Talking Transmissions-Impacts on Fuel Economy

Larkin O'Hern

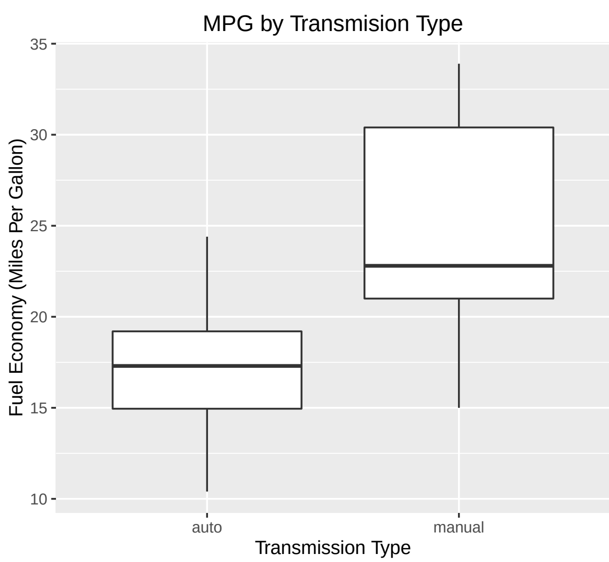
August 7, 2016

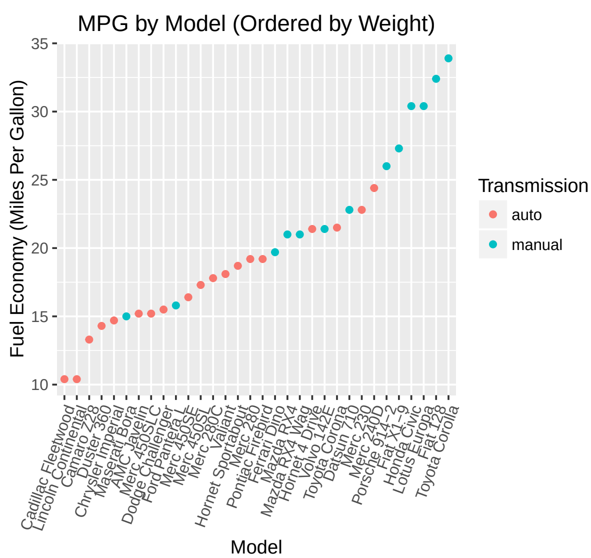
Executive Summary

This analysis explores the relationship between fuel economy (miles per gallon) and vehicle transmission type (automatic or manual) to determine:
Q1. Is an automatic or manual transmission better for MPG?
A. Manual transmissions are associated with better fuel economy for the vehicles in this data set.
2. What is the quantitative MPG difference between automatic and manual transmissions?
A. Once other factors are considered (vehicle weight and horsepower) the impact of transmission type is associated wth a 2.08 (95% CI -.736, 4.903) MPG increase in fuel economy-but that result is not statistically significant at either the .05 or .1 levels. This is likely because for the vehicles in the dataset, weight and horsepower overpower transmission type in determining fuel economy.

Exploratory Data Analysis

The mtcars data contains observations on 11 design and performance characteristics of 32 cars drawn from the 1974 Motor Trend US magazine. I will focus primarily on transmission type (manual vs automatic) and fuel economy (in miles per gallon). The cars in the data set are almost evenly split between manual (13) and automatic (19) giving plenty of observations for each. Fuel economy ranges from 10.4-33.9 MPG with a mean 20.09 and median 19.02. If we condition on transmission type, stark differences emerge.





From the histogram we see that cars with manual transmissions have greater median fuel economy and great variability. From the scatter plot we see that almost all of the low fuel economy cars are automatics while most of the high fuel economy cars are manual.

We can use a quick t test to determine whether the difference in means is statistically significant.

