Which is more Eco-Friendly: Manual or Auto?

Serkan Serttop

21 Jun 2014

Executive summary

In this article, we will be examining a question that has intrigued many drivers. Is there a difference in miles per gallon consumed by a car with respect to the fact that it has a manual or automatic transmission? This is the question we will be trying to answer. Our analysis suggests that cars with manual transmission tend to be more efficient than cars with manual transmission, but when the weight of a car reaches a threshold of somewhere around 3.15 - 3.4 thousand pounds, the advantage seems to be reversed and cars with manual transmission become more efficient.

Exploratory Analysis

```
data(mtcars)
str(mtcars)
```

Looking at the structure of the variables, several variables ('am', 'carb', 'cyl', 'gear', 'vs') lend themselves to be converted into factor variables. Although we will also build models on the original dataset.

Looking at the correlation matrix and the pairs' plots (Figure "Pairs" in the appendix), it is quite visible that there are many confounding variables. This is to be expected as many of the features of a car go hand in hand together with other features. Just to point out an example, horse-power is .785 correlated with the number of cylinders, which is hardly surprising. If a car needs more horse-power, its design would most likely require more cylinders.

```
cor(mtcars_org, use="pairwise.complete.obs", method="kendall")
```

Looking at the mpg column of the correlation matrix, transmission variable (am) has one of the smallest (absolute) correlation ratios. It is .469 for transmission variable while for cylinder variable it is -.79, which is the highest absolute value. But this does not necessarily mean that transmission variable is not significant, what it shows that there are other variables that have a lot more impact on the outcome of 'mpg' variable.

Our next analysis will be to build regression models and examine the marginal impact transmission variable affects the miles per gallon usage.

Regression Analysis

R contains a function called "step" that simplifies variable selection process. We will follow the guidelines presented in this pdf link http://www.stat.columbia.edu/~martin/W2024/R10.pdf to carry out our analysis. Results are published in the appendix.

```
null = lm(mpg ~ 1, data = mtcars);
full = lm(mpg ~ ., data = mtcars);

forward <- step(null, scope=list(lower=null, upper=full), direction="forward");
backward <- step(full, data=mtcars, direction="backward");
stepwise <- step(null, scope = list(upper=full), data=mtcars, direction="both");
```

Analysis of the forward and stepwise selection procedure ends up with "wt + cyl + hp + am" variables as significant with an AIC of 61.65, but it must be added that the "am" variable is added at the last step, thanks to an AIC difference of 0.01!

Analysis of the backward selection procedure ends up with the same variables, but this time adding the next most significant variable to the mix lowers the AIC by 0.41. Hence it can be argued that "am" variable has a lot more predictive power compared to the excluded variables.

Same procedure was also followed with the original dataset without variable factoring applied. Forward and stepwise selection procedures chose variables "wt + cyl + hp" with an AIC of 62.66, while the backward selection procedure ended with the variables "wt + qsec + am" with an AIC of 61.31. It is interesting to note that "wt + cyl + hp" are the same variables in our processed dataset that differed from our best model with an AIC of 0.01. We can infer that the factorization of variables seems to improve the significance of the transmission variable. Although "wt + qsec + am" variables selected by the backward selection procedure provides the lowest AIC value, we will prefer the dataset with the factored variables, which more accurately depicts our dataset. But we will run regression models on both variable sets for comparison.

```
model <- lm( mpg ~ wt + cyl + hp + am, data = mtcars )
```

Our chosen (with factored variables) model suggests that manual transmission increases miles per gallon consumed by a car by 1.81 miles. However, the result is not significant. Considering the importance of the weight variable in its impact to mpg, let's see if an interaction term between weight and transmission improves things.

```
model2 <- lm( mpg ~ wt + cyl + hp + am + wt:am, data = mtcars )
```

Now, Adjusted R-squared value increases from .84 to .856 and transmission variable becomes significant. The results are interesting. This suggests that cars with manual transmission drive 9.9 mpg more than cars with auto transmission when weight is zero, similar to the concept of intercept. But as weight increases, this difference lessens until a car reaches 3147 pounds (1000 * 9.8986 / -3.14499), after which threshold cars with automatic transmission seems to have better mpg.

Let's try the same process to the model with unfactored variables and original dataset.

```
model_org <- lm( mpg ~ wt + qsec + am, data = mtcars_org )
```

This model suggests that manual transmission increases miles per gallon consumed by a car by 2.94 miles and the result is significant.

```
model_org2 <- lm( mpg ~ wt + qsec + am + wt:am, data = mtcars_org )
```

Adjusted R-squared value increases from .834 to .88 and transmission variable is significant in both cases. The results are quite similar compared to previous 2 models. This model suggests that cars with manual transmission drive 14.08 mpg more than cars with auto transmission when weight is zero. But as weight increases, this difference lessens until a car reaches 3400 pounds (1000 * 14.079 / -4.141), after which threshold cars with automatic transmission seems to have better mpg. Plotting models in the appendix also show that the second and fourth rows that contain the interaction term has a more linear residuals vs fitted curves, which is a visual indication that the models are better fit when the interaction terms are included.

