Data Science Specialization: Regression Project

Stephen O'Connell

August 23, 2014

Executive Summary

As editors of the most prestigious automobile magazine, Wagon Wheels, we are always looking for innovative analysis that challenges conventional thinking in the auto industry. In the following analysis we attempt to challenge the conventional wisdom that a manual transmission provides better gas mileage than an automatic, considering all factors. For this analysis we are using the will known mtcars data set provide by the base R Open Source Statistical package. See Appendix A for a description of the data.

Our analysis begins with an exploration of the data set used in the analysis. From there we present a series models to provide a lens into the raw data in an effort to answer the primary question, which transmission is better for gas mileage, a standard or automatic. No analysis is complete until we have summarized our findings.

Exploratory Analysis

The exploratory plot in Appendix B. Pairs Plot shows the relationship between each variable in the data set. From this plot we can see there appears to be a positive correlation, .60, between mileage (mpg) and transmission type (AM), the trend also appears to favor the manual transmission with regard to high mileage. However, the number of cylinders (cyl), engine displacement (disp), and weight (wt) have positive correlation, each over .85.

Model Analysis

An initial model of MPG using all the factors in the raw data does not yield any compiling evidence that transmission type influences the MPG.

Using the correlation has a guide we created the following three models:

1. Model.1 = mpg vs. am, A comparison of mpg to transmission type.

2. Model.2 = mpg vs. am + wt, A comparison of mileage (mpg) to transmission type and weight (wt)
3. Model.3 = mpg vs. am + cyl + wt + disp + hp, A comparison of mileage (mpg) to transmission type, weight (wt), cylinders (cyl), engine displacement (disp), and horsepower (hp).

See the Appendix for model details.

In Model.1 we find that a manual transmission will add over 7 mpg compared to an automatic transmission, and the P value indicates this relationship is significant. However, as auto specialists we know that mileage (mpg) is influenced by other factors. In Model.2 we compare the factor with the highest correlation, weight (wt), with transmission type, to mileage. In this case we find that transmission actually reverses sign when combined with weight (wt) and has no significance with regard to mileage (mpg), given its P value of .98.

However, without considering the elements with the highest correlation to mileage (mpg) our analysis is incomplete. Our Model.3 includes the factors, weight (wt), cylinders (cyl), displacement (disp), and horsepower (hp). In this case the manual transmission does indicate an increase in mileage (mpg), however, in this model the increase is not significant.

Conclusions / Summary

So yes or no, does a manual transmission provide better mileage than an automatic. Maybe. There is some indication that a manual transmission will provide improved mileage, however, this increase is not significant. In our analysis a manual transmission will add up to 1.8 miles per gallon vs an automatic transmission, all this being consisder. The other factors of weight, horsepower, displacement and cylinders are bigger factors in determining mileage.

Appendix

A. Data Description:

Data

Description

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).

