Coursera Regression Models Project

Mario Segal

Executive Summary

The purpose of the Analysis is to explore the factors that drive fuel efficiency in motorcars using the Motor Trend data for 1973 and 1974 models available on the R dataset mtcars. In particular we wan to understad if automatic transactions are better or worse in terms of fuel efficiency

For this purpose I conducted some exploratory data analysis and then built several regressions models to quantify the difference in fuel economy between the transmission types and identify other factors that may impact fuel efficiency.

Initial Data Exploration and Analysis

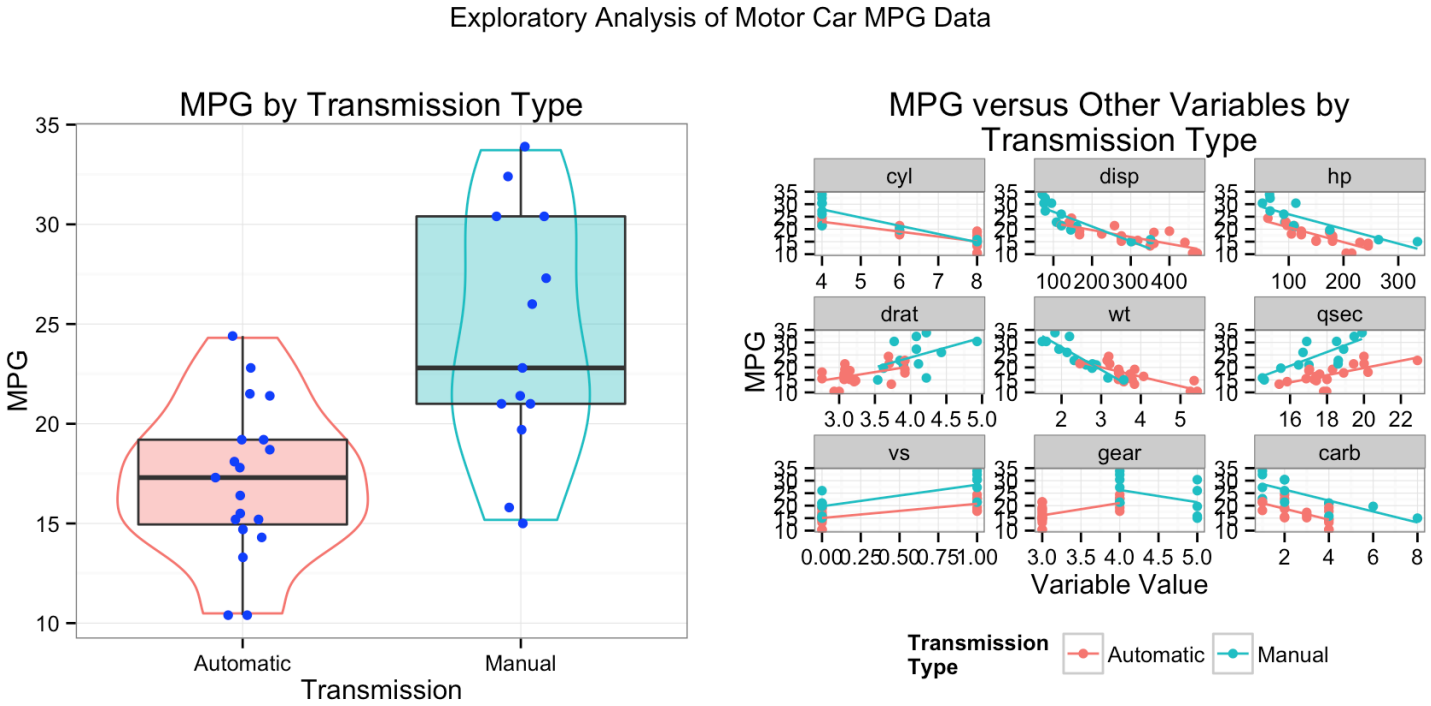


Figure 1 :The Figure above shows that transmission type appears to be related to fuel economy. The left panel shows that manual transmission cars generally have higher efficiency. The right panel shows that multiple other variables appear to be related to fuel efficiency as well, in particular it appears that some variables such as rear axle ratio and weight (drat and wt) separate manual and automatic cars well, and that the realtionships appear to be well approximated by linear relationships.

