Regression Models course project

Adrian Martin

March 31, 2016

Executive summary:

While the difference in the mean MPG between automatic and manual transmission cars is about 7.25, we found that about some of that difference is the result of differences in engine size and type, car weight, car gearing, and other factors. Changing a car from an automatic transmission to a manual transmission should increase your MPG between 0.7027569 and 11.3628762 MPG.

Exploratory analysis

What are the mean MPG values for stick shifts and automatic transmissions? Stick shift cars' average MPG is 24.3923077. Automatic transmission cars' average MPG is 17.1473684. So we'll expect to see some sort of MPG increase when moving from an automatic transmission to a manual. However what if those differences are really because manual transmission are in small light Japanese and Italian economy cars with small engines, and automatic transmissions are in big heavy German and American luxury cars? Many of these factors are highly correlated (or highly negatively correlated). For instance, the coorelation between a cars' MPG and its weight is `r cor(mtcars$wt, mtcars$mpg)`. So we'll have to include these factors in our analysis because they might be confounding our analysis.

So what factors should we include in our model to determine MPG? i.e., what factors other than transmission might influence MPG, and might confound our analysis? Data in the mtcars dataset includes cylinders (more cylinders will be correlated with engine displacement which results in lower MPG), displacement (larger displacement results in lower MPG), horsepower (creating more power results in less MPG), rear axle ratio (fewer driveshaft rotations necessary per mile results in higher MPG), weight (more weight means less MPG), quarter-mile time (a function of engine displacement, horsepower and weight, but should not directly impact MPG), "V" versus "straight" engine cylinder layout (should not impact MPG), transmission (the object of our study), number of forward gears (similar to rear axle ratio, fewer RPMs per mile will increase MPG) and number of carburetors (should have no impact on MPG).

We'll need to add interaction terms between transmission and weight, because automatic transmissions are heavier than manual transmissions. The boxplot in the appendix shows the large disparity in weight between automatic and manual tranmission cars.

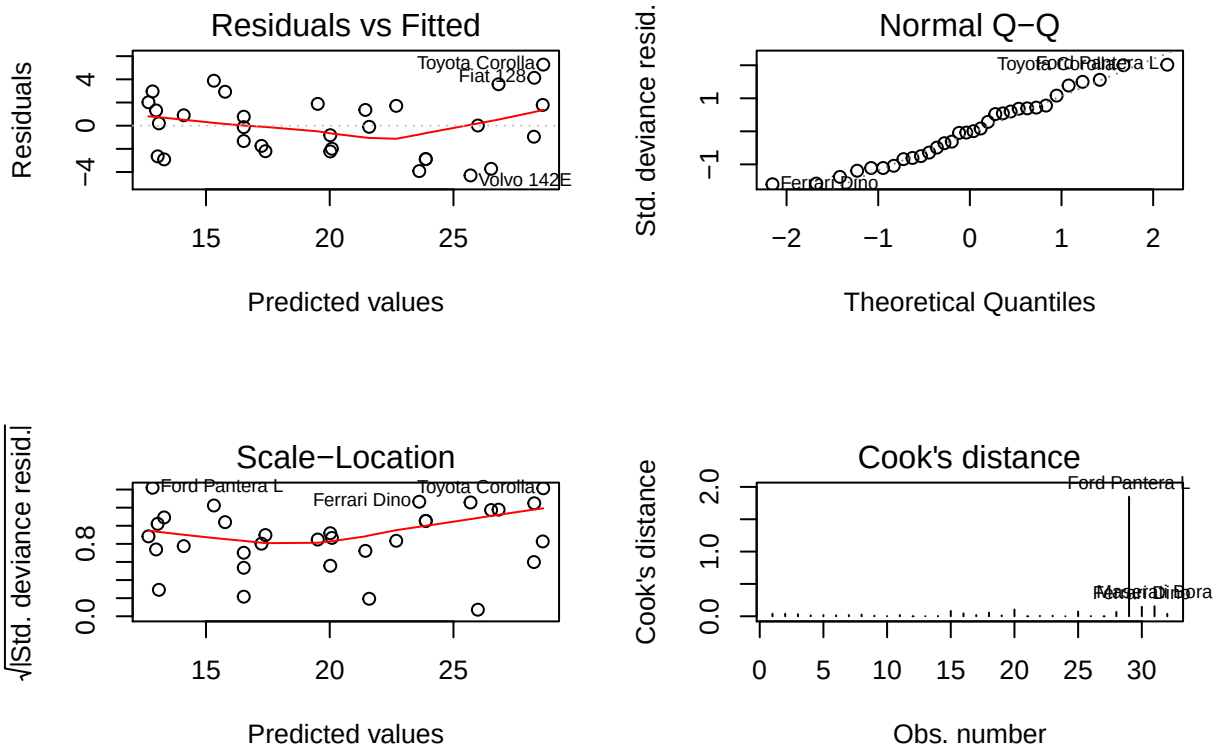
So let's look at how those variables look compared to MPG. The figure is presented in the appendix.

