On the Relationship between Automobile Features and Fuel Efficiency

dk mckinney

aug-14

Contents

Abstract

The purpose of this research is to identify any fuel efficiencies as measured by vehicle miles driven per gallon of fuel, based upon which of the two most common transmission types automatic and manual. The first phase of the project involves a screening of the data where I identify the extent of missing, inconsistent or anomalous values. The next phase was the preparation of visualization plots. The next phase was the creation of regression models to increase my understanding of the relationships between the different attributes. The final phase involved plotting residuals and performing model diagnostics and other post-hoc testing.

Executive Summary

The final regression model captured 89.46 percent of the total variance and 87.9 percent of the adjusted variance. The coefficents indicate that when the weight of the vehicle increases by 1,000 lbs, the mpg decreases by 3.176 for cars with automatic transmissions, and decreases by 6.099 for cars with manual transmissions. When the quarter-mile performance time increases by 1 second, the mpg increases 0.834 for cars with automatic transmissions and 1.446 for cars with manual transmissions. The answer to the question of whether an automatic or manual transmission is better for MPG would depend on the particular vehicle. The heavier the vehicle, the greater the MPG penalty for having a manual transmission. The quicker the vehicle, the greater the MPG advantage for the manual transmission. The conclusion of this analysis in one sentence is that the MPG variable is largely influenced by the interaction between weight, acceleration and transmission type. T-Testing indicated that the mean difference in MPG between the two transmission types is 7.245

```
##                            Coef.   S.E.  t val     p val
## Automatic:Weight          -3.1759 0.6362 -4.992 3.114e-05
## Manual:Weight             -6.0992 0.9685 -6.297 9.704e-07
## Automatic:Quarter_Mile_Time 0.8338 0.2602  3.205 3.458e-03
## Manual:Quarter_Mile_Time   1.4464 0.2692  5.373 1.121e-05
```

Table 1: Ordinary Least-Squares regression output predicting MPG from the interaction of transmission type, vehicle weight and quarter-mile time.

Analysis

This analysis is conducted at the behest of Motor Trend Magazine. Motor Trend is particularly interested in exploring the relationships between miles per gallon and the other feature attributes. Two primary points of interest are:

- Is an automatic or manual transmission better for MPG?
- What is the MPG difference between automatic and manual transmissions?

This analysis will use the mtcars data set to answer their questions using regression models and exploratory data analyses.

Table 2: List of variable names and descriptions.

The first step in this analysis was to load the data set and view the summary statistics. I found the data to be very clean and have no missing values. The cyl, vs, and am attributes were converted to categorial, or factor variables. The magnitude of hp is much higher than the other attributes so it was normalized. The disp attribute will not be normalized because it is highly correlated with cyl and will be dropped from the analysis.

The next step was to explore MPG graphically in isolation, and by relevant groups. These plots are located at section 5 of the analysis section.

T Testing

The next step was to test the null hypothesis that the MPG fuel efficiency is identical for automatic and manual transmission vechicles. I used the Welch test which assumes unequal variance because it is more conservative with respect to rejecting the null hypothesis. The test provided evidence to reject the null hypothesis of equal MPG for both transmission types.

```
##
##  Welch Two Sample t-test
##
## data:  automatic and manual
## t = -3.767, df = 18.33, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.28  -3.21
## sample estimates:
## mean of x mean of y
##     17.15     24.39
```

Table 3: T-test of the hypothesis that there is no MPG difference between types of transmissions.

Regression Analysis

My regression strategy was to begin with all attributes, then keep removing the non-factor attribute with the highest p-value until the adjusted R-squared value drops, or no removable attributes remain. I will then compare my best model with an auto-generated model. The best performing model using this strategy uses Transmission Type, Weight, and Horsepower as predictors. Transmission type is not statistically significant but weight and horsepower are.

```
##              Coef.  S.E.  t val     p val
## [Intercept] 28.505 3.3713  8.455 3.404e-09
## Manual       2.084 1.3764  1.514 1.413e-01
## Weight      -2.879 0.9050 -3.181 3.574e-03
## Horsepower  -2.570 0.6586 -3.902 5.464e-04
```

Table 4: Preliminary regression model showing transmission type not significant.

The best model created using a stepwise algorithm uses Weight and Quarter-Mile Time conditioned on Transmission Type.

```
##                               Coef.   S.E.  t val     p val
## [Intercept]                 13.9692 5.7756  2.419 2.259e-02
## Automatic:Weight            -3.1759 0.6362 -4.992 3.114e-05
## Manual:Weight               -6.0992 0.9685 -6.297 9.704e-07
## Automatic:Quarter_Mile_Time  0.8338 0.2602  3.205 3.458e-03
## Manual:Quarter_Mile_Time     1.4464 0.2692  5.373 1.121e-05
```

Table 5: Refined model showing interactions are significant among the three attributes.

ANOVA

I will run an ANOVA to compare the two competing models to see if they are significantly different.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + wt + hp
## Model 2: mpg ~ am:wt + am:qsec
##   Res.Df RSS Df Sum of Sq  F  Pr(>F)
## 1     28 180
## 2     27 119  1      61.6 14 0.00087 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 6: ANOVA of model comparision.

With a p-value of 9×10^{-4} , I reject the null hypothesis which states that the stepwise multivariate model is not significantly different from my hand-rolled model. This is the superior model.

Before I report the details of this model, it is important to check the residuals for any signs of non-normality and examine the residuals vs. fitted values plot to look for any signs of heteroskedasticity. These diagnostics are located in section 6 of this analysis.

Exploratory Plots

Plot 1: The histogram of average MPG for the entire sample is multi-modal indicating that the sample is drawn from at least two seperate distributions.

Plot 2: The histogram of average MPG by transmission type (left) and number of cylinders (right). The group means do not appear to be equal.

Residual Plots and Diagnostics

Conclusion

The final model captures 89.46 percent of the total variance and 87.9 percent of the adjusted variance. The coefficents indicate that when the weight of the vehicle increases by 1,000 lbs, the mpg decreases by 3.176 for cars with automatic transmissions, and decreases by 6.099 for cars with manual transmissions. When the quarter-mile performance time increases by 1 second, the mpg increases 0.834 for cars with automatic transmissions and 1.446 for cars with manual transmissions. This implies that if the car has slower acceleration speeds, other things held constant, cars with manual transmissions have larger mpg values. The conclusion of this analysis in one sentence is that the mpg variable is largely determined by the relationship between weight and accelaration in combination with transmission type.

References

Friedrich Leisch. Sweave, part I: Mixing R and Latex. R News, 2(3):28-31, December 2002.