

Auto-Manual Comparison for Regression Model course project

Mark Nicholls

16/05/2021

Executive Summary

I have been asked to explore the relationship between vehicle transmission and fuel economy (in Miles Per Gallon / MPG), in order to answer these questions:

- “Is automatic or manual better for MPG”
- “Quantify the difference between automatic and manual transmissions”

For this analysis I will use the mtcars data, a standard data set included with the R data analysis software. The help file for the data set provides this description of the data:

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

After some exploratory data analysis I decided to model how mpg changes by transmission and to compare this to a model which also takes into account vehicle weight, as I believed this to be an important factor.

My analysis shows that after accounting for vehicle weight, transmission has very little effect on mpg.

Analysis

Lets look at the first 3 rows of data data to see what variables are available

Table 1: mtcars dataset

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1

So we have MPG and transmission included in the data set as **mpg** *MPG Miles per (US) gallon* and **am** *Transmission (0 = automatic, 1 = manual)* respectively. I also believe that vehicle weight might also be in important factor that affects mpg so we'll look at that one as well: **wt** *Weight (1000 lbs)*.

I plotted a box and whisker plot of mpg by transmission, and also a scatter plot of mpg by weight, with a regression line added, see Appendix: Exploratory Data Analysis plots

A first glance there appeared to be a relationship between transmission and mpg. The box and whisker plot showed automatic has a lower mean and has less variance than manual. The scatter plot showed there is also a clear relationship between weight and mpg, the greater the weight, the lower the mpg, lets model both of these relationships.

Modelling

Model 1 is mpg by transmission, Model 2 is the same as Model 1 but with an added term for vehicle weight. I remove the intercepts from both so the estimates can be interpreted as the mean change in mpg for each term. Here are the estimates from the model.

Table 2: Model Estimates

	Model.1	Model.2
Automatic	17.15	37.32
Manual	24.39	37.30
Weight	NA	-5.35

I can see from this that Model 1 shows higher mpg for manual than automatic, but Model 2 shows that when you account for vehicle weight the mpg for each transmission type is very similar.

In order to quantify uncertainty we can look at the confidence intervals for the models. In table 3 we can see that high end of the manual interval is quite close to the low end of the automatic interval which means we can be only confident there is a small difference in mpg by transmission type.

Table 3: Model 1

	2.5 %	97.5 %
trans_Automatic	14.85	19.44
trans_Manual	21.62	27.17

Table 4 shows the confidence intervals for the second fit. Intervals for both transmission types are very similar, confirming our results from Model 1. And Weight has a much greater factor on vehicle mpg than transmission type, here we can be 95% certain there for each 1000lbs vehicle weight the mpg will reduce by between 6.965 and 3.741 miles per gallon.

Table 4: Model 2

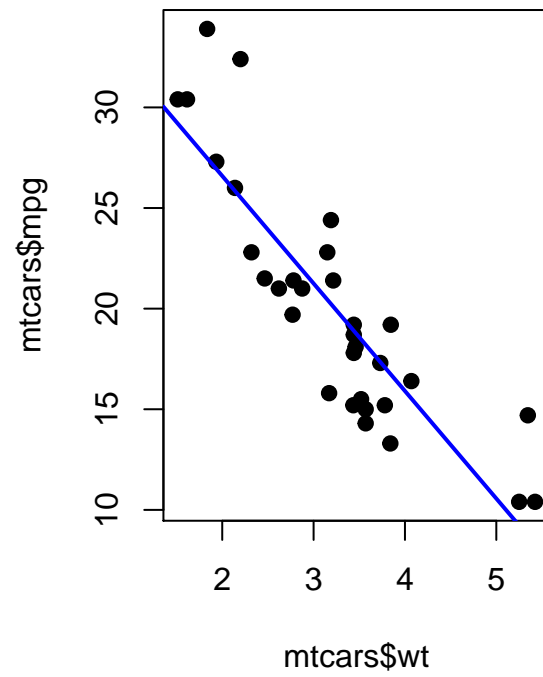
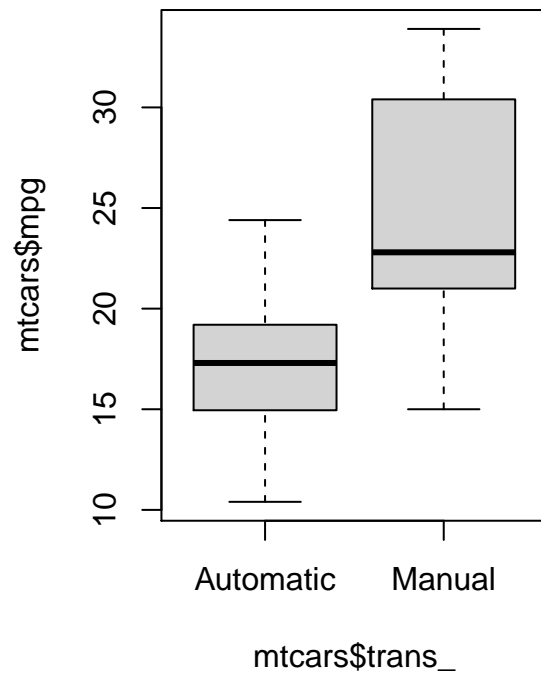
	2.5 %	97.5 %
trans_Automatic	31.07	43.57
trans_Manual	33.03	41.56
weight	-6.96	-3.74