Explanatory Analysis

As a first step in the explanatory analysis I built a liner model to explain MPG as a function only of the transmission type. This model by definition will predict the average for automatic as the intercept coefficient and the difference in the averages as the coefficient for manual cars. While this model determines a significant difference among transmission it is not very instructional as it only explains 34% of the variation in MPG.

The second step was to perform a full search across multiple linear models using the caret method “lmStepAIC” with otions forward, backward and stepwise. Both the backward and stepwise options resulted in a model explaining 83% of variation in MPG. The model summary is presented below:

Regression Model Coefficients and Significance

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.62	6.96	1.38	0.18
wt	-3.92	0.71	-5.51	0.00
qsec	1.23	0.29	4.25	0.00
am	2.94	1.41	2.08	0.05

The final model was evaluated with diagnostic plots (see appendix) and found to perform well, In particular the residuals vs. fitted plot does not show a clear pattern in the residuals which would indicate that they are related to the outcome variable. In addition the Normal Q-Q plot shows that the points lie mostly along the expected line indicting proper model fit. Finally the predictor versus residuals plots do not show a patterns ae expected.

To illustrate the model fit, Figure 2 of the actual MPG values versus their predicted values, which shows they are relatively close and that lower observed values tend to be related to lower predictions and higher observed values to higher predictions.

