Regression Models Course Project

Sofia Perez

06/04/21

Executive Summary

This report was prepared as a part of Regression Models Course by Johns Hopkins University.

The aim of the report is to explore the relationship between the performace of cars measured in miles per gallon (mpg) and a set of other variables from the dataset *mtcars*.

The analysis detailed below intends to demostrate that:

- Manual transmission cars have higher performance compared to automatic transmission cars.
- Only 35% of the variance in mpg can be attributed to transmission types. Other variables have a strong impact in cars performance and should be included in the model. These variable are: cylinders (cyl), power (hp) and weight (wt).

Data Analysis

The first step is to load the dataset and take a look at the data in it.

```
##                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
```

To compare performance in terms of transmission types we can first take a look at the means:

Manual transmission (am=1)

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

Automatic transmission (am=0)

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

From the boxplot shown in Appendix A1 it seems that manual transmission cars perform better. To confirm this, a t-test was conducted.

```
p_value<-t.test(mpg~am, mtcars)$p.value
```

As the p-value is ~0.0014, we can reject the null hypothesis and conclude, that manual transmission cars have higher performance in terms of mpg compared to automatic transmission cars. This statements assumes all other characteristics remain the same, so further analysis should be performed.

Regression Models

First we analize the simple linear regression:

```
fit<-lm(mpg ~ am, mtcars)
r_squared<-summary(fit)$r.squared
```

Considering that the R-squared value for this test is ~0.36. Which means that only 36% of the variance can be attributed to transmission types. This confirms that other variables should be taken into account in the study.

Mutivariable regression should be analized:

```
fit2<-lm(mpg ~ ., mtcars)
```

The function *step()* was used for choosing the best model. The function basically will remove variables and analize the AIC until it obtains the lower value.

```
summary(fit3)$coef
```

```
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## am1          1.80921138 1.39630450  1.295714 2.064597e-01
```

```
r_squared2<-summary(fit3)$r.squared
```

Considering that the R-squared value for this test is ~0.87 Which means that over 87% of the variance in mpg can be attributed to the variables: cylinders (cyl), power (hp), weight (wt) and transmission type (am).

Residual Analysis

When analyzing the “Residuals vs Fitted” plot in Appendix A3, it appears to be randomly scattered, meaning the residuals are homoscedastic. Also observing the “Q-Q” plot, it shows that most points fall in the line, meaning the residuals have a normal distribution.

Conclusion

Is an automatic or manual transmission better for MPG?

In terms of MPG, manual transmission cars are better. On average manual transmission cars can achieve a performace of 24.39mpg whereas automatic ones achieve only 17.15mpg. This statement assumes all other variables constant.

