

Fuel Efficiency and Transmission Type

Raghu

September 26, 2015

Executive Summary

Motor Trend is interested in exploring the relationship between a set of variable and miles per gallon (MPG), particularly interested in automatic transmission vs. manual transmission for better MPG. And this document explores “mtcars” data and performs necessary data analysis by quantifying the MPG difference between automatic and manual transmissions. This document determines the conclusion with a statistical evidence, by fitting multiple regression models and select appropriate model.

Exploratory Analysis

```
## Warning: package 'knitr' was built under R version 3.2.1
```

Looking at the variables in the question, amManual, mpg (figure 1 in Appendix), it seems obvious that manual transmissions have a higher average MPG than automatic transmissions. However, We build an exploratory linear model using every feature available in the data set (a “kitchen sink” regression):

Note; kitchen sink regression is a statistical regression which uses a long list of possible independent variables to attempt to explain variance in a dependent variable.

```
ksregression <- lm(mpg ~ ., mtcars)
summary(ksregression)$coef
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.30337416	18.71788443	0.6573058	0.51812440
cyl	-0.11144048	1.04502336	-0.1066392	0.91608738
disp	0.01333524	0.01785750	0.7467585	0.46348865
hp	-0.02148212	0.02176858	-0.9868407	0.33495531
drat	0.78711097	1.63537307	0.4813036	0.63527790
wt	-3.71530393	1.89441430	-1.9611887	0.06325215
qsec	0.82104075	0.73084480	1.1234133	0.27394127
vs	0.31776281	2.10450861	0.1509915	0.88142347
amManual	2.52022689	2.05665055	1.2254035	0.23398971
gear	0.65541302	1.49325996	0.4389142	0.66520643
carb	-0.19941925	0.82875250	-0.2406258	0.81217871

Above table lists the coefficients of the kitchen sink model. With so many variables and so few data points, the p-values for the coefficients are not statistically significant at $p < 0.05$; however, weight does have the most extreme absolute p-value by far of all the variables, and a plot of weight versus MPG indicates that we should account for weight. This makes sense physically; it takes more energy to move heavier objects. In fact, our initial understanding to the effect of transmission type on fuel efficiency might even be completely explained by the weight.

The kitchen sink regression has a positive coefficient for having a manual transmission (amManual) of 2.5202, indicating that having a manual transmission increases fuel efficiency by 2.5202 MPG when we hold the other variables constant, but the corresponding p-value is only 0.2340, which can be rejected at a significance level of 0.05. So it we cannot make a statement about a significant effect on fuel efficiency by the transmission type under this model.

Two Models of Weight and Transmission Type

For model selection, we will search for a model with a very low p-value with respect to the coefficients for transmission type, so we can be more confident about our conclusions. Based on our exploratoration, we will consider a linear model that only takes into account weight and transmission, and one with the interaction term added:

Model 1

```
fit <- lm(mpg ~ am + wt, mtcars)
summary(fit)$coef
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	37.32155131	3.0546385	12.21799285	5.843477e-13
amManual	-0.02361522	1.5456453	-0.01527855	9.879146e-01
wt	-5.35281145	0.7882438	-6.79080719	1.867415e-07

The above table lists the coefficients of the first model. **After accounting for weight discrepancies, it appears that having a manual transmission might slightly reduce fuel efficiency if we assume the same rate of change in fuel efficiency with respect to weight for the two kinds of transmissions. However, there is a very high p-value, so we would be wrong to make a such a claim on this data alone.** If it were significant, we would have looked at the amManual coefficient and say that, comparing cars of the same weight, the a manual transmission would have 0.02 MPG lesser fuel efficiency.

