Regression Sample

Ali Shinde

26/04/2021

Executive Summary

The main aim of this report is to explore the relationship between mpg(miles per gallon) and another set of variables in the mtcars dataset. The final target features are weight, quarter mile and the transmission on which these are classified. More the weight less is the mileage and opposite to that higher the quarter mile time higher is the mileage. Manual transmission cars give higher mileage than automatic since manual satisfies all the above requirements.

Exploratory Data Analysis

The question has directly stated that we need to find the relation between mpg and the transmission
From the Figure 1 in the appendix section we can see that the mpg values for manual transmission is quite higher than automatic transmission at first glance.
Also we can conclude that the median for manual transmission is also at a higher level than automatic transmission.
Hence we will continue the analysis on the assumption of the boxplot(Figure 1)

Regression Model Analysis

Choosing feauters

The conclusions from the models for this particular topic are very intersting. First the direct relationship was considered between mpg and transmission. The model gives a residual std error of 4.902 on 30 degrees of freedom and can explain ~34% of the variance in the model.
On the other hand when we consider all the features in the dataset the multivariate model gives a residual std. error of 2.65 on 21 degrees of freedom and can explain 80% of the variance. However only the weight parameter is significant at 0.05 significance.

```
mdl1 <- lm(mpg ~ am,mtcars)
mdl2 <- lm(mpg ~ .,mtcars)
```

Removal of unnecessary features

The VIF functions from the car package will help us to give us a headstart in eliminating variables and also will help in confirming.
In any case we have to eliminate those variables in which inter-related varuables that inflate the variance

and caunse instability in the model.
A vif above 10 is bad for the model.

```
vif(mdl2)
```

```
##       cyl      disp        hp      drat        wt      qsec        vs        am
## 15.373833 21.620241  9.832037  3.374620 15.164887  7.527958  4.965873  4.648487
##      gear      carb
##  5.357452  7.908747
```

As shown in this there are 3 variables which are related and are inflating the variance.

From this, the model with the lowest AIC was mpg ~ wt + qsec + am. The model has a residual std. error of 2.459 on 28 degrees of freedom and can also explain 83% of the variance. We will check the VIF for this model as well to confirm our earlier VIF

```
finalmdl <- lm(mpg ~  wt + qsec + am,mtcars)
vif(finalmdl)
```

```
##       wt     qsec       am
## 2.482952 1.364339 2.541437
```

The VIF has dropped down significantly and we can confirm that the features included in the model are uncorrelated.
In the Figure 2 of the Appendix we can see a graph of mpg and weight w.r.t transmission.

Residual Analysis and Diagnostics

Consider Figure 3 1. The Residuals vs. Fitted plot shows no consistent pattern, supporting the accuracy of the independence assumption. 2. The Normal Q-Q plot indicates that the residuals are normally distributed because the points lie closely to the line. 3. The Scale-Location plot confirms the constant variance assumption, as the points are randomly distributed.

```
summary(finalmdl)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## am           2.935837  1.4109045  2.080819 4.671551e-02
```