Usage

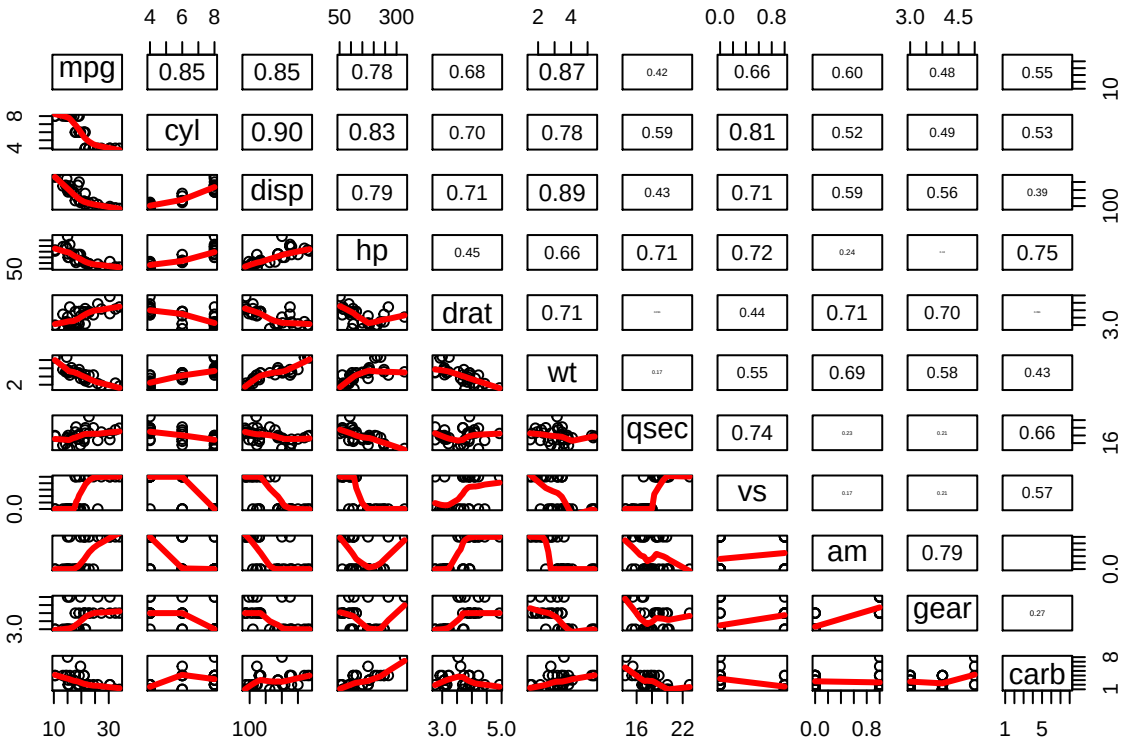
mtcars

Format

A data frame with 32 observations on 11 variables.

1. mpg Miles/(US) gallon
2. cyl Number of cylinders
3. disp Displacement (cu.in.)
4. hp Gross horsepower
5. drat Rear axle ratio
6. wt Weight (lb/1000)
7. qsec 1/4 mile time
8. vs V/S
9. am Transmission (0 = automatic, 1 = manual)
10. gear Number of forward gears
11. carb Number of carburetors

B. Pairs Plot



C. Models

Model 1, mpg vs. transmission type:

```
##
## Call:
## lm(formula = mpg ~ am, data = my.mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## ammanual        7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

Model 2, mpg vs. transmission type + wieght:

```
##
```

```
## Call:
## lm(formula = mpg ~ am + wt, data = my.mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.530 -2.362 -0.132  1.403  6.878
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.3216     3.0546   12.22  5.8e-13 ***
## ammanual     -0.0236     1.5456   -0.02     0.99
## wt           -5.3528     0.7882   -6.79  1.9e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.1 on 29 degrees of freedom
## Multiple R-squared:  0.753,  Adjusted R-squared:  0.736
## F-statistic: 44.2 on 2 and 29 DF,  p-value: 1.58e-09
```

Model 3, mpg vs. transmission type + cylindars + wieght + displacment + horsepower:

```
##
## Call:
## lm(formula = mpg ~ am + cyl + wt + disp + hp, data = my.mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -3.94  -1.33  -0.39   1.19   5.08
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.86428    2.69542   12.56  2.7e-12 ***
## ammanual     1.80610    1.42108    1.27    0.215
## cyl6        -3.13607    1.46909   -2.13    0.043 *
## cyl8        -2.71778    2.89815   -0.94    0.357
## wt          -2.73869    1.17598   -2.33    0.028 *
## disp         0.00409    0.01277    0.32    0.751
## hp          -0.03248    0.01398   -2.32    0.029 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.45 on 25 degrees of freedom
## Multiple R-squared:  0.866,  Adjusted R-squared:  0.834
## F-statistic:   27 on 6 and 25 DF,  p-value: 8.86e-10
```