After fitting the models I also looked at some residual and diagnostic plots (see Appendix: Residual / diagnostic plots)

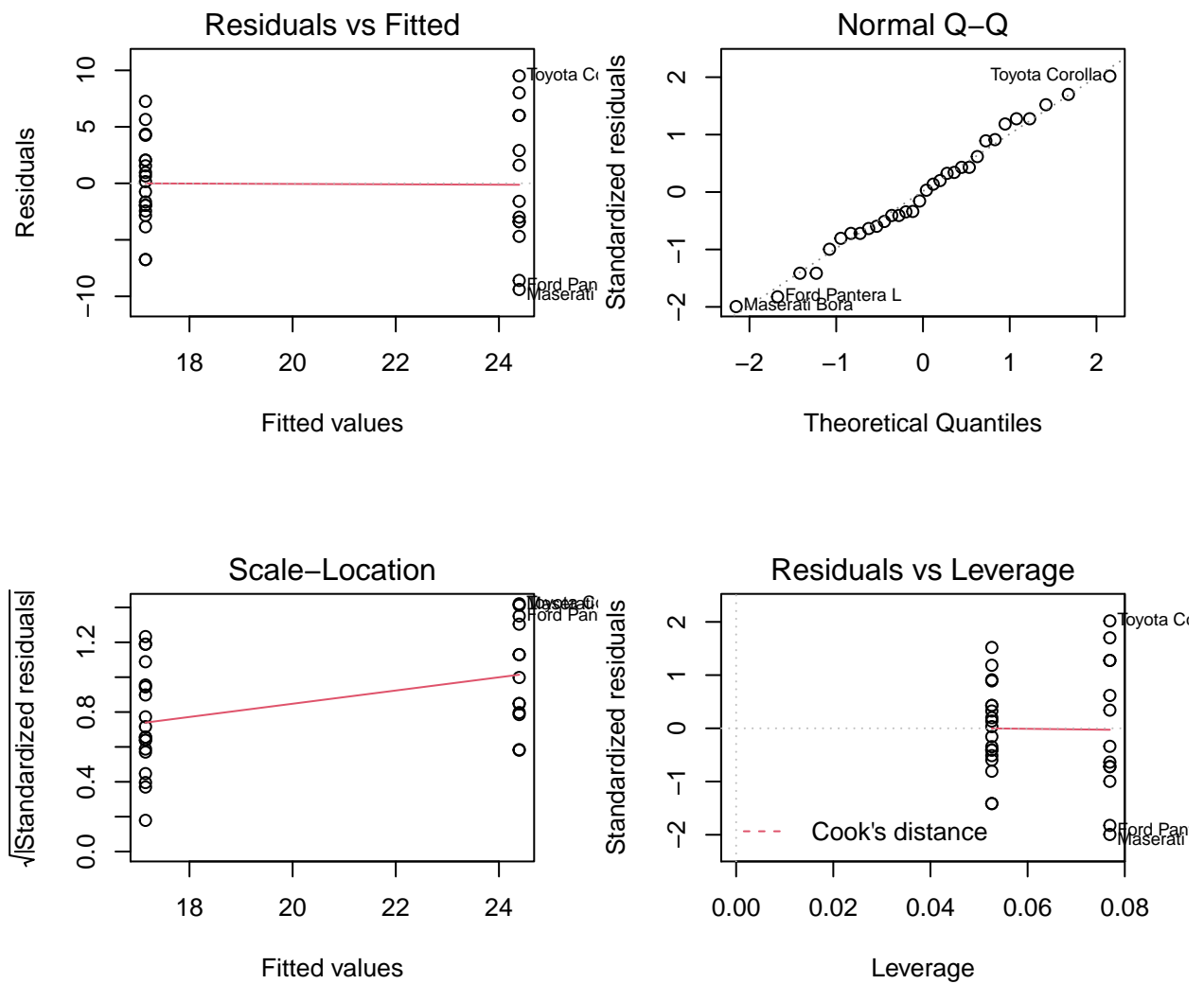
The Normal Q-Q plots show that the linear assumption of normal residuals holds true for both models.

