Comparison of Transmission Type vs. Fuel Economy

Lenny Fenster

Friday, November 21, 2014

Executive Summary

The context for this project is that of an employee working for *Motor Trend* magazine; leveraging the data from 1974 that is comprised of fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). We are asked to answer the questions:

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

We investigated this relationship using multiple different models; first analyzing the correlation between only *mpg* and *am (transmission)* and then exploring whether additional aspects might play a confounding role in any identified correlation. While it initially appeared that the transmission type has a direct relationship on *mpg* when observed in isolation, we identified that the *wt (weight)* and *cyl (number of cylinders)* were significant confounders to this effect when comparing all variables.

Cleaning the data

Observing the data, we see that the primary predictor we want to analyze, *am (transmission)* is best represented as a factor where *0* can be interpreted as *automatic* and *1* can be interpreted as *manual*. We converted that variables as well as a few other variables to factors where they are best represented by as a category as opposed to a continuous number.

```
data(mtcars)
mtcars$am <- factor(mtcars$am,labels=c("automatic","manual"))
mtcars$cyl <- factor(mtcars$cyl)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Exploratory Analysis

First, we analyzed the relationship between just *mpg* and *am*. We started with a t-test suggesting that the mean *mpg* is equivalent for both automatic and manual transmissions.

```
mpgamTtest <- t.test(mpg~am, mtcars)
```

The test results in a p-value of 0.0014 and confidence intervals of -11.2802 and -3.2097. The means for automatic transmissions show as 17.1474 while the mean for manual transmissions is 24.3923. Thus, we conclude from this t-test that there is a significant difference between the average *mpg* for cars with automatic transmissions and those with manual transmissions. We then conducted a linear regression analysis on just these variables to corroborate our observations.

```
mpgamFit <- lm(mpg~am, mtcars)
```

Through regression analysis, we confirm that transmission has a significant effect and that the default factor, automatic, is observed to have a mean of 17.1474 and manual transmission has a mean of 24.3923. This is easily observed in the box plot shown in *Appendix: Figure 1*. Analyzing just *mpg* and *am* in isolation of any other variable leads us to believe that cars with a manual transmission have better fuel consumption than those with an automatic transmission.

However, this analysis alone is not be enough to conclude the impact of transmission on fuel economy. Upon examining the residuals for this correlation, shown in *Appendix: Figure 2*, we did **not** observe the random plotting of points that we expected around the regression line. That is, there appeared to be some pattern to the residuals. This suggested that this correlation may be due to some other factor(s). ### Considering other models Therefore, we considered the possibility that other variables may be having a confounding effect on this correlation. We viewed the correlation among all variables in *Appendix: Figure 3* to gauge whether a confounding effect may indeed exist. By observing the numerical values and stretched correlation ellipses in this view, we detected strong correlations between many other variables. Thus, we considered the possiblity that other variables were having a confounding effect. We ascertained this numerically by first modeling all the variables against *mpg*. Then, we leveraged the stepwise modeling selection to determine the ideal predictors for the best model.

```
mpgallFit<-lm(mpg~., mtcars)
mpgBestFit<-step(mpgallFit)
```

```
summary(mpgBestFit)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.939 -1.256 -0.401  1.125  5.051
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  33.7083     2.6049   12.94  7.7e-13 ***
## cyl6         -3.0313     1.4073   -2.15   0.0407 *
## cyl8         -2.1637     2.2843   -0.95   0.3523
## hp           -0.0321     0.0137   -2.35   0.0269 *
## wt           -2.4968     0.8856   -2.82   0.0091 **
## ammanual      1.8092     1.3963    1.30   0.2065
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.866,  Adjusted R-squared:  0.84
## F-statistic: 33.6 on 5 and 26 DF,  p-value: 1.51e-10
```

This analysis suggested that *cyl*, *hp*, and *wt* represent significant confounders. Thus, we used the *anova* function to compare this model against that with just *mpg* and *am*:

```
anova(mpgamFit, mpgBestFit)
```

```
## Analysis of Variance Table
##
```

```
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df RSS Df Sum of Sq    F  Pr(>F)
## 1     30 721
## 2     26 151  4       570 24.5 1.7e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This comparison led us to reject the hypothesis that *am* alone is the cause for the correlation with *mpg* as we saw that *cyl*, *hp*, and *wt* had significant impact.

