RegressionModelsFinalProject

Tony Smith

March 20, 2021

Impact of Automatic and Manual Transmissions on MPG

Introduction

In this exploration, I will be looking to do two things:

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions.

Manual or Automatic?

First, let's get the data and look at what types of variables we are working with:

```
data(mtcars)
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

We can see that there are several other variables besides transmission type that may further impact mpg. We may not be able to come up with an answer without looking at some different combinations.

Looking at the graph in appendix 1, it seems pretty obvious that manual transmissions have better MPG overall. If we look at the numbers(appendix 2), we see that the average manual transmission has is over 7 mpg better than for automatic transmissions.

Quantitative Difference

While we can see the difference in overall averages from above, this is not taking into account other variables that affect how efficient a car is. Certainly, we would expect a heavier car to be less fuel efficient. Let's see what kinds of impact the other variables have on the change in mpg(appendix 3).

Looking at the coefficients, it seems that aside from the type of transmission the weight and quarter mile time may have significant impacts. I want to look at some nested models and see which combination of variables will give us the best result.

Looking at the result of the ANOVA analysis (appendix 4), we see that both weight and quarter mile time seem to be necessary inclusions when looking at the change in mpg. Let's look at the third model and see what it shows.

```
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ factor(am) + wt + qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## factor(am)1   2.9358     1.4109   2.081 0.046716 *
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

We can see that when we take weight and quarter mile time into effect, the difference in mpg for manual transmissions is less than what we saw when we just looked at transmission type. Here there is a change of 2.94 mpg. Looking at the R squared value, we see that this model accounts for about 85% of the variance, so there is still some variance left unexplained, but this model does a fairly good job.

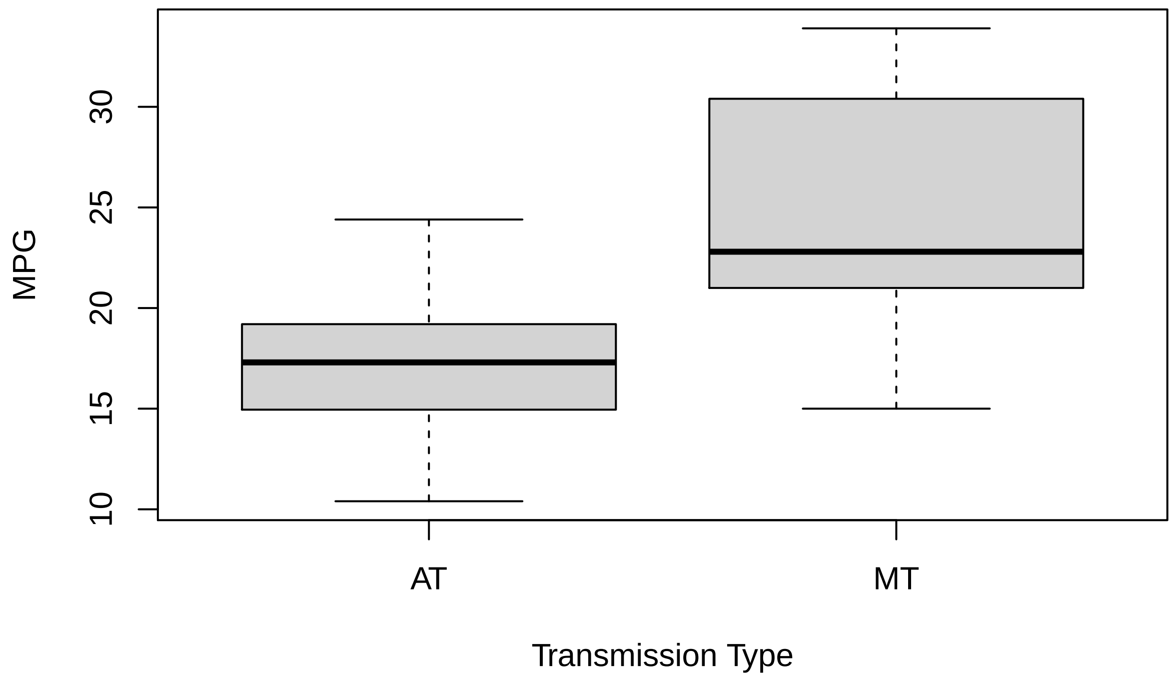
Summary

From our results, we see that when trying to quantify the difference in mpg for transmission type, we also need to take weight and quarter mile time into account. When we do that, we can determine that there is still an increase in mpg based on using manual transmissions instead of automatic transmissions, but it is impacted by at least two other variables.

Appendix

1. Graph of MPG vs transmission type

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <-c("AT", "MT")
plot(mtcars$am, mtcars$mpg, xlab="Transmission Type", ylab="MPG")
```



2. Summary of mpg for manual and automatic transmissions.

