Transmission Type and MPG

Lucas McLaughlin

Saturday, July 25, 2015

Summary

This document contains an analysis of the mtcars dataset included in R. In it, I seek to answer the following questions:

1. Is an automatic or manual transmission better for mpg?
2. Quantify the MPG difference between automatic and manual transmissions.

In short, it is easy to see that cars that manual cars get better MPG. Understanding why though is a more complicated matter. I believe the weight of the car is really the most important factor and through using linear models, predict that the true difference in MPG between an automatic and manual car, if all else in the car were the same, is somewhere between 3 and 7 mpg.

MPG vs. Transmission Type

An initial boxplot showing mpg for both automatic and manual cars (fig 1) shows that manual cars have much better MPG on average. Specifically, our observed difference is 7.2 MPG. This is helpful initial information, but this observed difference can only tell us so much. The transmission system of a car effects things like car weight, rear axle ratio, and number of forward gears. These things have effects on engine size, horsepower, and so on. To figure out which variables are really causing this noticable change in MPG, let's examine what variables are correlated with MPG and transmission type and then fit a linear model using that information.

Characteristics of Automatic Cars

Something we can to do explore the relationship of these variables is simply to look at correlation values (fig 2). Values closer to -1 and 1 indicate a stronger relationship. We can then plot some of the things that are different between automatic and manual cars (fig 3). Automatic cars are heavier, have larger engines, more horespower, and weaker rear axle ratios. In addition, the strongest variable associated with mpg is car weight (fig 4), so it's possible that automatic cars simply weigh more, thus requiring larger engines and getting lower mpg. It's also interesting to note that weight has a greater effect on manual cars than automatic cars. Linear models are ill-equiped to describe causation, but let's see if we can find a good linear model that takes into account these variable relationships.

Fitting the Best Linear Model

Now, we will consider three linear models that describe the relationship between transmission and fuel efficiency. The models are as follows:

1. MPG predicted by transmission type only.
2. MPG predicted by all variables available.
3. MPG predicted by transmission, weight, and their continuous interaction.

Interpreting Coefficients

Here's what we can conclude from interpreting the coefficients of our models.

Model 1 This model (fig 5) shows that the average MPG for automatic cars is 17.1 and that the average MPG for manual cars is 17.1 + 7.2, or 24.3. This is essentially the imperical average that we previously observed, but this model now allows us to create a confidence interval. Under this model, we can say with 95% confidence that the true difference in MPG between an automatic and manual car lies within the interval (3.7, 10.8). This is done by taking our 7.2 and adding/subtracting two times the standard error.

Model 2 This model (fig 6) takes into account all varaibles available. Under this model, our estimate for the difference in MPG between a manual and automatic car is much lower at 2.6 and our confidence interval is (-1.4, 6.6) I believe this lower value is caused by two things. Firstly, the model takes into account counfounding variables that really may be having a greater impact on on MPG, thus taking emphasis off the transmission variable. Second, the model introduces greater variance in our regression variables due to the larger number of predictor variables included.

Model 3 This final model (fig 7) seeks to describe how weight affects mpg for both automatic and manual cars. This model shows us that automatic cars have a -3.8 decrease in MPG for every 1000lb increase in car weight, whereas manual cars have a -3.8 + (-5.3), or -9.1 decrease in MPG for every 1000lb increase in car weight. In other words, manual cars handle car weight increase poorly, but since manual cars tend to be lighter they still come out ahead in MPG.

Diagnostics

Lastly, let's briefly look at some residual plots and variance diagnostics of our models to help diagnose their efficacy. We'll specifically look at models 1 and 2 since those try and quantify the MPG difference by transmission type. We don't see much variation in these residual plots (fig 8), which is a good thing. In model 1, we can see that there is less variation around the lower mpg, which would correspond to automatic cars and greater variance around the higher mpg, which would correspond to manual cars. This is interesting, but not necessarily a sign of bad model fit. In model 2, there isn't much of a pattern evident.

Let's also look at our R^2 values for models 1 and 2. This is a way of describing how much of the total variance in MPG was explained by our model. For model 1, the R^2 value was 0.36 and for model 2 it was 0.88. Thus, it's clear that our second model explains much more of the variation than model 1.