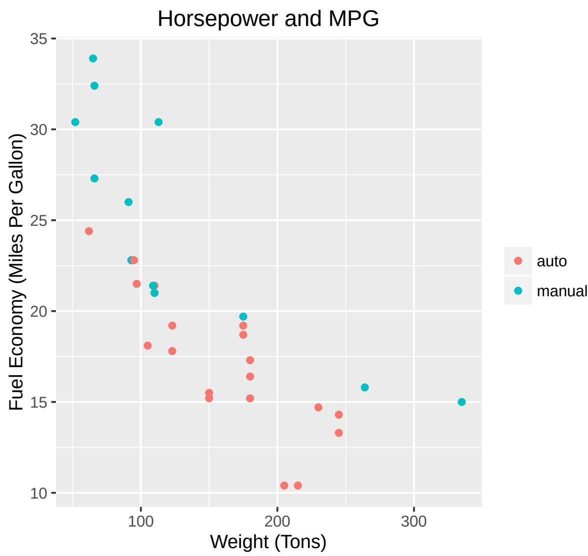
Zero is not included in our confidence interval (-11.2801944, -3.2096842) and our p-value (0.0013736) is below .05 so transmission type appears to make a statistically significant difference in fuel economy-at least when it is the only predictor evaluated.

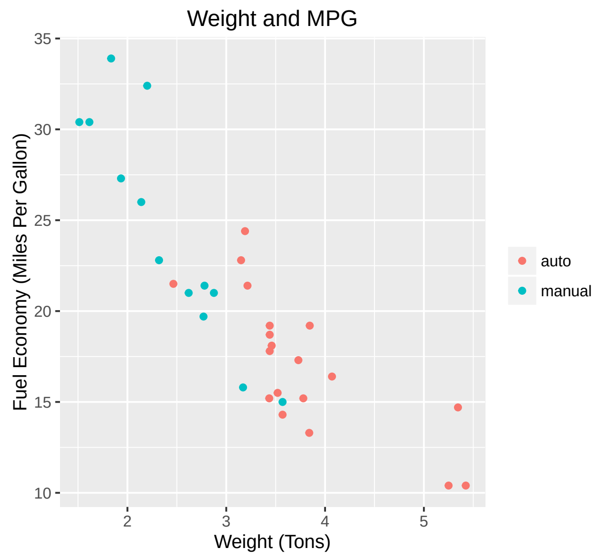
Model Selection

To quantify the effect of transmission type on fuel economy we will fit a series of linear models of increasing complexity. The simplest model examines transmission type as the only predictor of mpg.

This basic model suggests that transmission type is a significant predictor and the coefficient for transmission type indicates an increase of 7.25 MPG for vehicles with manual transmissions versus those with automatic transmission-all else being equal. Expanding our point prediction to encompasses a 95% confidence interval gives us an expected increase of 3.4 to 10.84 MPG in cars with manual transmissions.

However, transmission type explain only about 35% of the variance in MPG-so there are likely other significant variables that could add explanatory power to our model. And importantly, all else is not equal. The data set does not include the same model vehicle with different transmission types. Instead it contains different models where transmission type is correlated with other factors impacting fuel economy-especially weight and horsepower. The distribution of these confounders across models leave very few "like to like" comparisons and increased weight and horse power are both associated with decreased fuel economy AND automatic transmissions as you can see from the plots below.





I next fit five additional models each with one additional predictor between predictor (nested models) and compare them using ANOVA The results are included in the appendix. Vehicle weight and horse power add predictive power-the other variables do not.

I've re-fit the model using the significant predictors to generate a new estimate of the impact of transmission type after accounting for the other key variables.Only weight and horsepower were statistically significant and when they are included transmission type is no longer significant (p=.141). A 1 ton increase in vehicle weight is associate with a loss of 2.87 MPG and each additional unit of horsepower is associated with a loss of .037 MPG. Manual transmissions are associated with a 2.08 increase in MPG-but again the result is not statistically significant. The new model explains 82% of the vairance in MPG. Additional information on the model including confidence intervals for each predictor is available in the appendix.

Residual Diagnostics

To determine the appropriateness of the model we can examine the residuals it generates to verify that they conform to our assumptions of linearity, normality,and homoscedasticity-see residual plots in the appendix. Plotted against fitted values the residuals show no pattern suggesting our linear relationship has captured the systemic variance. Plotted against fitted values the scaled residuals show constant variance (homoskedasticity). The only hang up is the Q-Q Normal plot. Our residuals are not normally distribute which could be cause for further investigation. There are also three potential outliers that future analysis should investigate.

