

Modeling Fuel Consumption in ('73-'74) Automobiles: Automatic vs. Manual Transmission.

Mark Susol, mock writer for US Motor Trend Magazine

21 January, 2015

In this analysis, we aim to examine the fuel consumption of 1973–74 model automobiles as it relates to different aspects of automobile design and performance. The main factor around which we will categorize the analysis will be the type of transmission, automatic or manual. Dependency on the remaining design aspects will be further evaluated to look to quantify the differences observed, if any, in fuel consumption between transmission types.

EXPLORATORY DATA ANALYSIS

The data is part of [R's datasets package](#) that 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). The type of transmission, Automatic vs Manual, will be used as a Factor variable in the analysis. We are interested in modeling how the remaining design aspects possibly affect the fuel consumption, characterized by the Miles per Gallon (MPG).

First, a **Variance Inflation Factor (VIF)** test is performed to quantify multicollinearity, and evaluate how these independent variables affect the coefficient estimates. The model will be fitted using all the independent variables with MPG as the factor variable. The higher results of the VIF should provide for better candidates for forming a model.

##	cyl	disp	hp	drat	wt	qsec	vs
##	15.373833	21.620241	9.832037	3.374620	15.164887	7.527958	4.965873
##	am	gear	carb				
##	4.648487	5.357452	7.908747				

From the results, it shows that independent variables CYL, DISP, HP and WT should be further evaluated for inclusion in our model. For the sake of simplicity in this exercise, CYL & DISP will be dropped in favor of HP as a measure of the “power” to the road of the automobile. Out of being purely curious about this design aspect, QSEC will be considered for regression as well. A plot of the possible independent design aspects WT, HP, & QSEC (**Figure 1**) validates the possible inclusion of these variables for consideration.

Plotting the single regression of MPG on each of the possible variables (**Figure 2**) for modeling shows that in fact we have reason to investigate further. We see that there is a clear difference in fuel consumption for each variable on its own. It is also important to recognize the interaction between the factor variable, transmission type, and the independent variables WT & QSEC. This is shown by the difference in intercepts AND slopes, where as the slope for HP remains the same regardless of transmission type.

MODEL SELECTION

Several model were considered where the independent variables weight (WT), horse power (HP), and quarter mile acceleration (QSEC) were added in succession to the model. The ANOVA test (**Figure 3**) shows that the model that includes WT + HP + QSEC is a significant model for consideration. However the model is only marginally improved by adding QSEC to the model.

It is important to note that from revisiting **Figure 2** and evaluating the slope and intercept as a factor of transmission type, we see that there is an interaction between the factor and the variables (WT & QSEC). So the model will have to include the interaction variables **factor(am) wt** & **factor(am) qsec**, if these independent variables are included in the model.

The correlation between the possible independent variables are presented below. QSEC appears to be uncorrelated to either WT or HP. Addition of QSEC to the model will negatively impact the model, as it might be affecting WT & HP's influence on MPG. QSEC will be removed from the suggested regression model.

```
## cor(wt, hp) cor(wt, qsec) cor(hp, qsec)
## 1 0.6587479 -0.1747159 -0.7082234
```

The independent variables WT & HP with interaction variable **factor(am) wt** will be used in the regression model with the transmission type (am) as the factor. A summary of the linear model in the Appendix (**Figure 4**) shows the model stats as well as the coefficients for the model. A final check on the residuals (**Figure 5**) shows that we have a normal distribution and non-Heteroskedacity.

The regression model selected will have the formula:

$$factor(am) + wt + factor(am)wt + hp - 1$$

CONCLUSION

We have concluded that there is in fact a quantifiable difference in fuel consumption between manual and automatic transmissions in the models studied. The model selected was determined to have a $R^2 = 0.9895427$ and a $p - value = 7.6771838 \times 10^{-26}$ which suggests that it accounts for the significant variance in the data. Which transmission is better for fuel consumption, depends on the model and specific design aspects involved.

The model can be presented as:

$$E[mpg]Automatic = 30.95 + -2.52 * wt + -0.03 * hp$$

$$E[mpg]Manual = 42.5 + -6.52 * wt + -0.03 * hp$$

We can infer from this model that:

- for every 1000 lb increase in weight, we expect a 2.52 (6.52) decrease in MPG for an Automatic (Manual) transmission, holding all other variables constant.
- for every 33.3 increase in hp, we expect a 1.0 decrease in MPG for either transmission, holding all other variables constant.

If you classify the models into **class A:** lighter, sport (high hp) & **class B:** heavier, work (low hp), then the model will predict that:

- **class A:** mpg (wt=2800 lbs, hp=300) = 32.894 Automatic :: 33.244 Manual
- **class B:** mpg (wt=5000 lbs, hp=150) = 22.85 Automatic :: 14.4 Manual

APPENDIX

Figure 1: Plot of possible independent design aspects.