Conclusion

In conclusion, it is clear that manual cars get better gas mileage, and I believe that car weight is the biggest reason why. In addition, if I were to quantify this change in MPG, I would do so by taking the overlapping section of my confidence intervals and giving a final estimate of (3.7, 6.6).

Appendix

Figure 1

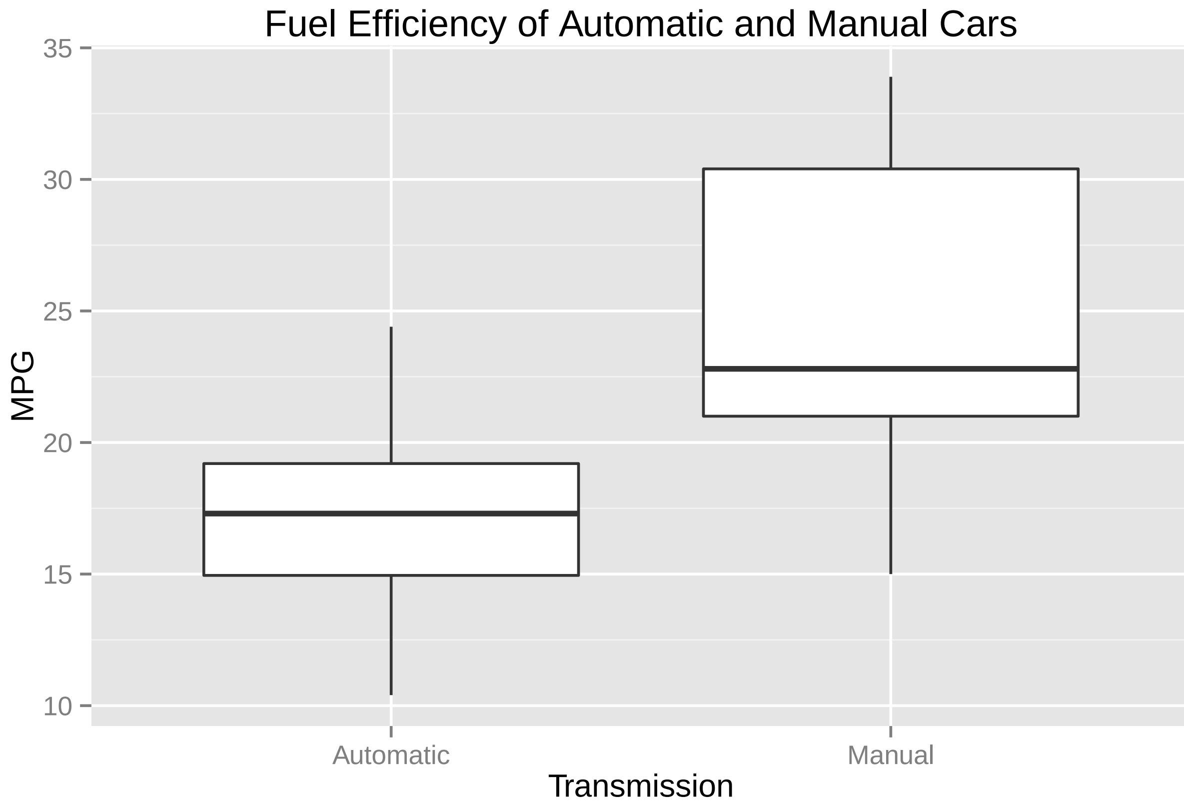

Figure 2 (Variables correlated with mpg and transmission)

	Transmission	MPG
mpg	0.5998324	1.0000000
cyl	-0.5226070	-0.8521620
disp	-0.5912270	-0.8475514
hp	-0.2432043	-0.7761684
drat	0.7127111	0.6811719
wt	-0.6924953	-0.8676594
qsec	-0.2298609	0.4186840
vs	0.1683451	0.6640389
am	1.0000000	0.5998324
gear	0.7940588	0.4802848
carb	0.0575344	-0.5509251

