Exploring Effects of Transmission Type on Automobile Miles per Gallon Rates

Alex Hudson

2019-03-15

Executive Summary

Our goal in this analysis is to determine whether any significant difference in vehicle miles per gallon exists based primarily on whether it uses an automatic transmission or a manual transmission, and if so, to quantify that difference. In the analysis, we consider several possible variables, but we ultimately settle on transmission type, vehicle weight, and vehicle horsepower as our primary measures. We determine that there is a relationship between a vehicle's miles per gallon and its transmission type. However, it is dependent upon the weight of the vehicle. For the mean vehicle weight in the data, approximately 3200 pounds, our model estimates that if weight and horsepower are held constant, then a vehicle with a manual transmission will have about 1.14 fewer miles per gallon than a vehicle with an automatic transmission. However, for every 1000 pounds of weight added, a vehicle with a manual transmission will have about another 4.09 miles per gallon fewer than a vehicle with an automatic transmission, provided that both vehicles have identical weight and horsepower.

Exploratory Analysis

Exploratory analyses were conducted primarily considering relationships between miles per gallon and other predictors available while highlighting transmission type throughout. Transmission type remained the primary predictor due to the particular interest in any effect it may have. Of particular interest were apparent relationships between miles per gallon and each of the individual predictors weight and horsepower. Overall, miles per gallon generally decreased as weight increased, but separating the data by transmission type revealed a potential interaction effect; the slope of the decrease in miles per gallon appears to be different between vehicles with manual transmissions and vehicles with automatic transmissions. Meanwhile, miles per gallon generally decreased with horsepower as well. While transmission type makes it appear as if there is a difference in a linear relationship, closer examination of the plot shows that there may instead be a nonlinear relationship between miles per gallon and horsepower.

Model Selection

Our model selection process involved creating a set of nested models and employing ANOVA to compare the nested models in succession. Every model was tested with miles per gallon as the response, the first model included only an intercept, and the order in which variables were added to successive models is as follows: transmission type, centered vehicle weight and weight/transmission interaction, number of cylinders, linear centered horsepower, and quadratic centered horsepower. Beyond this point, adding more predictors did not yield significant results. Further, upon analysis of the last model, it appeared as if adding horsepower as a predictor removed the significance of number of cylinders, and so number of cylinders was removed from the final model.

Final Model

Our final linear model is as follows. Note that the variables for weight and horsepower have been centered prior to creation of the model. We ultimately included transmission type, weight, the interaction between transmission type and weight, and both linear and quadratic forms of horsepower in the final model.

$$
\begin{aligned}
\text{(Miles per Gallon)} = 19.21 & \\
& + (-1.14) \times \text{(Transmission Type: Manual)} \\
& + (-2.00) \times \text{(Weight)} \\
& + (-11.98) \times \text{(Horsepower; Linear)} \\
& + (7.25) \times \text{(Horsepower; Quadratic)} \\
& + (-4.09) \times \text{(Transmission Type: Manual)} \times \text{(Weight)} \\
& + (\text{Error} : \epsilon \sim N(0, \sigma^2))
\end{aligned}
$$

Meanwhile, our 95% confidence intervals for the intercept and each of the coefficients are as follows. While the confidence interval for the main effect of transmission type contains zero, the presence of the interaction term, whose confidence interval does not include zero, warrants the inclusion of the main effect.

Coefficient	2.5 %	97.5 %
(Intercept)	17.966889	20.458251
(Transmission Type: Manual)	-4.011824	1.734046
(Weight)	-3.579452	-0.417556
(Horsepower; Linear)	-18.891483	-5.062936
(Horsepower; Quadratic)	2.109163	12.393403
(Transmission Type: Manual) (Weight)	-6.745988	-1.440691

Model Diagnostics

Diagnostics performed for the final model include residuals versus fitted values, a normal quantile-quantile plot for the residuals, a plot of the square root of the absolute value of the standardized residuals versus fitted values, and a plot of residuals versus leverage. The plots suggested the possibility of two very slight outliers, but these were investigated and nothing anomalous was found in the data. Both were somewhat lighter vehicles with unusually high miles per gallon, but they were not so unusually high as to warrant concern about the model as a whole.