All of the plotted factors appear to have a significant relationship or correlation with MPG. So let's add those variables to our model one by one, and look at how much variance increases. We can also compare the residuals between the models, and whether the coefficient for manual transmission is significant at the .05 level. We'll include interaction terms between the covariates as well. We'll also run one model with no interaction terms. Then we'll look at the variance of the residuals and the significance level for the transmission coefficient, and pick which model to use based on that.

```
##          model residuals variance         significance
## 1       fitInt          0.4079764    0.924278069888958
## 2  fitallnoInt          4.7871560    0.522409454691813
## 3         fit4          3.3543805    0.345187119294808
## 4         fit5          1.7690292     0.23262597297758
## 5        fitwt          5.6909305    0.143554634706011
## 6         fit6          0.4744863    0.141239487582963
## 7      fitgear         18.3871720   0.0690008186507085
```

```
## 8        fithp          7.9172270   0.0602899817221691
## 9         fit3          6.7327517   0.0291859784717124
## 10      fitcyl          7.7115860   0.0181760532676255
## 11     fitdisp          7.6320770  0.00460053245424212
## 12        fit1         23.2547290 0.000285020743935068
```

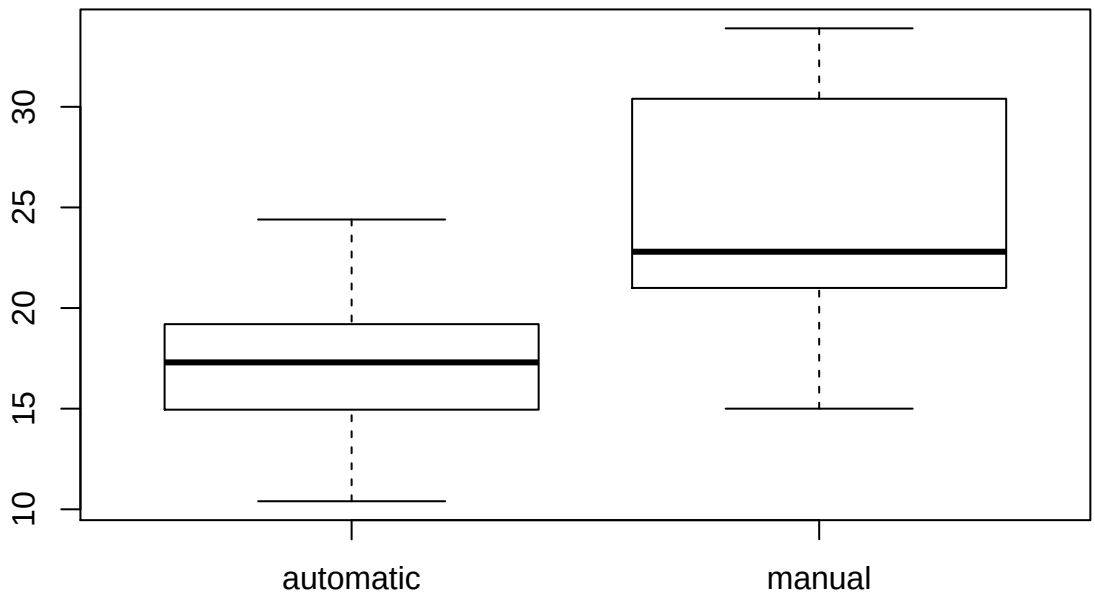
The model "fit3" is the best by our standards. Of the models with significance at under .05, it has the lowest residuals. Let's look at the residuals.

