Effect of Automatic or Manual Transmission on Car Fuel Consumption

Author: Ziv Carthy

Executive Summary

The magazine AutoTrends US created a dataset describing 32 automabiles according to 10 attributes. Bassed on this data set we were asked in investigate whether transmission type has an effect on fuel consumption, and by how much.

Based on linear regression and ANOVA analyses of several model alternatives, we have fitted the followoing model: Fues conumption("mpg") ~ transmission type ("am") + weight ("wt"). Our regression analysis of the model showed that *transmission type does not have an effect on fuel consumption*. In contrast, the weight attribute has very significant effect, whereby an increase of 1000 pound in weight reduces mpg by 5.35. Weight seem to be the key factor for predicting mpg, while other attributes might have effect, but to a lesser extent. As said, whether it is an automatic or manual transmission does not have an effect.

The Reasearch Questions

Based on 1974 **Moto Trends US** dataset we where asked two related questions: 1. Is an automatic or manual transmission better for MPG ?
2. What is the quantitative difference in MPG difference between automatic and manual transmissions.

Exploratory Data Analsis

While the questions revolve around the type of transmission, the data includes several other variables that may influence the fuel consumption, such as weight, engine displacement, number of engin cylinders, number of carburators, etc. Chart 1 in the Appendix provides initial look at how Mile Per Galon (mpg) is effected by these variables, and how these variables relate to each others.

What emerges from this chart is that transmission (marked as"am") shows *some* obserable impact on fuel consumption, whereby manual transmissin cars show higher "mpg". BUT - relationship seem to be weak, as the points' spread is large and there seem to be a lot of overlap between the categories (0=automatic, 1= manual).

Moreover, it is clear from the chart that *other* variables have strong impact on mpg. Particularily weight("wt") has a clear visual invesre impact: lower "wt" cars have higher "mpg". Similarily, engine size - manifested by engine displacement ("disp"), and number of cylinders ("cyl") - has this inverse effect.

Importantly, transimission type has some interaction with those influencial variables. For example, it shows that in our sample automatic vehicles were havier than the manual ones. Therefore, when we test the impact of transmission on mpg, we have to account for those variables as well!

Regression Modeling

It make sense that weight will impact mpg because intuitively more energy is required to move around higher weight. Also - the larger the engine size it burns more fuel. Causality link is there, so it make a lot of sense to include it. In addition to "wt", we're interested to model engine size, by using displacement (but not cylinders due to redundancy). We'll construct three linear regression models and test which is the best one

to explain fuel consumption. The models are:
model 1: mpg ~ transmission type ("am")
model 2: mpg ~ transmission type ("am") + weight ("wt")
model 3: mpg ~ transmission type ("am") + weight ("wt") + engine displacement ("disp")

Analysis

ANOVA for Model Selection: We compared the three models using ANOVA, and the ouput is shown in Chart 2 in the Appendix.
Based on the ANOVA, *model 2 has the best fit* of the three, because the F statistic for model 2 is high (50), which means that model 2 fits better than 1. In contrast, F statistic for model 3 was not statistically significant. Hence we seleted model 2.

Compate Betas for Model Selection: Chart 3 shows the beta of the transmissin type for each model. In model 1, beta=7.2 with very low p.value. So, model one predicts that cars with manual transmissions drive 7.2 miles more per galon than automatic cars. However, in model 2, we added *weight* as a regressor, the transmission beta becomes -.0.02 with high p.value. In other words - the transmission type variable lost its impact on MPG altogether.

Interpretation of Coefficients - Selected Model Summary of model 2 is shown in Chart 5 below. The beta of the transmissin variable is -0.02 with p.value=0.988, so its not significant both from a scientific and from statistical perspectives. In this model, the transmission variable has no impact on MPG !

The wt variable, on the other hand, has a beta of -5.35 with very low p.value. The model asserts that increase of 1000 pounds in car weight, results in a decrease of 5.35 miles per galon.

Selected Model: Residual Analysis Chart 6 below shows a few residual plots. According to the Residuals v. Fitted we can't detect a pattern that might hint for missing variables or heteroskedasticity. The residuals are normal, based on the NOrmal Q-Q plot.

Conclusions

Based on our statistical models we conclude that transmission type *does not* impact fuel consumption in any significant way, whereas the meaningful exmplanation of fuel consumption is done by the vehicle weight.