Required Assumptions

As with all multiple linear regression models, some assumptions were required. The model residuals are assumed to have mean zero, have constant variance, and be independent and identically distributed. The use of the model assumes the existence of linear relationships between the response variable and each of the predictors used. At no point during the exploratory analysis, model selection, or model diagnostics did these assumptions appear to be violated in any significant way.

Conclusions

Using multiple linear regression, we conclude that the data support the existence of a relationship between transmission type and miles per gallon for automobiles. That relationship is dependent upon the weight of the vehicle and, in general, as the weight of the vehicle increases, the miles per gallon for vehicles with automatic transmissions decreases more slowly than the miles per gallon for vehicles with manual transmissions, provided that all other predictors are held fixed. For lighter vehicles, this may mean that manual transmissions are associated with greater miles per gallon, while for heavier vehicles, this may mean that automatic transmissions are associated with greater miles per gallon.

Appendix

Data

The data are stored in the `mtcars` data set in R. From the documentation for 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)." In the data, `mpg` records the approximate miles traveled per US gallon of fuel, `cyl` records the number of cylinders in the engine, `disp` records the displacement of the engine in cubic inches, `hp` records the gross horsepower of the engine, `drat` records the rear axle ratio, `wt` records the vehicle weight in units of 1000 pounds, `qsec` records the time in which the vehicle can travel one quarter of a mile, `vs` records whether the cylinders of the engine are arranged in a V shape (0) or a straight line (1), `am` records whether the transmission is automatic (0) or manual (1), `gear` records the number of forward gears in the transmission, and `carb` records the number of carburetors.