Appendix

ANOVA comparison of Trial Models:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + hp
## Model 4: mpg ~ am + wt + hp + vs
## Model 5: mpg ~ am + wt + hp + carb
## Model 6: mpg ~ am + wt + hp + cyl
## Model 7: mpg ~ am + wt + hp + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 70.7228 5.074e-09 ***
## 3     28 180.29  1     98.03 15.6648 0.0004947 ***
## 4     27 168.96  1     11.33  1.8101 0.1896852
## 5     27 174.93  0     -5.96
## 6     27 170.00  0      4.93
## 7     27 179.91  0     -9.91
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Final Model Summary:

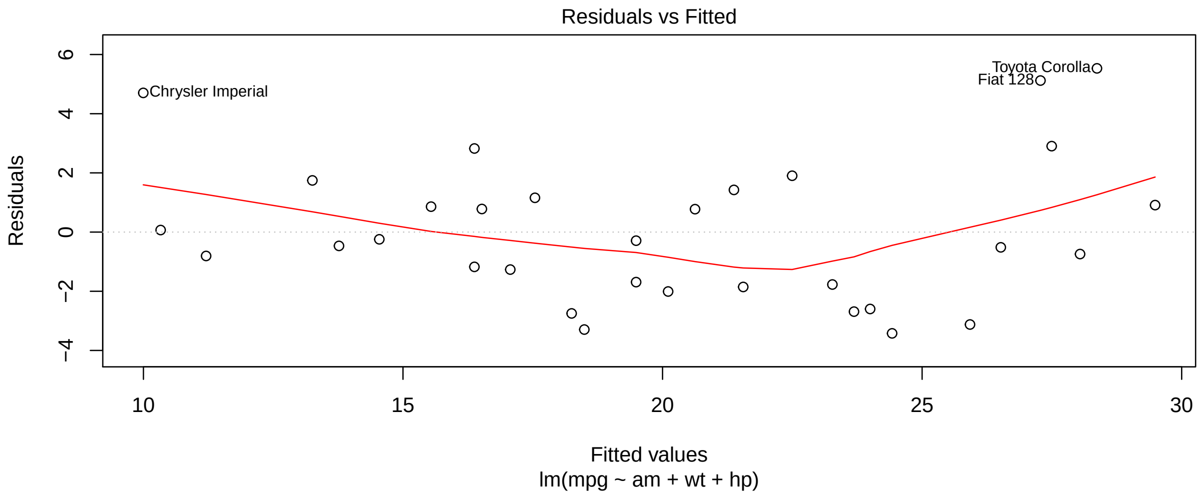
```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## ammanual     2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## 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
```

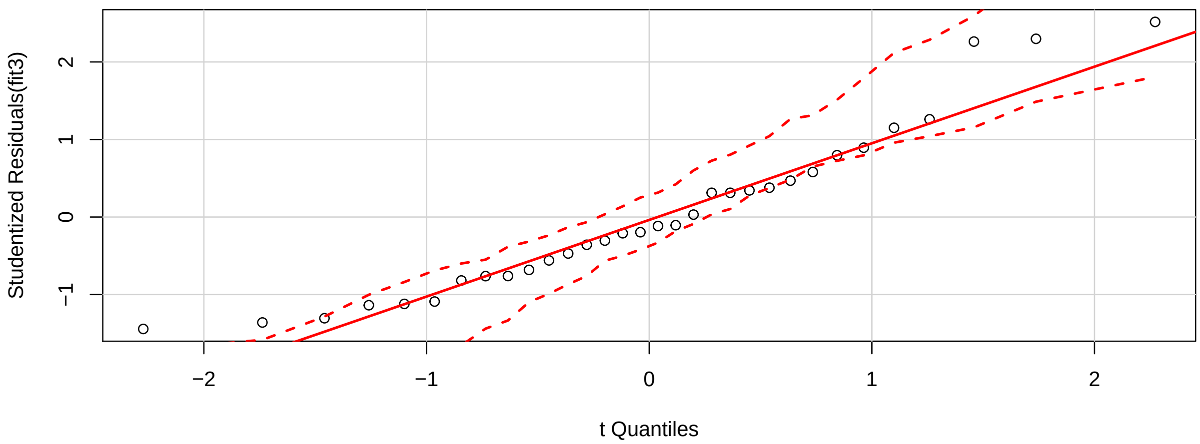
Confidence Intervals:

