Effects of Transmission Type on Fuel Efficiency

Regression Models Course Project

Brian Gulbis

October 03, 2016

Executive Summary

This analysis used data from the Motor Trend Car Road Tests data set to determine which transmission type is better for a vehicle's fuel efficiency and to quantify the difference. An exploratory analysis identified several differences in key variables thought to impact fuel efficiency between cars with Manual and Automatic transmission types. Three regression models were created: a univariate model (Model A), a model developed through nested model testing (Model B), and a model using all variables in the data set as predictors (Model C). Through comparison of these models, it was determined that the model developed through nested model testing (Model B) resulted in the best fitting model with the lowest variance. The results of this model showed that transmission type did not significantly predict a car's miles per gallon, but rather, the vehichle weight, number of cylinders, and gross horse power are the factors which should be used to predict the vehicle's fuel efficiency.

Introduction

This analysis will examine the effects of different transmission types (automatic vs. manual) on an automobile's fuel efficiency. Additionally, using regression models, the effect of transmission type on fuel efficiency will attempt to be quantified. The data for this analysis will come from the Motor Trend Car Road Tests data set, which contains information on 32 automobile models from 1973-74, extracted from the 1974 *Motor Trend* magazine.

Exploratory Analysis

First, an exploratory analysis was performed. There were 10 variables which may impact the fuel efficiency of the vehicle in this data set. The primary variable of interest in this analysis is the transmission type of each car, either Automatic or Manual. There are 19 cars with an Automatic transission and 13 with a Manual transmission in the data set. An exploratory plot of vehicle fuel efficiency, in miles per gallon (mpg), for cars with each type of transmission suggests that cars with a Manual transmission may have greater fuel efficiency than cars with an Automatic transmission (see Figure 1 in the appendix).

However, this difference can be misleading, since there are several other variables which may be impacting fuel efficiency and potentially skewing the data towards favoring a Manual transmission. Some other factors which might impact fuel efficiency include the number of cylinders in the engine, gross horse power, and weight of the vehicle. Looking at plots of these factors suggests that they are not equally distributed between cars with an Automatic transmission and those with a Manual Transmission in this data set (see Figure 2 in the appendix). Further inferential analysis using t-tests show that the variables weight and number of cylinders are significantly different between the two transmission types (p-values 10^{-5} and 0.002, respectively), while horse power is not (p-value 0.22).

Regression Models

To further evaluate the effect of transmission type on fuel efficiency, three different regression models will be fitted to predict the outcome of miles per gallon:

1. Model A will fit a single univariate model containing just the transmission type as the predictor
2. Model B will include transmission type along with those factors (weight, number of cylinders, and horse power) which are suspected of skewing the data as predictors (note: nested model testing using an ANOVA test indicated that these three factors significantly added to the model)
3. Model C will include all 10 variables in the data set as predictors

```
model.a <- lm(mpg ~ factor(am), data=mtcars)
model.b <- lm(mpg ~ factor(am) + wt + factor(cyl) + hp, data=mtcars)
model.c <- lm(mpg ~ ., data=mtcars.factored)
```

In Model A, having a Manual transmission resulted in an increase of 7.24 miles per gallon over an Automatic transmision, with a 95% confidence interval of 3.64 to 10.85. This appears to be a compelling answer to the question of which transmission type is better for fuel efficiency, however, as discussed above, there are several other factors which may impact the final results that are not being accounted for in this model. Looking at a plot of the residuals, there is a clear linear pattern to the residuals, which suggests that this may not be the best model fit (see Figure 3 - Model A Residuals vs. Fitted in the appendix). Furthermore, the R-squared of this model is quite low at 0.36 and the variance is quite large (4.902), providing additional support to the conclusion that this is not the best model to use.

In Model B, having a Manual transmission resulted in a smaller 1.81 increase in miles per gallon compared with an Automatic transmission, holding all other predictors constant. However, there is more uncertainty in this model, as the 95% confidence interval (-1.06 to 4.68) crosses 0, meaning that there is no significant difference in fuel efficiency between the transmission types. The other variables included in this model (weight, number of cylinders, and horse power), were all found to be significant predictors of the vehicle's fuel efficiency. A plot of the residuals for this model does not demonstrate a pattern (see Figure 3 - Model B Residuals vs. Fitted in the appendix), and the residuals follow a normal distribution (see Figure 3 - Model B Normal Q-Q in the appendix). The R-squared for Model B is 0.866, which implies this is a better model than Model A.