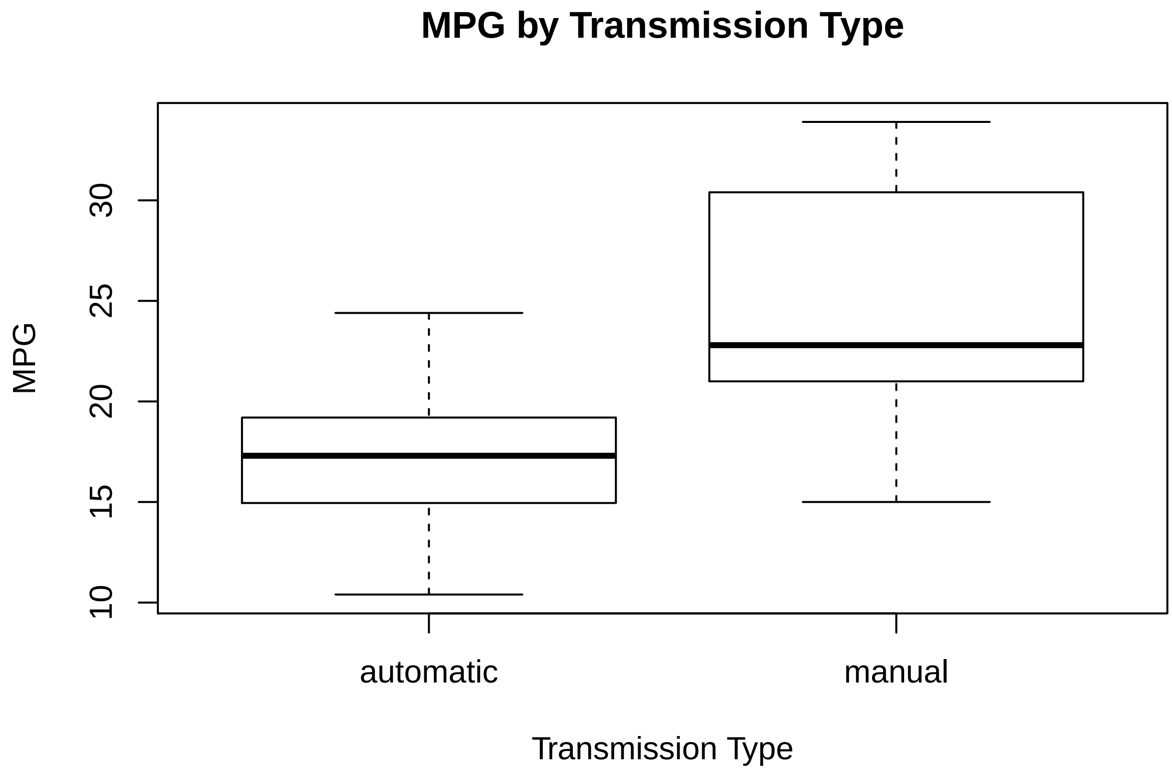
Again, we examined the residuals to determine whether a random set of plots around the regression line for this new model. The residual plot depicted in *Appendix: Figure 3b* did indeed show that this random pattern. Additionally, we observed the Normal Q-Q plot, Scale-Location, and Residuals versus Leverage plots illustrating the pattern we expect to see to support this regression analysis.

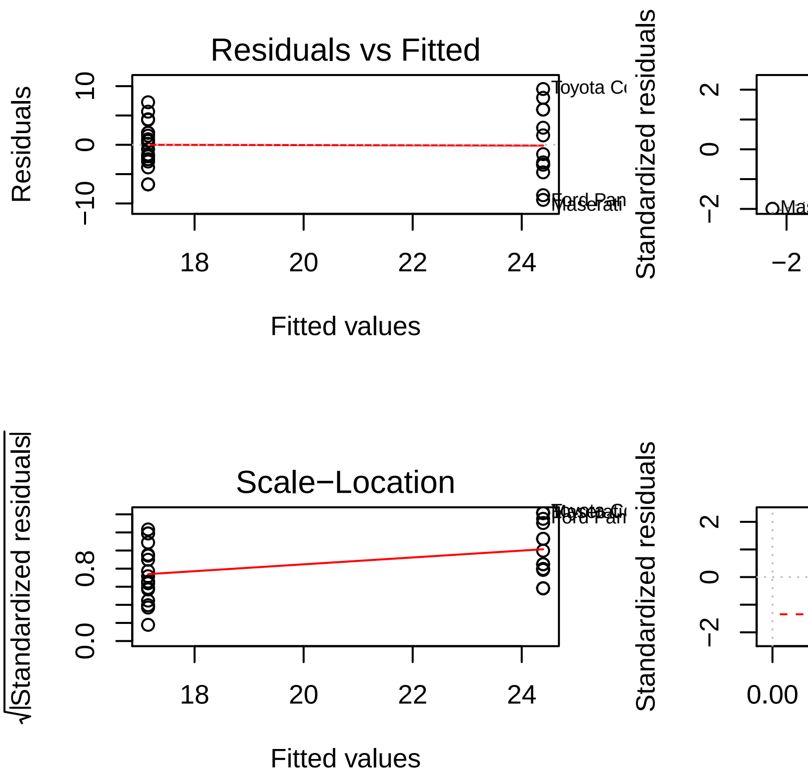
Conclusion

We conclude that while transmission plays a role in the fuel economy for a vehicle, it is not the only variable that has this impact. The weight and the number of cylinders also impact the fuel economy for a vehicle and must be considered together with the transmission when trying to predict the *mpg* that particular car will get.