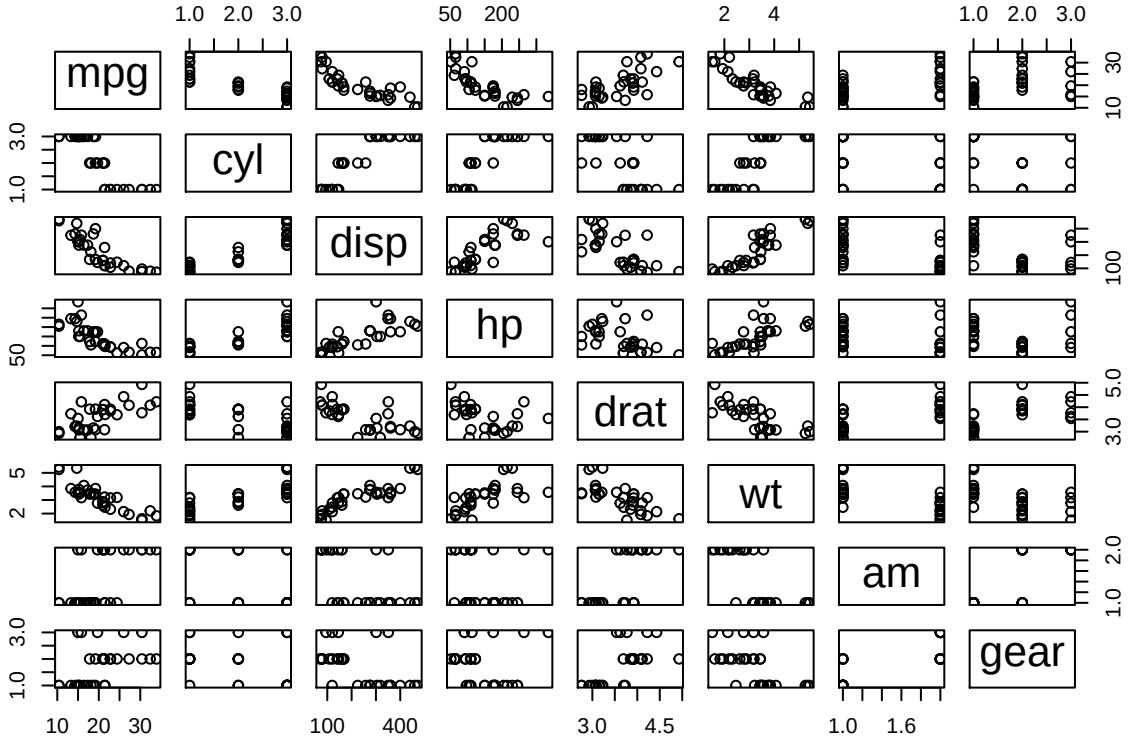


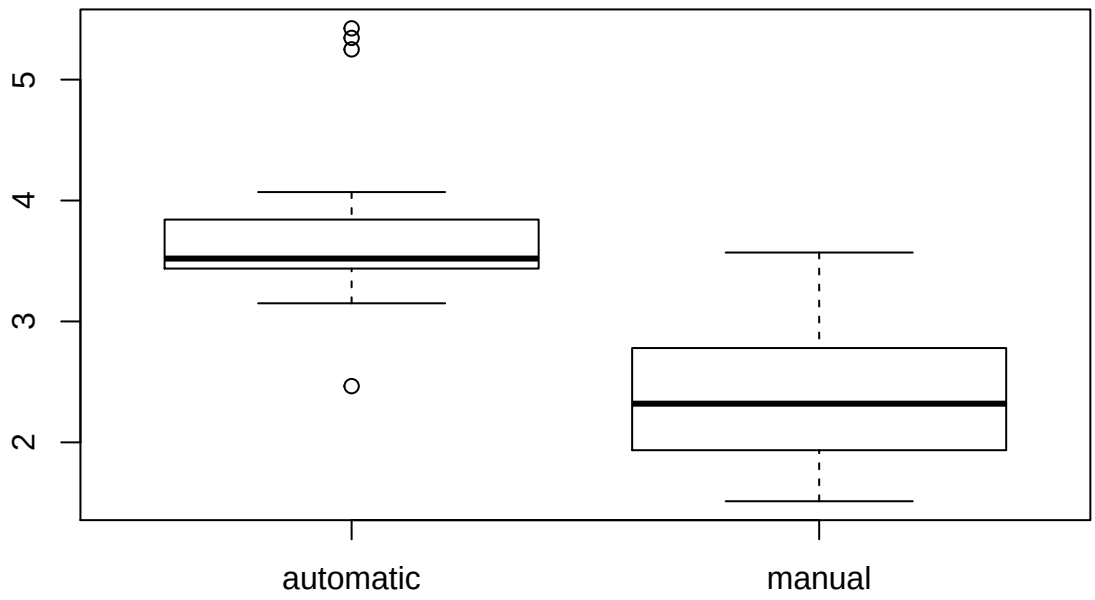
The errors appear normally distributed. The most influential outlier appears to be the Ford Pantera.

