Analysis of the Relationship between MPG and Type of Transmission

Cindy Lim

Saturday, May 23, 2015

Executive Summary

This report seeks to explore the relationship between a set of variables and miles per gallon (mpg). Weight of car, quarter mile time and type of transmission (automatic or manual) were found to be significantly associated with the outcome variable, mpg.

Summary of data

The set of variables which we consider for exploring the relationship with mpg are the number of cylinders (cyl), engine displacement (disp), gross horsepower (hp), rear axle ratio (drat), weight of car (wt), quarter mile time (qsec), vehicle shape (vs), transmission type (am), number of forward gears (gear) and number of carburetors (carb).

From the summary plots given in Appendix 1, we can observe that there seems to be a linear relationship between mpg and cyl, disp, hp, wt and carb. In addition, it also seems that there is a difference in mpg for different vehicle shape and type of transmission.

Analysis

For the analysis, we fit regression models to identify variables which are associated with mpg. A variable with a p-value of less than 0.05 is considered to be significant. If there are more than one variable which are significantly associated with mpg, we will proceed to fit a multivariable regression model with all the significant variables. We will use the stepwise selection method to retain significant variables in the multivariable regression model, and remove variables which are not. From the multivariable model, the variable with the largest p-value will be removed first, and the model refitted with the remaining variables. The process of removing the variable with the largest p-value will be carried out iteratively till the variables left in the multivariable model are all significant.

Results from regression models

We fitted regression models for each variable with mpg separately. The p-values for all the variables fitted were less than 0.05, indicating that all the variables were significantly associated with mpg. In particular, we show the results of the regression model with type of transmission as the explanatory variable in the table below.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	17.1474	1.1246	15.25	0.0000
am	7.2449	1.7644	4.11	0.0003

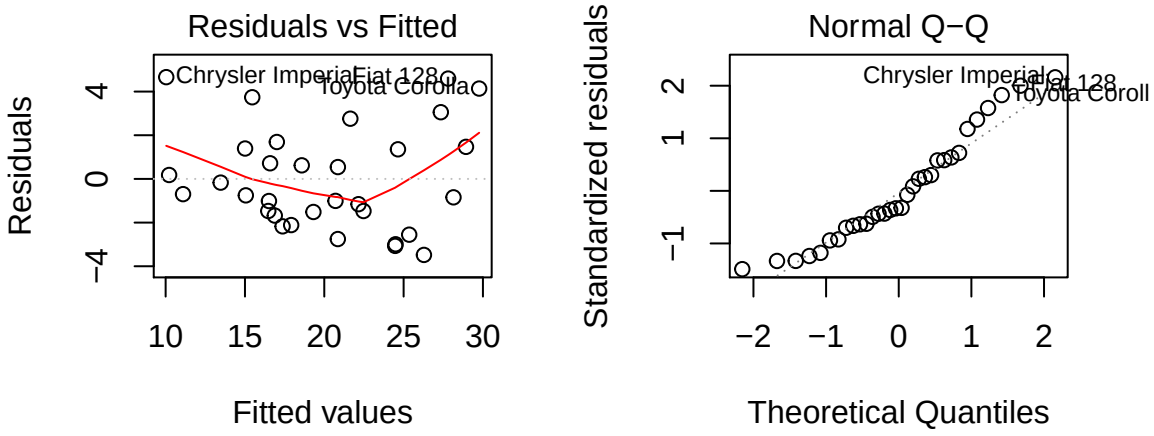
From the results, we estimate an expected 7.24 increase in mpg when comparing a manual transmission car to an automatic transmission car.

We fit a multivariable regression model next, following the steps outlined previously to only retain significant variables in the final model. The final multivariable regression model had the variables car weight (wt), quarter mile time (qsec) and type of transmission (am) in the model. The coefficients from the model with the associated p-values are shown in the table below.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.6178	6.9596	1.38	0.1779
wt	-3.9165	0.7112	-5.51	0.0000
qsec	1.2259	0.2887	4.25	0.0002
am	2.9358	1.4109	2.08	0.0467

From the results, we estimate an expected -3.92 decrease in mpg for every 1000 lbs increase in car weight, while holding the remaining variables constant. We also estimate an expected 1.23 increase in mpg for every second increase in quarter mile time, while holding the remaining variables constant. For type of transmission, we estimate an expected 2.94 increase in mpg for manual transmission cars when compared to automatic transmission cars, while holding the remaining variables constant.