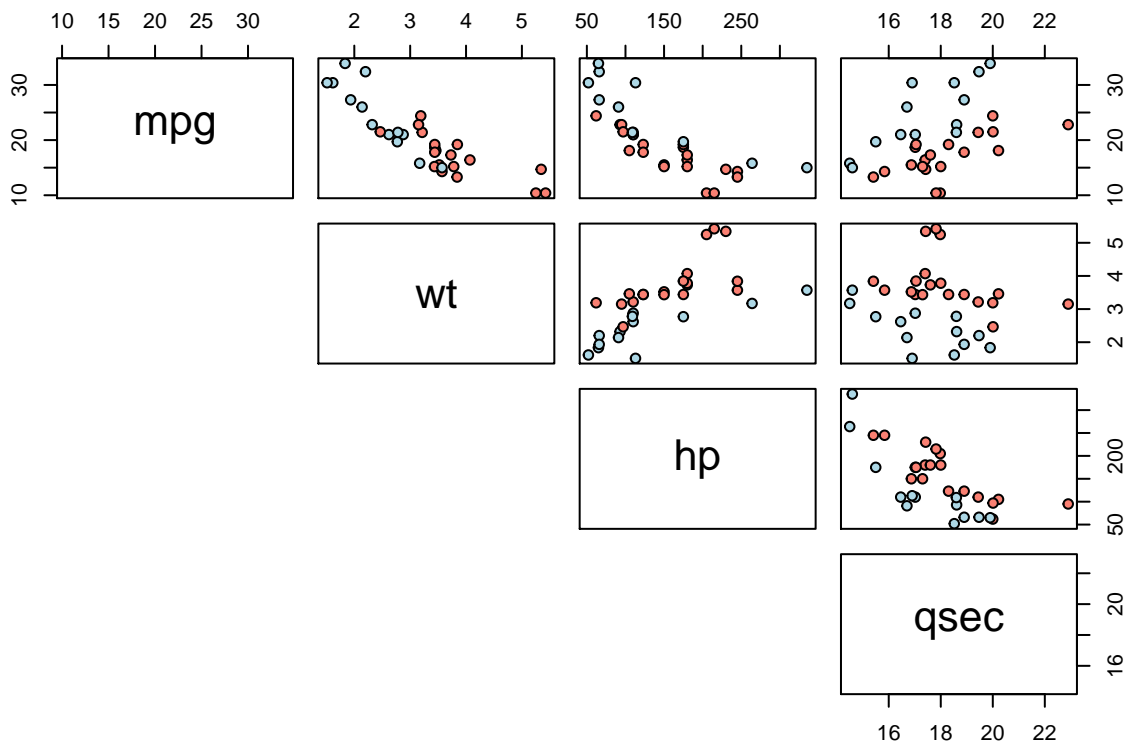


Figure 2: Regression of MPG on selected design aspects.