Supporting Tables

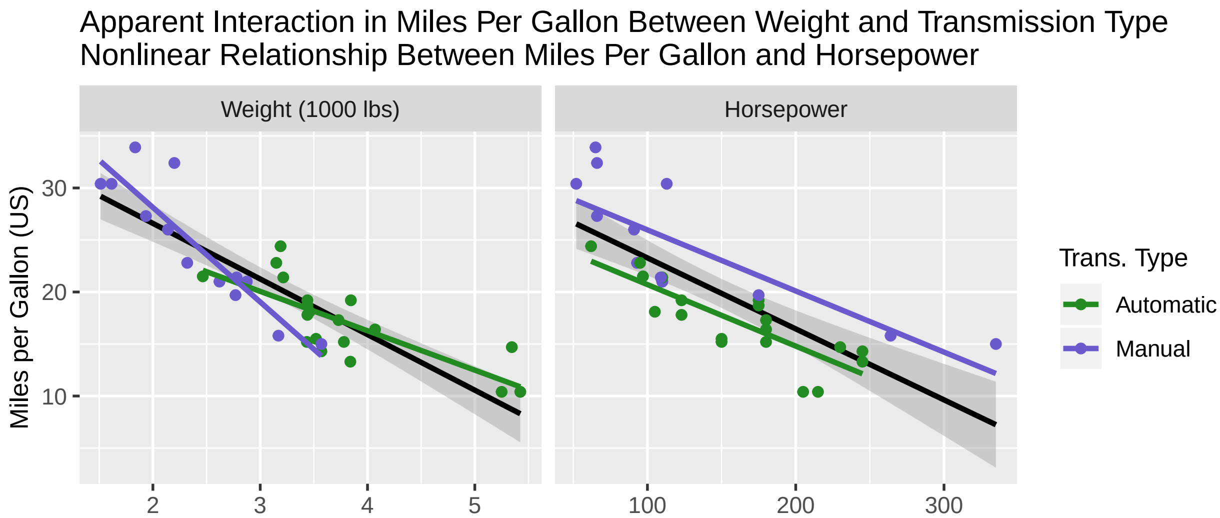
Model Selection ANOVA Table

```
## Analysis of Variance Table
##
## Model 1: mpg ~ 1
## Model 2: mpg ~ amf
## Model 3: mpg ~ amf * wtc
## Model 4: mpg ~ amf * wtc + cyl
## Model 5: mpg ~ amf * wtc + cyl + hpc
## Model 6: mpg ~ amf * wtc + cyl + poly(hpc, 2)
##   Res.Df     RSS Df Sum of Sq       F    Pr(>F)
## 1     31 1126.05
## 2     30  720.90  1    405.15 91.9765 7.406e-10 ***
## 3     28  188.01  2    532.89 60.4877 2.635e-10 ***
## 4     27  138.51  1     49.50 11.2379  0.002552 **
## 5     26  133.93  1      4.58  1.0391  0.317796
## 6     25  110.12  1     23.81  5.4042  0.028504 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

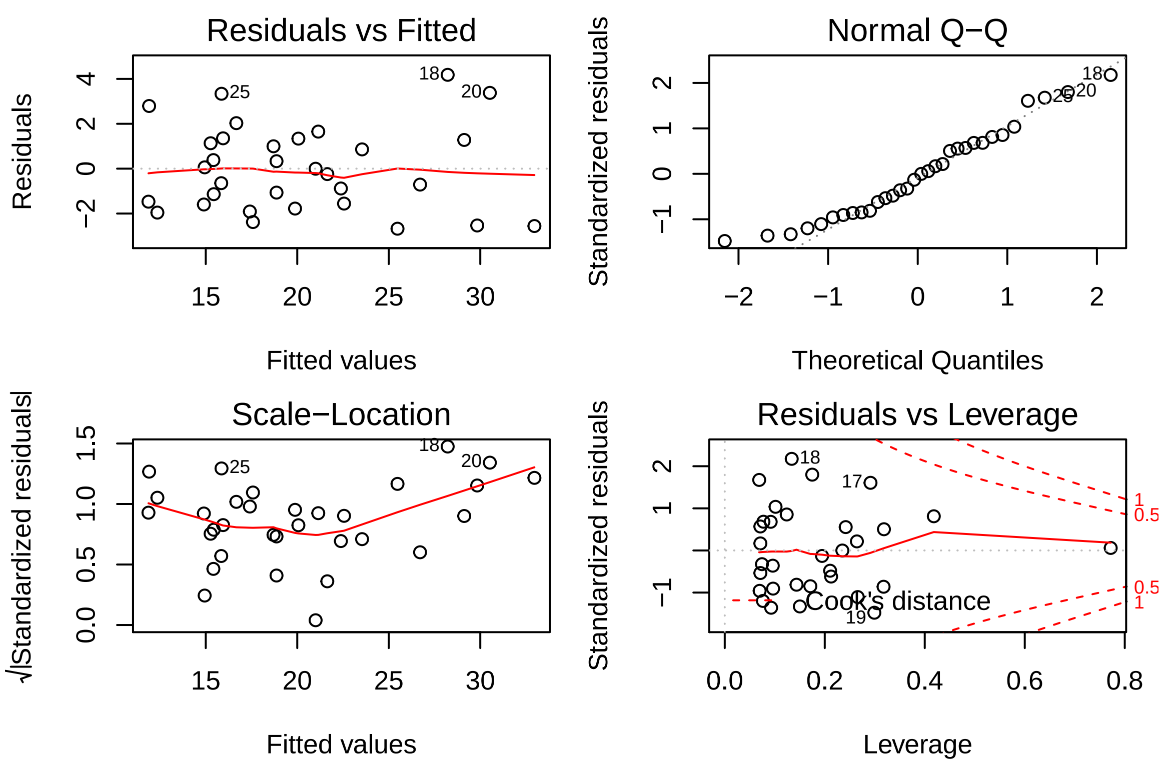
Model Coefficient Summary Information

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	19.212570	0.6060148	31.7031391	0.0000000