Model 2

```
interaction_fit <- lm(mpg ~ am * wt, mtcars)
summary(interaction_fit)$coef
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	31.416055	3.0201093	10.402291	4.001043e-11
amManual	14.878423	4.2640422	3.489277	1.621034e-03
wt	-3.785908	0.7856478	-4.818836	4.551182e-05
amManual:wt	-5.298360	1.4446993	-3.667449	1.017148e-03

The above table lists the coefficients of the second model. **With a p-value of 0.001, we can say that, for a given weight, when we go from an automatic transmission to a manual transmission, fuel efficiency degrades on average 5.2984 MPG faster.** The corresponding plot in is in figure 3 in Appendix. The p-value for the difference in the F statistic from ANOVA between the two models is 0.0010171, indicating that we need to include the interaction term.

Diagnostics and Discussion

```
press <- resid(interaction_fit) / (1 - hatvalues(interaction_fit))
top <- head(press[order(-abs(press))], n = 1)
top
```

Fiat 128
6.668901

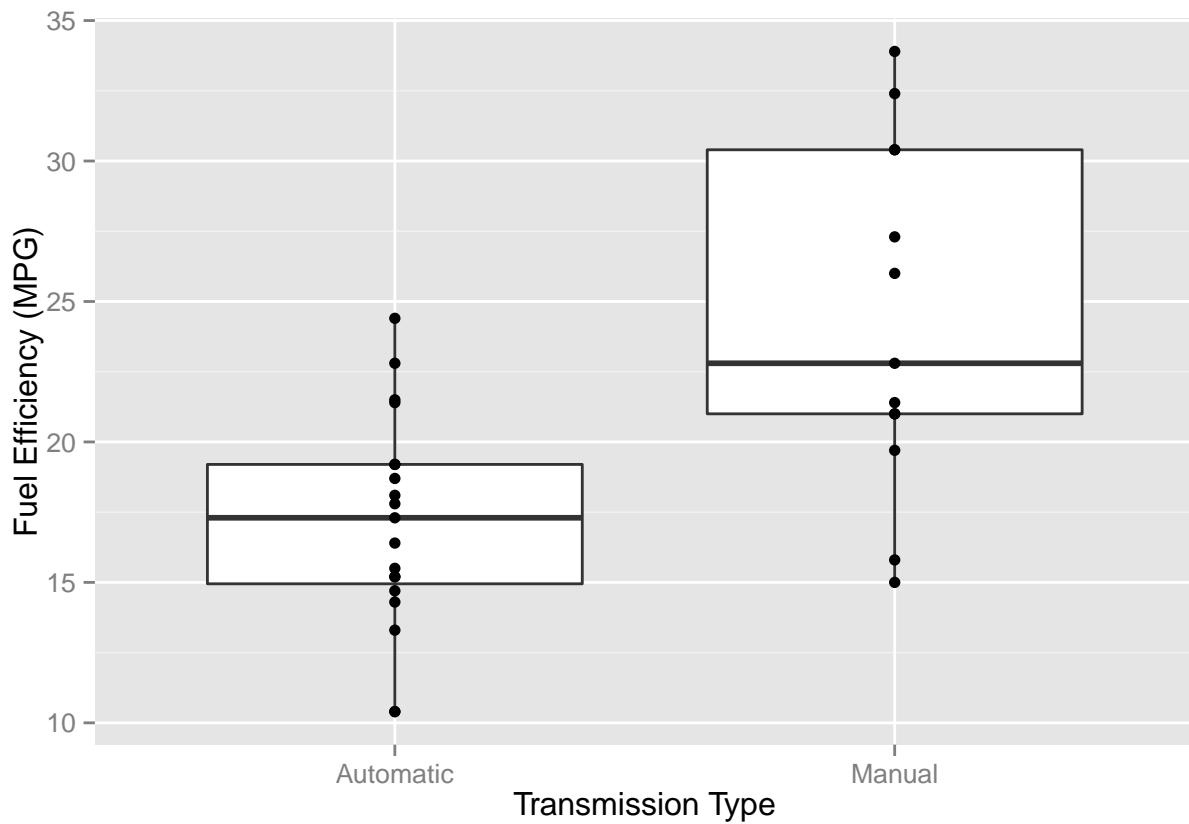