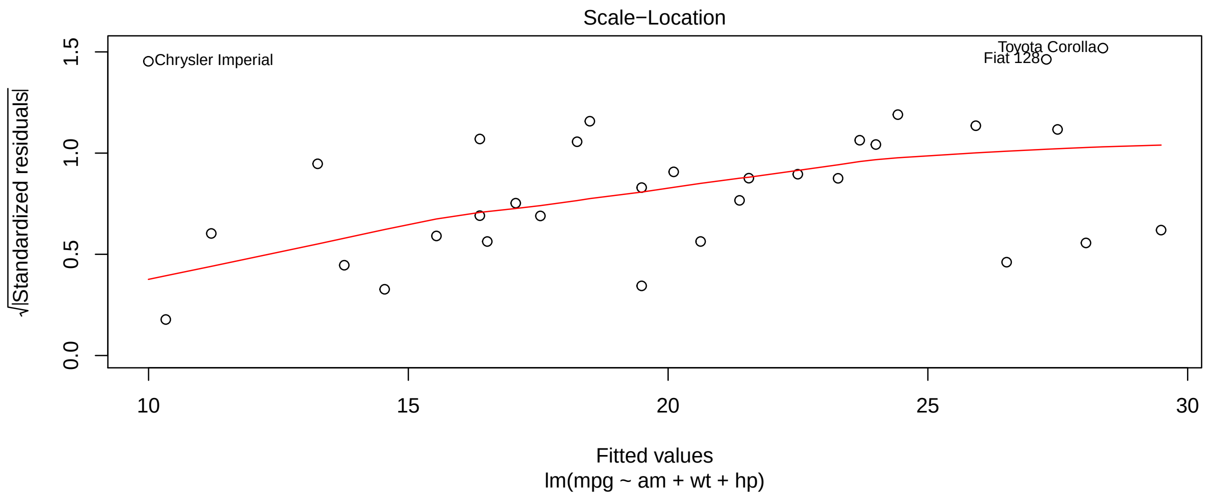
```
##                  2.5 %      97.5 %
## (Intercept) 28.58963286 39.41611738
## ammanual    -0.73575874  4.90317900
```

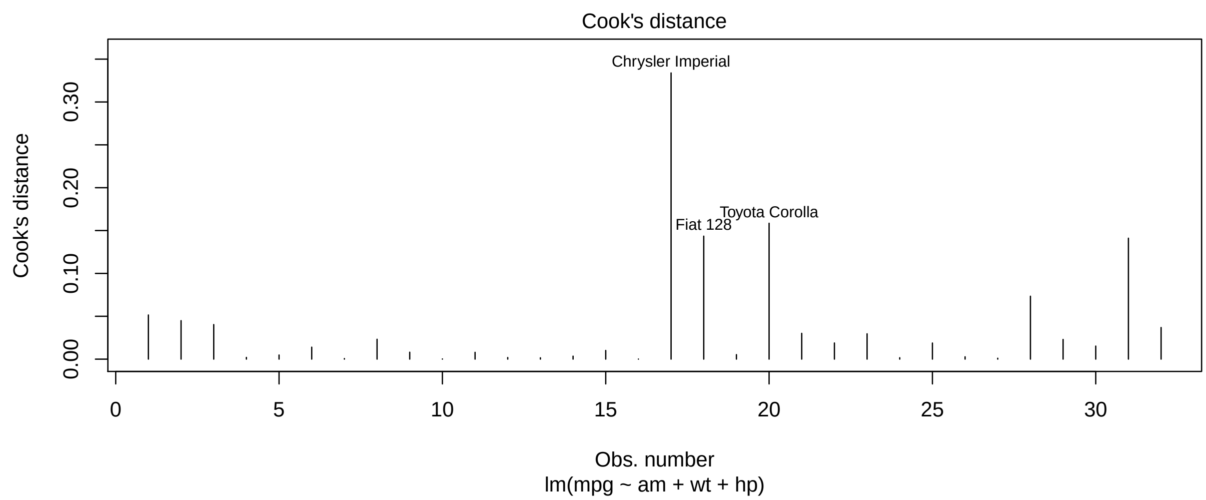
```
## wt          -4.73232353 -1.02482730
## hp          -0.05715454 -0.01780291
```

Residual Plots:









####Sources: In addition to course material I utilized R in Action by Robert Kabacoff and R Graphics Cookbook by Winston Chang.