Quantify the MPG difference between automatic and manual transmissions

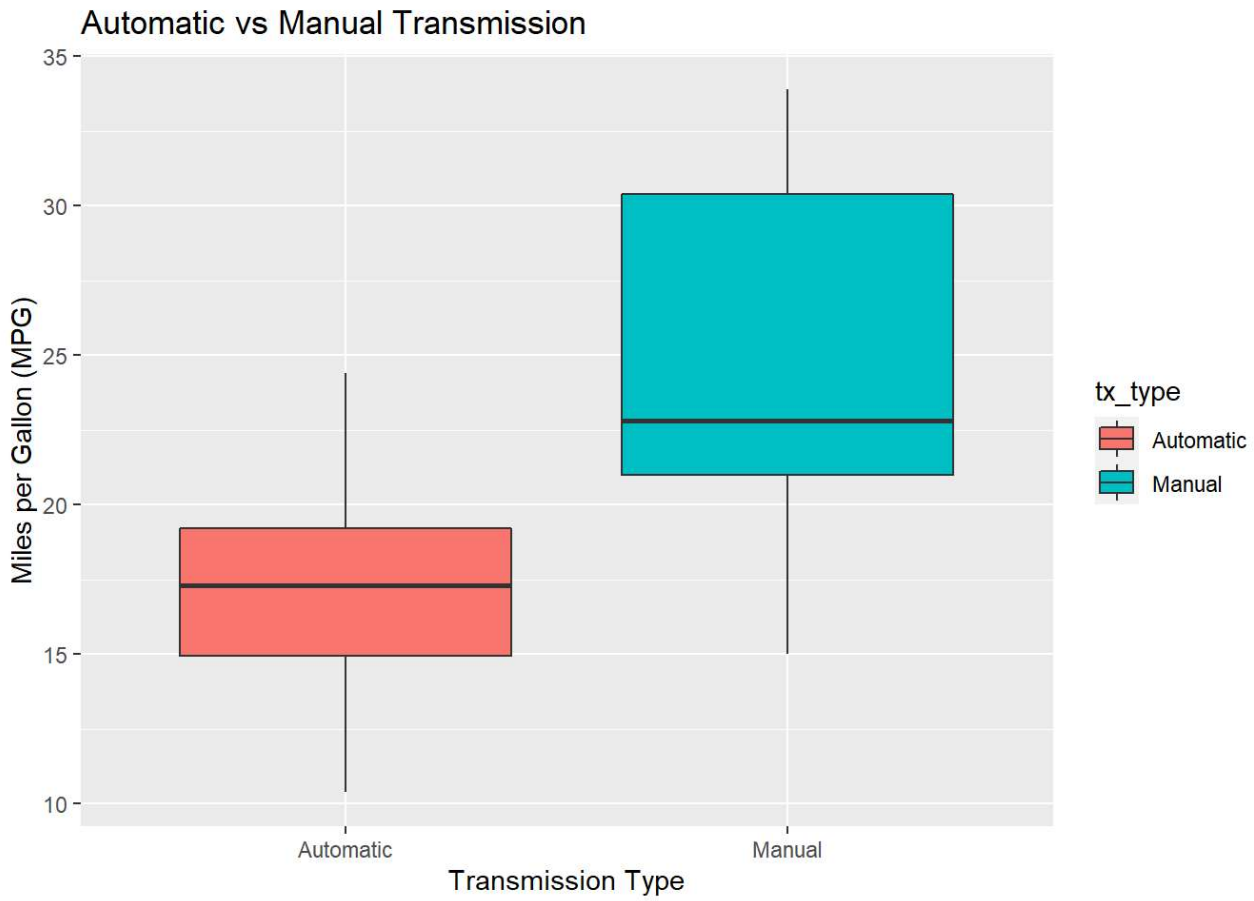
Further analysis shows that when incluing other variables the manual transmission car advantage drops to 1.81 on average.

Based on the analysis performed we conclude that:

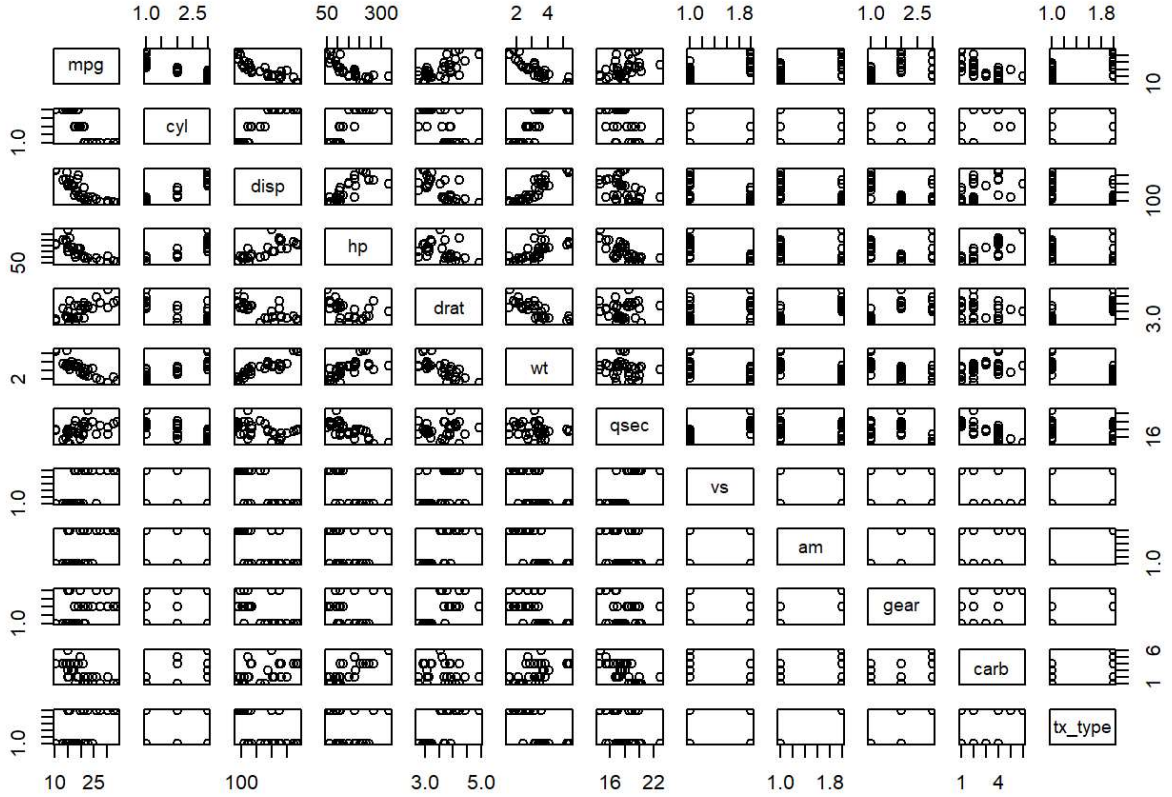
- The best model considers the variables cylinders (cyl), power (hp), weight (wt) and transmission type (am), and explains 87% of the variability in mpg.
- MPG change negatively with cylinders. For 6 cylinders, on average decreases -3.03 miles per galon. For 8 cylinders, on average decreases -2.16 miles per galon.
- MPG changes negatively with power, on average decreases -0.03 miles per galon.
- MPG changes negatively with weight, on average decreases -2.5 miles per galon.
- Manual transmission is 1.81mpg better than automatic transmission.

