Motor Trend mpg (Arguing with HNDR)

Barret Miller

Executive Summary

In this study, I performed exploratory data analysis and built two models that analyze and attempt to answer the question of what impact the transmission type (manual vs. automatic) of a car has on its miles per gallon (mpg) efficiency. I used the mtcars dataset that comes prepackaged with R. This study is of particular interest to me, because my long-time friend who will be named only as HNDR, has long argued the superiority of manual transmissions. I tend to prefer the easygoing comfort (and free hands) offered by the modern technology of automatic transmissions, but HNDR insists that I am wasting money on gas and lowering my testosterone levels by missing out on the thrill of shifting gears myself. While this dataset does not currently offer the possibility of answering the latter, the former question, I hope, can finally be laid to rest as a result of this study.

Exploring the Data Visually

If I just plot a boxplot of mpg vs. transmission type, I can get an idea of the center and variability of the data. In Figure 1 in the appendix, I have plotted boxplots for the transmission types, a blue line at the mean of manual transmission mpg (24.39), and a red line at the mean of automatic transmission mpg (17.15). So, this looks to be a very large marginal effect. The difference in means tells me I can achieve 7.24 more miles per gallon by driving a manual. Things are looking good for ole HNDR so far (ˆˆˆˆ)(-.-)(ˆˆˆˆ).

Simple Linear Regression (SLR) Analysis

I can run a simple linear regression analysis to capture the marginal effect of transmission type on mpg outcome. I will treat automatic transmissions as my reference group, and measure the marginal effect of having a manual transmission. After fitting my linear model, my intercept term β_0 is the mean of my reference group of automatic transmissions (17.15). I show the same marginal increase of 7.24 for my manual transmission β_1 coefficient as before just looking at averages, because that what this linear model is doing when you are predicting an outcome of two groups with a single dummy variable. It is also noteworthy that I have a very significant p-value of 0.000285 for the coefficient of transmission type. This suggests that the effect is real. See Figure 2 in the appendix for the linear model summary output.

Residuals

A good way to understand if our model explains the outcome well is to look at the residuals and their variation. If there is a lot of variation left over in the mpg data after adjusting for the transmission type, then that suggests there is more to the story. The are other variables that I need to consider to explain more of the variation. In this case, my R^2 value is only 0.3598, so about 64% or so of the variation is unexplained by transmission type. One good thing to do in order to hone in on your residuals is to plot them versus your predictor. In Figure 3, I've plotted the residuals of the mpg against the transmission type. The fact that there is still so much of the variation in mpg left unexplained, especially within the manual transmission group, tells me that maybe there is another variable that is correlated with transmission types that is really a better predictor of mpg. I need to do further analysis of other variables in the dataset to see if adjusting for any of them will change the effect of transmission type.

Multivariable Regression Analysis

To start off, I'd like to get a visual of the relationships between the different variables. I ran a pairwise scatterplot of the mtcars dataset for Figure 4. Specifically, I am looking for anything that may be correlated negatively with mpg and which correlates positively with manual vs. automatic transmission. If I look at the relationship of wt (weight) to mpg, it is obviously negatively correlated: as weight goes up, you get less mpg. Interestingly, lower weights also seem to be associated with manual transmissions. Manual is represented by level 1, so it's higher on the y axis with a line sloping down toward the automatics' level 0 as weight increases. The same effect versus am appears with the disp, hp, and cyl variables, however, those variables also appear to be positively correlated with weight, so I won't include those in my model. In fact, most things that appear to be positively correlated with weight show the same splitting of the transmission types where the 1 (manual) is associated with lower values of those variables, and 0 (automatic) is associated with increasing values. All of those variables that positively correlate with weight also negatively correlate with mpg. With that in mind, I am going to try a multivariate model $mpg = \beta_0 + \beta_1(am = "Manual") + \beta_2 wt + \beta_3(am = "Manual") * wt$. I am centering the wt variable, so the intercepts will be interpretable as the the mpg for the average cars within the transmission groupings.