On the Residuals vs Fitted plot for Model 2 the red dotted line is not very horizontal. This suggests that the relationship might be non-linear and a curve might fit these data better. Also the Residual vs Leverage for Model 2 shows some outliers, this suggests further investigation might be worthwhile to see if there is more information we could include to produce a more accurate model.

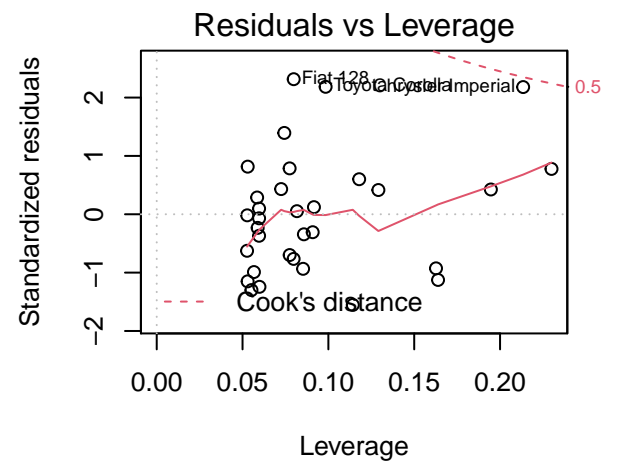
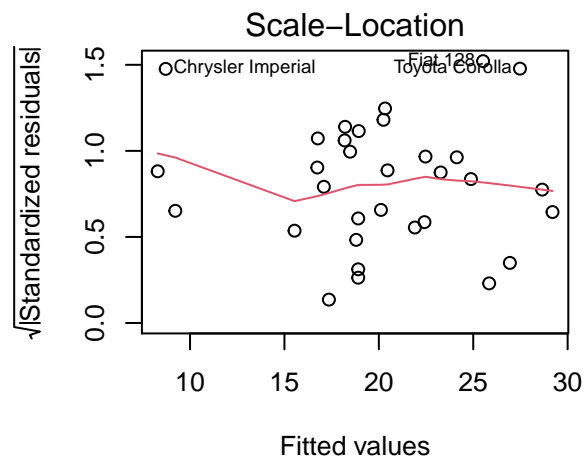
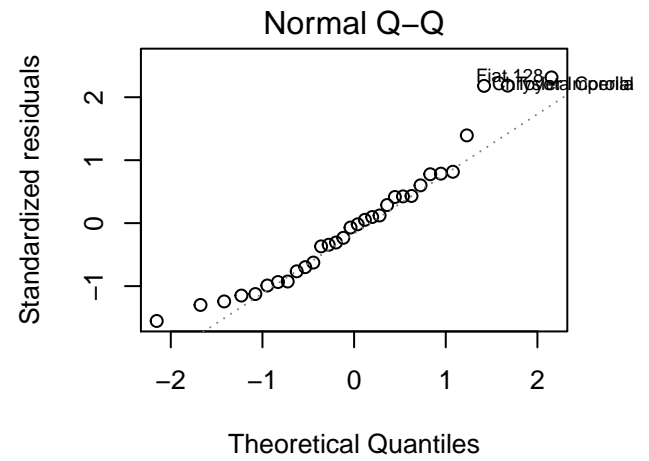
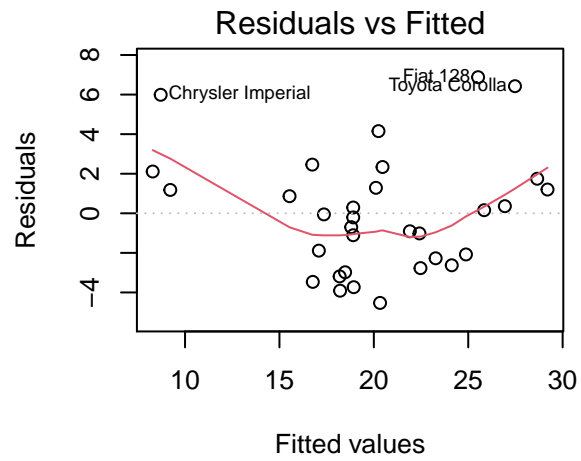
Exploratory Data Analysis Plots



Residual / Diagnostic plots



Model 1



Model 2