Impact of Transmission Type on Fuel Economy

jrfoster

Executive Summary

This analysis seeks to contrast the impact of automatic and manual transmissions on fuel economy and to quantify that impact. The analysis utilizes data from the 1974 Motor Trend US magazine on 32 automobiles. The analysis concludes that if only transmission type is considered then a manual transmission provides a 7.245 mpg increase. However, that model also suggests that transmission type alone cannot account for this entire increase with the model having adjusted $R^2 = .3385$.

Using ANOVA, several models are compared, informed by (Hocking 1976), (Henderson and Velleman 1981) and (US EPA 1976). The analysis concludes that a more appropriate model includes the vehicle's weight and quarter mile time. Using this model and holding all other variables constant, transmission type provides an increase of 2.94 mpg. This model has adjusted $R^2 = .834$.

Exploratory Data Analysis

The data were extracted from the 1974 Motor Trend US magazine, and comprise fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models) and therefore consist of 32 observations of 11 variables. The following table lists the variables and their classes.

mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
numeric	numeric	numeric	numeric	numeric	numeric	numeric	numeric	numeric	numeric	numeric

Note that the data contain several numeric variables that are, by nature, discrete and will therefore be converted to factors: these include cyl, vs, am, gear and carb.

In the Appendix, Figure 1 shows a pair-wise plot for the raw data which also contains a correlation matrix. From the plots we can observe fairly strong, but curved, relationships to mpg with displacement, horsepower and weight, but there may be other factors at play as well, such as quarter mile time and the rear axle gear ratio. From the correlation matrix, we can see that there are a number of highly correlated variables: cylinder and displacement are highly correlated, which makes sense because displacement is a measure for all cylinders, weight is highly correlated with cylinders, which makes sense because more cylinders mean larger engines.

Simple Inference on Direct Relationship

We first perform a Welch two-sample t-test with $\alpha = .05$ with the null hypothesis, H_0, that the mean mpg of cars with automatic transmissions is equal to the mean mpg of cars with manual transmissions. We also generate a simple linear model for `mpg ~ am`. The following table contains the results of the t-test and selected attributes from the linear model.

Conf Int	Deg F	Tabulated t-value	t-Statistic	p-value	Mean (auto)	Mean (man)	R Sq
[-11.2802, -3.2097]	18.33	2.0982	-3.7671	0.0014	17.1474	24.3923	0.3598

From the test we can see that $p < .05$, the t-statistic is greater than the tabulated t-value and the confidence interval does not contain zero, so there appears to be sufficient evidence at $\alpha = .05$ to reject the claim that the mean mpg of cars with automatic transmissions is equal to the mean mpg of cars with manual transmissions. From the linear model, we can see that cars with manual transmissions have an increase of 7.245 miles per gallon. However, we can also see that $R^2 = .36$, which means that this model doesn't entirely explain the increase in mpg.

The Search For A Better Model

Examining the correlation matrix yields a number of variables that are all highly correlated, although some are highly correlated with one another (such as disp and cyl). (US EPA 1976) noted that vehicle weight and displacement were the most significant factors affecting fuel economy. We also note from (Henderson and Velleman 1981) that wt is likely the best single predictor of mpg, so we start by adding wt. (Hocking 1976) settled on a model that also included qsec (which itself is highly negatively correlated with hp and cyl). So we start with transmission type alone, then add in weight and subsequently other predictors based on correlation and prior research and then compare the resulting models using ANOVA.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt
## Model 3: mpg ~ am + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 73.203 2.673e-09 ***
## 3     28 169.29  1    109.03 18.034 0.0002162 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this we can see our p-values are extremely low in the final model `mpg ~ am + wt + qsec`, which corresponds with (Hocking 1976). We also check each model for multiple predictors that are collectively correlated by examining the Variance Inflation Factor. Below are the results of this analysis.

```
##       am       wt
## 1.921413 1.921413
```

```
##       am       wt     qsec
## 2.541437 2.482952 1.364339
```

Based on this analysis, we agree with Hocking's model of `mpg ~ am + wt + qsec` and choose this as our final model. The following tables contain the coefficients for this model.

	Estimate
(Intercept)	9.617781
am1	2.935837
wt	-3.916504
qsec	1.225886

The following table shows the confidence intervals for this model. Note that none of the intervals for the variables contain zero.

