Is Automatic or Manual Transmission Better for MPG?

Steve Senior

Executive Summary

This project tries to answer the question: "is an automatic or manual transmission better for mpg?" and to quantify the mpg difference between automatic and manual transmissions. It uses the mtcars dataset. Fitting a simple linear model with only the type of transmission (manual versus automatic) as predictor and mpg as the response suggests that there is a statistically significant relationship between type of transmission and mpg (with manual cars getting more miles per gallon). However, when I adjust for the other variables in the dataset, this relationship is no longer statistically significant. So it is not possible to say whether automatic or manual transmission is better for mpg. The expected difference in mpg between automatic and manual cars *holding all other variables constant* is 2.52 mpg (manual cars have the higher mpg).

Exploratory Data Analysis

To begin with it is worth looking at the distribution of mpg whether mpg is different for cars with manual and automatic transmissions. The mean mpg for manual cars is 24.3923 and the mean mpg for automatic cars is 17.1474. Plot 1 in the annex shows a boxplot of miles per gallon for manual and automatic transmission cars. Both the plot and means suggest that manual cars get more miles per gallon than automatics.

However, the mtcars dataset contains 10 other variables, many of which could be related to mpg and which could vary systematically with the type of transmission. A quick way to look at this is to do a pairs plot of the data like plot 2 in the annex. Looking at the top row only, you can see that lots of the other variables appear to correlate with mpg, and vary with the type of transmission. Fortunately, that's why we have regression models!

Fitting Regression Models

To answer the questions above, the response variable is miles per gallon (mpg). MPG is a continuous variable which is bounded from below (you can't have negative mpg, even in a humvee). However the lowest mpg in the data set is `10.4`, so the lower bound doesn't affect the data too badly. As a result, I'm only going to use simple linear regression, rather than any generalised linear model that could better model the constraints on the data. I have also not considered interaction variables: the high number of possible interactions means I'd never fit it all on to two pages.

Model 1: Single Predictor Variable

The simplest model is to use am as a single predictor variable (plus an intercept term).

```
fit1 <- lm(mpg ~ am, data = mtcars)
summary(fit1)$coeff
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.134e-15
## ammanual       7.245      1.764   4.106 2.850e-04
```

This model gives a statistically significant relationship between am and mpg. The intercept term gives the expected mpg for an automatic transmission car (i.e. am = 0 is the reference level). The estimated increase in mpg for manual cars is 7.2449. A 95% confidence interval for this increase is 3.6415, 10.8484. Note that this confidence interval does not include the null value of 0. The p-value associated with the am variable is $2.8502 \times 10\text{-}4$ So *if* we were only considering the am variable, we would reject the null hypothesis that there is no relationship between transmission type and conclude that manual transmissions get higher mpg than automatics. *However*, as noted above, it looks like we need to control for a number of confounding variables like weight, number of cylinders, engine size and so on.

Model 2: All predictor variables

The next step is to include all the other variables in the model as predictor variables:

```r
fit2 <- lm(mpg ~ ., data = mtcars)
summary(fit2)$coeff
```

