Does transmission type predict fuel economy?

Executive Summary

We are interested to know if cars with manual transmissions are more fuel efficient than cars with automatic transmissions. We examine data from R's mtcars data set, which provides 11 variables of data for each of 32 automobiles as they appeared in Motor Trend Magazine (US Edition) in 1974. We discover collinearity among several of the variables.

Although the sample size is rather small, a variety of cars are represented, including small economy cars, large luxury sedans, and high performance sports cars. All vehicles are either 1973 or 1974 models. Vehicles with manual transmissions from the sample enjoy better fuel economy, 7.245 more miles per gallon among our sample vehicles. However, that doesn't tell the whole story, oversimplifying things. By considering multiple models, we find that differences in fuel efficiency are better explained using variables other than transmission type.

Methodology

After loading the data, we rename some of the predictors with descriptive names. We also convert variables carb, cyl, and trans to factor variables.

We start with some exploratory analysis. In particular, we suspect collinearity among some of our predictors. As expected, we find that engine displacement tends to increase with the number of cylinders, horsepower tends to increase with displacement (and, unsurprisingly, with number of cylinders), and as vehicle weight increases, vehicles tend to be equipped with more powerful engines. None of these observations are surprising. What is somewhat surprising is that heavy cars tend to be equipped with automatic transmissions, while lighter cars are typically equipped with manual transmissions. We see that cars with manual transmissions tend to have smaller, lower powered engines with fewer cylinders than their automatic transmission equipped counterparts. *See Figure 1 in Appendix.*

We also see that, as expected, fuel efficiency decreases as weight increases. Also, fuel efficiency decreases as horsepower increases, as cylinder count increases, and as displacement increases. *See Figure 2 in Appendix.*

A full model suggests that weight, displacement, and horsepower are the most significant indicator variables, so we start by performing simple regression on these, yielding R^2 values of 0.75, 0.72, and 0.60, respectively. We also perform a simple regression on differential ratio and transmission type, which gives an R^2 value of 0.46 and 0.34 (adjusted R^2), respectively. While all intercepts and coefficients are statistically significant, transmission type explains less variance in fuel efficiency than any of weight, displacement, horsepower, or differential ratio.

There are over 1000 possible models to test, so we eliminate some predictors from consideration. All eight cylinder engines have a V configuration (i.e., they are V8 engines), and all but one each of four and six cylinder engines have a straight configuration. Most of the automatic transmissions are three speed, with only four having four speed (all Mercedes make) and no five speed. Similarly, all manual transmissions are either four or five speed. Rear differential ratio and number of carburetor barrels tend to increase and decrease, respectively, with engine cylinder count. Quarter mile time is not relevant to most consumers, and it is not reliably provided by manufacturers. We disregard all of these variables and focus on number of cylinders, engine displacement, horsepower, weight, and transmission type. There are 26 possible multiple regression models remaining, disregarding interactions. (There are $2^5 = 32$ subsets of five predictor variables, but we are not interested in the empty subset or the 5 singleton subsets; 32 - 6 = 26.) We check all 26 models and compare their adjusted R^2 values and the predictive adjusted residuals sum of squares (PRESS), which we use as selection criteria to identify candidate models. We then perform nested likelihood ratio tests for these candidate models, as appropriate.

Table 1: Selected models with selection criteria

	Model	Adjusted R^2	PRESS
Model 1	mpg ~ trans + cyl + disp + hp + wt	0.834	234.915
Model 2	mpg ~ trans + cyl + hp + wt	0.840	223.469
Model 3	mpg ~ cyl + disp + hp + wt	0.830	229.334
Model 4	mpg ~ cyl + hp + wt	0.836	220.551
Model 5	mpg ~ cyl + wt	0.820	233.050

Results

We identify five models of interest, based on our selection criteria. Model 1 is our new full model (after discarding all but five predictors from consideration), and Models 2, 4, 5 provide the highest adjusted R^2 and lowest PRESS values for reduced models of their respective sizes (*See Table 1*). We also show Model 3, which has an adjusted R^2 value almost as high as the full model, as well as a lower PRESS value. Only Models 1 and 2 utilize the transmission type as an indicator. In neither model does the coefficient for manual transmission have a p-value less than 0.2, so at the 95 percent confidence level, we cannot reject a null hypothesis that that coefficient is equal to zero. Note that there is no coefficient for an automatic transmission; its effect, along with the four cylinder effect, is included in the intercept.

