Linear Regression Course Project

Cameron Charness

September 4, 2018

Executive Summary

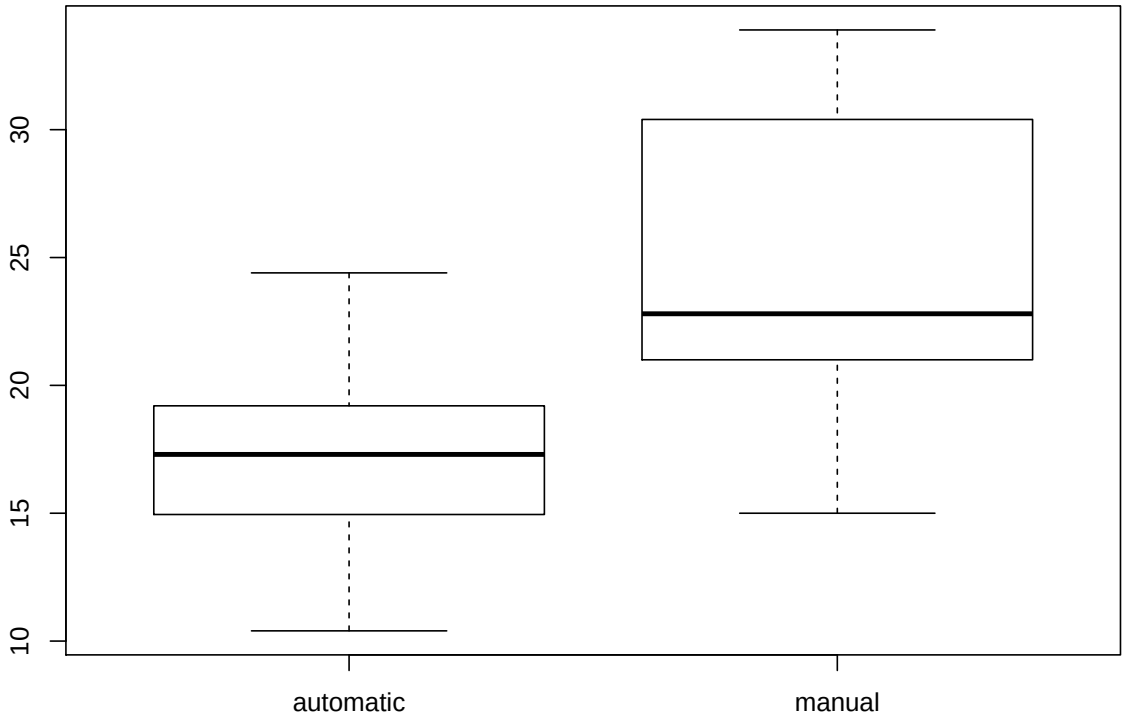
For this project, we are tasked with determining the effect, if any, of transmission type on fuel economy for vehicles in the 'mtcars' dataset of cars tested by Motor Trend magazine in 1974. Examining only fuel economy and transmission type seems to indicate a relationship between these two variables, with manual transmissions delivering higher fuel economy than automatics. However, when we consider other factors, especially vehicle weight and horsepower, this strong relationship vanishes and we see that transmission type is not a significant predictor of fuel economy.