(Transmission Type: Manual)	-1.138889	1.3976619	-0.8148528	0.4225581
(Weight)	-1.998504	0.7691197	-2.5984306	0.0152241
(Horsepower; Linear)	-11.977209	3.3637435	-3.5606785	0.0014534
(Horsepower; Quadratic)	7.251283	2.5016037	2.8986536	0.0075163
(Transmission Type: Manual) (Weight)	-4.093339	1.2904941	-3.1719164	0.0038621

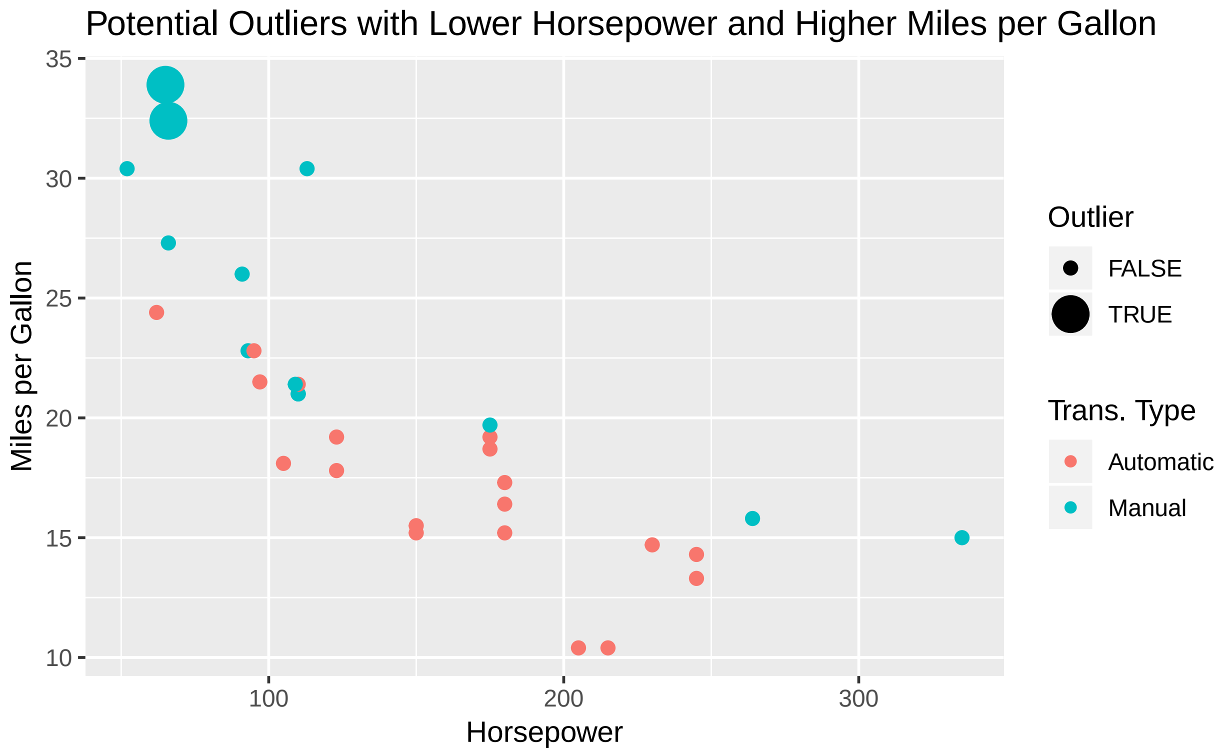
Supporting Figures



Exploratory Plots



Model Diagnostic Plots



Outlier Identification from Diagnostic Plots

Setup Code

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(knitr)
library(kableExtra)
library(stringr)
data(mtcars)
mtcars <- mtcars %>%
    mutate(name = rownames(mtcars),
           wtc = wt - mean(wt),
           hpc = hp - mean(hp),
           amf = factor(ifelse(am == 0, "Automatic", "Manual")))
```