Appendix - Graphs and Compute Output

Chart 1 - Plot of Multivariable Pairs

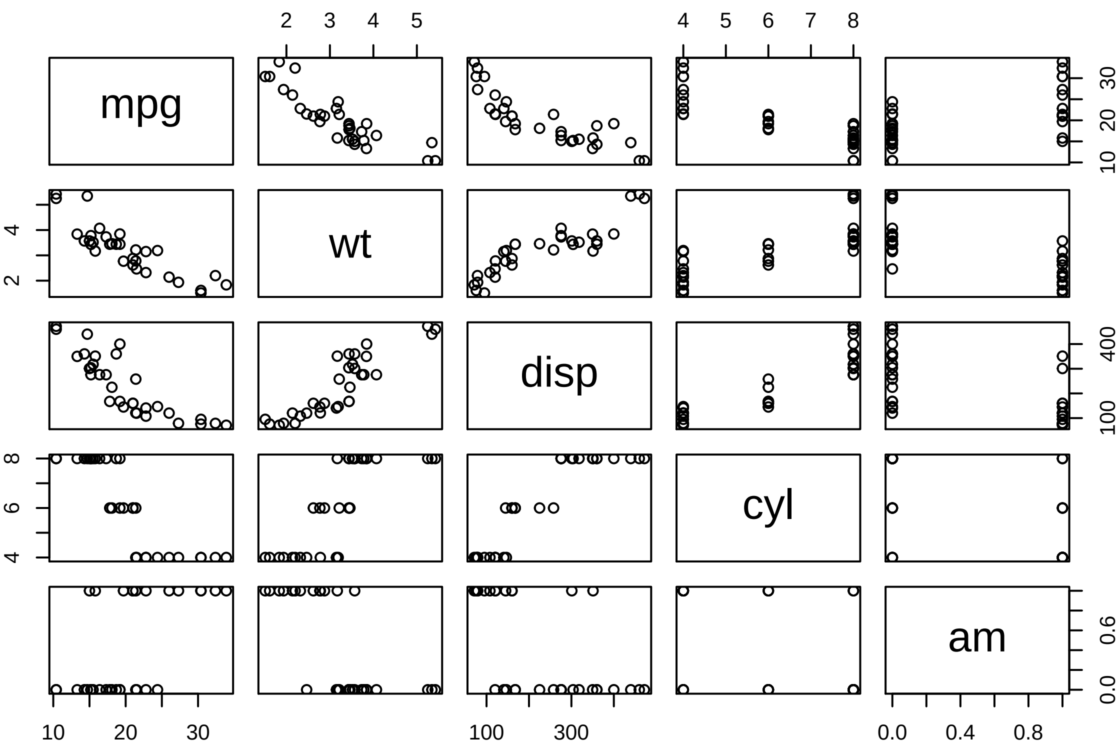


***Chart 2* - Compare Automatic ("0") to Manual("1") Impact on MPG v. Weight**

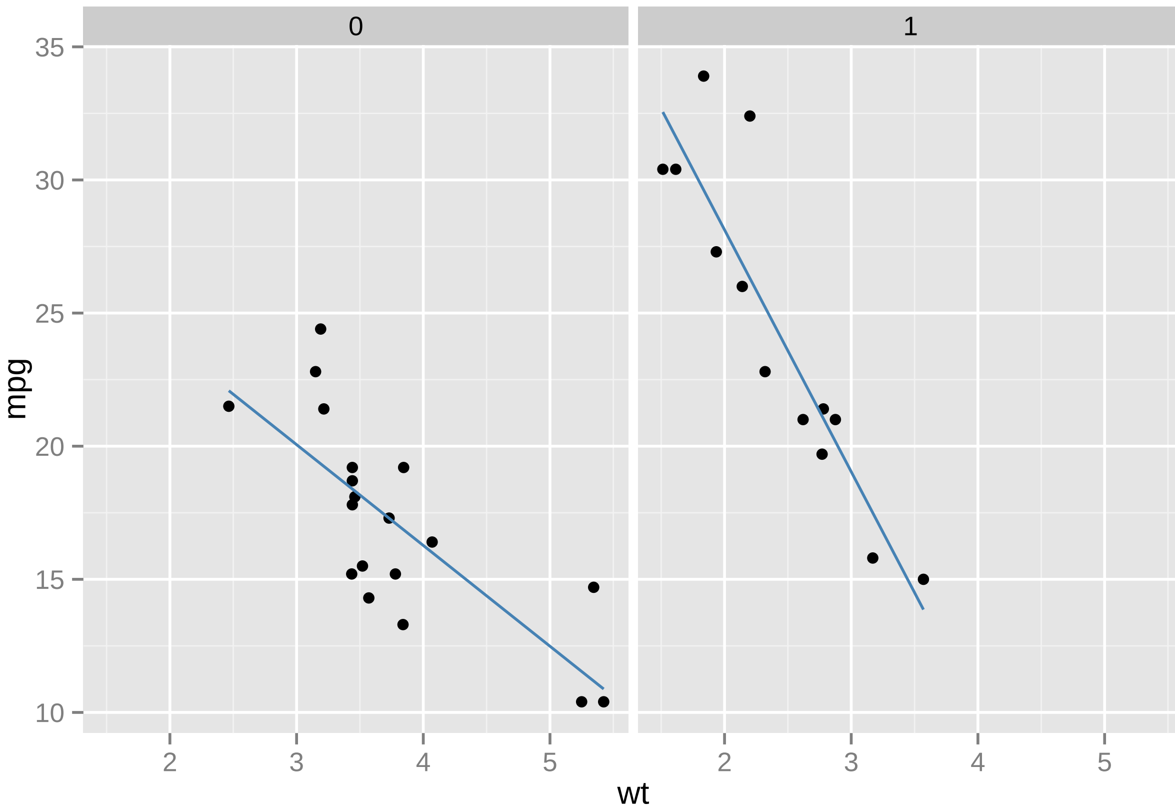


***Chart 3* - ANNOVA to FIT the Right Model**

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt
## Model 3: mpg ~ factor(am) + wt + disp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 50.2610 1.032e-07 ***
## 3     28 246.56  1     31.76  3.6072   0.06788 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

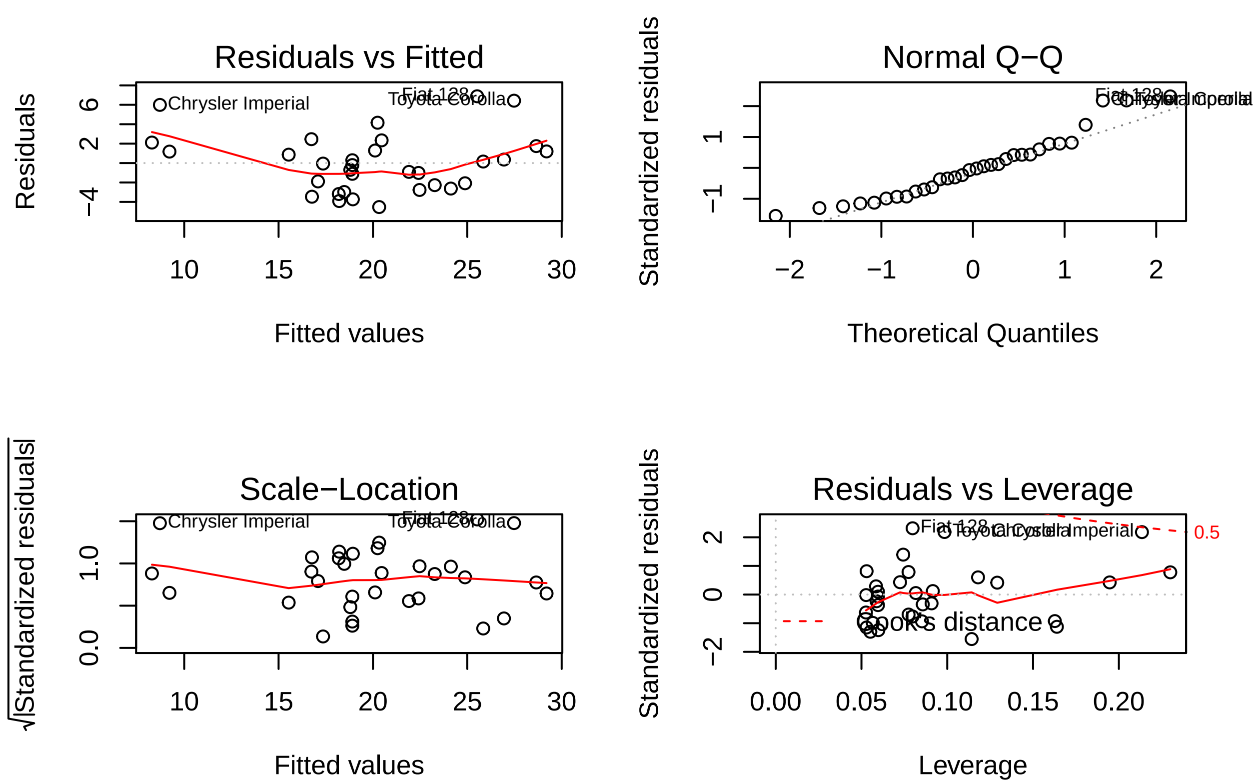
***Chart 4* - Compare Transmission's Impact on MPG ~ WT**

```
##         Estimate Std. Error     t value     Pr(>|t|)
## [1,]  7.24493927   1.764422  4.10612698 0.0002850207
## [2,] -0.02361522   1.545645 -0.01527855 0.9879145855
```

Chart 5 - Regression Summary of the Fitted model

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.32155    3.05464  12.218 5.84e-13 ***
## factor(am)1 -0.02362    1.54565  -0.015    0.988
## wt          -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

Chart 6 - Residual Charts



Link to .Rmd File

To get the markdown file that created this .pdf file, please click here