The Impact that the Type of Transmission has on Fuel Efficiency

Trenton Potgieter

Saturday, April 18, 2015

Executive Summary

One of the key criteria in any vehicle purchasing decision is fuel economy, or how many Miles per Gallon (MPG) the vehicle gets. There are various additional factors that need to be taken into consideration when assessing the effects on fuel-economy. Among these is the type of car transmission the vehicle is equipped with. Which of the two types of transmissions (**manual** or **automatic**) impacts fuel economy more? This report will answer this question by analyzing what impact each type of transmission has well as a multitude of other factors has Miles per Gallon (MPG).

Initial Analysis and Data Transformations

The data for this report comes 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). The data is made up of **32** separate motor vehicles, with **11** different features. Since the focus of this report is on the MPG and Transmission type, there are **19** cars with **Automatic** transmissions and **13** with **Manual** transmissions.

The plot in Appendix A: Box Plot shows the distribution of the transmission types for the various vehicles and their respective MPG. As can be seen, the vehicles with a Manual transmission have an average of **24.39** MPG, while the vehicles with an Automatic transmission have an lower average of **17.15** MPG. Since there are other variables that can influence this conclusion, the rest of this report will quantify these more.

Since there are no missing values or outliers within the data set, but there are a number of variables that have a number of categories. So to prepare the data for a better analysis, we factorize the variables into their individual categories and provide better names for the type of transmission.

Regression Analysis and Model Selection

We run an initial regression analysis to answer the original question of just how closely correlated the type of transmission is to MPG.

```
Call:
lm(formula = mpg ~ am, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-9.3923 -3.0923 -0.2974  3.2439  9.5077

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   17.147      1.125  15.247 1.13e-15 ***
amMan          7.245      1.764   4.106 0.000285 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 4.902 on 30 degrees of freedom
Multiple R-squared:  0.3598,    Adjusted R-squared:  0.3385
F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

As can be seen there is a significant correlation between the **Manual** transmission and **MPG**. But the adjusted R^2 is low (**0.34**) showing that the coefficient is not necessary the best predictor. To find a better fitting model, we make use of Step-wise Regression. The `step()`[1] function in `R` makes use of a search algorithm[2] to find the best model. In this case we will use both a Forward selection as well as Backward elimination for model selection.

```
Call:
lm(formula = mpg ~ cyl + hp + wt + am, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-3.9387 -1.2560 -0.4013  1.1253  5.0513

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
cyl6        -3.03134    1.40728  -2.154  0.04068 *
cyl8        -2.16368    2.28425  -0.947  0.35225
hp          -0.03211    0.01369  -2.345  0.02693 *
wt          -2.49683    0.88559  -2.819  0.00908 **
amMan        1.80921    1.39630   1.296  0.20646
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.41 on 26 degrees of freedom
Multiple R-squared:  0.8659,    Adjusted R-squared:  0.8401
F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

The output shows the coefficients that have the most impact on MPG. Using automatic methods provides the best linear function, because the adjusted R^2 for the best model is much higher (**0.84**).

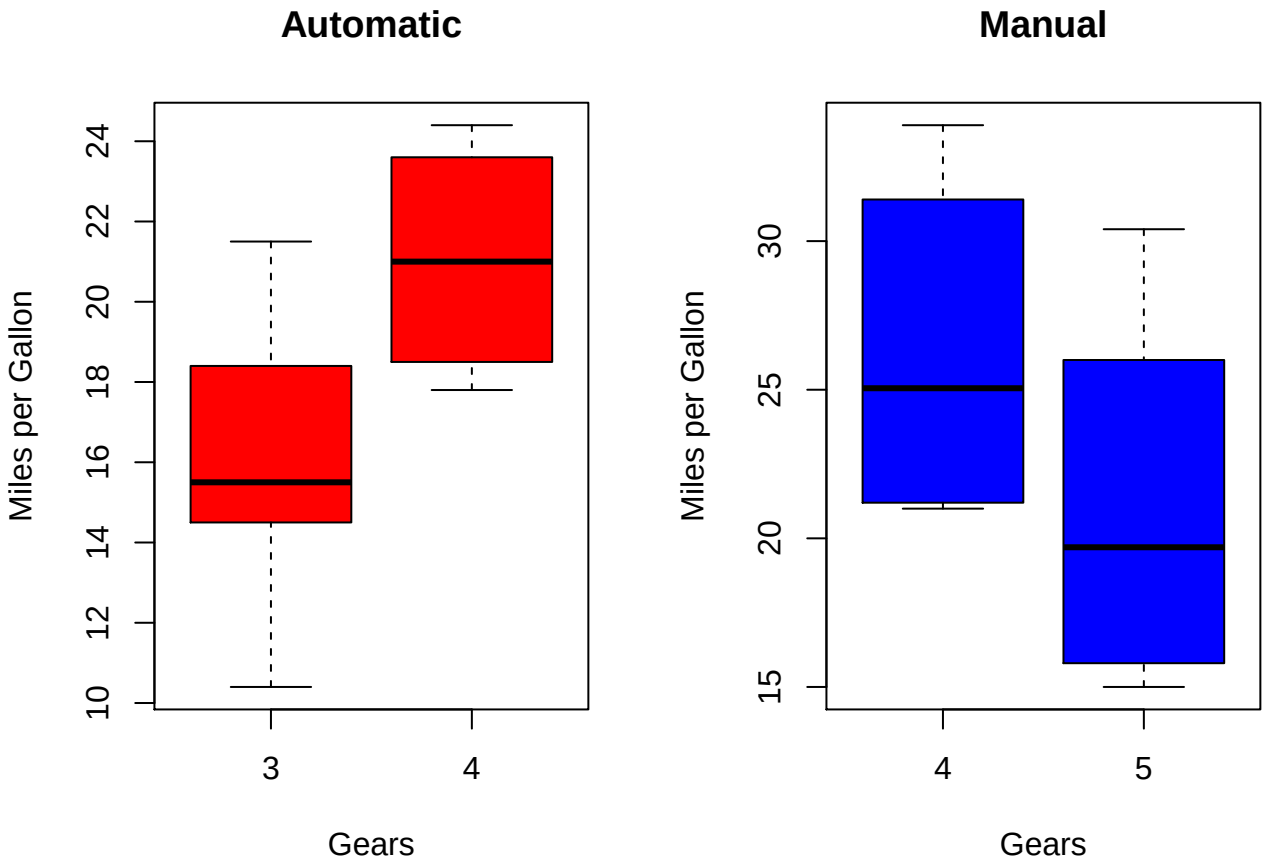
Added to this is the fact that not only are the residuals randomly centered around 0 throughout the range of fitted values (see Appendix B: Residual Plot), but they are well within the 95% confidence intervals of the plot, see Appendix C: Q-Q Plot. To this end we can deduce that `cyl`, `hp`, `wt` and `am` are the best factors affecting `mpg`.