Outline

- Executive Summary
- Exploratory Data Analysis
- Model Selection
- Conclusions

Exploratory Data Analysis

Fig. 1: Fuel economy (miles per gallon) by transmission type



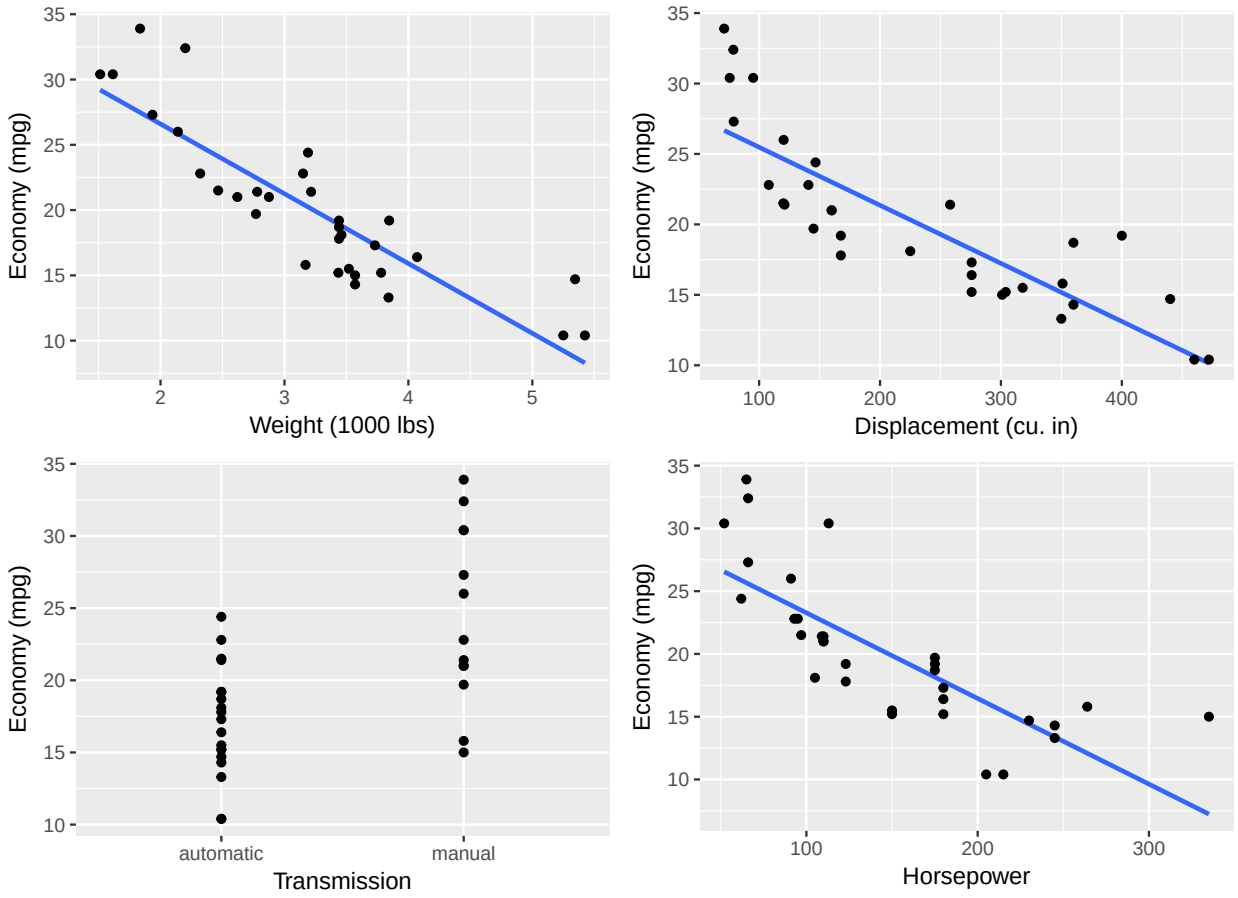


Fig. 2: Fuel economy by several factors

In order to visualize the relationship between fuel economy and transmission type, we can begin with a simple box-and-whisker plot (*Figure 1*). This will give us a sense of the distribution of fuel economy (mpg) results for each transmission type. As the plot shows, there appears to be a pretty significant difference between transmissions, with the manual group showing a higher mean mpg. So, can we simply fit a linear model here and call it a day? Unfortunately, fuel economy depends on a number of other factors, and we can't make definitive conclusions on the effect of transmission type until we explore these as well. *Figure 2* gives a multipanel plot depicting the relationships between fuel economy and 4 variables (engine displacement, weight, transmission type, and horsepower). We can clearly see that all 4 seem to have an effect on fuel economy. Thus, we will need to apply some regression modeling magic to disentangle these effects and see the true relationship between transmission type and fuel economy.

Model Selection

We should be able to get away with a relatively simple linear model to analyze this relationship. We can use the "step" function in R to help us select an appropriate model. Step will allow us to iterate from a model including only the transmission type to one including several variables and their interactions and evaluate which model is best (by minimizing information loss as measured by the AIC statistic).

Final Model
Our step function returns a model with the below parameters:
Model formula: mpg ~ am + hp + wt + gear + hp:wt
Model r^2: 0.8745088
Model AIC: 146.7103929

Coefficients

The coefficients on each term of the model are as follows:

```
##                Estimate  Std. Error    t value     Pr(>|t|)
## (Intercept) 46.60216891 5.455579183  8.5421121 5.062049e-09
## ammanual    -1.60680583 1.705156128 -0.9423218 3.547020e-01
## hp          -0.13510670 0.027756632 -4.8675467 4.766880e-05
## wt          -8.61890220 1.773990752 -4.8584820 4.882354e-05
## gear         1.45513184 0.928238915  1.5676264 1.290598e-01
## hp:wt        0.03171113 0.008681237  3.6528355 1.148172e-03
```

We can interpret the coefficients of the model as follows:
Intercept- estimate of mpg for a car with hp=0,wt=0,gear=0
ammanual- increase in mpg going from automatic to manual transmission, holding other variables constant
hp- increase in mpg per additional horsepower, holding other variables constant
wt- increase in mpg per additional 1000 lbs vehicle weight, holding other variables constant
gear- increase in mpg per additional gear, holding other variables constant
hp:wt- interaction between horsepower and weight

Conclusions

Just looking at the p-values for the coefficients, we can see that many other factors seem to have a more significant impact on fuel economy than the transmission type does. Horsepower and vehicle weight especially seem to be more significant. We can look at a 95% confidence interval on the 'ammanual' coefficient to be a little more rigorous about it.

The endpoints of our confidence interval are: -5.1118044, 1.8981928. Since this interval includes 0, we fail to reject the null hypothesis and can't say definitively whether transmission type has any significant effect on fuel economy.