Conclusion

We examined the dataset from few different points of view, and the results indicate that cars with manual transmission seems to be more eco-friendly than those with automatic transmission. But there also seems to be a weight threshold between 3.15 - 3.4 thousand pounds, after which cars with manual transmission start becoming less efficient than cars with automatic transmission. There may be other reasons than the type of transmission for why this happens, and it may merit further analysis.

Appendix

Details of Variables

Model Summary - Factorized Variables, wt + cyl + hp + am

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 33.70832    2.60489 12.9404 7.733e-13
## wt          -2.49683    0.88559 -2.8194 9.081e-03
## cyl6        -3.03134    1.40728 -2.1540 4.068e-02
## cyl8        -2.16368    2.28425 -0.9472 3.523e-01
## hp          -0.03211    0.01369 -2.3450 2.693e-02
## ammanual     1.80921    1.39630  1.2957 2.065e-01
```

Model Summary - Factorized Variables, wt + cyl + hp + am + wt:am

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 30.65247    2.90991  10.534 1.111e-10
## wt          -2.20687    0.85205  -2.590 1.578e-02
## cyl6        -2.38062    1.37364  -1.733 9.540e-02
## cyl8        -2.89911    2.19666  -1.320 1.989e-01
## hp          -0.01782    0.01484  -1.200 2.412e-01
## ammanual     9.89860    4.28571   2.310 2.945e-02
## wt:ammanual -3.14499    1.58475  -1.985 5.828e-02
```

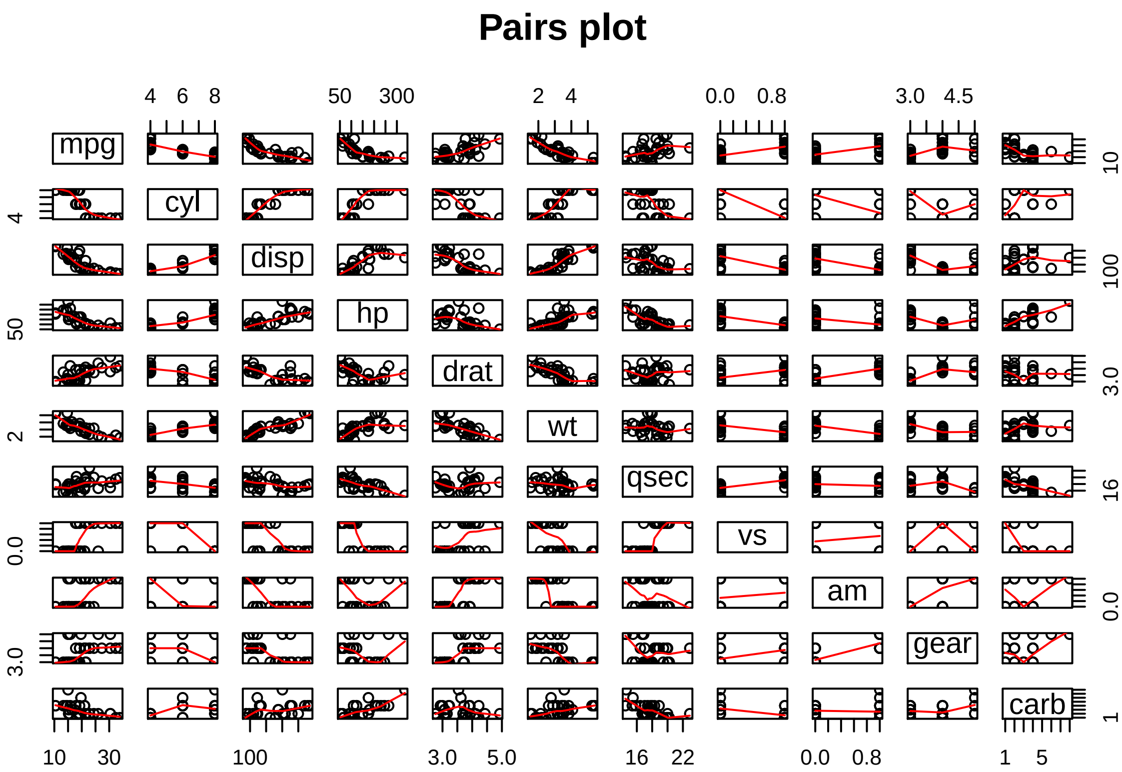
Model Summary - Original Dataset, wt + qsec + am