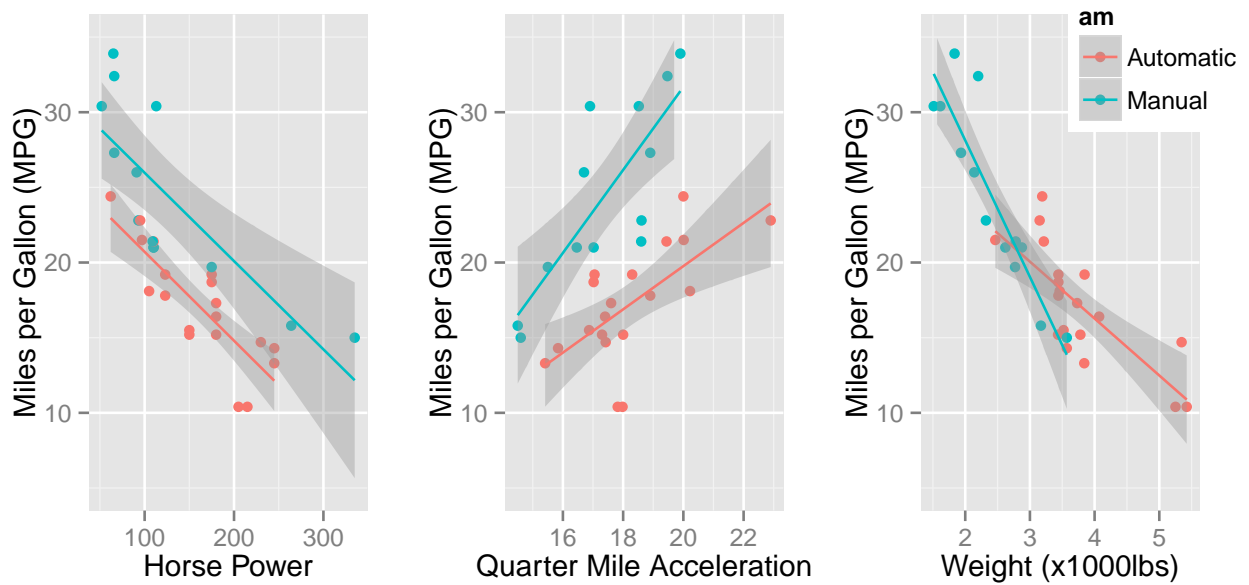


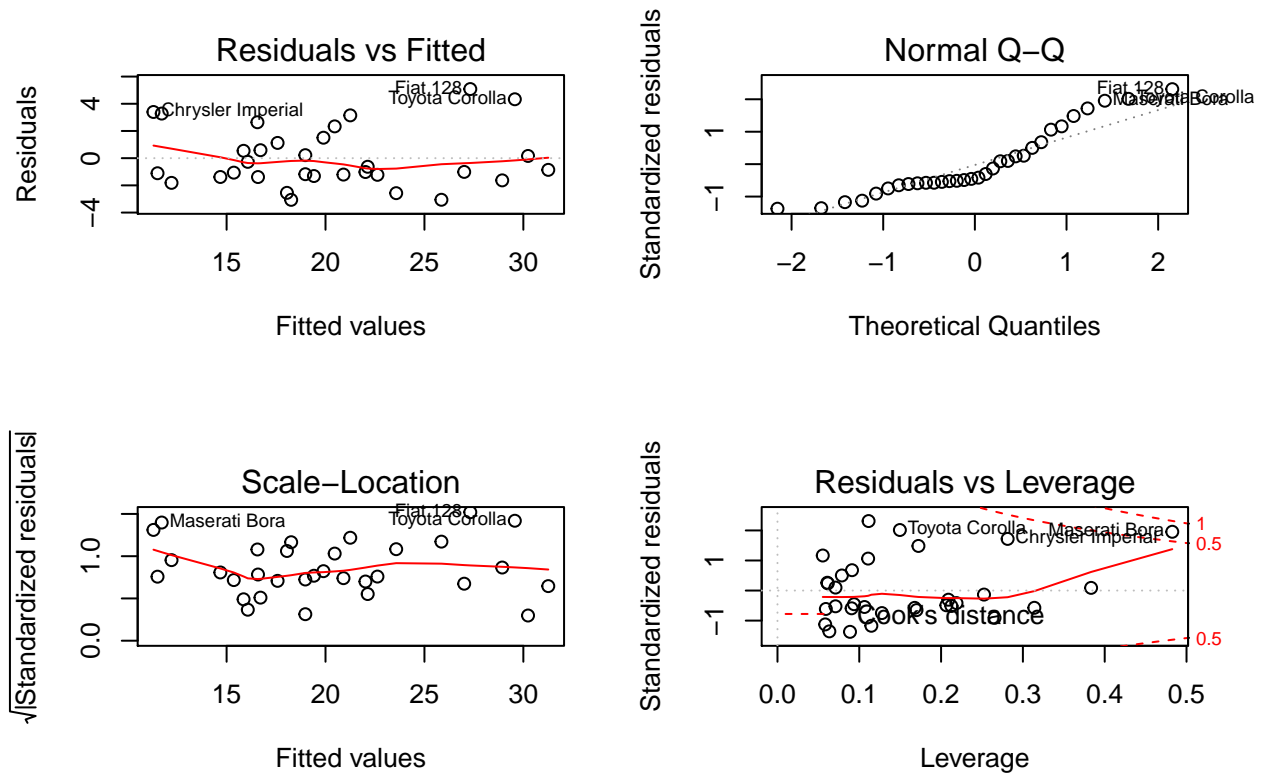
Figure 3: Analysis of Variance in Possible Models.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am) + wt - 1
## Model 2: mpg ~ factor(am) + wt + factor(am):wt - 1
## Model 3: mpg ~ factor(am) + wt + hp + factor(am):wt - 1
## Model 4: mpg ~ factor(am) + wt + hp + qsec + factor(am):wt + factor(am):qsec -
##      1
##      Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1          29 278.32
## 2          28 188.01  1    90.312 19.3850 0.0001752 ***
## 3          27 146.84  1    41.163  8.8353 0.0064510 **
## 4          25 116.47  2    30.374  3.2598 0.0552079 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 4: Summary of our Selected Regression Model.

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt + factor(am) * wt + hp - 1,
##     data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.0639 -1.3315 -0.9347  1.2180  5.0822
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## factor(am)Automatic  30.947333   2.723411  11.363 8.55e-12 ***
## factor(am)Manual     42.502146   3.039657  13.983 6.96e-14 ***
## wt                   -2.515586   0.844497  -2.979  0.00605 **
## hp                    -0.026949   0.009796  -2.751  0.01048 *
## factor(am)Manual:wt  -3.577910   1.442796  -2.480  0.01968 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.332 on 27 degrees of freedom
## Multiple R-squared:  0.9895, Adjusted R-squared:  0.9876
## F-statistic: 511 on 5 and 27 DF, p-value: < 2.2e-16
```

Figure 5: Residuals for our Selected Regression Model.



Source File for this report found at <https://github.com/msusol/JHUIDataScience>

© Mark Susol, completed for a Coursera Class Assignment in [Regression Models](#)