Residual and diagnostic plots



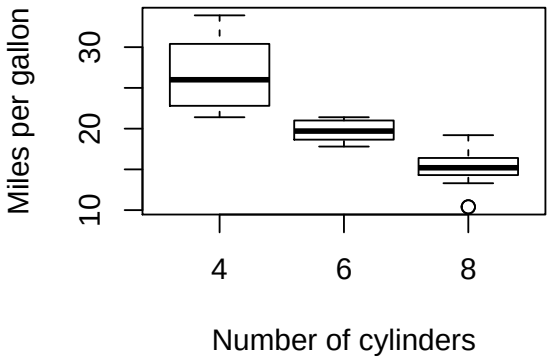
plot-1.pdf
The residuals versus fitted plot for the multivariable regression model indicate that there is some correlation between the residuals and fitted values. The normal Q-Q plot seem to deviate from normality at the ends.

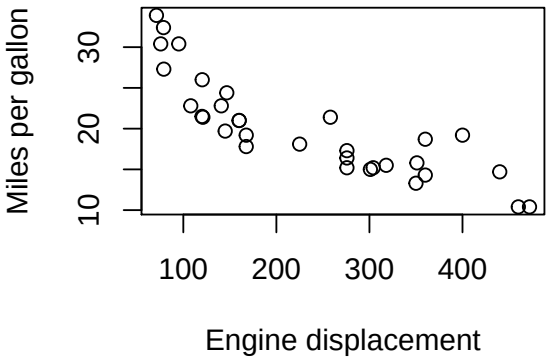
Conclusion

In this study, we have explored the relationship between mpg and a set of variables, and have found that the weight of car, quarter mile second and type of transmission to be significantly associated with mpg. From our analyses, type of transmission was not the only variable that was associated with mpg, and thus proceeded to fit a multivariable regression model. However, caution needs to be taken when interpreting the results. The residual and diagnostic plots indicate that other models could also be explored.

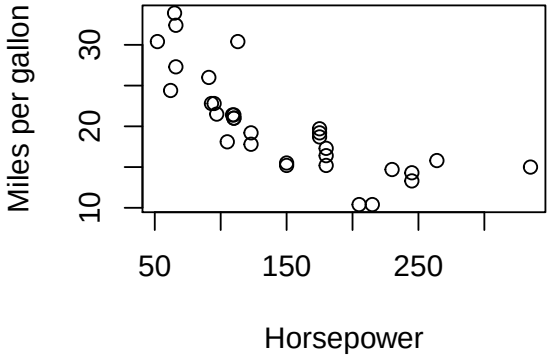
Note: This report was authored in R Markdown and compiled to pdf via knitr). To view the raw source, please visit the GitHub repo associated with this project at the following link:https://github.com/cindylim/MotorTrend

