No detectable association between transmission type and MPG

Julian Trachsel

April 19, 2021

Executive Summary

In this dataset, manual trasmission cars have a lower mean MPG, however when we consider the information in the other covariates we are unable to identify any meaningful effect of transmission type on fuel economy (MPGs). Other variables, particularly weight seem to be much more important for estimating fuel economy.

"Is an automatic or manual transmission better for MPG"

basic EDA

If we just compare the fuel economy (Miles per gallon: MPG) of automatic and manual transmission vehicles without considering any of the other available covariates in this data set, it appears as though manual transmission vehicles have better fuel economy (appendix figure 1).

However, if we use the pairs() function to look at relationships between other variables and our variable of interest, we observe that there are several variables that appear to have strong relationships with our varible of interest (MPGs) (see Appendix figure 3).

It appears as though `wt`, the weight of the car, and `disp`, a measure of the size of the engine are also important, we also consider the variable `drat` or rear axel gearing as it measures something different than weight and engine size. Other variables appear to be correlated, such as `hp` and `disp` and so have been omited.

We will fit several linear models using `am`, `disp`, `wt`, `drat`, and use an anova method to check whether adding each variable produces a model that better explains the data:

model fitting and selection

```r
fit0 <- lm(data=mtcars, mpg ~ 1)
fit1 <- lm(data=mtcars, mpg ~ 1 + factor(am))
fit2 <- lm(data=mtcars, mpg ~ 1 + factor(am) + disp)
fit3 <- lm(data=mtcars, mpg ~ 1 + factor(am) + disp + wt)
fit4 <- lm(data=mtcars, mpg ~ 1 + factor(am) + disp + wt + drat)
anova(fit0, fit1, fit2, fit3, fit4)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ 1
## Model 2: mpg ~ 1 + factor(am)
## Model 3: mpg ~ 1 + factor(am) + disp
## Model 4: mpg ~ 1 + factor(am) + disp + wt
## Model 5: mpg ~ 1 + factor(am) + disp + wt + drat
##   Res.Df     RSS Df Sum of Sq       F    Pr(>F)
## 1     31 1126.05
```

```
## 2     30  720.90  1    405.15 44.9233 3.400e-07 ***
## 3     29  300.28  1    420.62 46.6380 2.458e-07 ***
## 4     28  246.56  1     53.73  5.9571    0.0215 *
## 5     27  243.51  1      3.05  0.3383    0.5657
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that each additional variable except drat produces a significantly better model. Because of this we choose the `fit3` model, we exclude the `drat` variable because it does not significantly improve the model and it is desirable to have this simplest model possible.

model coeficients