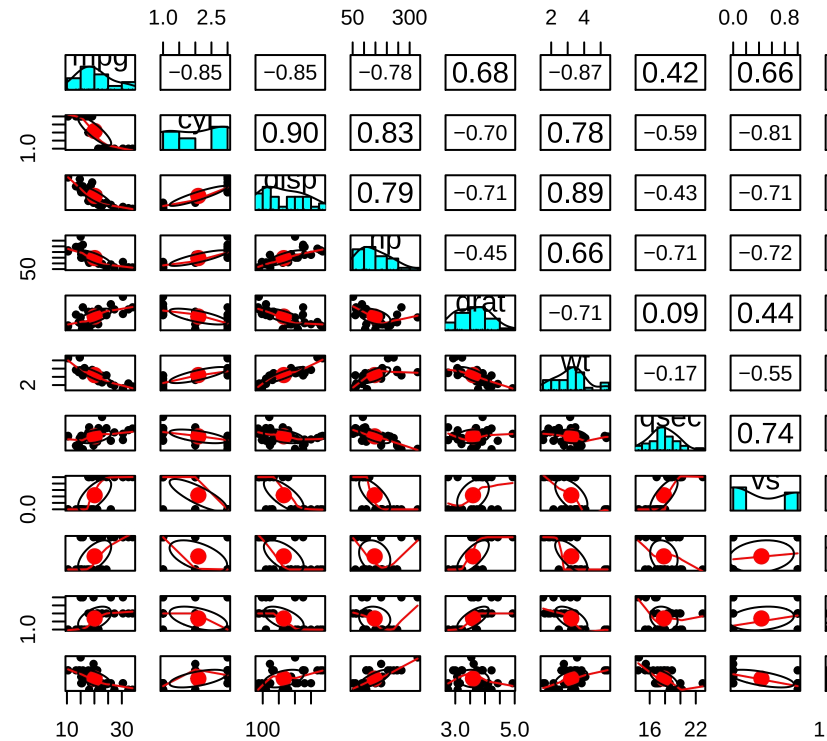
Appendix

Figure 1: Boxplot of MPG by Transmission Type

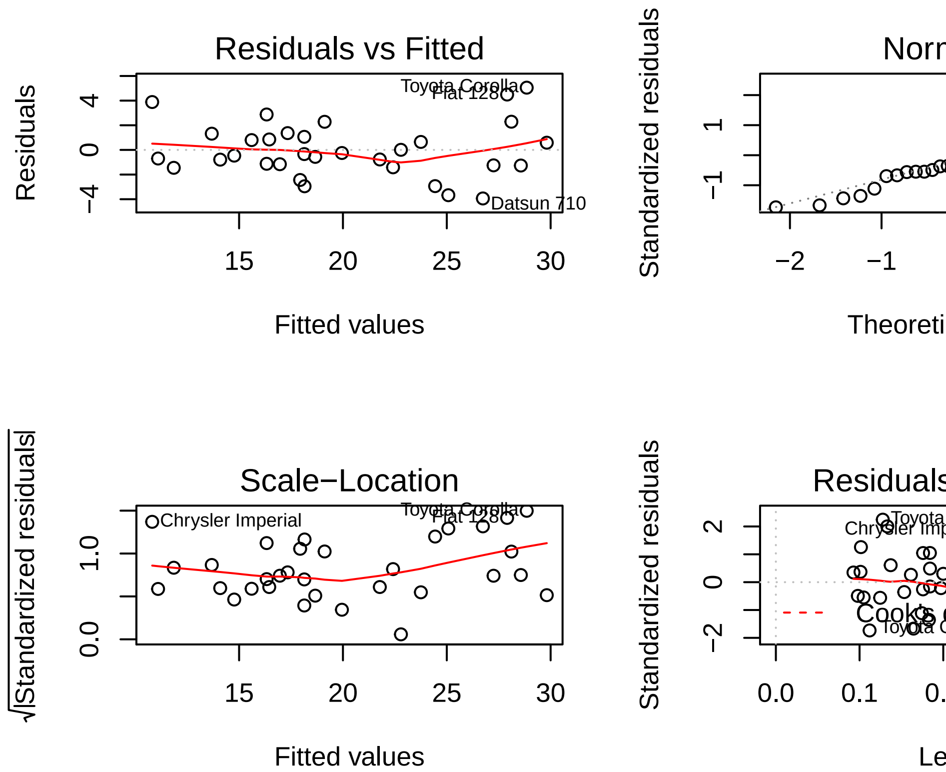




###Figure 2: Residual plot for just Transmission Type



###Figure 3: Scatterplot matrix for all variables



###Figure 4: Residual plot for Best Model