In Model C, having a Manual transmission also resulted in a non-signficant increase of 1.21 miles per gallon compared with an Automatic transmission, holding all other predictors constant. This model contains the largest amount of uncertainty, demonstrated by a wider 95% confidence interval (-5.64 to 8.06). A plot of the residuals for this model does not demonstrate a clear pattern (see Figure 3 - Model C Residuals vs. Fitted in the appendix), and the R-squared is 0.893, which implies this is a better model than Model A and essentially the same as Model B (although the adjusted R-squared is actually lower than Model B, 0.779 vs. 0.84, respectively). However, including all variables raises concerns about overfitting the model which could result in variance inflation, and in fact, Model C has more variance than Model B (2.41 vs. 2.833, respectively).

Conclusions

In this analysis of 3 regression models, only the univariate regression model (Model A) suggested that transmission type may have any effect on vehicle fuel efficiency. However, considering the poor fit of the univariate model, use of a multivariate model is recommended. When comparing the two multivariate models (Models B and C), both models found a small, non-significant effect of transmission type on fuel efficiency and had a similar R-squared. Since Model C contains all variables as predictors there is concern for overfitting and variance inflation with this model, as shown by the higher variance in Model C compared to Model B. There is more uncertainty in Model C, shown by the wider 95% confidenc interval. From this, it can be concluded that Model B is the best fitting model. The variables included in Model B were determined by using nested model testing.

In response to the two questions being evaluated, using the best fitting model (Model B), there is no significant difference in miles per gallon between vehicles with an Automatic or Manual transmission (95% confidence interval -1.06 to 4.68). According to this model, vehicle weight, number of cylinders, and gross horse power are significant predictors of a vehicle's fuel efficiency, while transmission type is not.

Appendix

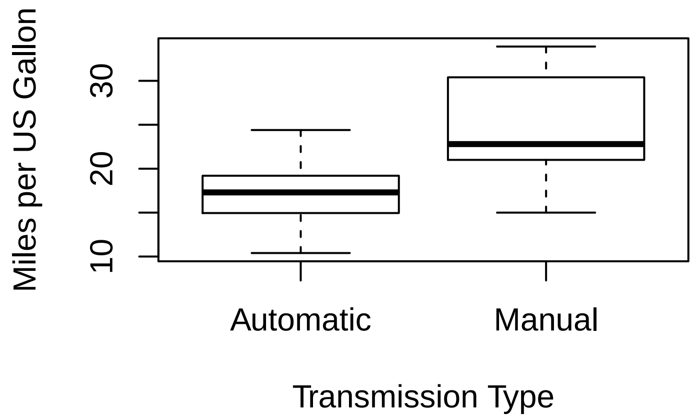

Figure 1: Comparison of fuel efficiency by transmission type

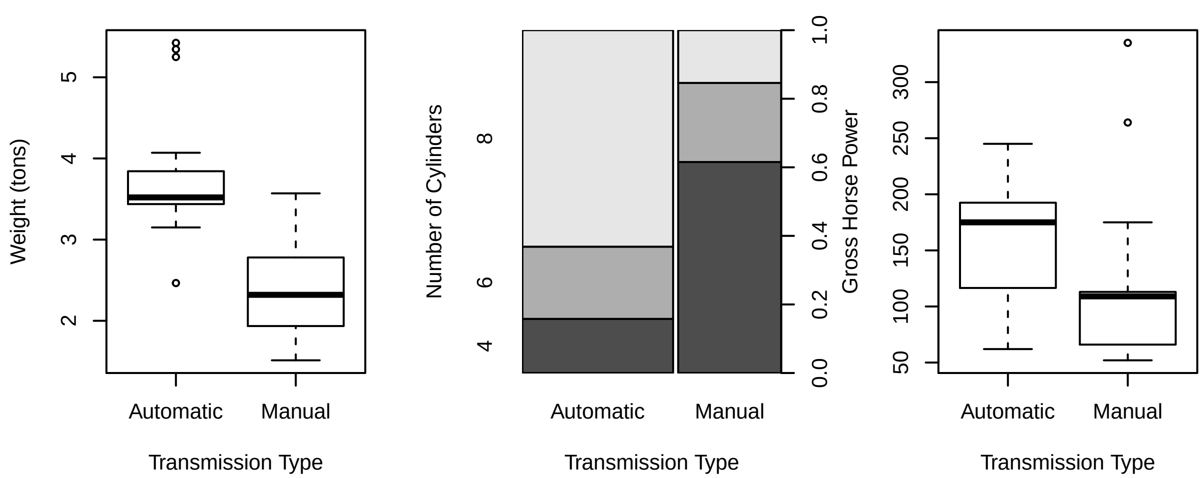

Figure 2: Exploration of other variables impacting fuel efficiency by transmission type

```
## Warning: not plotting observations with leverage one:
##   30, 31
```