```
##
## Call:
## lm(formula = mpg ~ 1 + factor(am) + disp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4890 -2.4106 -0.7232  1.7503  6.3293
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.675911   3.240609  10.700 2.12e-11 ***
## factor(am)1  0.177724   1.484316   0.120   0.9055
## disp        -0.017805   0.009375  -1.899   0.0679 .
## wt          -3.279044   1.327509  -2.470   0.0199 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.967 on 28 degrees of freedom
## Multiple R-squared:  0.781,  Adjusted R-squared:  0.7576
## F-statistic: 33.29 on 3 and 28 DF,  p-value: 2.25e-09
```

If we examine the model coeficients we observe that the coeficient for having a manual transmission is positive, but very small and the standard error is relatively large. Additionally the T-test indicates this coeficient is not significantly different from 0.

"Quantify the MPG difference between automatic and manual transmissions"

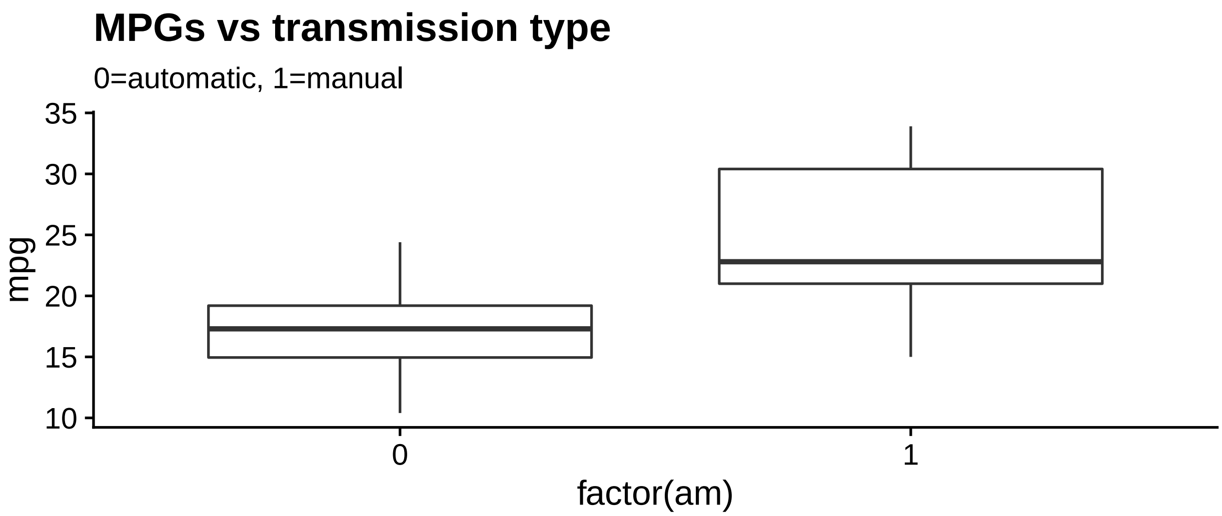
According to this data and the final linear model that we fit, we estimate manual transmission cars have 0.177724 higher MPGs relative to automatic transmission cars holding `disp`, and `wt` constant. However, the 95% confidence intervals for this estimate is quite wide and contains 0 (Appendix figure 2).

Looking at these confidence intervals, we can appreciates that weight appears to have the largest influence on MPGs in this dataset, The heavier the car, the worse the MPGs. For every 1000 lbs a car weighs we can expect a reduction of 3.28 mpgs.

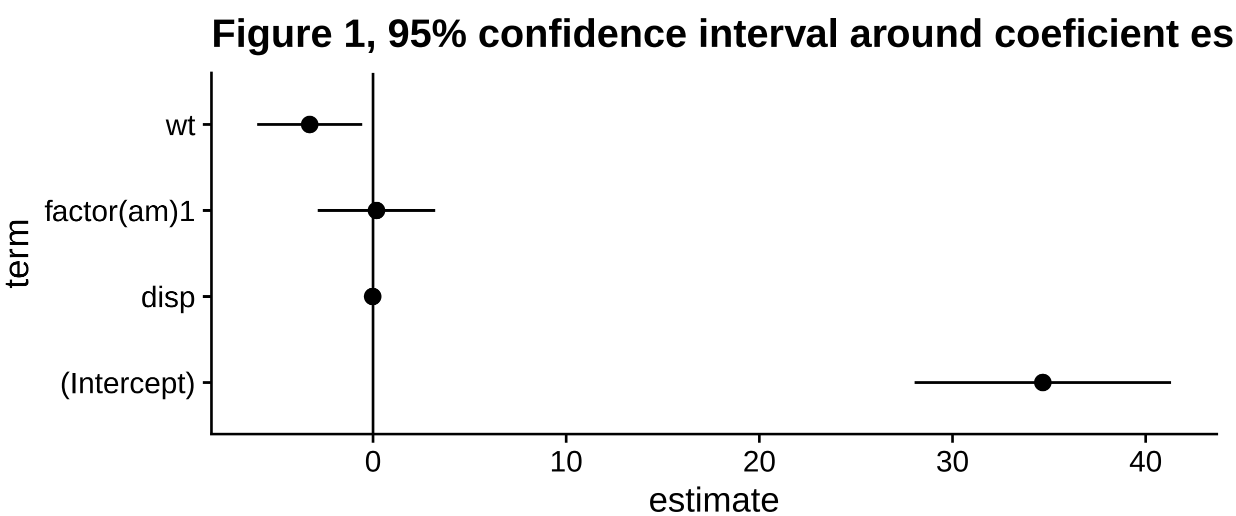
However, this model is not perfect. The sample size is quite small and the model diagnostic plots (appendix figure 4) indicate that our residuals may not be normally distributed. In particular the qqplot suggests that these data may violate the assumption of normally distributed residuals. A larger sample of cars may help to aleviate some of these issues.

Appendix

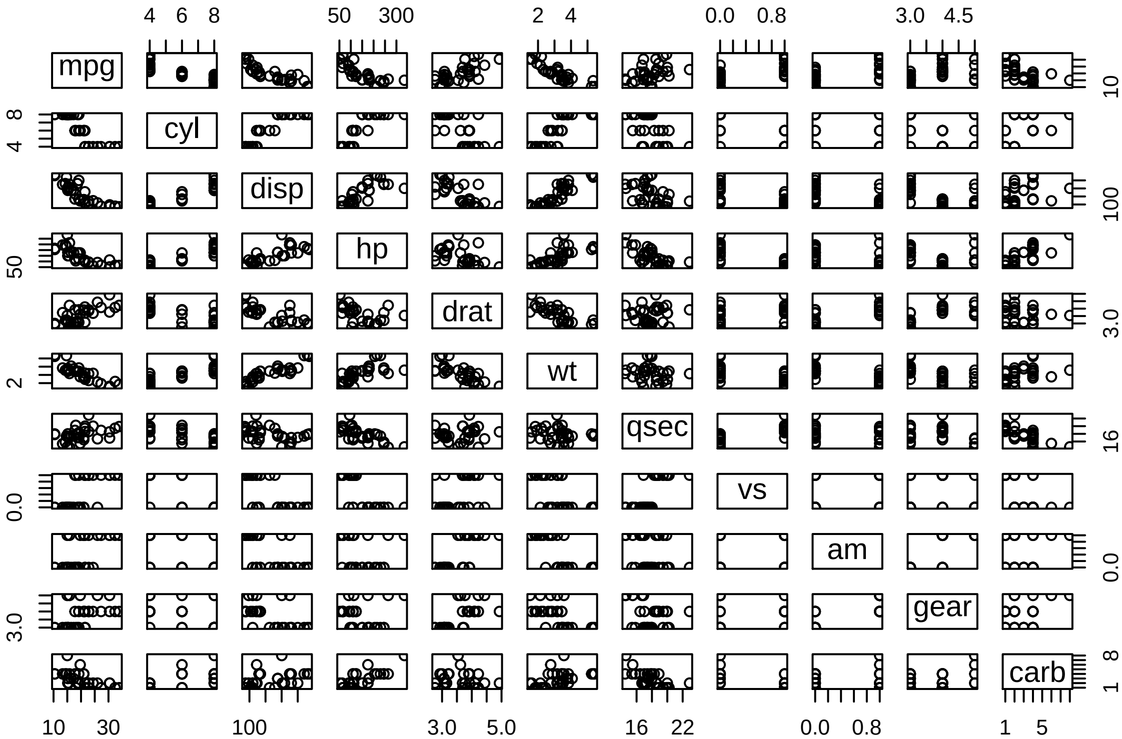
Appendix figure 1



Appendix figure 2



Appendix figure 3



Appendix figure 4

Model Diagnostic plots