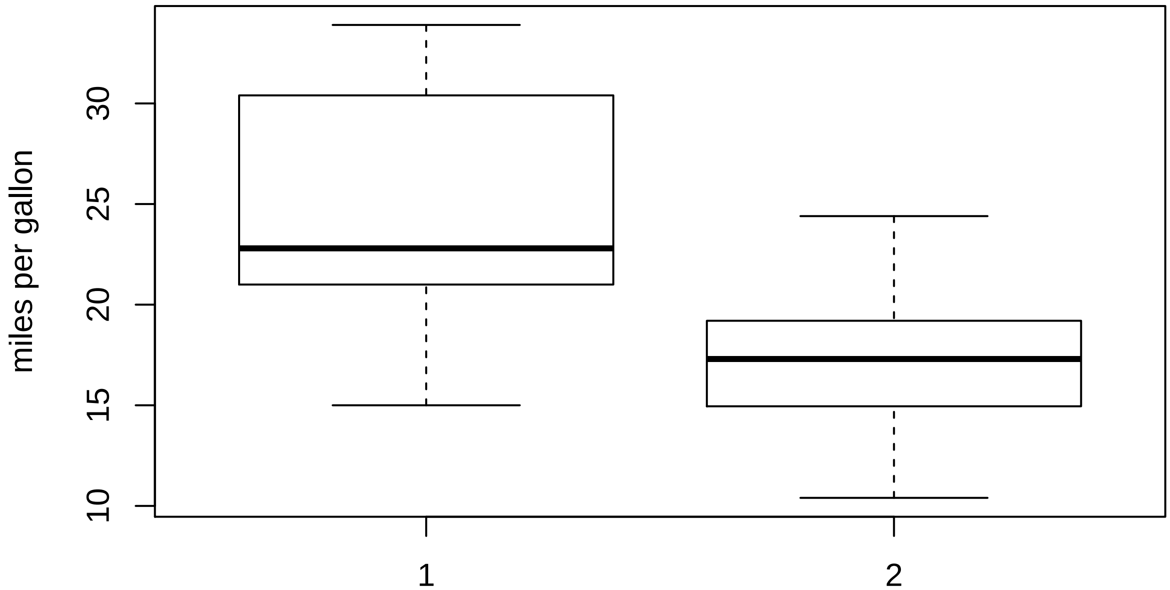
```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788  0.6573  0.51812
## cyl         -0.11144    1.04502 -0.1066  0.91609
## disp         0.01334    0.01786  0.7468  0.46349
## hp          -0.02148    0.02177 -0.9868  0.33496
## drat         0.78711    1.63537  0.4813  0.63528
## wt          -3.71530    1.89441 -1.9612  0.06325
## qsec         0.82104    0.73084  1.1234  0.27394
## vs           0.31776    2.10451  0.1510  0.88142
## ammanual     2.52023    2.05665  1.2254  0.23399
## gear         0.65541    1.49326  0.4389  0.66521
## carb        -0.19942    0.82875 -0.2406  0.81218
```

From this, we can see that *holding the other variables in the dataset constant*, the expected change in mpg between manual and automatic transmissions is 2.5202 (the expected mpg is still higher for manual transmission cars). A 95% confidence interval is -1.7568, 6.7973. Note that the confidence interval now contains the null value of zero and the associated p-value is 0.234, which is greater than the benchmark cut-off of 5%. So by adjusting for all the other variables, we can no longer reject the null hypothesis that 'true' expected change in mpg between automatic and manual transmission cars is zero. As a result, although we can quantify the difference, we can't be certain whether automatic or manual cars are better for mpg. It's interesting to note that the overall model achieves a greater level of signficance, while none of the individual predictor variables achieves the benchmark 5% level of significance.

Plot 3 in the annex gives a series of diagnostic plots for this model. Of note the first and third panels suggest that the residuals don't vary in any obvious systematic way with the fitted values (though there may be something odd going on in the middle of the fitted values range). The second panel shows that the residuals are approximately normally distributed. Further diagnostics could be investigated by looking at measures of leverage (hat values) and measures of how much each data point is driving the coefficients (dfbetas, dffits and cooks.distance). Just looking at plot 3, the Ford Pantera L, Merc 230 and Chrysler Imperial look like they could be exerting excess influence. Re-fitting the model excluding these (and possibly other data points) would be one way to check that they don't change the conclusions that I've reached here.