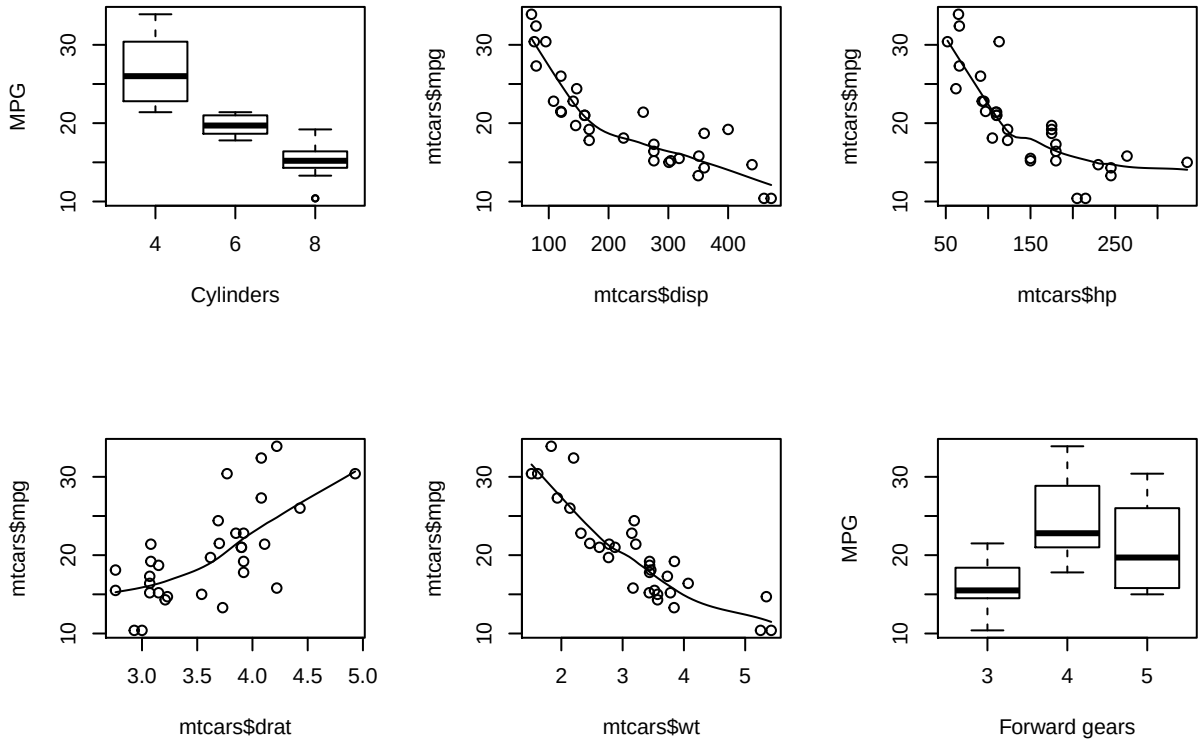
Looking at the factor variable for transmission: automatic, manual shows us that "automatic" is the default factor level. So the model coefficient will look at the change in MPG when moving from automatic transmission to manual. That coefficient is 416.8875519, meaning that the expected effect of moving from an automatic transmission to manual, even when correcting for engine cylinders, displacement, horsepower, rear axle ratio, weight, and number of forward gears, is 416.8875519. Creating a 95% confidence interval for the coefficient, we just take the standard error from the summary of the model (`r summary(fit3)$coefficients["ammanual",2]`) and multiply it by the t quantile. The result is that we can be 95% confidence that the real effect of moving from an automatic transmission to a manual transmission is to gain between 0.7027569 and 11.3628762 MPG.

Appendix


10
15
20
25
30
automatic
manual


mpg
cyl
disp
hp
drat
wt
am
gear


5
4
3
2
automatic
manual


MPG
10
20
30
4
6
8
Cylinders
mtcars$mpg
10
20
30
100
200
300
400
mtcars$disp
mtcars$mpg
10
20
30
50
150
250
mtcars$hp
mtcars$mpg
10
20
30
3.0
3.5
4.0
4.5
5.0
mtcars$drat
mtcars$mpg
10
20
30
2
3
4
5
mtcars$wt
MPG
10
20
30
3
4
5
Forward gears