Regression Class Project

Winston Saunders *November 2014*

Executive Summary

This analysis looks at factors affecting the efficiency (**mpg**) of cars in the Motor Trend data set. Particularly, it addresses the hypothesis that manual transmissions are more efficient than automatic transmissions.

The conclusion of the analysis is that we cannot reject the null hypothesis. There is no significan difference, for the cars in this data set, in the miles per gallon of automatic versus manual transmissions.

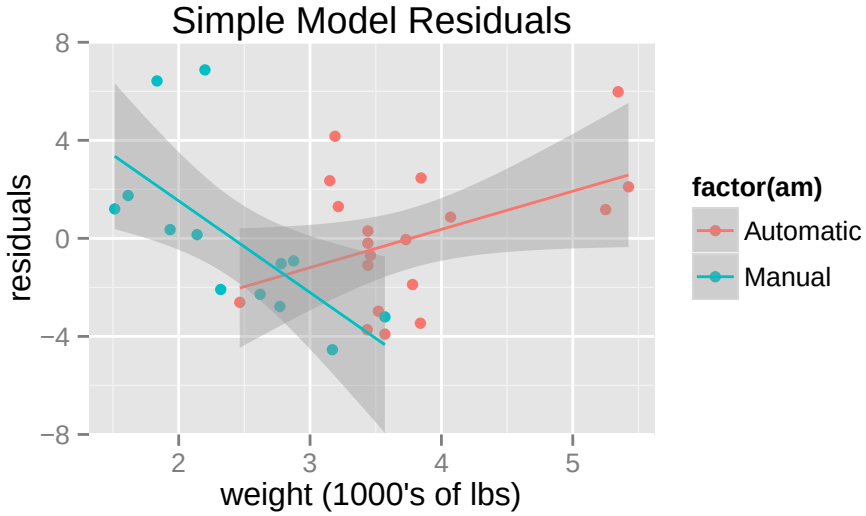
Exploratory data analysis

There are data for 32 cars in the **mtcars** data set representing *mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb*. Of interest here are: **mpg** - Miles/(US) gallon, **disp** - Displacement (cu.in.), **wt** - Weight (lb/1000), **am** - Transmission (0 = automatic, 1 = manual).

A exploratory graph (shown in the Appendix) of the influence of three variables, **wt**, **disp**, and **am** on **mpg** show depenencies on each.

A simple linear model is insufficient

A model of mpg fitted against weight accounts for about 75.3% of the overall variation. However, there are are systematic variations unaccounted for in the model. This can be seen by the below plot of the residuals, factored for automatic versus manual transmission.

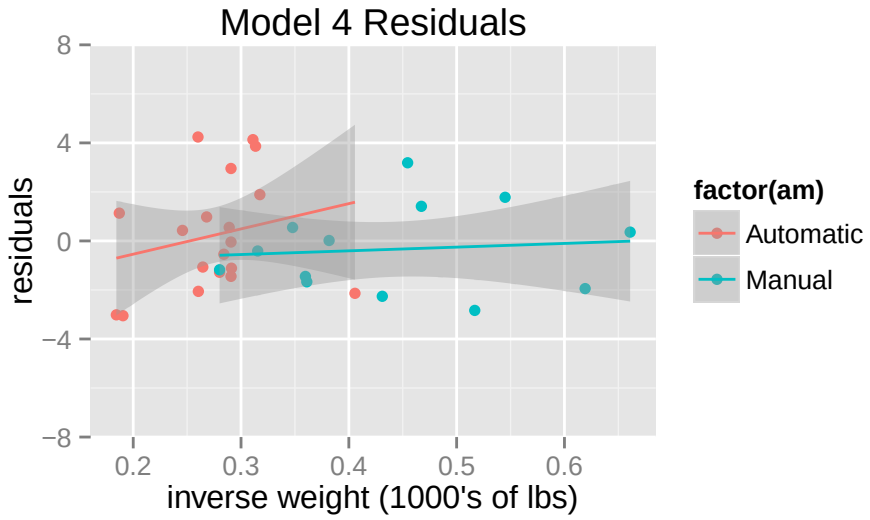


The residuals for manual transmission, with $> 95\%$ confidence, show a systematic dependency not accounted for in the model.

An improved model: including interaction terms and matching the model to physics.