```
summary(mtcars[mtcars["am"] == "MT",]["mpg"])
```

```
##       mpg
##  Min.   :15.00
##  1st Qu.:21.00
##  Median :22.80
##  Mean   :24.39
##  3rd Qu.:30.40
##  Max.   :33.90
```

```
summary(mtcars[mtcars["am"] == "AT",]["mpg"])
```

```
##       mpg
##  Min.   :10.40
##  1st Qu.:14.95
##  Median :17.30
##  Mean   :17.15
##  3rd Qu.:19.20
##  Max.   :24.40
```

3. Model on all the variables

```
fullfit <- lm(mpg~. -1, data=mtcars)
summary(fullfit)
```

```
##
## Call:
## lm(formula = mpg ~ . - 1, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## cyl  -0.11144    1.04502  -0.107   0.9161
## disp  0.01334    0.01786   0.747   0.4635
## hp   -0.02148    0.02177  -0.987   0.3350
## drat  0.78711    1.63537   0.481   0.6353
## wt   -3.71530    1.89441  -1.961   0.0633 .
## qsec  0.82104    0.73084   1.123   0.2739
## vs    0.31776    2.10451   0.151   0.8814
## amAT 12.30337   18.71788   0.657   0.5181
## amMT 14.82360   18.35265   0.808   0.4283
## gear  0.65541    1.49326   0.439   0.6652
## carb -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.9895, Adjusted R-squared:  0.984
## F-statistic: 179.8 on 11 and 21 DF,  p-value: < 2.2e-16
```

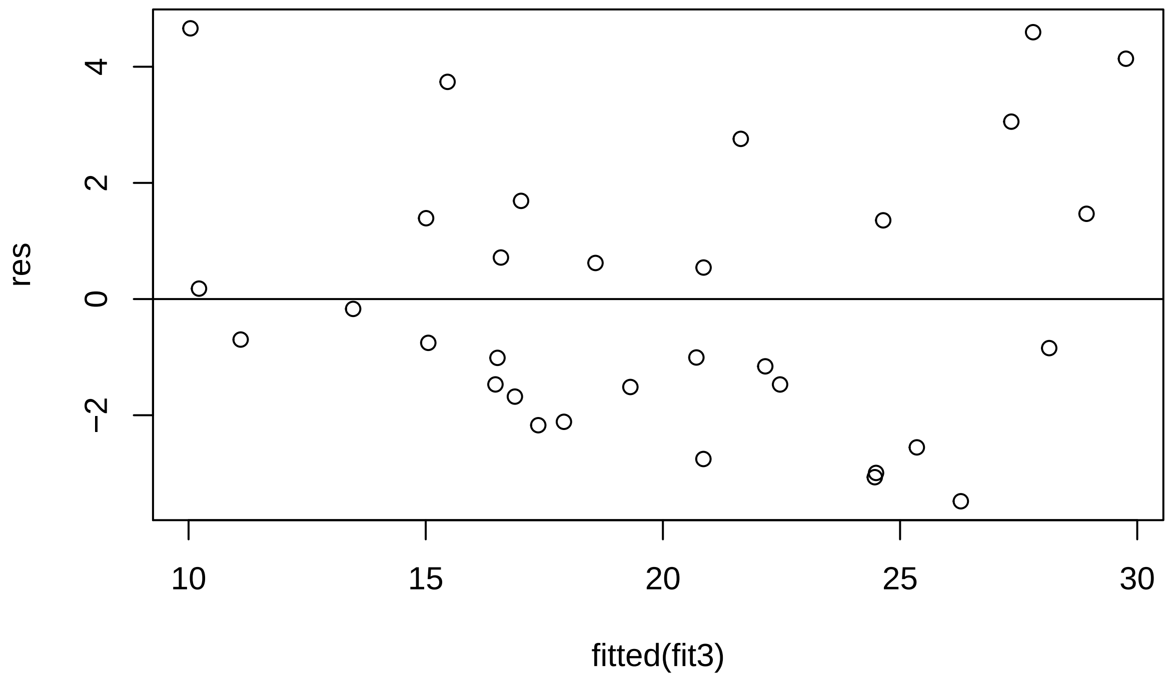
4. ANOVA of the three models with transmission, weight and quarter mile time

```r
fit1 <- lm(mpg ~ factor(am), mtcars)
fit2 <- lm(mpg ~ factor(am) + wt, mtcars)
fit3 <- lm(mpg ~ factor(am) + wt + qsec, mtcars)
anova(fit1, fit2, fit3)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ factor(am)
## Model 2: mpg ~ factor(am) + wt
## Model 3: mpg ~ factor(am) + wt + qsec
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 278.32  1    442.58 73.203 2.673e-09 ***
## 3     28 169.29  1    109.03 18.034 0.0002162 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. Residual plot of final model

```r
res <- resid(fit3)
plot(fitted(fit3), res)
abline(h=0)
```


res
fitted(fit3)
4
2
0
−2
10
15
20
25
30