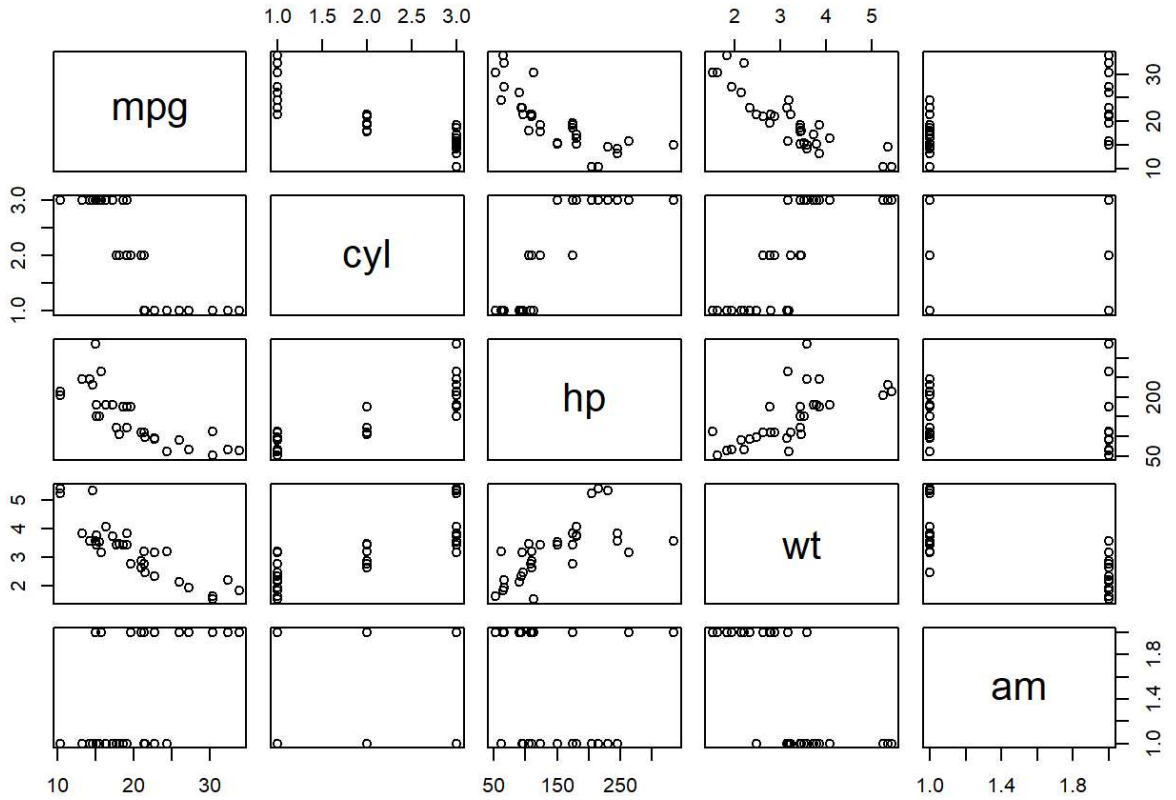
Appendix

A1 - Box Plot



A2 - Scatter Plots





A3 - Residual Plot