After comparing several models, the model **mpg** ~ **am** + **1/ wt** + **1/disp** improves the analysis with greatest parsimony. The graph below shows that residuals for the model do not show strong systematic variation with **am**.

The inverse relationship on **wt** and **disp** is what one would expect from the "physics" of the problem since both mass and engine displacment would be expected to relate proportionally to energy and fule consumption.



Using "Model 4" below was found to produce the best results with least complexity.

```
f4<-lm(mpg~ factor(am) + invwt+invdisp, data=mtcars2)
```

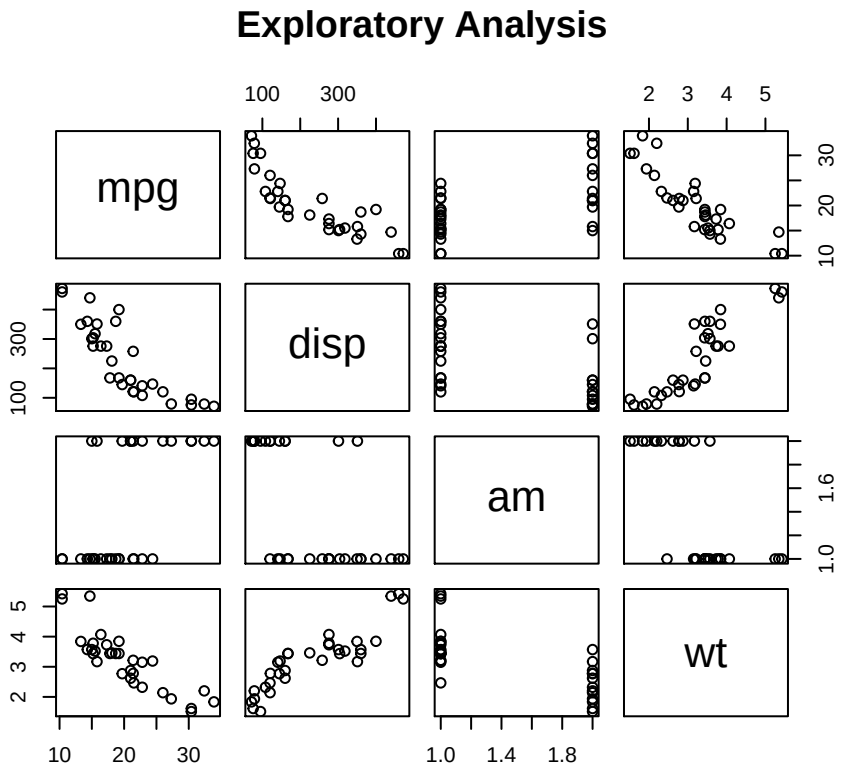
This model accounts for 88.2% of the variation, a signifcant improvement from the simple model above.

Based on the fit coefficients, a manual transmission has -1.16 lower miles per gallon, however, the standard error 1.11 has a t-statistic of -1.04 results in a p-value of 0.31 which is above our tolerance of 0.05. Hence, we reject the null hypothesis.
Per this analysis there is no statistically significant difference in mpg for manual versus automatic transmissions.