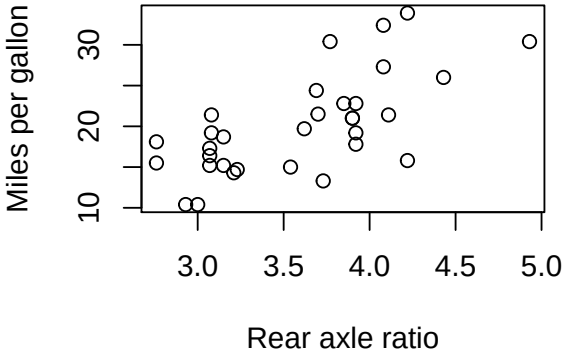
Appendix 1: Exploratory Analysis Plots



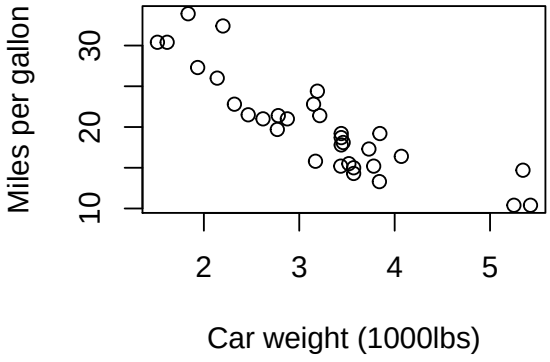


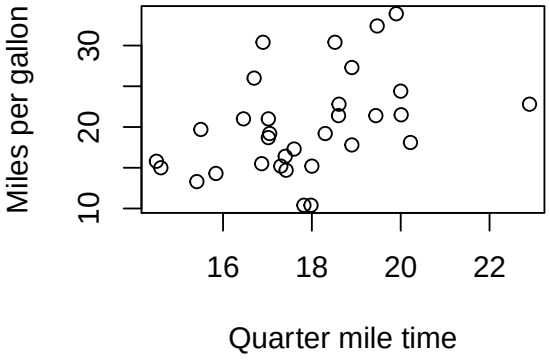
plot-1.pdf



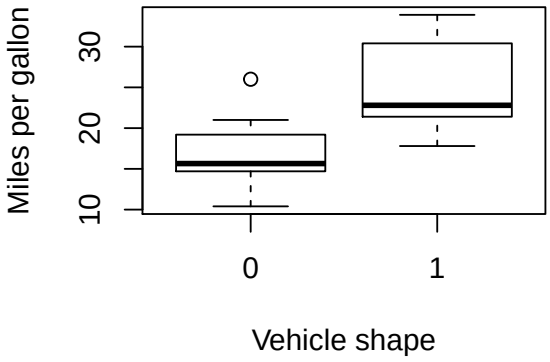


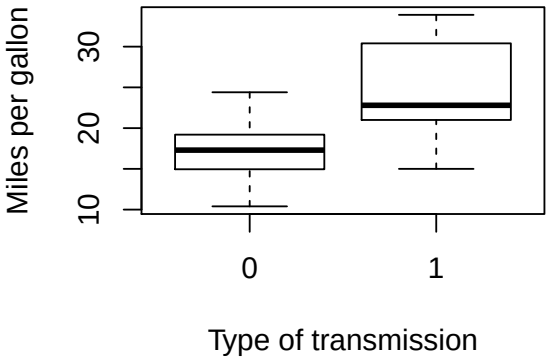
plot-2.pdf



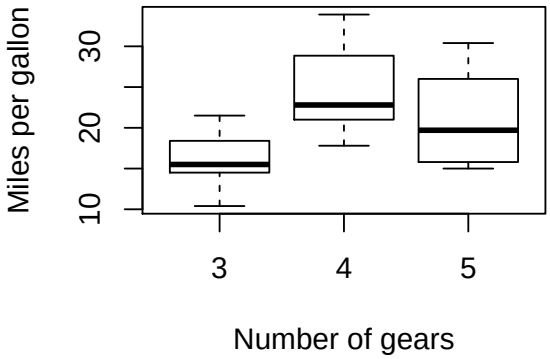


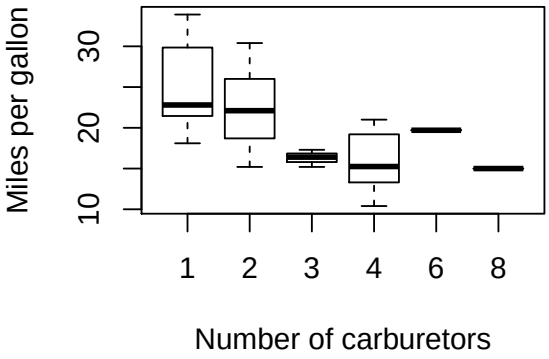
plot-3.pdf





plot-4.pdf





plot-5.pdf