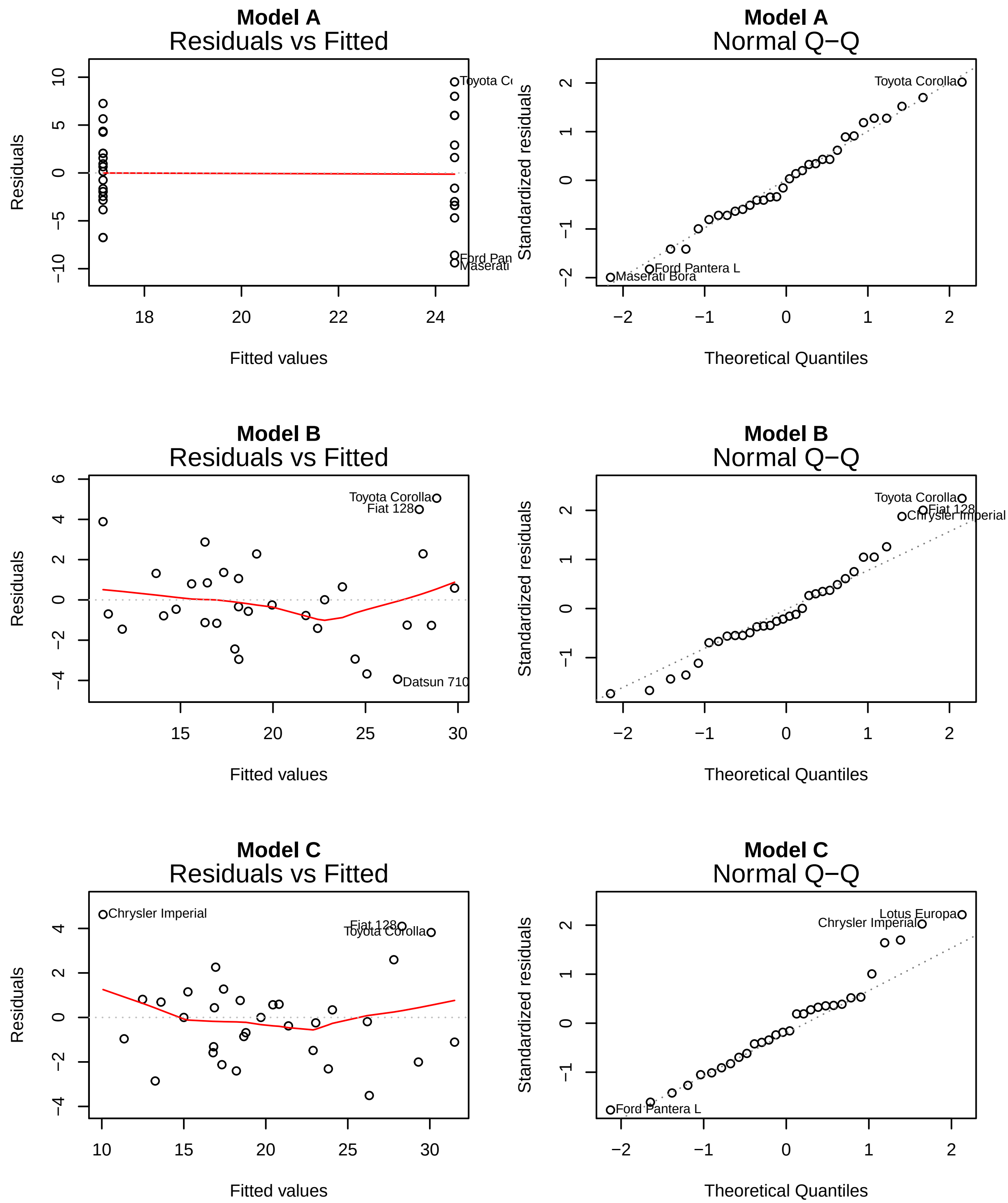

Figure 3: Diagnostic Plots for 3 Regression Models