	2.5 %	97.5 %
(Intercept)	-4.6382995	23.873860
am1	0.0457303	5.825944
wt	-5.3733342	-2.459673
qsec	0.6345732	1.817199

Figure 2 in the Appendix shows all the diagnostic plots for the final model.

Appendix - Tables and Figures

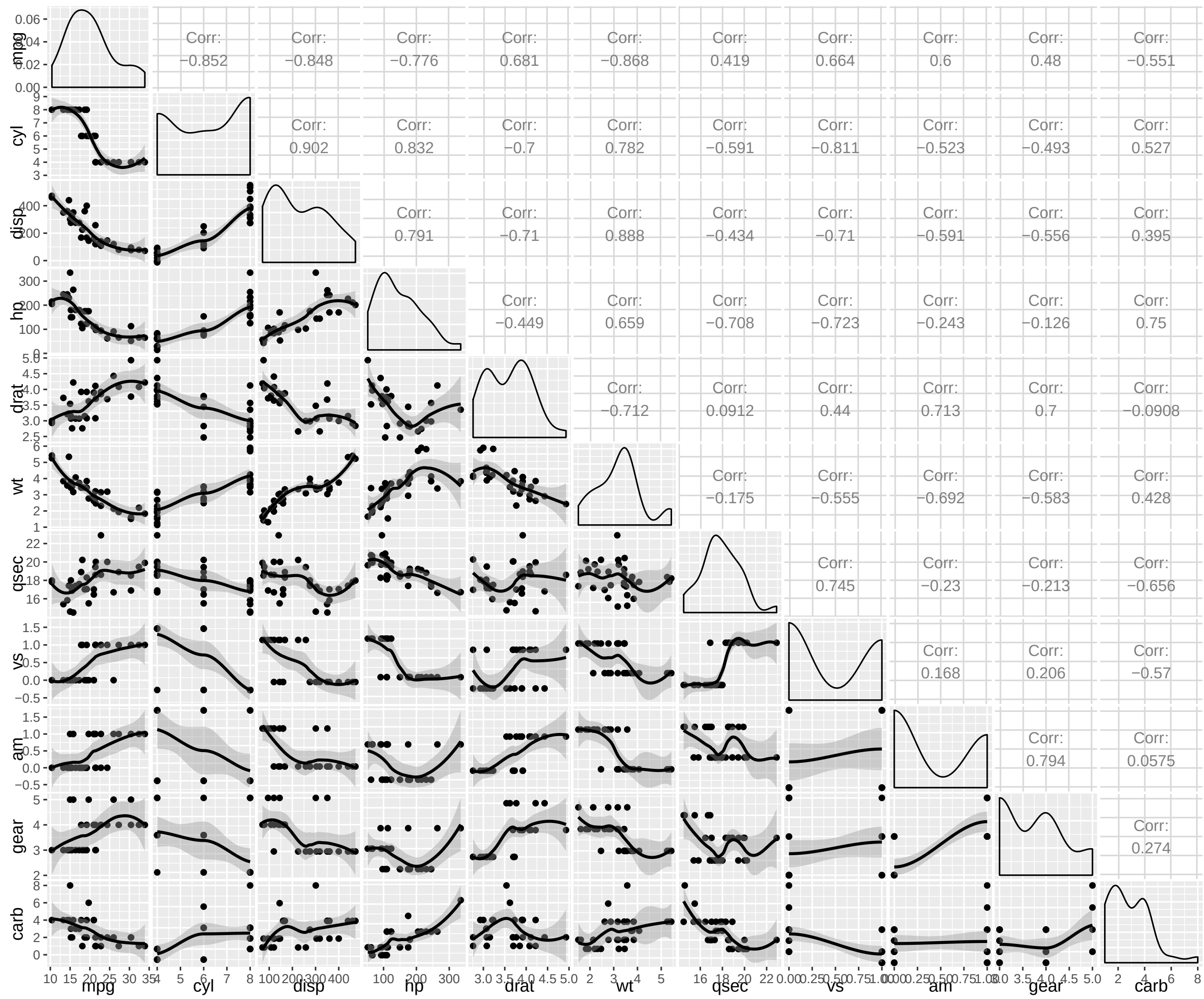

Figure 1: MTCars Pair-Wise Plots and Correlation Matrix

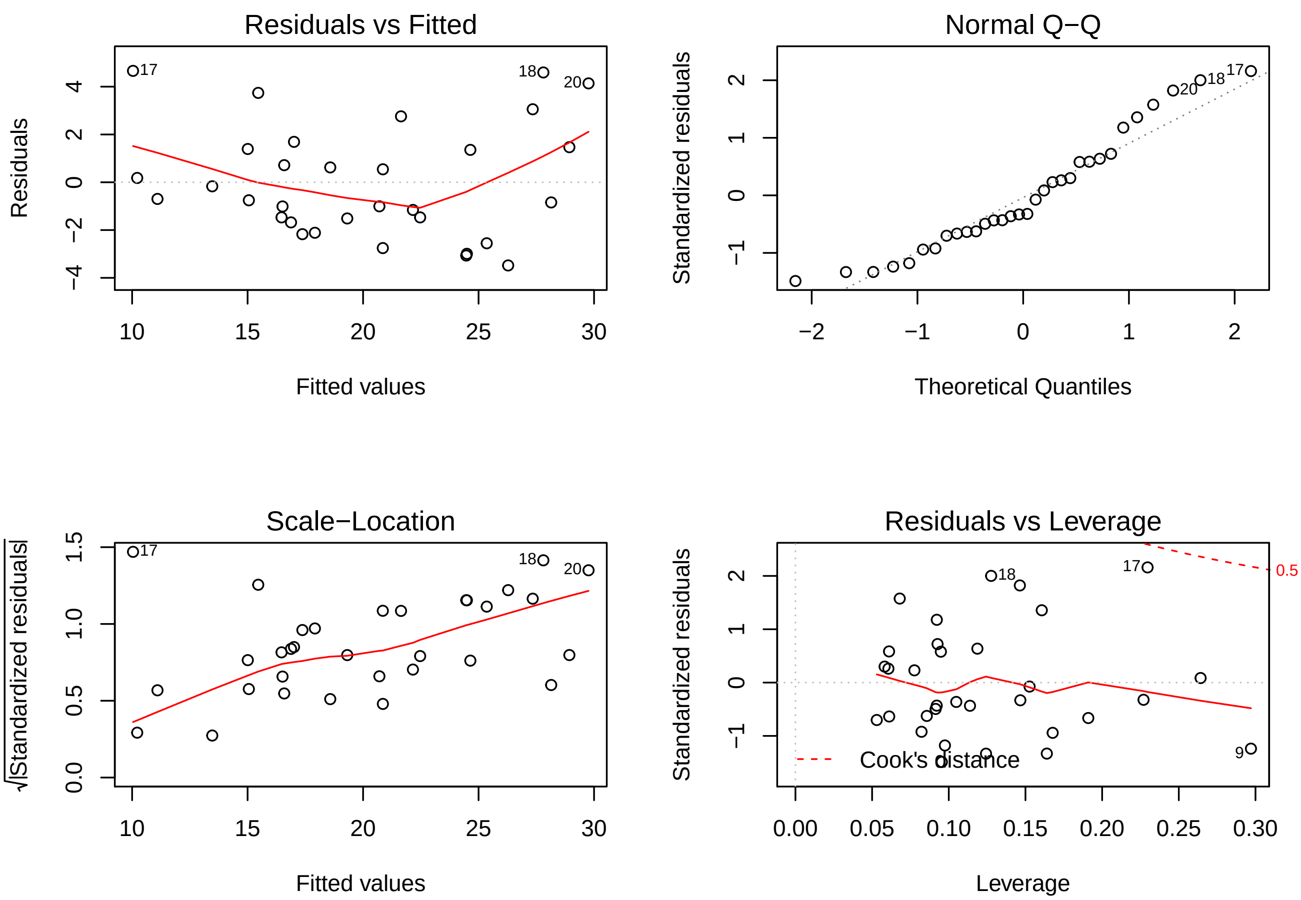

Figure 2: Diagnostic Plots for Final Regression Model

References

Henderson, Harold V, and Paul F Velleman. 1981. "Building Multiple Regression Models Interactively." *Biometrics* 37 (2): 391–411.

Hocking, R R. 1976. "A Biometrics Invited Paper. the Analysis and Selection of Variables in Linear Regression." *Biometrics* 32 (1): 1–49.

US EPA. 1976. "Factors Affecting Automotive Fuel Economy." United States Environmental Protection Agency Office of Air; Waste Management. October. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=9100LCHQ.txt.