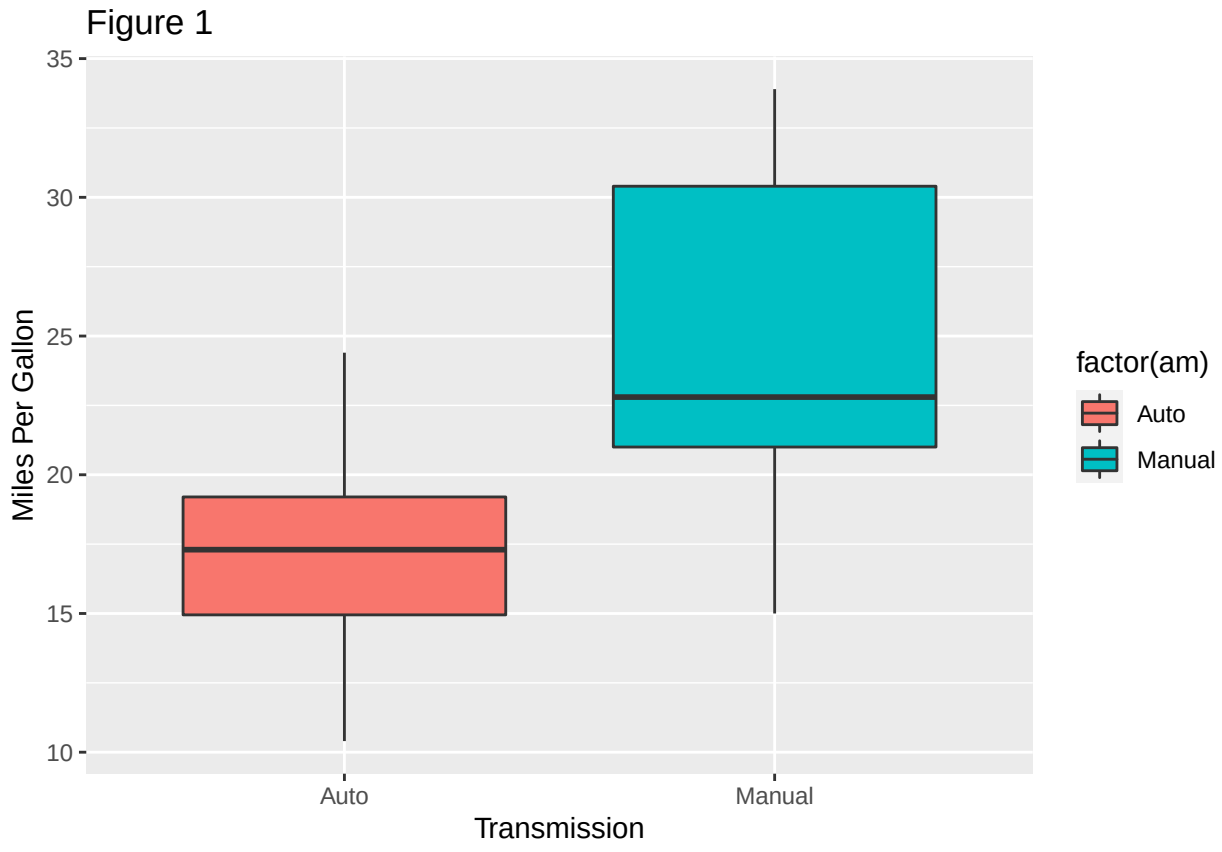
Conclusion

1. All in all it seems that manual transmission is better than automatic considering the negative sign of weight that manual cars are lighter in comaprison and give better mileage but slower than automatic in a quarter mile.

2. The coef of the model are w.r.t automatic transmission i.e there is an increase of mpg from automatic to manual transmission. And the increase in approximately 2.935 miles per gallon.

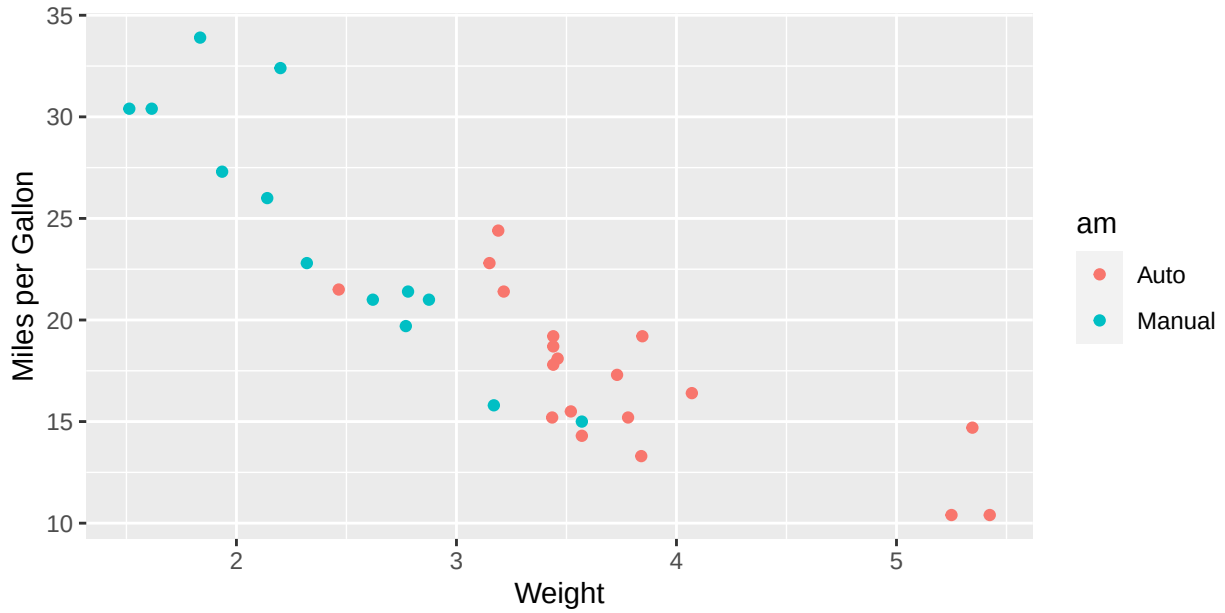
Appendix

```
mtcars$am <- factor(mtcars$am)
levels(mtcars$am) <- c("Auto","Manual")
g <- ggplot(mtcars,aes(x=am,y=mpg)) + geom_boxplot(aes(fill = factor(am)))
g + labs(x="Transmission", y = "Miles Per Gallon",title = "Figure 1")
```




```
ggplot(mtcars,aes(wt,mpg, col = am)) + geom_point() + labs(x = "Weight",y = "Miles per Gallon", title =
```

Figure 2




```
par(mfrow=c(2,2))
plot(finalmdl)
```