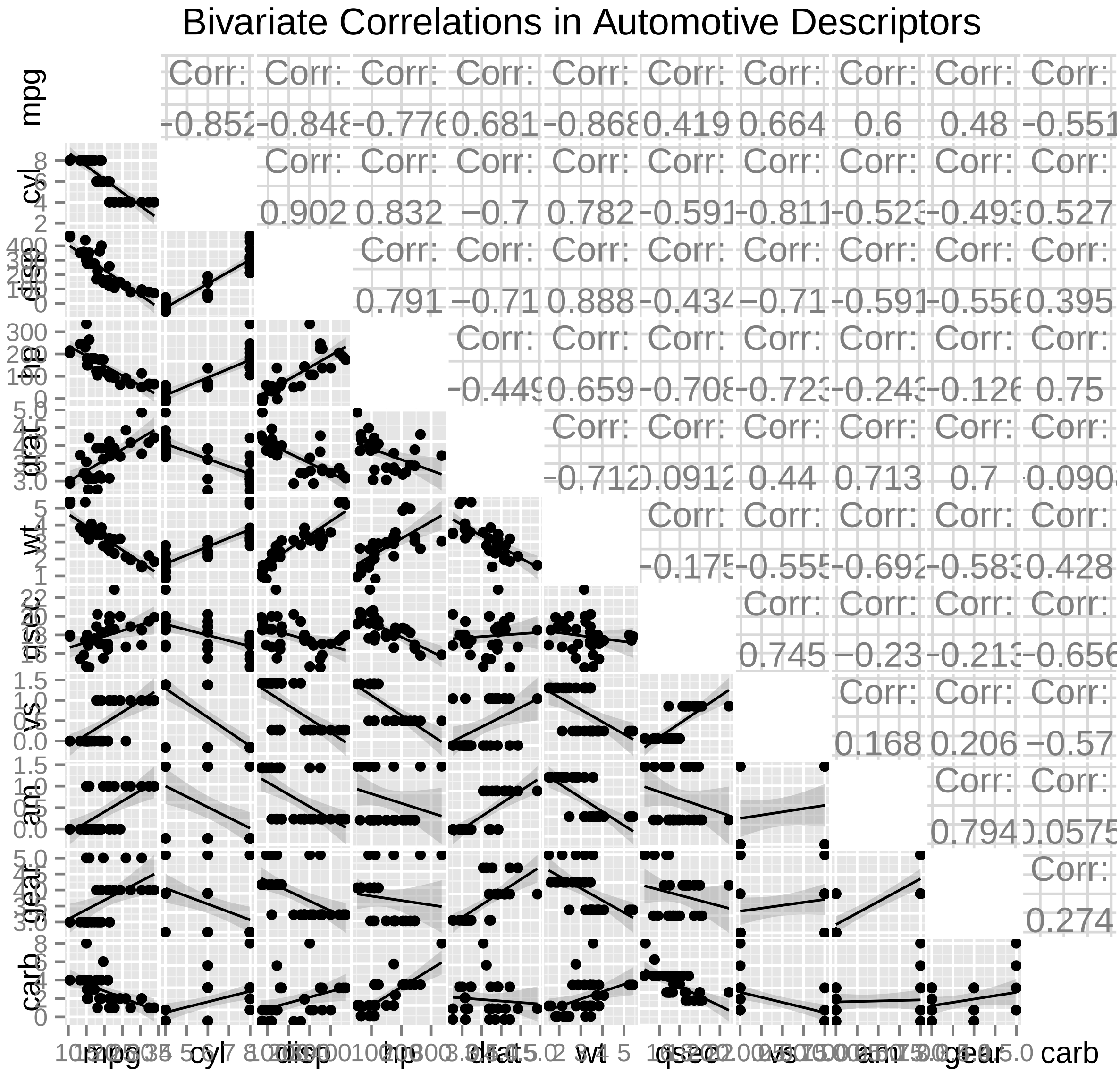
This boxplot compares fuel economy by transmission type. Manual transmissions have a higher average, and the standard deviations of manual and automatic transmissions do not overlap. Manual transmissions appear to have better fuel economy, but of course, this plot does not take into account other factors.



2. Exploratory analysis - pairwise plots

The following pairwise plots show the bivariate correlation between various factors. When building a regression model, it is best to avoid using predictors that are strongly correlated. This pairwise plot depicts correlation visually by plotting data in the lower left side of the diagonal, as well as displaying the calculated correlation in the upper right side of the diagonal.

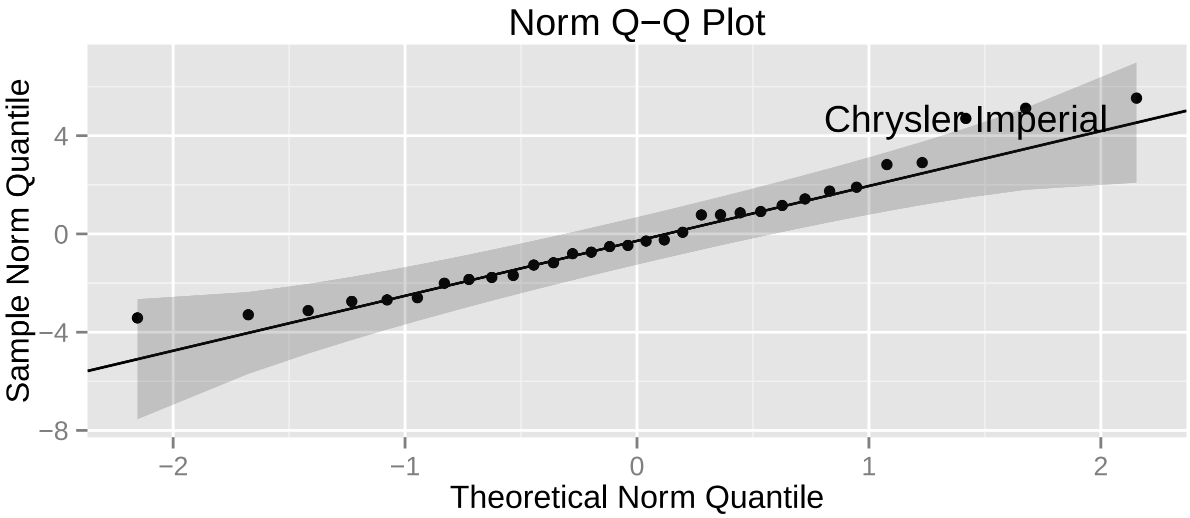
These panel plots and correlation values can be used in selecting predictors for regression models. For example, there are strong correlations between the number of cylinders, engine displacement, and horsepower. These correlations are not surprising, because larger engines develop more power by using more cylinders. When building a regression model, it is best to avoid choosing highly correlated predictors, however.



3. Residual plot

The residuals fitted to the model follow a normal distribution. In fact, there is only one outlier from a 95% confidence range, labeled on the plot.[1]

[1]The `gg_qq` function used here is customized from *Foo0*'s post on StackOverflow: http://stackoverflow.com/questions/4357031/qqnorm-and-qqline-in-ggplot2


Norm Q−Q Plot
Chrysler Imperial
Sample Norm Quantile
4
0
−4
−8
−2
−1
0
1
2
Theoretical Norm Quantile