While Model 2 is the best model, based on adjusted R^2 and PRESS vales, a likelihood ratio test indicates that the transmission type and horsepower are not necessary. A second nested likelihood ratio test indicates that displacement, which appears in Model 3, is also not necessary. Based on these likelihood ratio tests, Model 5 is the best choice, despite having adjusted R^2 and PRESS values that are slightly lower and higher, respectively, from those of our other candidate models. Model 5 does have the advantage of using only two predictor variables, cylinder count and weight. We present Model 5 as our final model (ignoring units):

$$mpg = 33.9908 - 4.2556 \times cyl6 - 6.0709 \times cyl8 - 3.2056 \times wt + \varepsilon$$

Thus, a four cylinder vehicle can be expected to get 33.9908 miles per gallon, less 3.2056 miles per gallon per 1000 pounds in weight. An additional loss of 4.2556 or 6.0709 miles per gallon can be expected if the vehicle is equipped with a six or eight cylinder engine, respectively, holding weight constant.

Conclusion

After looking at 26 models, we cannot conclude that transmission type is an important factor for fuel economy. While vehicles from our sample that are equipped with automatic transmissions get more miles per gallon of fuel, on average, compared to those equipped with automatic transmissions, other things, such as engine cylinder count and vehicle weight, seem to be more important factors for determining fuel economy.

Discussion

Manual transmissions are often used for two reasons, performance and economy. Our sample data contain numerous examples of both. If we focus only on cars designed for economy, we might reveal a statistically significant effect on fuel economy by transmission type.

It's also important to note that we are analyzing data that are over forty years old. Automotive technology has made major advances since then, especially in terms of fuel efficiency. We must be careful to avoid extrapolating our conclusions beyond the time period from which our data come. It would not be unreasonable to hypothesize that with today's technology, vehicles with computer controlled automatic transmissions can provide superior fuel efficiency compared to similar vehicles equipped with manual transmissions.

Appendix

Figure 1: HP vs. Disp, Cyl, and Wt; Disp vs. Cyl.