Figure 3

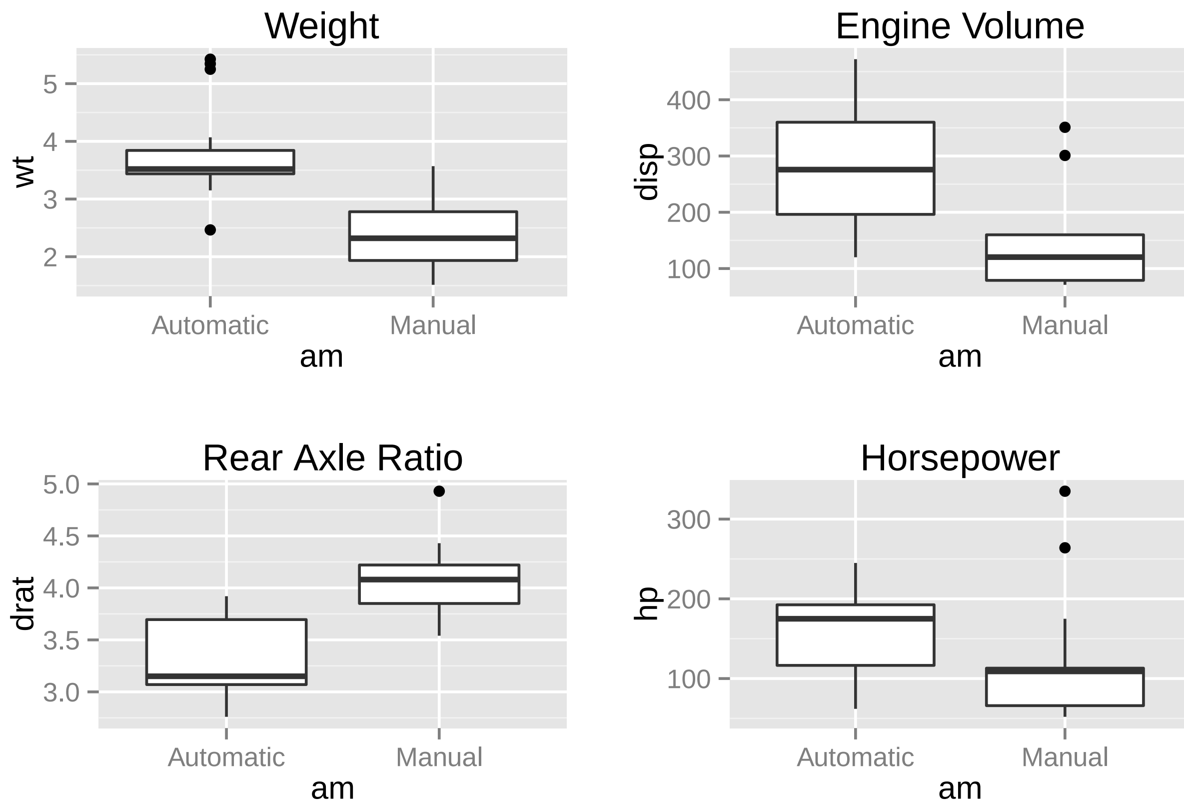
Weight
wt
5
4
3
2
Automatic
Manual
am
Engine Volume
disp
400
300
200
100
Automatic
Manual
am
Rear Axle Ratio
drat
5.0
4.5
4.0
3.5
3.0
Automatic
Manual
am
Horsepower
hp
300
200
100
Automatic
Manual
am