Appendices

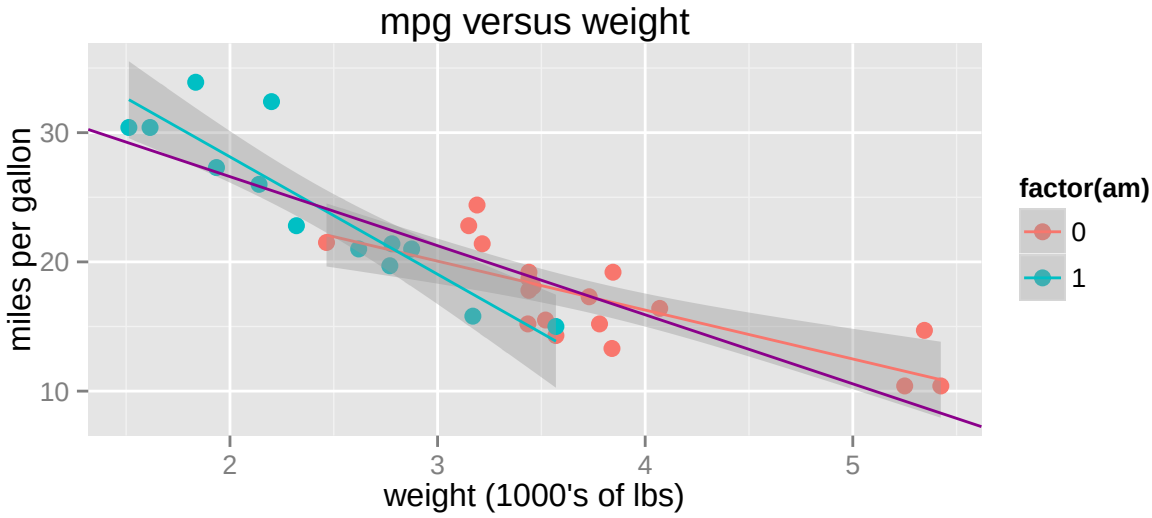
Exploratory analysis

This graph looks at the dependency of mpg on displacement (**disp**), automatic versus manual transmission(**am**), and vehcile weight (**wt**) in 1000's of lbs.



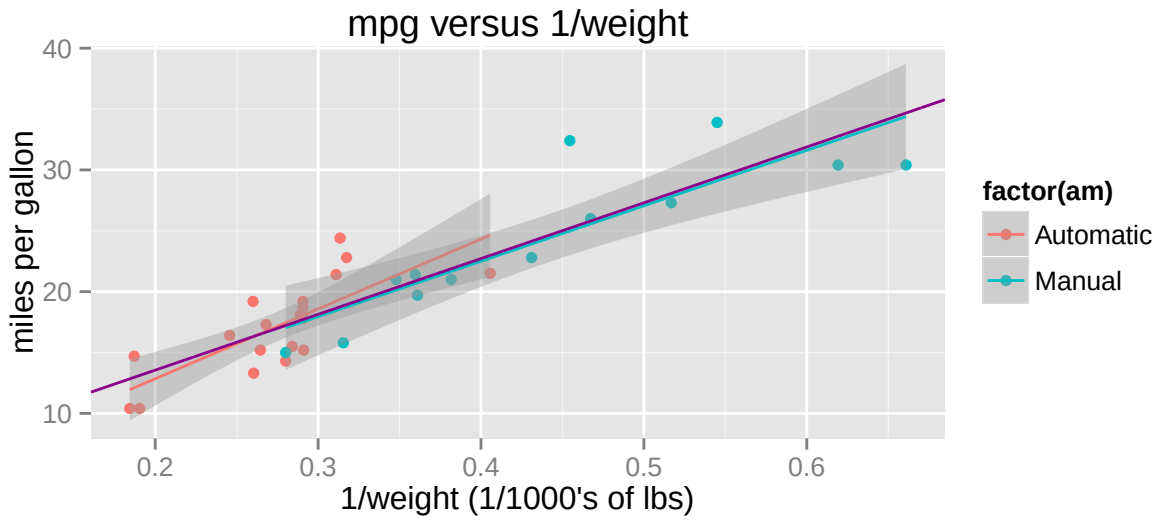
Plot of a linear model of mpg~wt

A fit of mpg versus weight shows a correlation. Fits for different transmission types do not line up well with the fit of the overall data (shown in magenta).



Plot of mpg~inverse weight

A fit of mpg versus inverse weight. Fits of the data for different transmissions types line up better with the overall fit (shown in magenta), showing how an improved "physical" interpretation of the model improves overall fit.



Comparison of the simple linear versus Model 4

The Anova shows Model 4 significantly reduces the variances.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ factor(am) + invwt + invdisp
##   Res.Df    RSS Df Sum of Sq     F    Pr(>F)
## 1     30 278.32
## 2     28 132.58  2    145.74 15.39 3.098e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 4

Summary plot for "Model 4" show the residuals follow a normal distribution to good degree, though some points still could be investigated further.

```
##
## Call:
## lm(formula = mpg ~ factor(am) + invwt + invdisp, data = mtcars2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9946 -1.5665 -0.1164  1.2809  4.0822
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)         7.478      1.604   4.663 6.97e-05 ***
## factor(am)Manual   -1.160      1.114  -1.042 0.306389
## invwt              18.638      8.090   2.304 0.028873 *
## invdisp          1117.183    248.214   4.501 0.000108 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.176 on 28 degrees of freedom
## Multiple R-squared:  0.8823, Adjusted R-squared:  0.8696
## F-statistic: 69.94 on 3 and 28 DF,  p-value: 4.023e-13
```