Automatic vs Manual Transmisson for better MPG

Sarah Massengill

2/19/2021

Executive Summary

Using the data provided in the mtcars dataset, the answer to the question as to which transmission is better for miles per gallon, MPG, depends on two factors, the weight of the car and the top speeds of the car. For cars that are below 3.2 tons, the manual transmission achieves between 3 and 9 mpg better than the automatic cars with the same peak speed capabilities. Over 3 tons and automatic transmission cars report better MPG than manual transmissions, and for the heaviest cars in the dataset this could mean about 7.5 mpg better. The bottom line is that light slow cars do better with a manual transmission and heavy fast cars do better with automatic transmissions. Transmission is not the only or even the main indicator for gas efficiency. Further analysis can be read below.

Exploratory Analysis and Model Search

model 1: $mpg = b_0 + b_1(am) = 17.1 + 7.2(am)$

A simple linear model that predicts mpg based on type of transmission (am: auto = 0,man = 1) suggests a linear relationship between transmission and MPG. On average, switching from an automatic to a manual transmission increases the mean miles per gallon by 7.2 mpg on average, with a 95% confidence interval of [3.64,10.84]. Since zero is *not* in the interval for b_1 we could conclude that there is a statistically significant difference in MPG for automatic and manual transmissions. Unfortunately, based on the R^2 statistic this linear model only describes about 36% of the variation in MPG results.

It is interesting to note that this simple linear model appears to satisfy the assumptions of linearity, constant variance and normally distributed residuals even if it does not quite answer if an automatic is better than a manual, but it eludes to manual being better.

After plotting the MPG with respect to weight (wt) and differentiating by color the manual and automatic transmissions show a relationship to weight. The lighter weight vehicles in the data set are mostly manual while the heavier vehicles are automatic. The weight of the vehicle is also linearly related with MPG in that the heavier vehicles correlate to less MPG. There are 18 automatic vehicles over three metric tones and one below, while two manual vehicles over and 11 under. This split makes it difficult to separate the effects of each from each other without an interaction term.

model 2: $mpg = b_0 + b_1(am) + b_2(wt) = \begin{cases} 31.32 - 5.4(wt) & \textbf{if } am = \textbf{auto} \\ 37.34 - .02(1) - 5.4(wt) & \textbf{if } am = \textbf{man} \end{cases}$

model 3: $mpg = b_0 + b_1(am) + b_2(wt) + b_3(am)(wt) = \begin{cases} 31.4 - 3.8(wt) & \textbf{if } am = \textbf{auto} \\ 31.4 + 14.89(1) - 3.8(wt) - 5.3(1)(wt) & \textbf{if } am = \textbf{man} \end{cases}$

The first of the two models estimates the average effects of transmission on the mileage while holding the weight constant. The second model does the same thing but also adds an effect modifier term.

- b_0: average MPG for an automatic (am = "auto") when the weight of the vehicle (wt) is zero

- b_1 average change in MPG when switching from an automatic to a manual(am="man") controlling for weight (wt = constant)
- b_2 average change in MPG with a 1-metric ton increase in the weight (wt) of the vehicle controlling for transmission type.
- b_3 effect modifier in model 3 alters the effect of the manual transmission on mpg based on the weight.

Graphically, the first model will have the same slope but the y intercepts change if b_1 is non-zero, the second model allows for a different slope and a different intercept between the two lines that represent the transmission groups.

In Model 2, the p-value for b_1 is much greater than our .05 allowance which suggests that there is no statistical significance in the two transmission types. Even if it did fall with in the significance threshold, a 0.02 MPG loss is not a significant enough difference when deciding which is better for MPG. Adding the weight (wt) variable to the model also loses some of the linearity between the predictors (am & wt) and the predictions (mpg). This is a bit concerning, with such a large change to b_1 between model 1 & 2, model 3's interaction term might work better to shed some light on the interaction between wt and am and reduce the non-linearity.

Model 3's interaction term of -5.3 has an interaction term with a p-value of 0.001. This suggests that the weight is a significant effect modifier of transmission type on MPG. The appendix includes a graph of the two lines. The model suggests that effect of switching to a manual is an average increase of 14.9 MPG if the car weighs nothing, but cars weigh something and this makes no sense. But moving to a one metric ton vehicle will give the manual an average of 9.6 mpg over the automatic of the same weight. At about 3 metric tons, there is not much actual difference between the two transmissions' MPGs, with an automatic getting approximately 20 MPG on average and a manual getting about 19 MPG. The heavier the vehicle the more the automatic transmission outperforms on average a manual of the same weight. The interaction term b_3 has a confidence interval of [-8.3, -2.3]. This means that with 95% confidence we can conclude that the actual interaction between the effects of weight and transmission is on this interval.