Figure 4

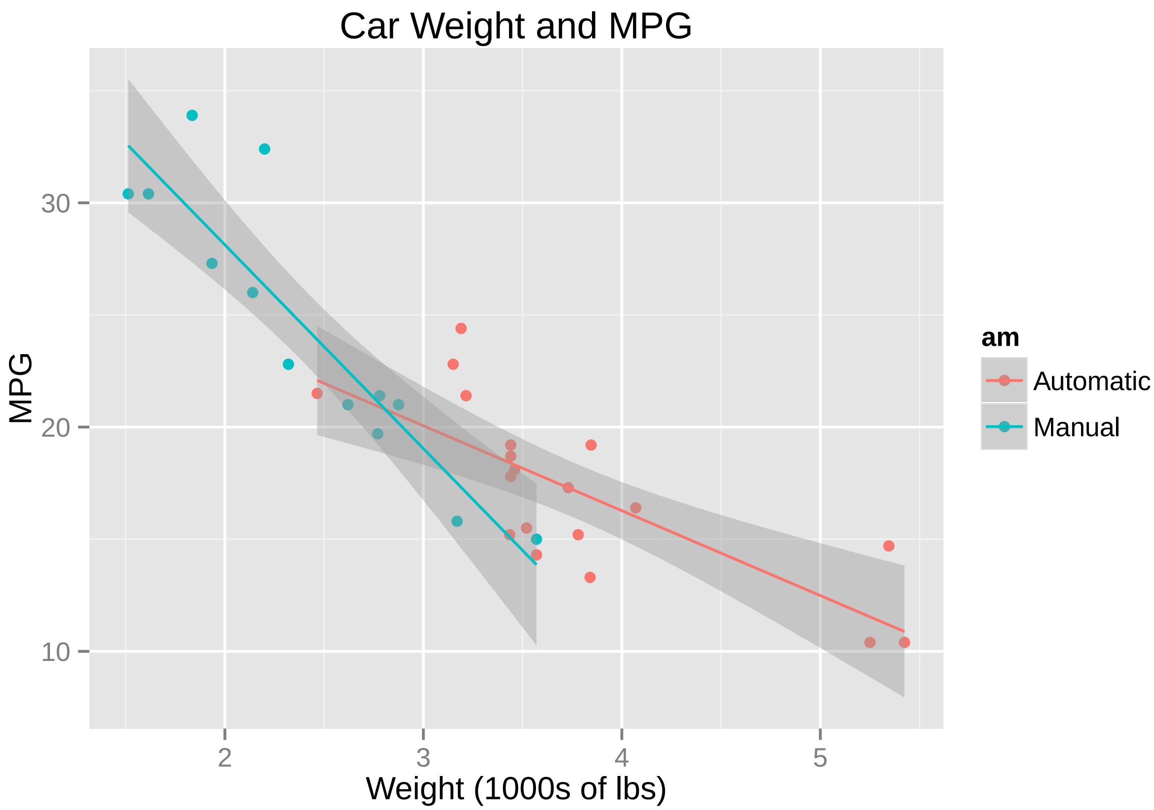

Figure 5 (Model 1)

```
fit1 <- lm(mpg ~ am, data = mtcars)
fit2 <- lm(mpg ~ ., data = mtcars)
fit3 <- lm(mpg ~ am + wt + am * wt, data = mtcars)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.147368	1.124602	15.247492	0.000000
amManual	7.244939	1.764422	4.106127	0.000285

Figure 6 (Model 2)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.8198433	16.3060232	1.0928381	0.2874542
cyl6	-1.6603067	2.2622966	-0.7339032	0.4715245
cyl8	1.6374398	4.3157345	0.3794116	0.7083808
disp	0.0139124	0.0174018	0.7994830	0.4334036
hp	-0.0461284	0.0271202	-1.7008869	0.1044619

	Estimate	Std. Error	t value	Pr(>|t|)
drat	0.0263503	1.6764895	0.0157175	0.9876155
wt	-3.8062476	1.8466431	-2.0611712	0.0525285
qsec	0.6469571	0.7219502	0.8961242	0.3808461
vsStraight engine	1.7473869	2.2726721	0.7688689	0.4509559
amManual	2.6172655	2.0047494	1.3055325	0.2065309
gear	0.7640292	1.4566802	0.5245003	0.6056959
carb	0.5093512	0.9424418	0.5404590	0.5948487

Figure 7 (Model 3)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	31.416055	3.0201093	10.402291	0.0000000
amManual	14.878422	4.2640422	3.489276	0.0016210
wt	-3.785907	0.7856478	-4.818836	0.0000455
amManual:wt	-5.298361	1.4446993	-3.667449	0.0010171

Figure 8 (Residual Plots)