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 and leveraging the data from 1974 that includes fuel consumption and 10 other aspects of automobile design and performance for 32 automobiles (1973-74 models). We were 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 variables might have a confounding effect. While it appeared that the transmission type had a direct relationship on *mpg* when observed in isolation, we further observed that *wt (weight)* and *cyl (number of cylinders)* were significant contributors to the fuel economy.

Cleaning the data

Observing the data, we saw that the primary predictor we wanted to analyze, *am (transmission)* was best represented as a factor where *0* was interpreted as *automatic* and *1* as *manual*. We converted that and other variables to factors where they were best represented 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 resulted in a p-value of 0.0014 and confidence intervals of -11.2802 and -3.2097. The means for automatic transmission was 17.1474 and the mean for manual transmission was 24.3923. This t-test indicated 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 confirmed that transmission had a significant effect and that the averages for both transmission types were exactly the same as calculated using the t-test. This was also observed in the box plot shown in *Appendix: Figure 1*. Analyzing just *mpg* and *am* in isolation of any other variable led us to believe that cars with a transmission type alone can be used to predict the fuel consumption for a vehicle.

However, this analysis alone was not be enough to conclude the impact of transmission on fuel economy. Upon examining the residuals for this model, 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 the correlation may be due to some other factor(s).

Considering other models

Therefore, we considered the possibility that other variables may have 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 observed strong correlations between many variables. Thus, we considered the possiblity that other variables could have a confounding effect. We ascertained this numerically by first modeling all the variables against *mpg* and then.using the stepwise modeling selection function to determine the ideal predictors for the best model.