Looking at the residual plots of Model 3 shows that some of the linearity lost in model 2 is recovered. Performing an analysis of variance test (anova) on model 2 vs 3 results in a significant improvement (p-value = 0.001) in model 3 when adding the interaction term. The adjusted R^2 value for model 3 is 81.5%, meaning the model explains 81.5% of the variance in the model.

Using the search method provided by the R function step() to find a model that fits better than model 3, we get Model 4, which includes the variable qsec, the car's quarter mile time.

model 4

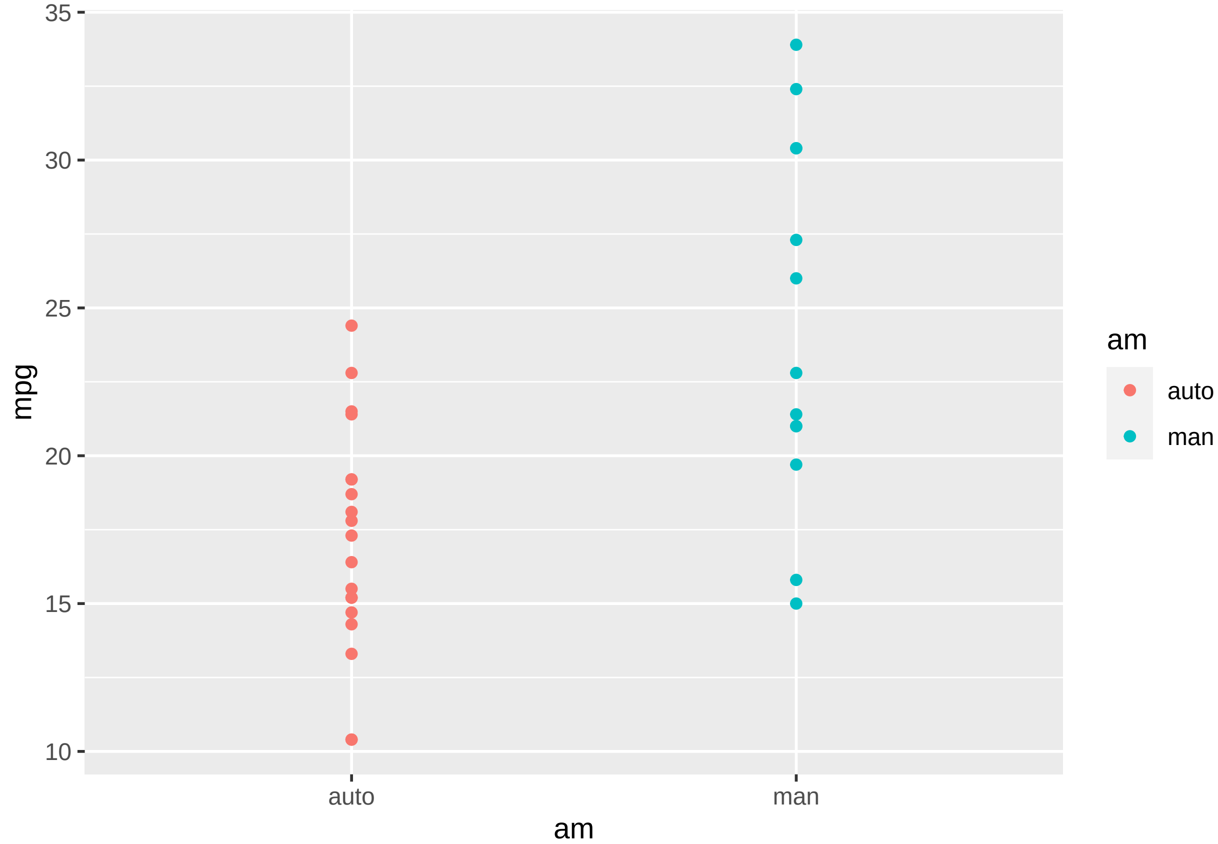
$$mpg = b_0 + b_1(am) + b_2(wt) + b_3(qsec) + b_4(am)(wt)$$