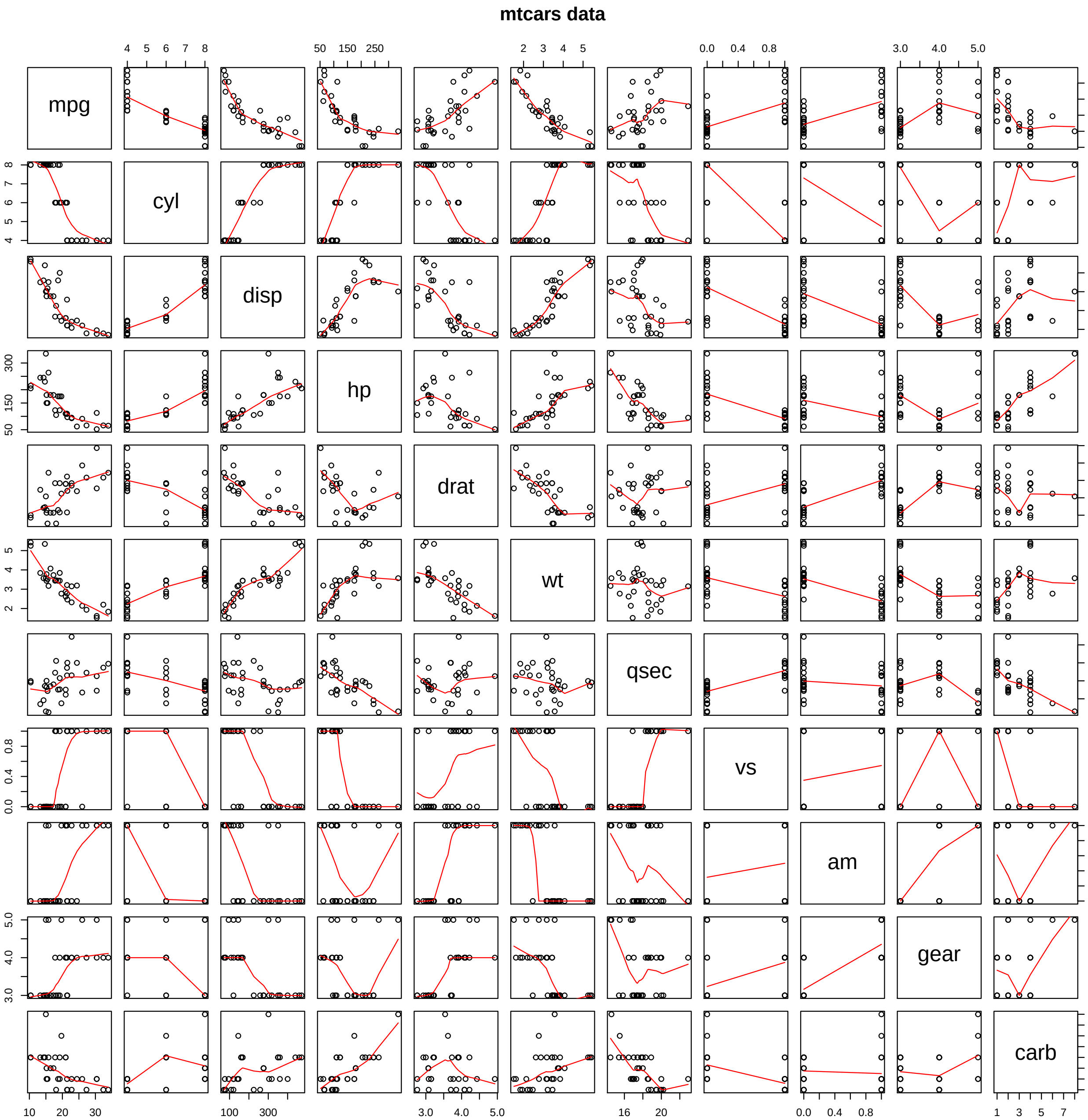
Annex: Figures

Plot 1: MPG in automatic versus manual transmission cars