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

This analysis showed other variables *(cyl6, cyl8, hp, wt, ammanual)* representing significant confounders with coefficients of *(-3.0313, -2.1637, -0.0321, -2.4968, 1.8092)*. 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
```

With this model comparison, we rejected the hypothesis that *am* alone impacted *mpg* as we saw that *cyl*, *hp*, and *wt* also had significant impact.

We examined the residuals again to determine whether a random set of plots existed around the regression line for this new model. This time, the residual plot depicted in *Appendix: Figure 3b* showed the random pattern we expected. Additionally, we observed the Normal Q-Q plot, Scale-Location, and Residuals versus Leverage plots illustrated the pattern we expected 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 drives 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 a particular car will get.

Appendix

Figure 1: Boxplot of MPG by Transmission Type

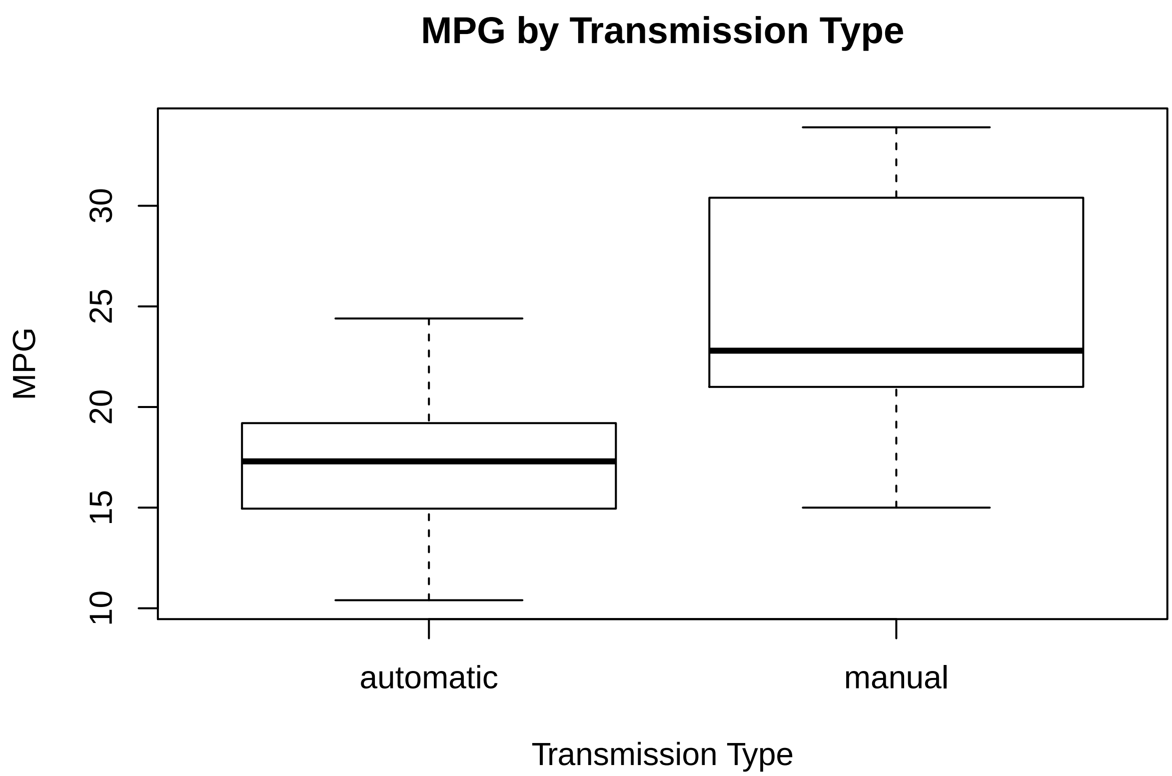

Figure 2: Residual plot for just Transmission Type

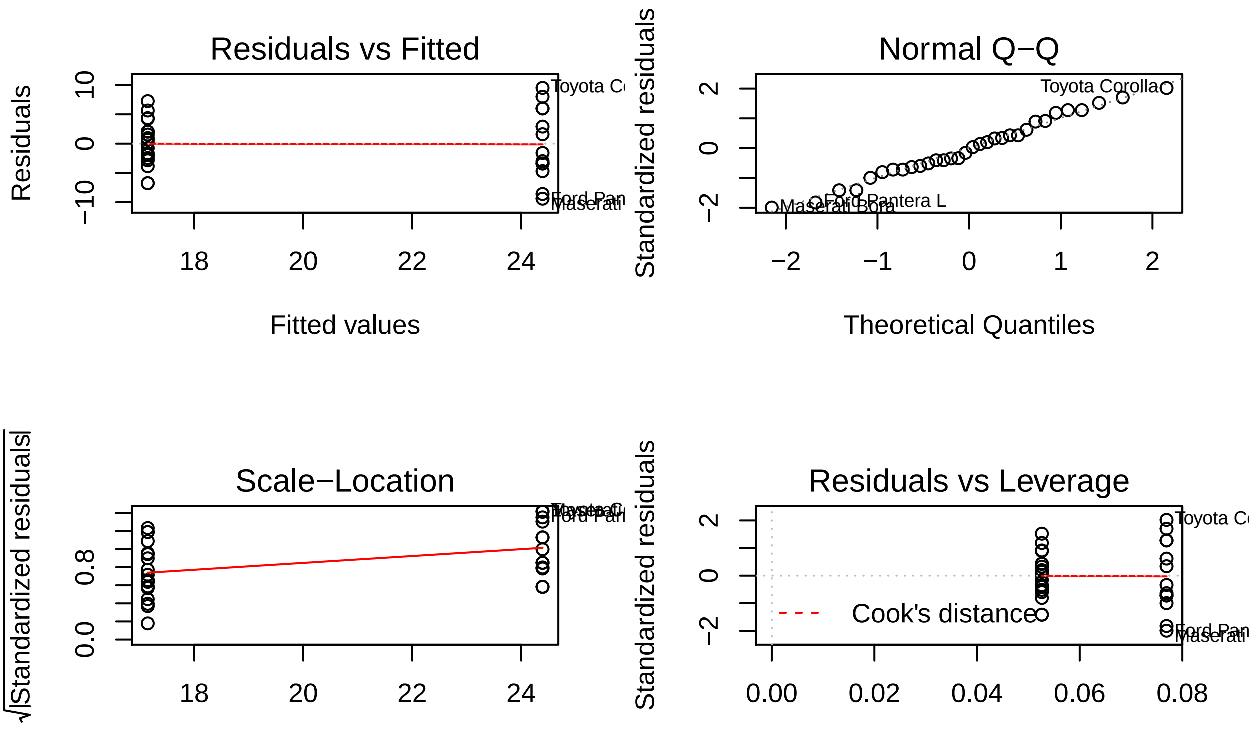

Figure 3: Scatterplot matrix for all variables

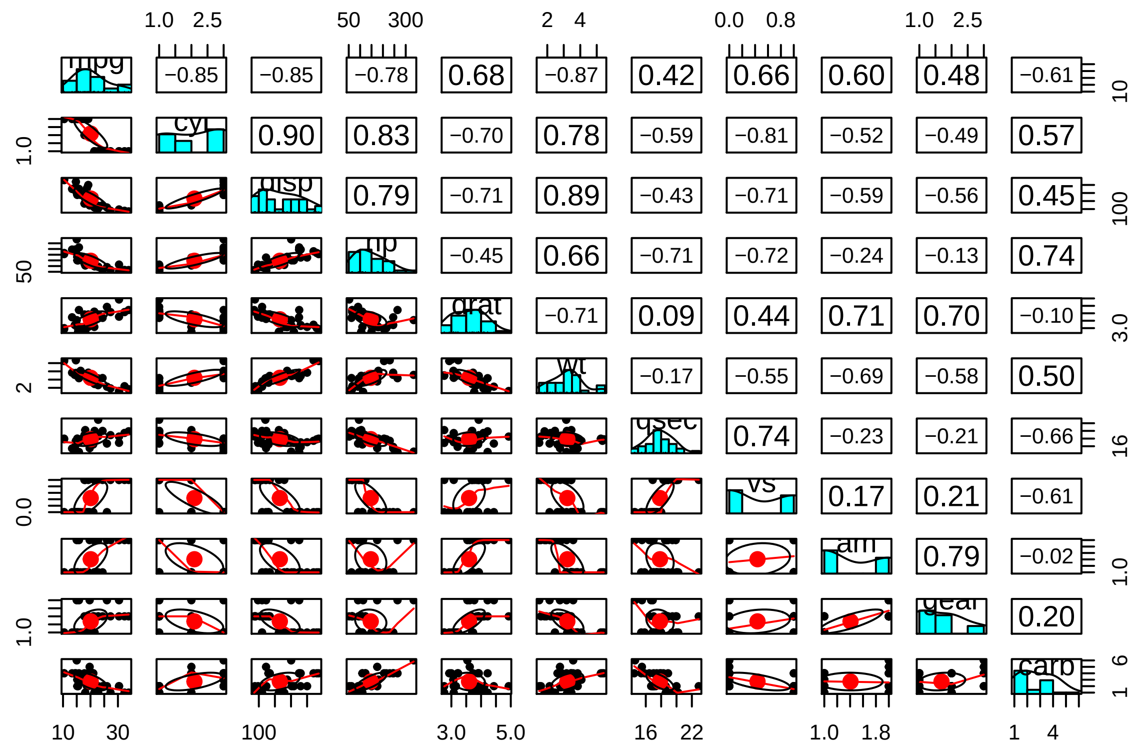

Figure 4: Residual plot for Best Model