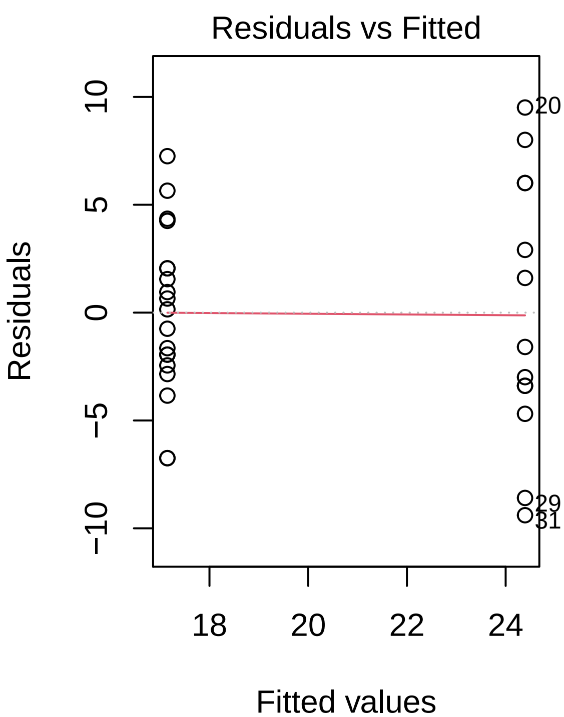
$$= \begin{cases} 9.7 - 2.9(wt) + 1.0(qsec) & \text{if } am = \text{auto} \\ 9.7 + 14.1(1) - 2.9(wt) + 1.0(qsec) - 2.9(1)(wt) & \text{if } am = \text{man} \end{cases}$$

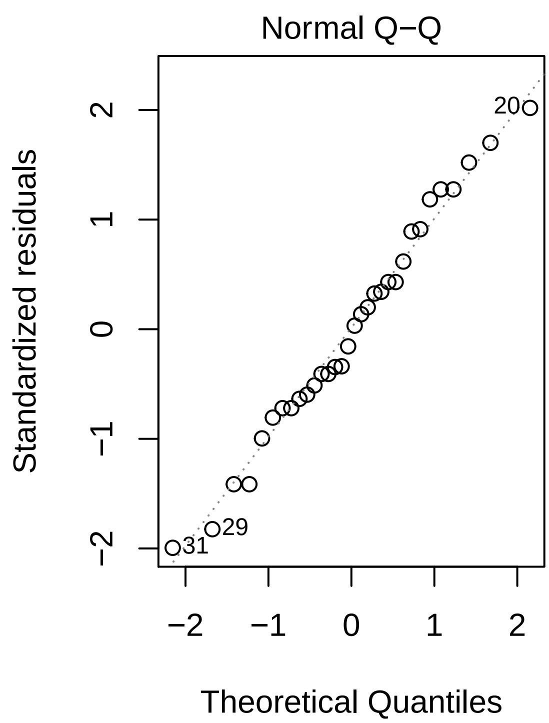
This suggests a positive correlation between MPG and peak speeds of a car. An anova tests suggests it reduces the residual sum of squares by another 70. The adjustted R^2 value is at 88%. The interaction term in model 3 is still significant as well as all the others. Model 4 only changes the quantities slightly as reported in model 3. The one new piece of information is that vehicles with the ability to reach top speeds tend to have worse MPG, so slower is better. The residual plot gains more linearity. This is a better model all around, but it does make it harder to interpret the coefficients.

