

The Effect of Transmission Type on Automobile Fuel Consumption

Jon Cusick

Executive Summary

Main Questions:

1. Is automatic or manual transmission better for mpg?
2. Quantify the difference between the two types of transmission.

Analysis

This analysis will make use of the `mtcars` dataset, a collection of fuel usage for 32 cars, each of which is a 1973-1974 model. An initial exploratory plot (figure A1) shows a high collection of automatic cars at lower MPG levels while manual vehicles are distributed somewhat more evenly at mid to high levels of MPG. A regression model will be used to

To begin, a simple linear model was fit using all of the available data of the `mtcars` set as predictors of MPG. There are likely unnecessary variables and colinearity in this model, but it is a good starting point to get a general understanding of the dataset. The variable `am` is a binary indicator, with 1 representing manual transmissions and 0 representing automatic transmissions. The slope estimate provided currently indicates that there is about a 2.5 MPG increase when driving a manual car. However, the only significant variable in this model, as described by the p-value is `wt`, the weight of each car.

term	estimate	std.error	statistic	p.value
(Intercept)	12.3033742	18.7178844	0.6573058	0.5181244
cyl	-0.1114405	1.0450234	-0.1066392	0.9160874
disp	0.0133352	0.0178575	0.7467585	0.4634887
hp	-0.0214821	0.0217686	-0.9868407	0.3349553
drat	0.7871110	1.6353731	0.4813036	0.6352779
wt	-3.7153039	1.8944143	-1.9611887	0.0632522
qsec	0.8210407	0.7308448	1.1234133	0.2739413
vs	0.3177628	2.1045086	0.1509915	0.8814235
am	2.5202269	2.0566506	1.2254035	0.2339897
gear	0.6554130	1.4932600	0.4389142	0.6652064
carb	-0.1994193	0.8287525	-0.2406258	0.8121787

However, it is always possible that there is a unseen influence from other variables in the dataset acting on the coefficients for the simple model above. A common tool used to assess the impact of this effect, colinearity, is analyzing the variance inflation factor (VIF). The VIF describes the change in variance in a regression model due to colinearity from a given variable. When calculating the VIF for the above model, it suggests that some variables should indeed be culled from this model in order to better understand transmission's influence on mpg. A rule of thumb is to discard variables where $VIF > 5$.

cyl	15.373833
disp	21.620241
hp	9.832037
drat	3.374620
wt	15.164887

qsec	7.527958
vs	4.965873
am	4.648487
gear	5.357452
carb	7.908747

This results with only the variables **drat** (rear axle ratio), **vs** (engine type), and **am** (transmission type). Vehicle weight, **wt**, was also considered because of it's initial significance value from above. When looking at the VIF from a model using only these four variables, the values are all well below 5. The final linear regression model was an additive model using the variables chosen above. A summary table from this model can be seen below and the residual plot is can be found in the appendix (figure A2).

term	estimate	std.error	statistic	p.value
(Intercept)	8.326573	6.016660	1.383919	0.1773164
drat	1.985086	1.882902	1.054270	0.3007724
vs	6.235194	1.421413	4.386616	0.0001479
am	4.668725	1.837723	2.540495	0.0168985

	2.5 %	97.5 %
(Intercept)	-3.9979958	20.651141
drat	-1.8718632	5.842035
vs	3.3235608	9.146828
am	0.9043208	8.433129

The residual errors do not display any obvious pattern suggesting heteroskedasticity or any other gross underlying failure of the model. Interpreting the summary statistics indicates that with a 95% confidence rate, manual transmission cars have an improvement of 4.67 miles per gallon (+/- 3.76) over automatic transmission cars.

Appendix A - Figures

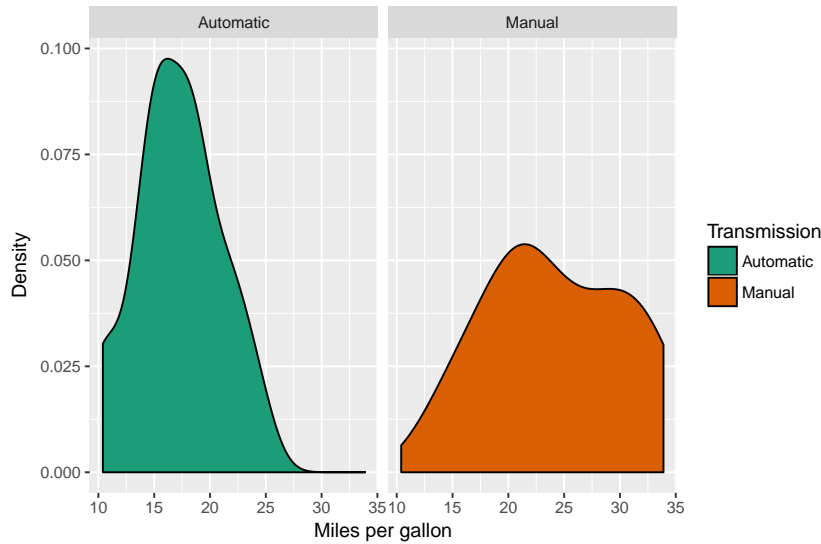


Figure A1. Distribution of MPG average for 32 vehicles by transmission type

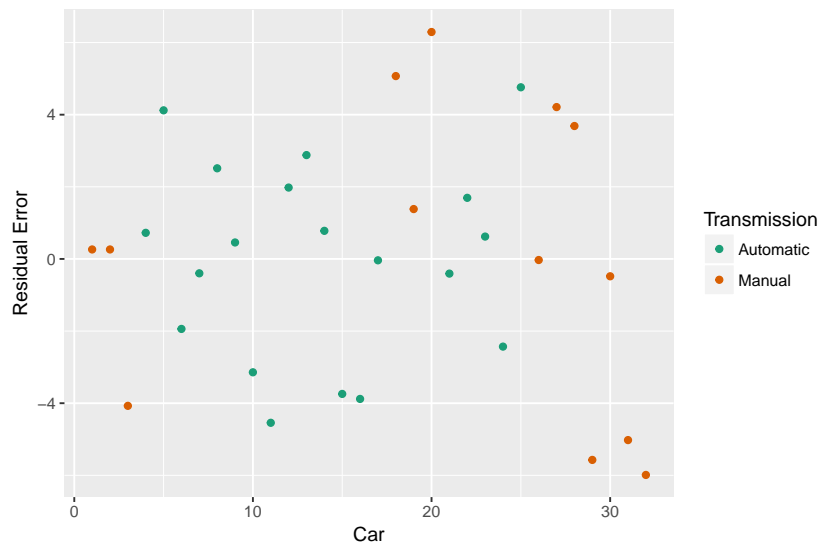


Figure A2. Linear model residuals by transmission type