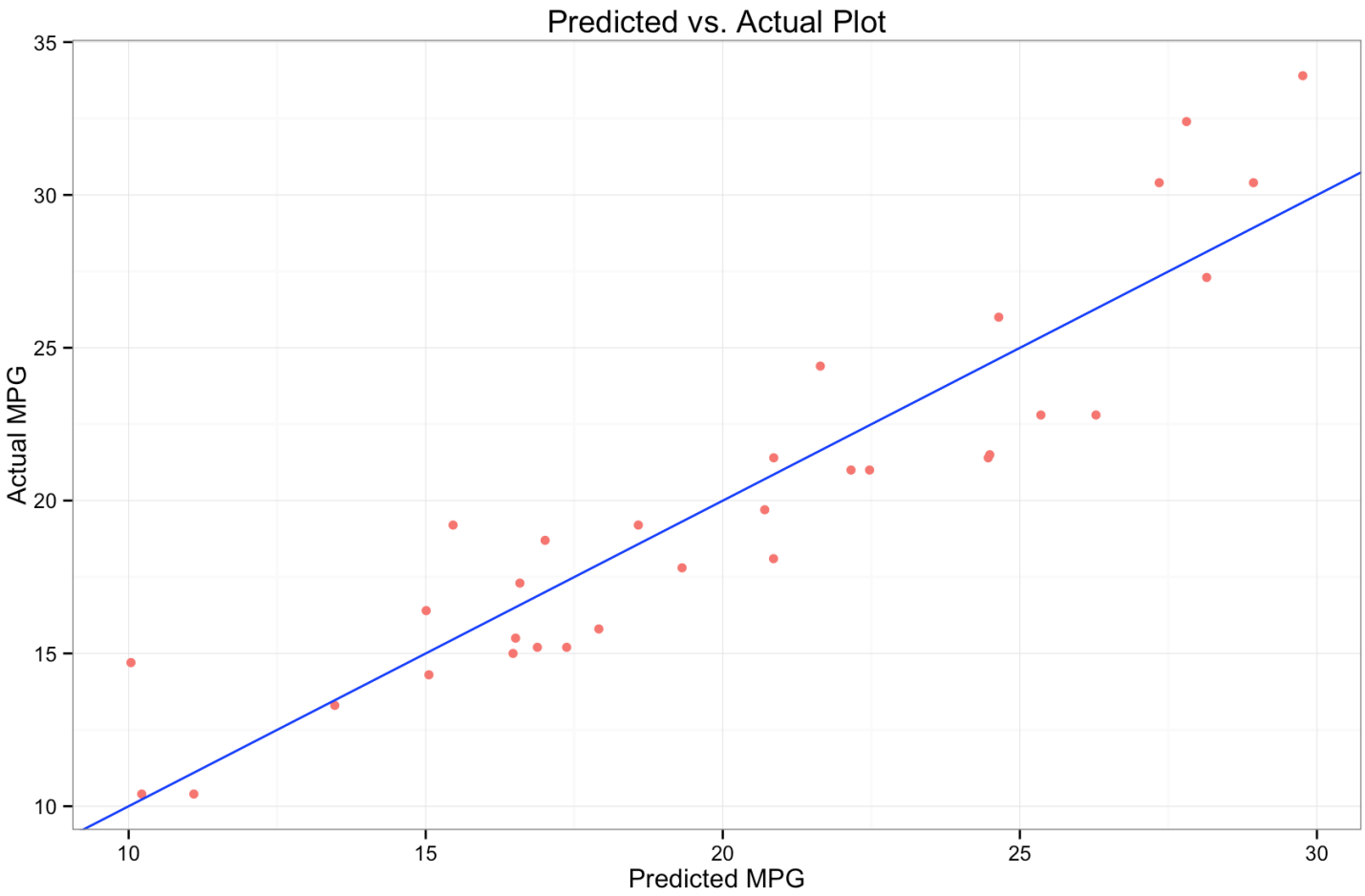
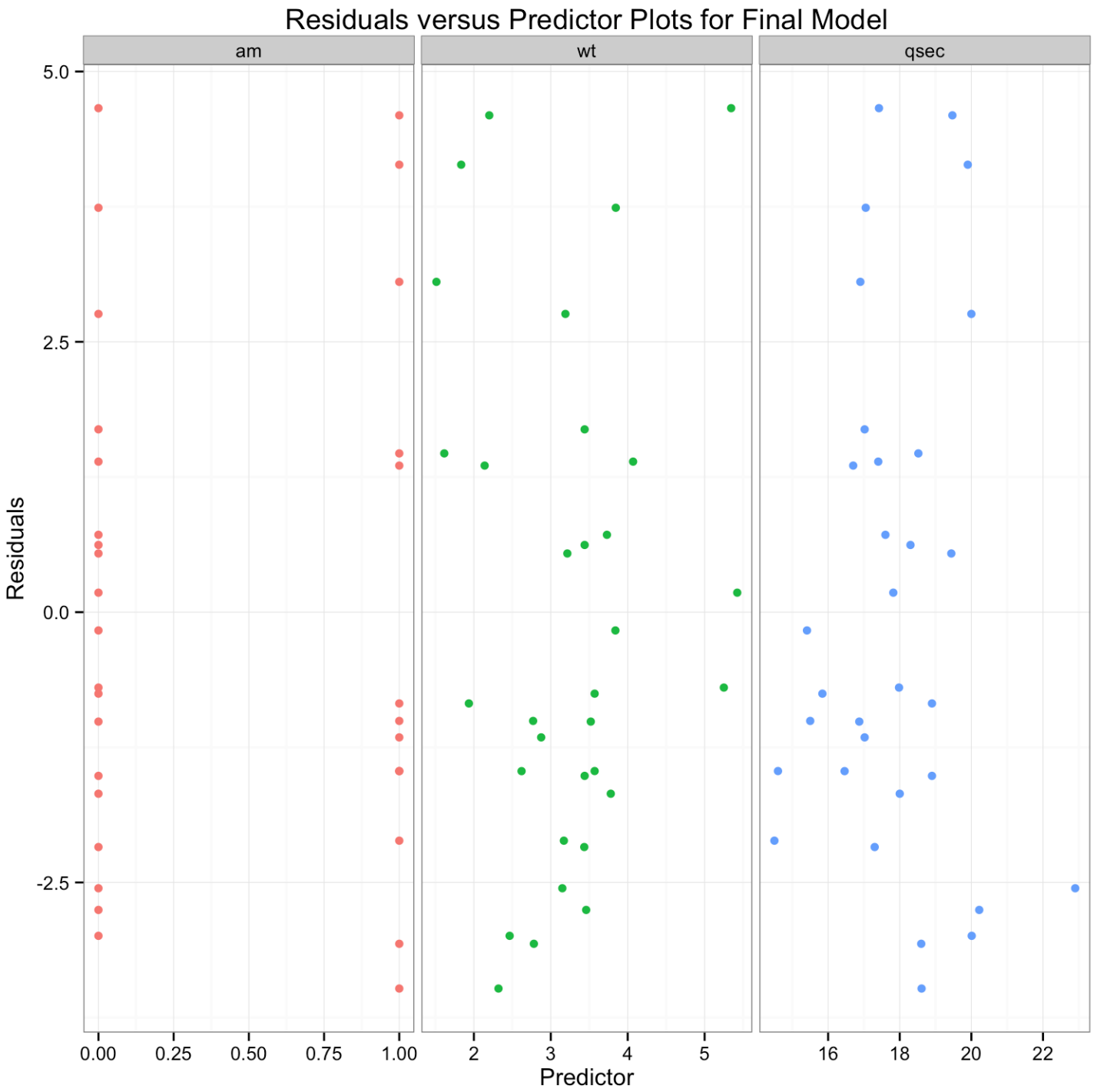
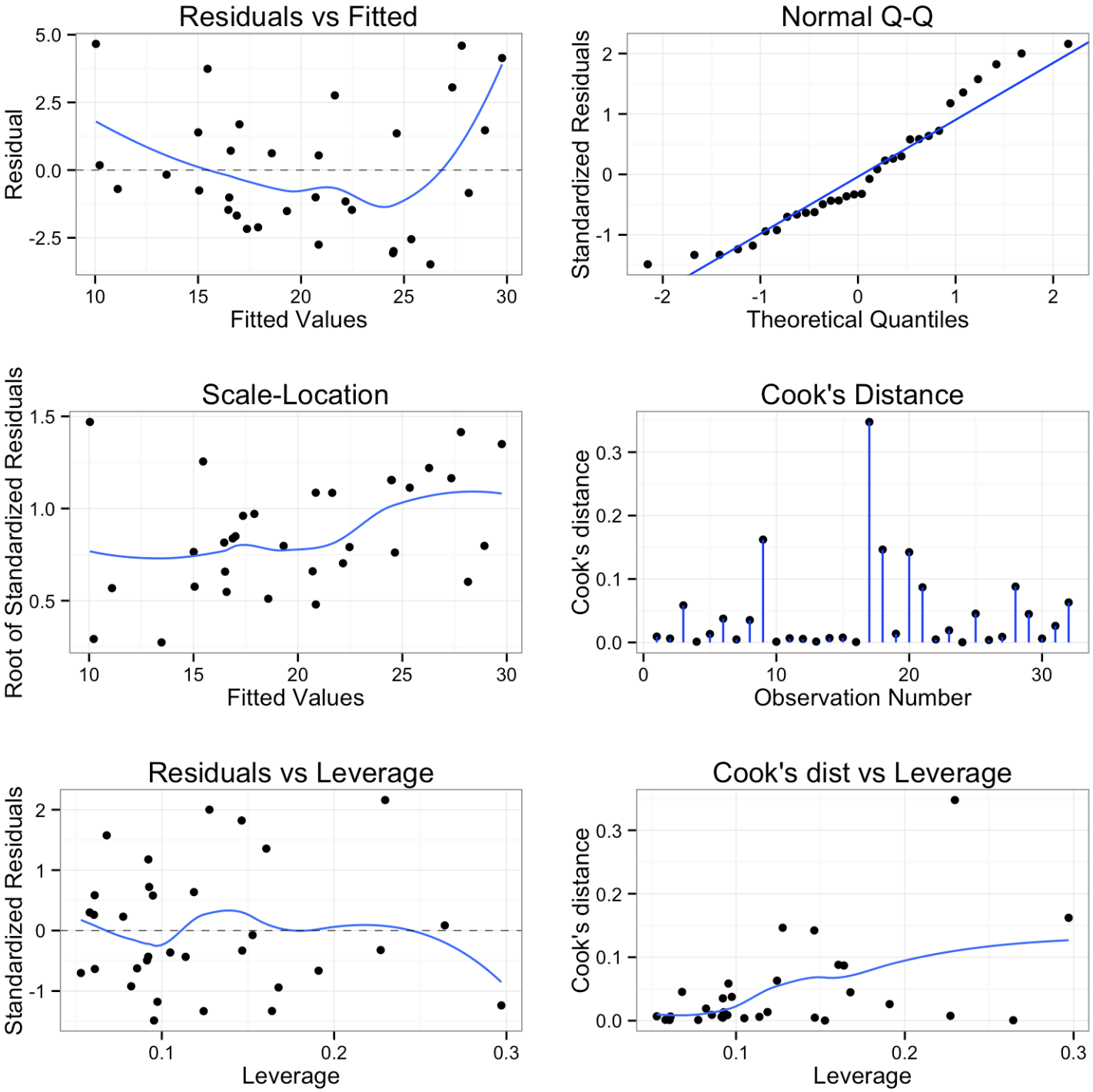


Figure 2 :The Figure above shows that the predicted MPG values are close to the actual MPG values and fall along the line defining perfect prediction.

Appendix



Appendix 1 :Appendix 1 shows plots of the residuals versus the predictors in the final model.



Appendix 2 :Appendix 2 shows traditional diagnostic plots for the final model.