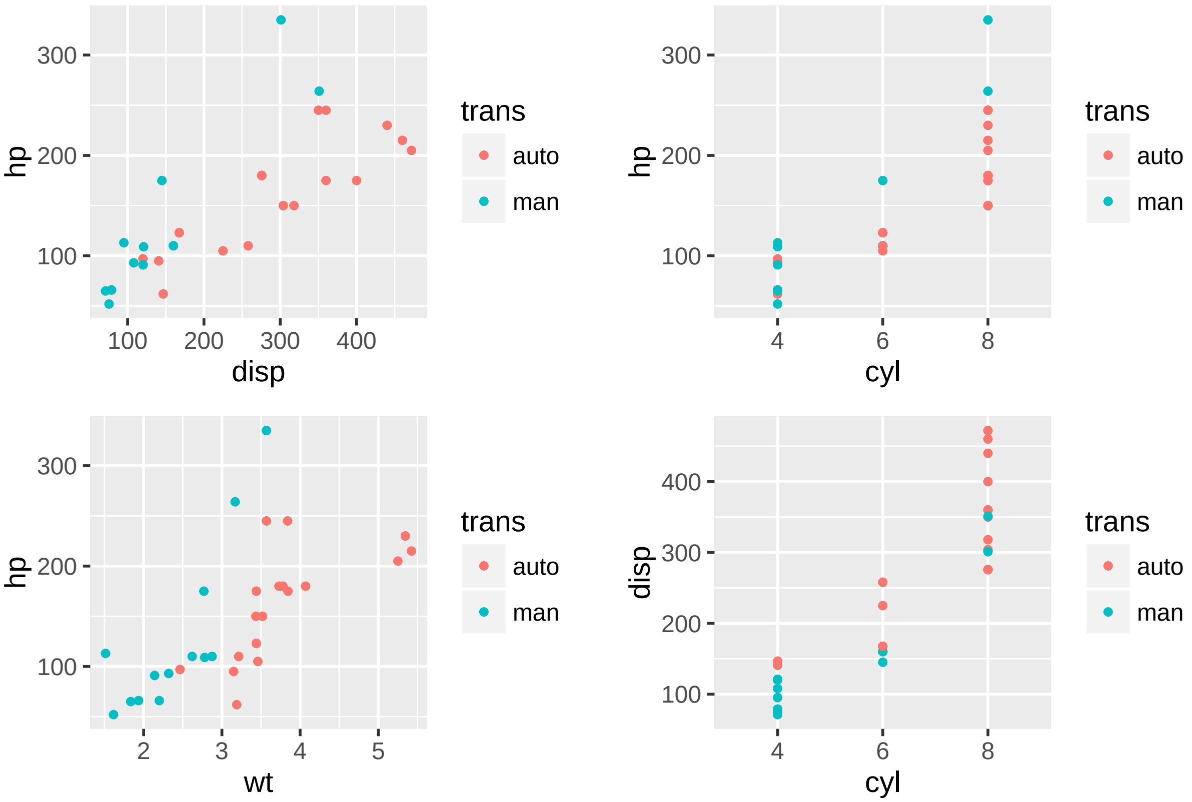
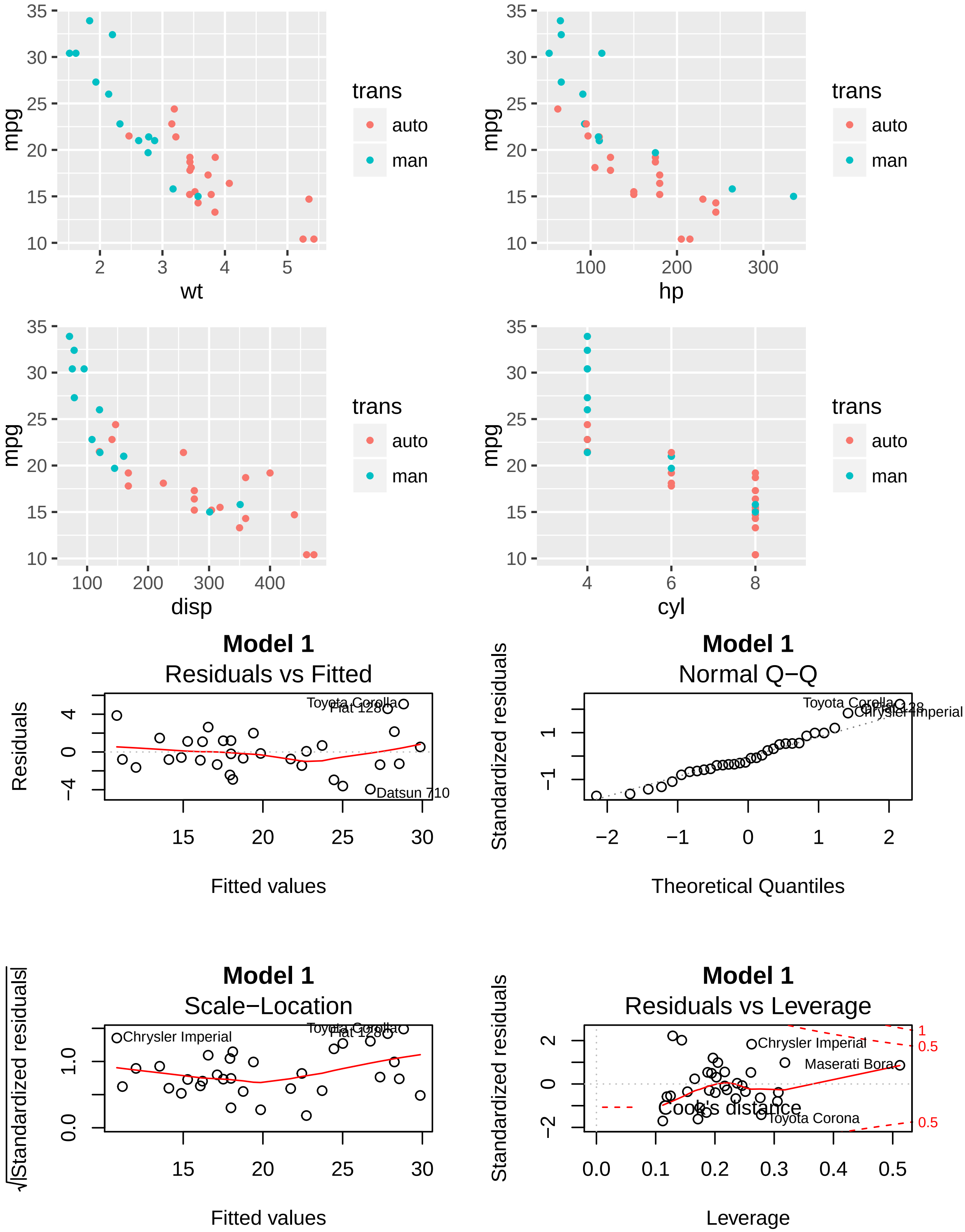
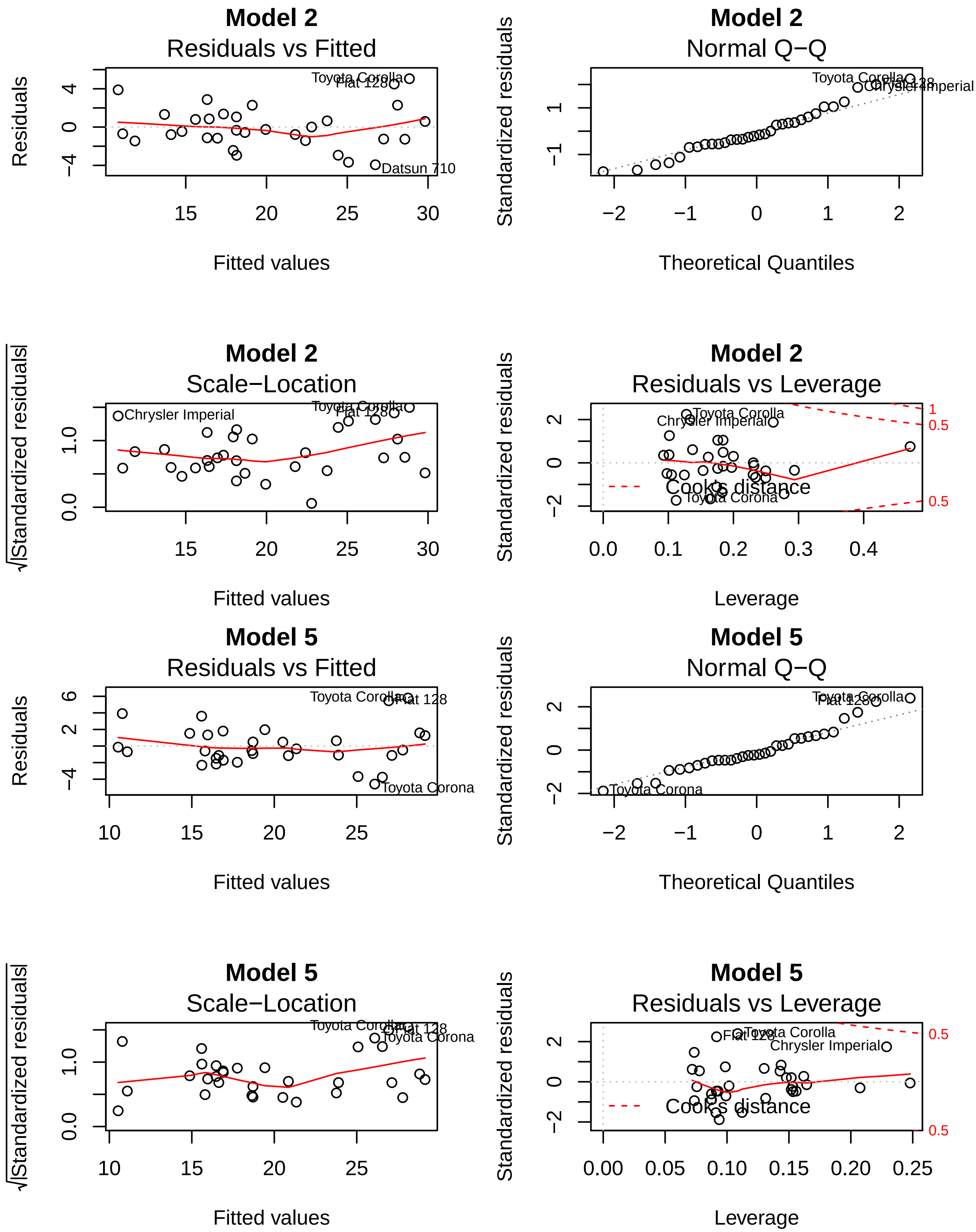


Figure 2: MPG vs. Wt, HP, Disp, and Cyl





Model 2
Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Datsun 710
Model 2
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Chrysler Imperial
Model 2
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Model 2
Residuals vs Leverage
Standardized residuals
Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
1
0.5
0.5
Model 5
Residuals vs Fitted
Residuals
Fitted values
Toyota Corolla
Fiat 128
Toyota Corona
Model 5
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Toyota Corolla
Fiat 128
Toyota Corona
Model 5
Scale−Location
√|Standardized residuals|
Fitted values
Toyota Corolla
Fiat 128
Toyota Corona
Model 5
Residuals vs Leverage
Standardized residuals
Leverage
Fiat 128
Toyota Corolla
Chrysler Imperial
Cook's distance
0.5
0.5