Appendix

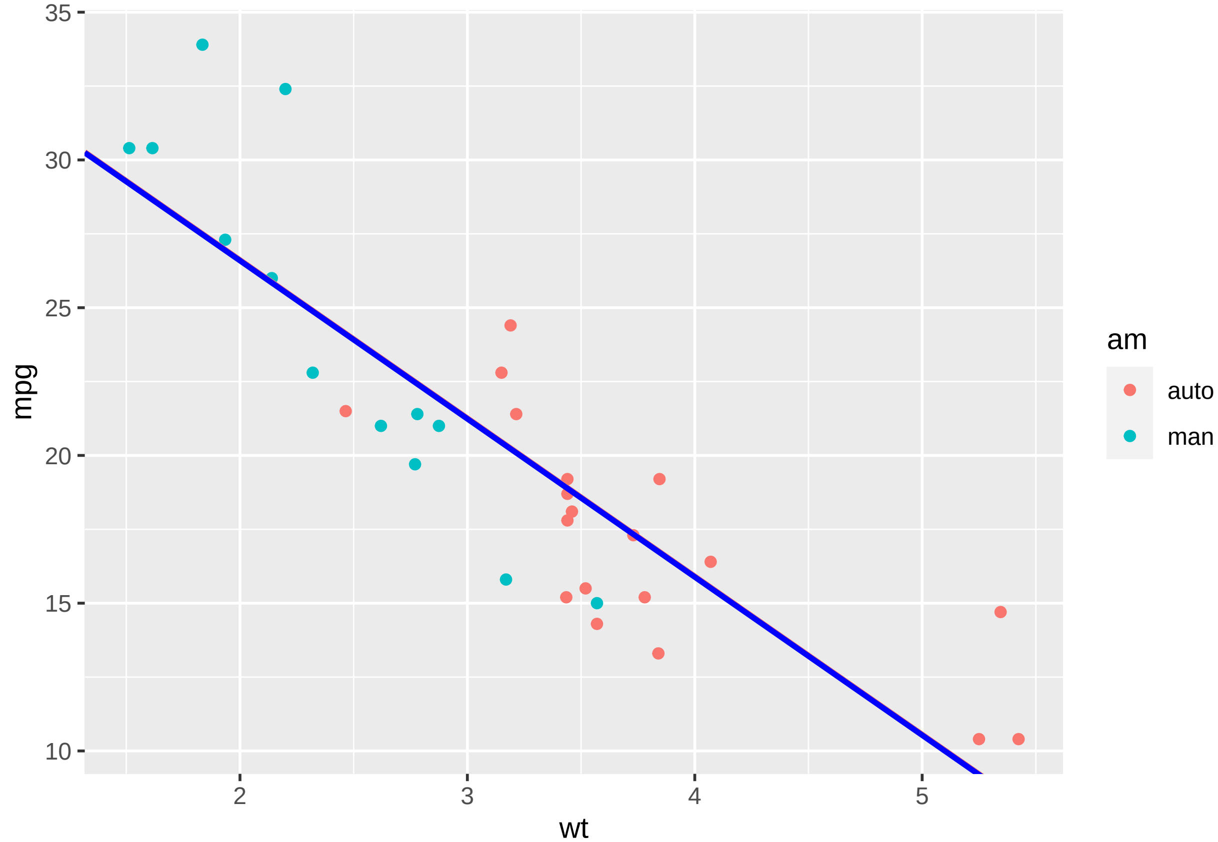
Model 1 Plots

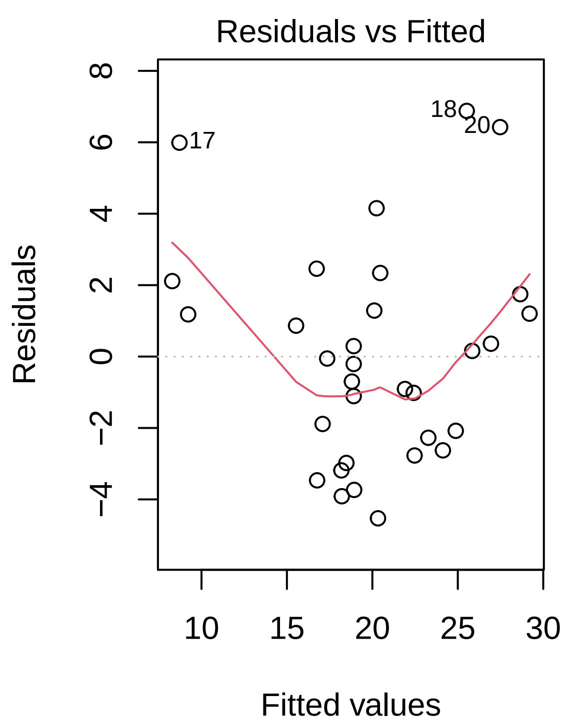


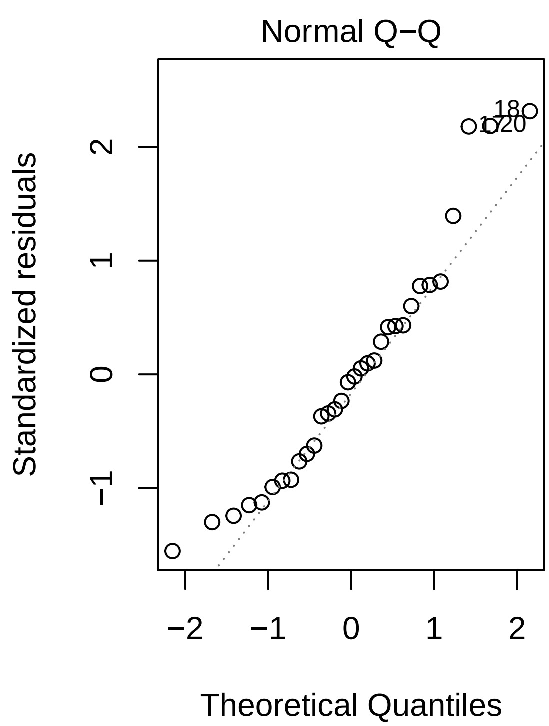