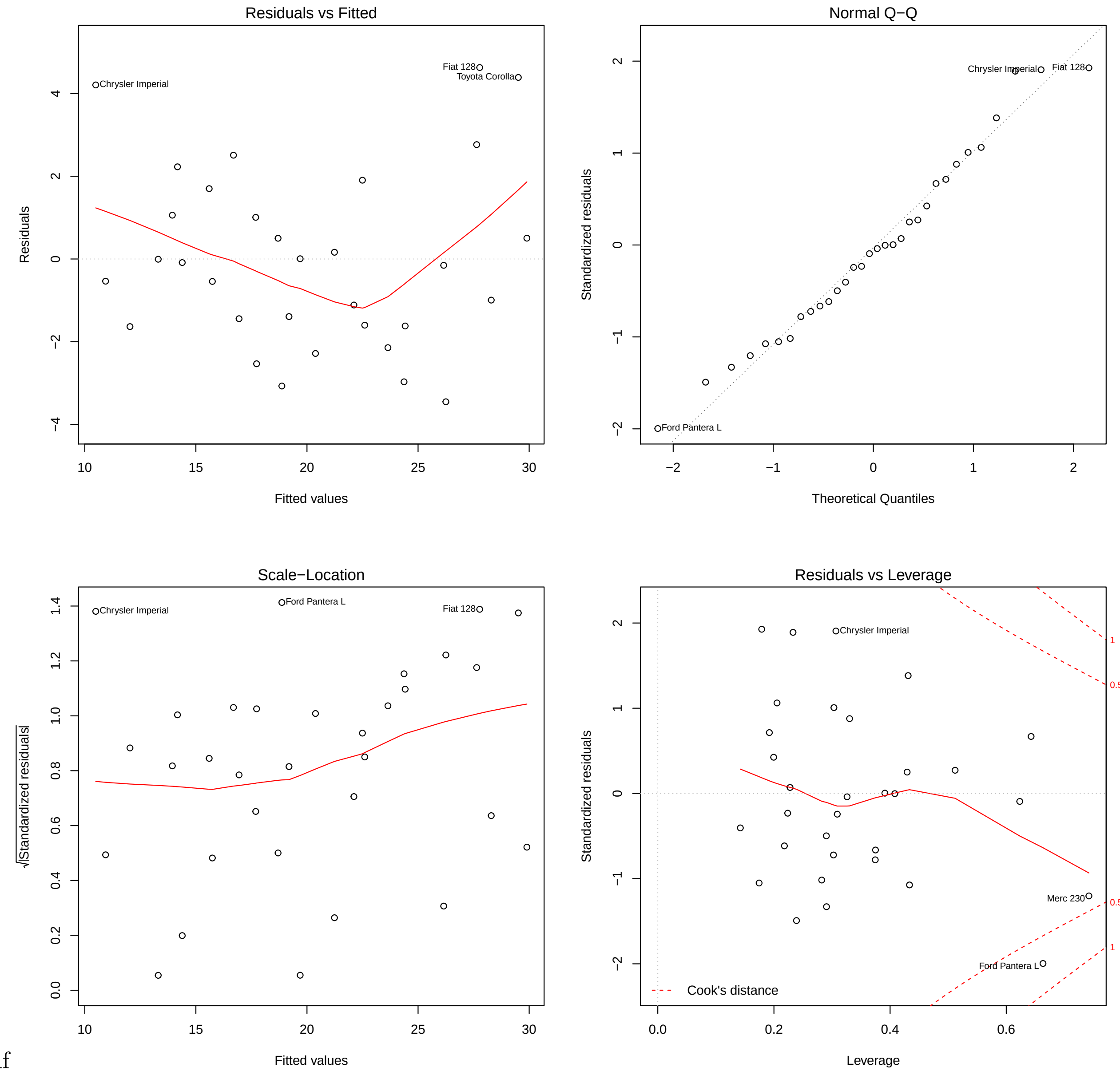
1.pdf

Plot 2: Pairs plot of variables in the mtcars dataset



2.pdf

Plot 3: Diagnostics plot of model 2



3.pdf