Figure 5 is my vindication from HNDR. The coefficient β_0 shows that the average weighted automatic has an expected mpg average of 19.24. The intercept for the average weighted manual car, β_1 (-2.17), is actually negative now, saying that the average manual transmission adjusted for weight is that much lower than the automatic. This is a reversal of the marginal effect from earlier. However, this effect is not significant, so I can't reject the null hypothesis. What are significant, though, are the coefficients for the slopes. β_2 has a value of -3.79, which means that for automatics, for every 1000lb increase in weight, the mpg goes down by -3.79. β_3, also significant, shows a decrease of -5.3 on top of that of automatics, for a total of decrease of -9.09 per 1000lb increase in weight. Also, note that the R^2 value is much higher at 0.833, suggesting I have explained much of the mpg variance in the data with this model.

Finally, I take a look at the 95% confidence intervals for this model in Figure 6. The lower and upper bounds for the mean automatic mpg are 17.73 and 20.74. Interestingly, you can see that the bounds for the manaul transmission's effect on top of the mean for automatics (-5.07 and 0.74) include 0, so we can't reject the null hypothesis. We must accept the reasonable possibility that manual transmissions have no effect on mpg. The confidence intervals for β_2 and β_3 are also shown.

Conclusion

Our results show that even though there appears to be a significant marginal effect of the transmission type on mpg efficiency giving manual transmissions an edge, in fact this is a myth. Manual transmissions seem to be associated with lighter vehicles, and weight is the most crucial factor. HNDR has been living a lie. Manuals may be cool, but a manual transmission in and of itself is not good for gas milage. If an automatic and a manual vehicle have the same weight, the transmission type is (statistically) insignificant in its effect on mpg. Furthermore, as you increase the weight of a vehicle, manual tranmissions tend to have a steeper decrease in their mpg efficiency. So, when HNDR loads up all those kegs in the back of his manual blue Chevy S-10, that gas milage is likely going to down more than if that S-10 were an automatic.

Appendix

Figure 1:

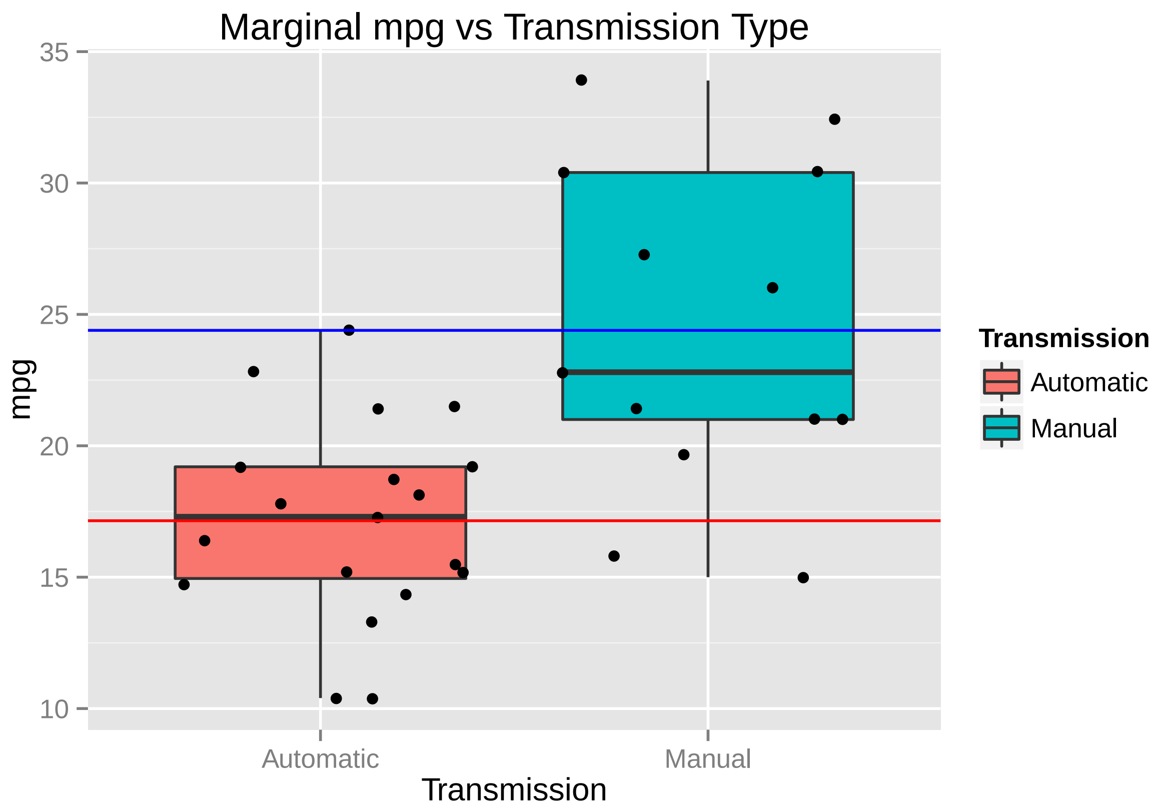

Figure 2:

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## amManual       7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

Figure 3:

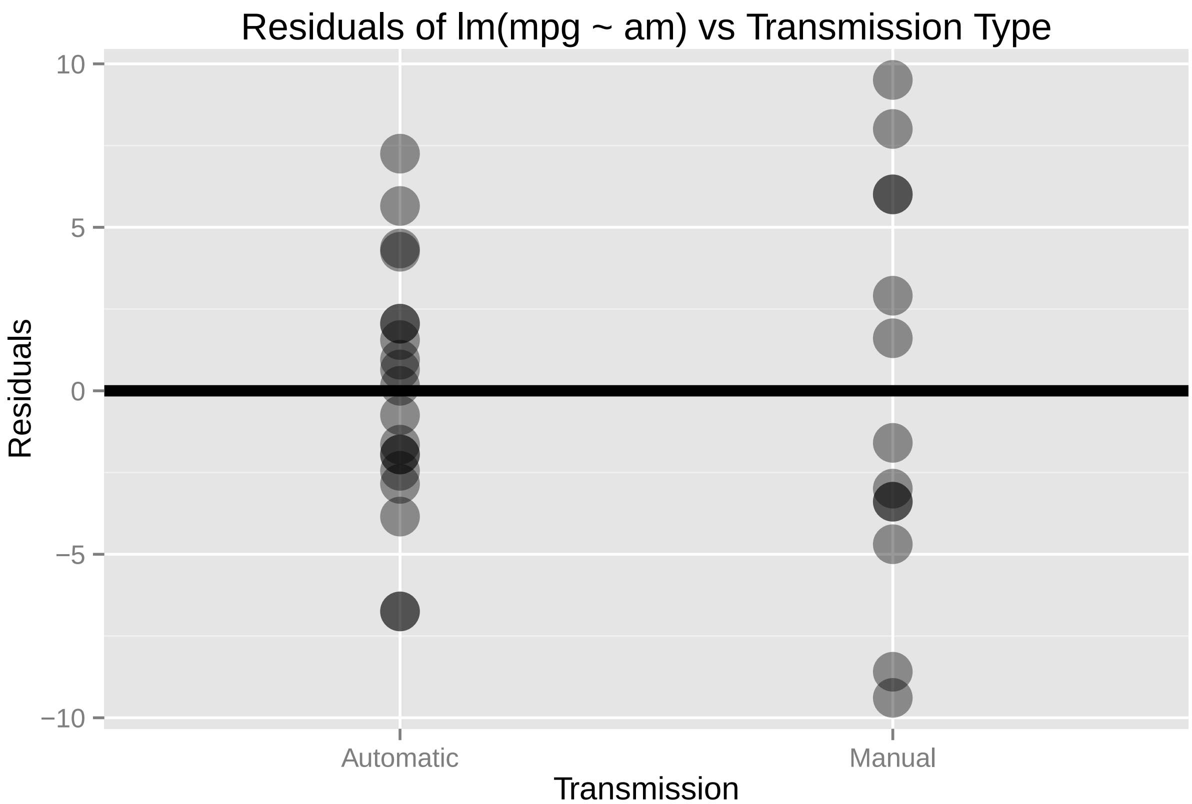

Figure 4:

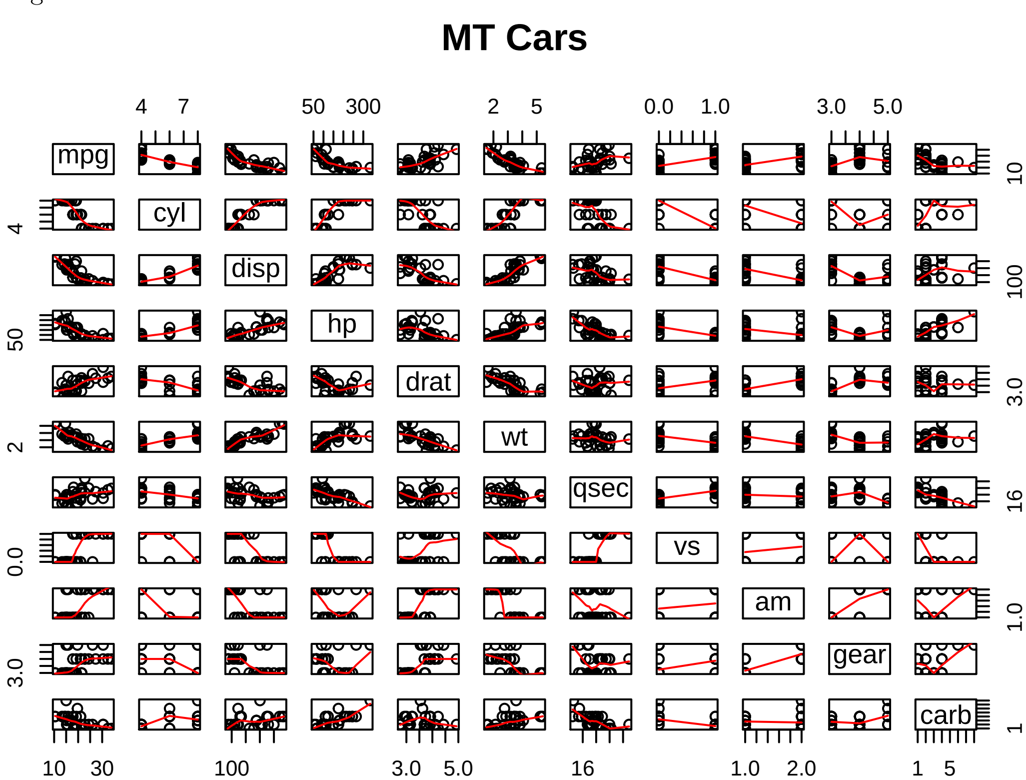

Figure 5:

##

```
## Call:
## lm(formula = mpg ~ am * I(wt - mean(wt)), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6004 -1.5446 -0.5325  0.9012  6.0909
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                19.2358     0.7357  26.147  < 2e-16 ***
## amManual                   -2.1677     1.4189  -1.528  0.13779
## I(wt - mean(wt))           -3.7859     0.7856  -4.819 4.55e-05 ***
## amManual:I(wt - mean(wt))  -5.2984     1.4447  -3.667  0.00102 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.591 on 28 degrees of freedom
## Multiple R-squared:  0.833,  Adjusted R-squared:  0.8151
## F-statistic: 46.57 on 3 and 28 DF,  p-value: 5.209e-11
```

Figure 6:

```
##                               2.5 %     97.5 %
## (Intercept)               17.728862 20.7428263
## amManual                  -5.074184  0.7387288
## I(wt - mean(wt))          -5.395234 -2.1765810
## amManual:I(wt - mean(wt)) -8.257693 -2.3390281
```