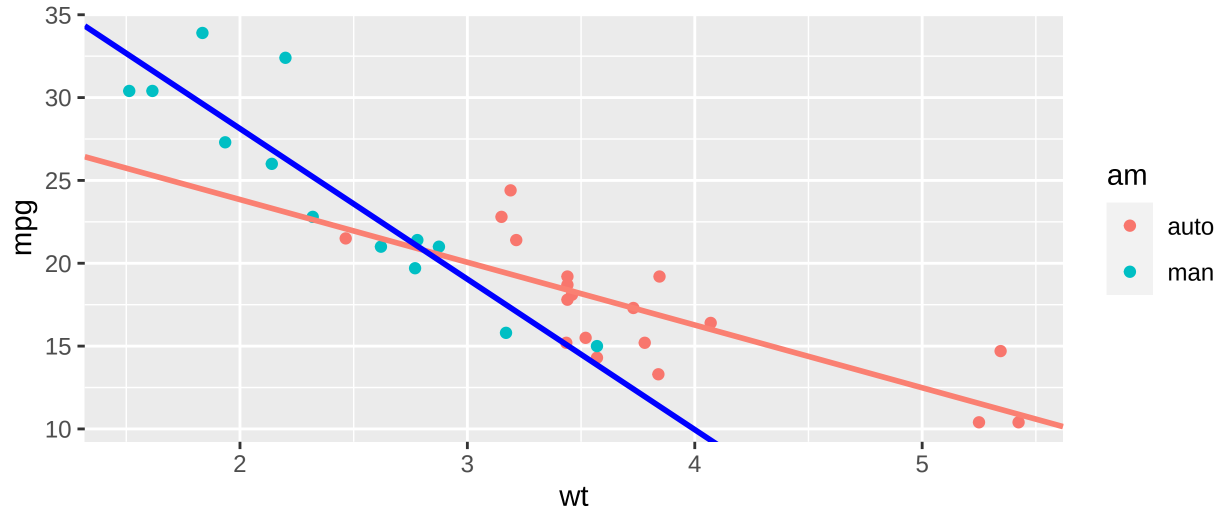
Model 2 Plots

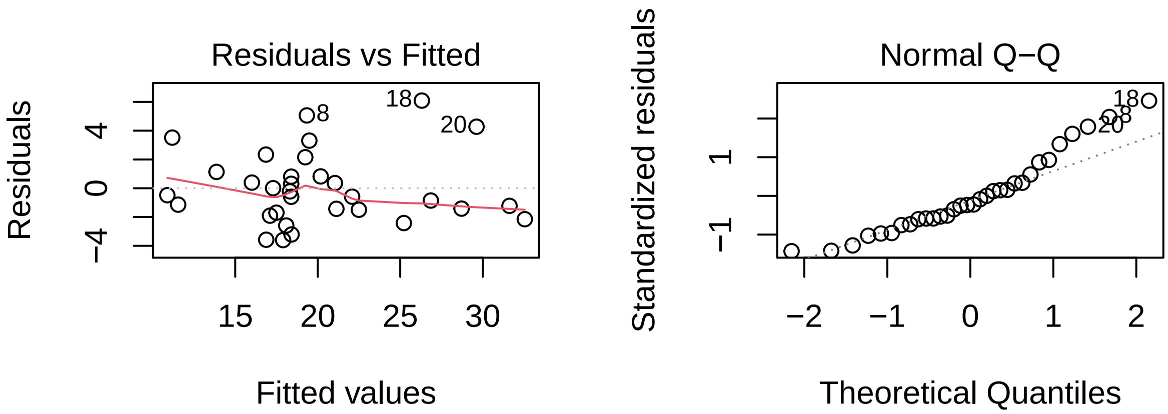






Model 3 Plots





Model 4 Residual and QQ Plot