Conclusion

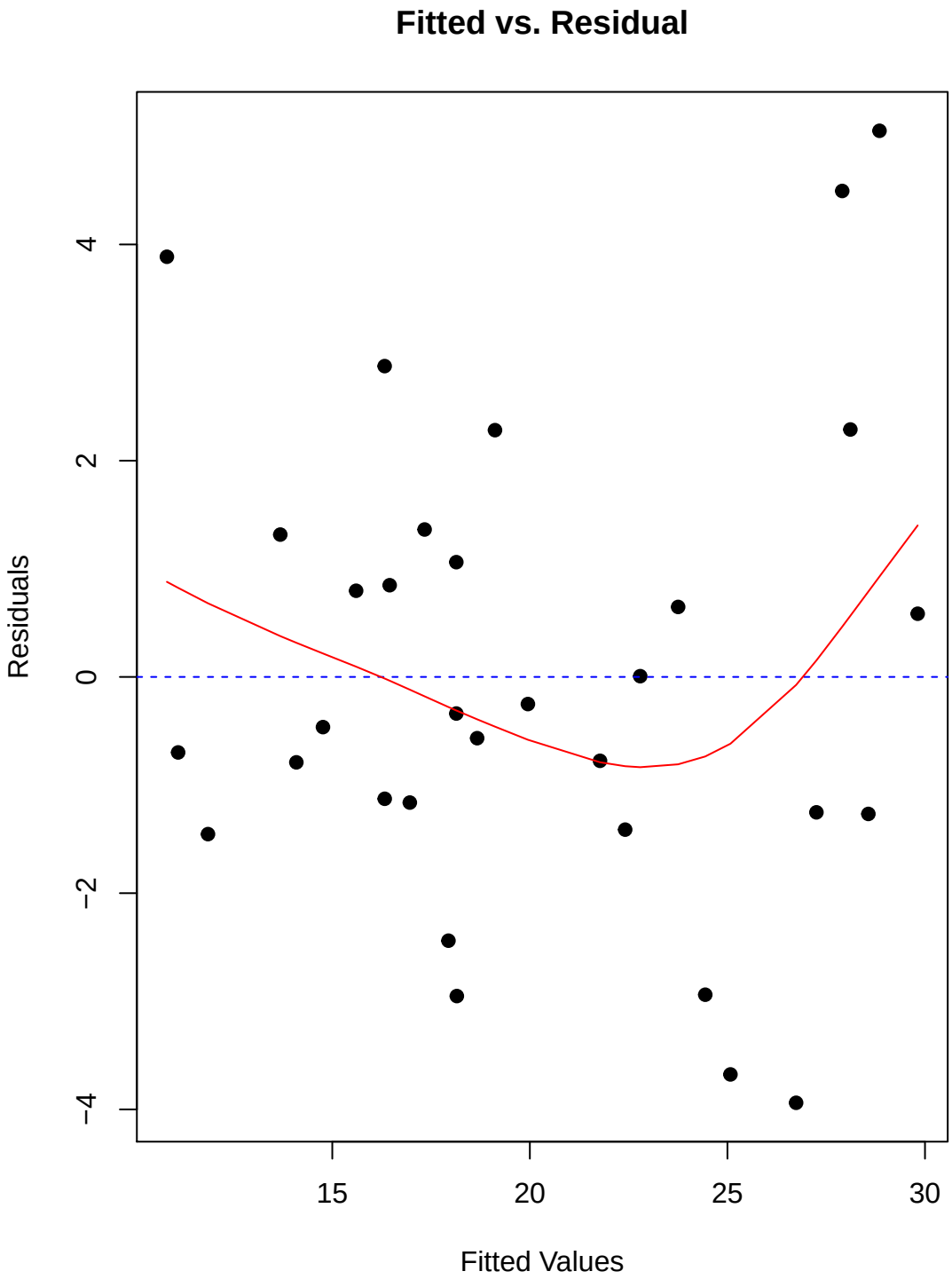
The data shows when making a decision to buy an economical vehicle, one should not only buy a vehicle with a **Manual** transmission, but as this report shows, having **6** or **8** cylinders, the **Horsepower** and the **Weight** of the vehicle must be considered as well.

[1] http://www.stat.columbia.edu/~martin/W2024/R10.pdf
[2] http://web.mit.edu/r_v3.0.1/lib/R/library/stats/html/step.html

Appendix A: Box Plot



Appendix B: Residual Plot



Appendix C: Q-Q Plot