The Effect of Transmission Type on MPG

Paul Martinez

04/10/2023

Executive Summary

The following analysis examines the relationship between transmission type and MPG. This analysis answers the following two questions:

- "Is an automatic or manual transmission better for MPG"
- "Quantify the MPG difference between automatic and manual transmissions"

Data Upload and Cleaning

I convert some variables to factor to facilitate with the analysis.

```
library(ggplot2)
library(dplyr)
data(mtcars)

mtcars2<-mtcars


mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("automatic", "manual")
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$vs <- as.factor(mtcars$vs)
levels(mtcars$vs) <- c("V", "S")
```

Exploratory data analyses

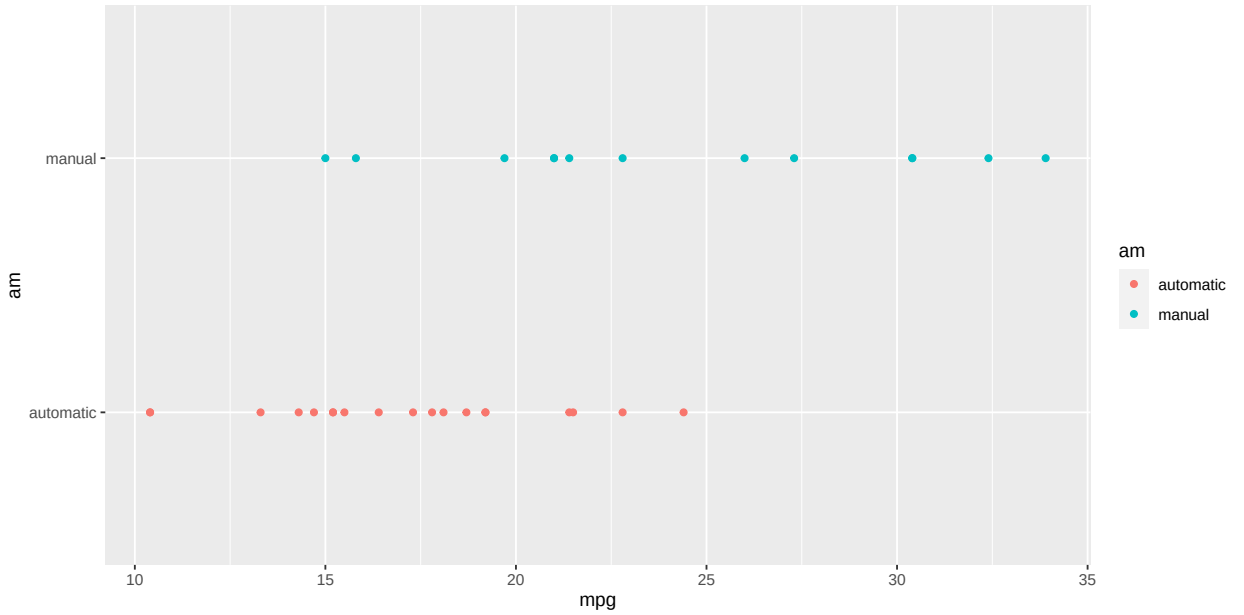
Next, I explore the data.

```
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  V    manual    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  V    manual    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  S    manual    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  S automatic    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  V automatic    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  S automatic    3    1
```

The following scatterplot examines the relationship between transmission type and MPG. From these visualization we can see that there is a significant difference between automatic and manual transmissions. It appears manual transmissions are slightly more efficient when it comes to MPG. We will conduct further testing.

```
# Figure 1
ggplot(data = mtcars,    # set data
       mapping = aes(     # map aesthetics to column values
                x = mpg,           # map x-axis to age
                y = am,          # map y-axis to weight
                color = am)
)+     # map color to age
        geom_point()
```



Prior to creating a model, I wanted to explore if other variables also had a similar relationship with MPG. After conducting the Pearson correlation test below, we can see that the weight(wt) of the car and how many cylinders also had an extremely significant relationship between how much MPG is used and those variables.

```
# Correlation Test
cors <- cor(mtcars2$mpg, mtcars2)
orderedCors <- cors[,order(-abs(cors[1,]))]
orderedCors
```

```
##        mpg         wt        cyl       disp         hp       drat         vs         am       carb
##  1.0000000 -0.8676594 -0.8521620 -0.8475514 -0.7761684  0.6811719  0.6640389  0.5998324 -0.5509251  (
##       qsec
##  0.4186840
```

Model selection

To start I test the question "Is an automatic or manual transmission better for MPG" using a linear regression model. In this model I only include MPG as my dependent variable and transmission type as my independent

variable. The results show that manual transmission in comparison to automatic yields a 7.24 positive MPG. This result was statistically significant at a p-value less than 0.000.