Figure 3 in Appendix is a plot of the residuals versus the fitted values under the interaction model, as well as a Normalized Q-Q plot. We see that we have divergent behavior in the tails, and there is almost certainly other variables that could be involved. Finally, if we look at the PRESS residuals to find the car that deviated the most from the interactions model when held out of the model was Fiat 128; it would be worth investigating if there are any particular features of this car that make it stand out.

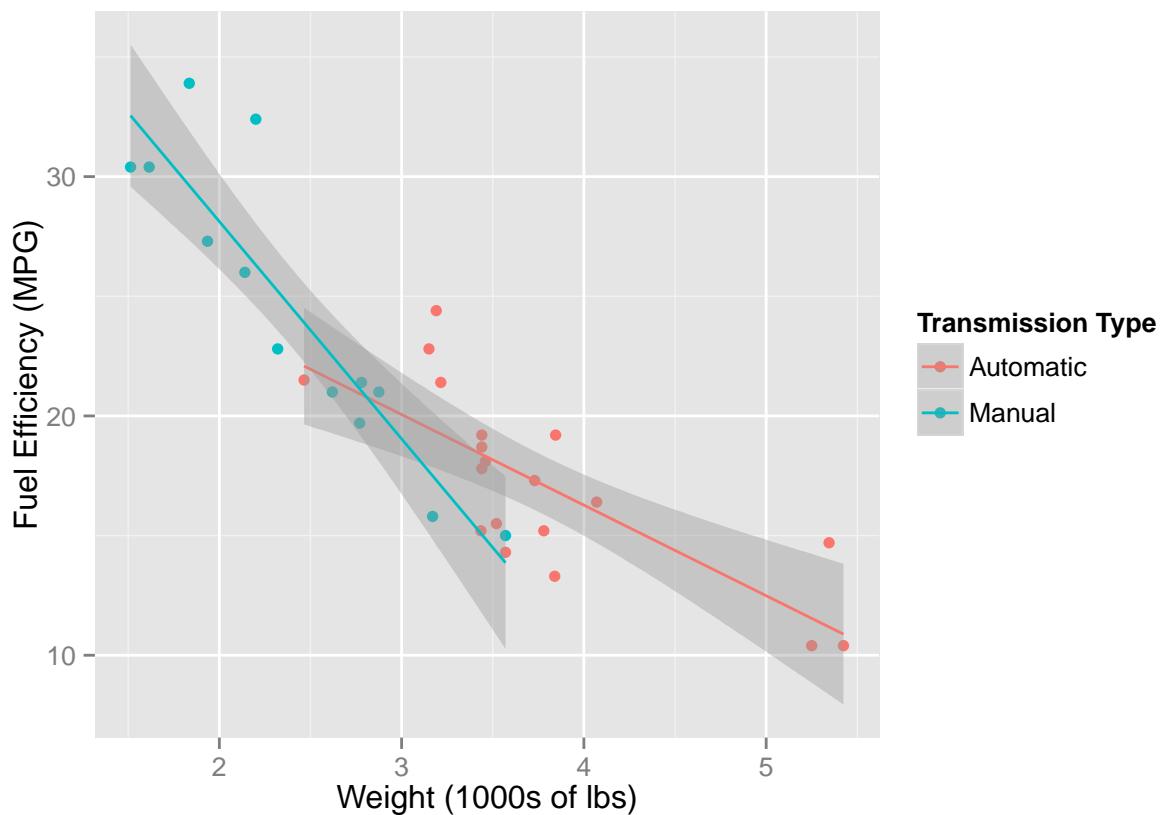
Since weight has such an important impact on MPG, and there was relatively little overlap in weight distributions between the two groups (see figure 2 in Appendix), the conclusions we make are highly based on the model we use.

Appendix

```
Warning: package 'ggplot2' was built under R version 3.2.1
```

```
Warning: package 'plyr' was built under R version 3.2.1
```





Warning: package 'gridExtra' was built under R version 3.2.1