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    9.618     6.9596   1.382 1.779e-01
## wt            -3.917     0.7112  -5.507 6.953e-06
## qsec           1.226     0.2887   4.247 2.162e-04
## am             2.936     1.4109   2.081 4.672e-02
```

Model Summary - Original Dataset, wt + qsec + am + wt:am

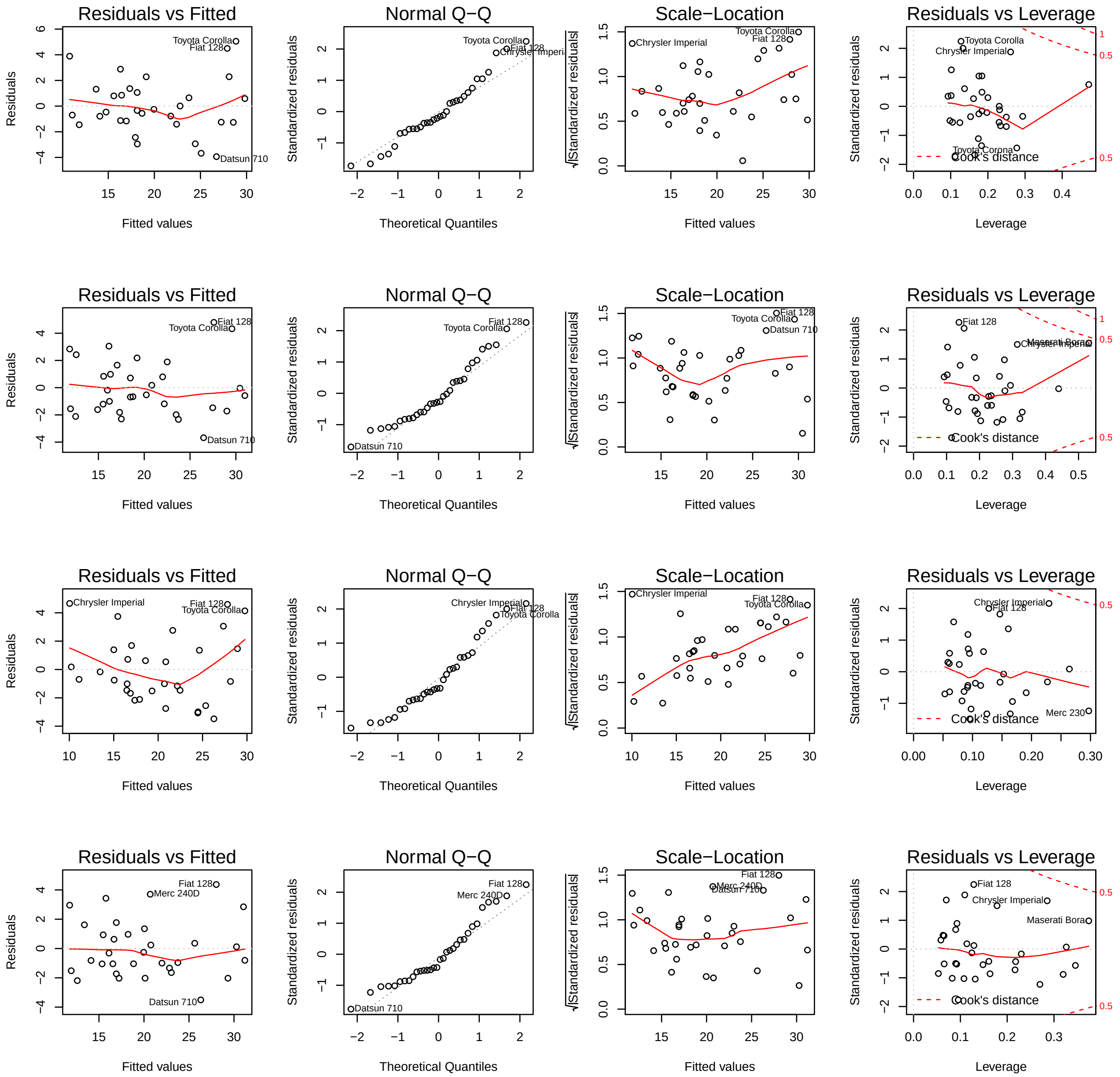
```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)    9.723      5.899   1.648 0.1108925
## wt            -2.937      0.666  -4.409 0.0001489
## qsec           1.017      0.252   4.035 0.0004030
## am            14.079      3.435   4.099 0.0003409
## wt:am         -4.141      1.197  -3.460 0.0018086
```

Figure - Pairs



Plotting Models

```
par(mfrow = c(4, 4));
plot(model);plot(model2);plot(model_org);plot(model_org2);
```


Residuals vs Fitted
Normal Q−Q
Scale−Location
Residuals vs Leverage
Residuals
Fitted values
Standardized residuals
Theoretical Quantiles
√|Standardized residuals|
Leverage
Cook's distance
Toyota Corolla
Fiat 128
Datsun 710
Chrysler Imperial
Toyota Corona
Maserati Bora
Merc 240D
Merc 230