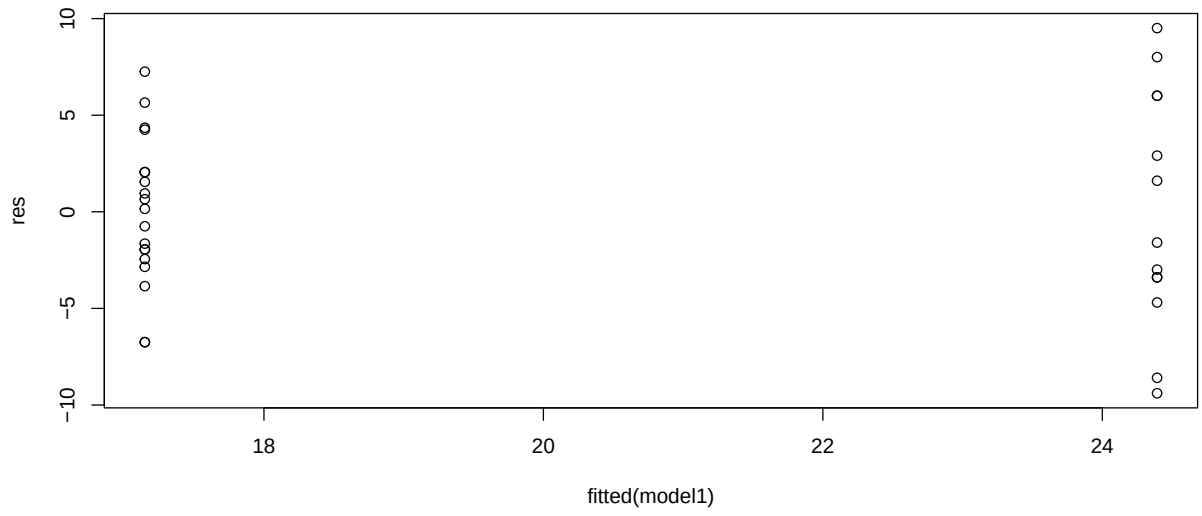
```
# MODEL 1- TRANSMISSION TYPE

model1 <- lm(mpg ~ am, mtcars)
summary(model1)
```

```
##
## 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
```

```
res <- resid(model1)

#Residual Plot for Model 1
plot(fitted(model1), res)
```



Next, while Model 1 showed promising results I wanted to explore what would happen if we add additional factors. Holding constant all other variables in the dataset, transmission type was not statistically signficant

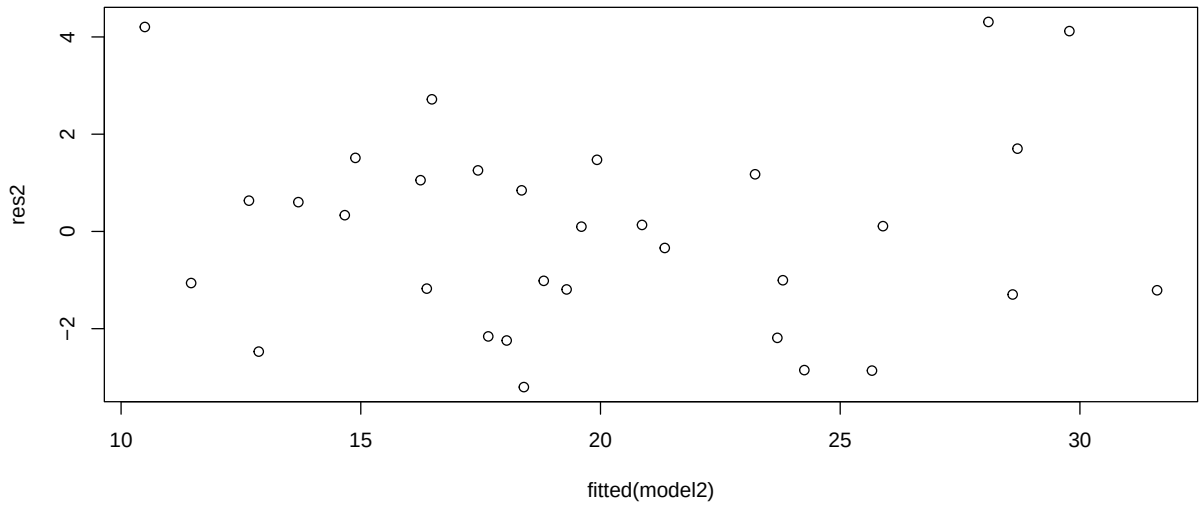
in model 2. In fact, the only two variables that showed to have a statistically significant value at p>0.05 were the Weight of the car and gross horsepower, while holding all other variables constant.

```
# MODEL 2-All VARIABLES
model2 <- lm(mpg ~ ., mtcars)
summary(model2)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2015 -1.2319  0.1033  1.1953  4.3085
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.09262   17.13627   0.881   0.3895
## cyl6        -1.19940    2.38736  -0.502   0.6212
## cyl8         3.05492    4.82987   0.633   0.5346
## disp         0.01257    0.01774   0.708   0.4873
## hp          -0.05712    0.03175  -1.799   0.0879 .
## drat         0.73577    1.98461   0.371   0.7149
## wt          -3.54512    1.90895  -1.857   0.0789 .
## qsec         0.76801    0.75222   1.021   0.3201
## vsS          2.48849    2.54015   0.980   0.3396
## ammanual     3.34736    2.28948   1.462   0.1601
## gear4       -0.99922    2.94658  -0.339   0.7382
## gear5        1.06455    3.02730   0.352   0.7290
## carb         0.78703    1.03599   0.760   0.4568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.616 on 19 degrees of freedom
## Multiple R-squared:  0.8845, Adjusted R-squared:  0.8116
## F-statistic: 12.13 on 12 and 19 DF,  p-value: 1.764e-06
```

```
res2 <- resid(model2)

#Residual Plot for Model 2
plot(fitted(model2), res2)
```



Conclusion

In conclusion, if we only considered Model 1 the answer to the question would be absolutely transmission type matter, when considering MPG. However, if we account for other factors, we can see that it is not a statistically significant factor in comparison to other variables like the Weight of the car or the horsepower.