Regression Models Course Project

Samir N. Hag Ibrahim

7/13/2020

EXECUTIVE SUMMARY

The magazine known as Motor Trend (MT) covers the automobile industry with an interest in seeing the relationship of MPG (miles per gallan) and a number of variables. The following questions needed to be Answered:

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

The following analysis represents the approach used to answer these two questions. From EDA, Manual transmission cars have significantly ($p<0.05$) higher average miles per gallon (MPG) compared with Autmatic transmission cars (i.e. Manual transmission is better for MPG). Furthermore, the difference between the automatic and manual transimssions was quantified using first Simple Linear Regression (SLR) analysis using type of transmission as a predictor for MPG. SLR only explained 34% of the variations in MPG data ($R^2 = 0.338$) and the model showed that with manual transmission, mpg increased by 7.245 compared with automatic transmission keeping the other variables constant. On the other hand, Multiple Regression (MR) analysis showed better model fit with $R^2 = 0.833$ (i.e. explaining 84% of the variations in the data). Weight "wt", acceleration "qsec" and manual transmission were the only significant variables that affacts mpg ($p<0.05$). ML model also confirmed that manual transmission is significnatly higher mpg (2.9358 mpg) than automatic transmission keeping the other variables constant.

1- DATA LOADING AND EVALUATION

Mtcars data first loaded and the following table shows the first few rows of the data.

Looking at the data structre we observe that mtcars data consist of 11 variables and 32 observations. All variables are numeric variables and there is 0 missing values in the dataset. The transmission variable was transformed into factor variable with two levels "Auto" and "Manual".

2- EXPLORATORY DATA ANALYSIS

From EDA, there are many possible relationships between the variable in the dataset as shown in the correlation plot (*see appendix 1 for the code*). The boxplot figure (*see appendix 2*) shows a comparison between "Auto" and "Manual" transmission for mpg.

```
##
##  Welch Two Sample t-test
##
## data:  auto_cars$mpg and manual_cars$mpg
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

The boxplot showed that Automatic cars has mean mpg of 17.1 compared with 24.4 for Manual cars, indicating that Manual cars has higher mpg. From the t.test these means are significantly different with p-value = 0.001, hence rejecting the null hypothesis and concluding that *Manual transmission is better for mpg* (see appendix 3 for the code).

3- QUANTIFYING THE RELATIONSHIPS

In order to quantify the relationship between mpg and transmission, we needs to perform first simple linear regression (SLR) analysis, the fowllowing table shows the regression coeficients.

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04
```

From SLR analysis, the slope for automatic and manual transmissions were 17.147 and7.245 respectively. Manual cars showed a higher mpg rate by 7.24 mpg. However, R^2 equals 0.338 i.e. only 34% of the variations in the data were explained by *am* as a variable. More variable need to be included in th model.

4- MULTIPLE REGRESSION (MR) ANALYSIS

```
##              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
## amManual     2.935837  1.4109045  2.080819 4.671551e-02
```

MR better fits the data and explained 84% of the variations within the dataset ($R^2 = 0.834$) with weight "wt", acceleration "qsec" and Manual transmission "amManual" as the only variables significant variables. Manual transmission has an uplift of mpg higer than that for the Automatic transmission keeping the other variables constant.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + qsec + am
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.90
## 2     28 169.29  2    551.61 45.618 1.55e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

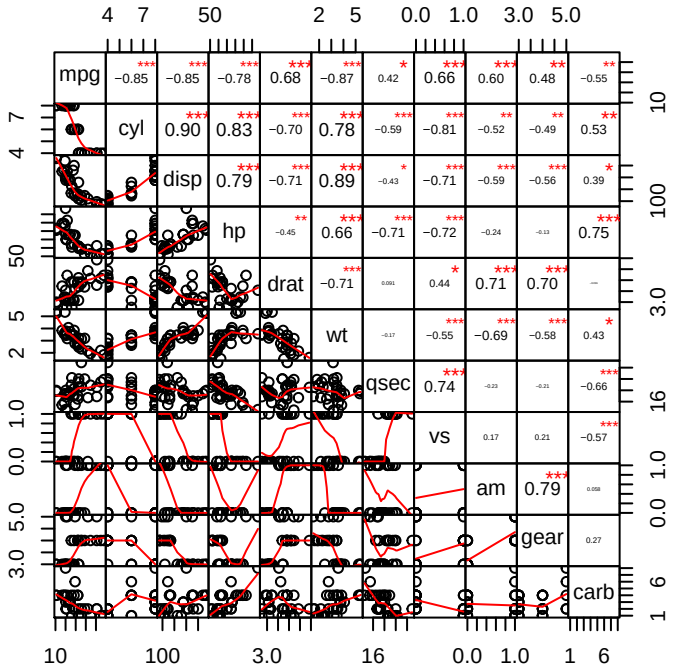
According to Anova test, there is a significant difference between the two model, therefore, rejecting the null hypothesis that other variables have no significant effect on mpg performance. In addition, the diagnostic plots (*see appendix 5*) showed that the residuals' distribution in ML model (fit2) is much better than in SLR model (fit1) and hence the results from fit2 was accepted.

5- CONCLUSION

Non we can conclude that manual cars are better for the mpg, and it has more compared with the automatic cars keeping the other variables fixed

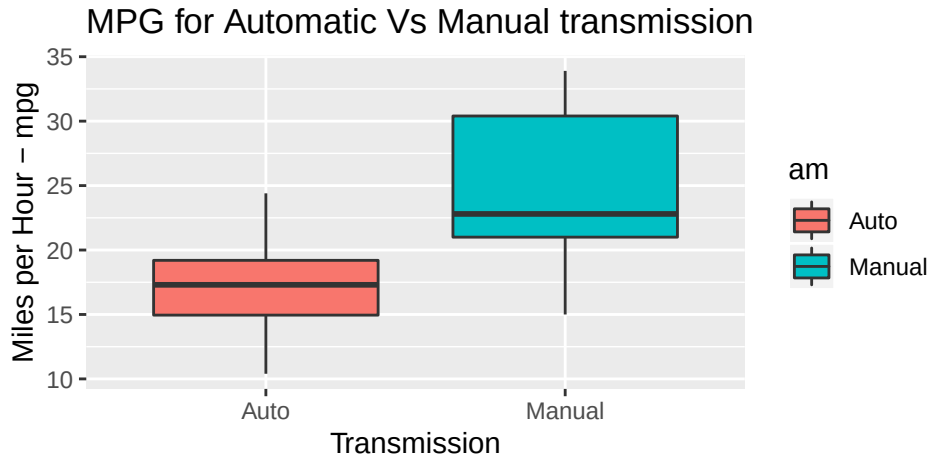
##APPENDIX ### 1- Correlation plot

```
data(mtcars)
chart.Correlation(mtcars, histogram = FALSE, pch = 12)
```



2- Code for boxplot MPG for Auto vs Manual

```
ggplot(mtcars, aes(am, mpg))+
        geom_boxplot(aes(fill= am))+
        ggtitle("MPG for Automatic Vs Manual transmission")+
        labs(x = "Transmission", y = "Miles per Hour - mpg")
```



2- Code for mpg means.

```
aggregate(mpg~am, mtcars, mean)
```

3- Code for t.test for the mpg means

```
auto_cars <- subset(mtcars, am == "Auto") ; nrow(auto_cars)
manual_cars <- subset(mtcars, am =="Manual"); nrow(manual_cars)
# nrow() just to check the subset is correct
```

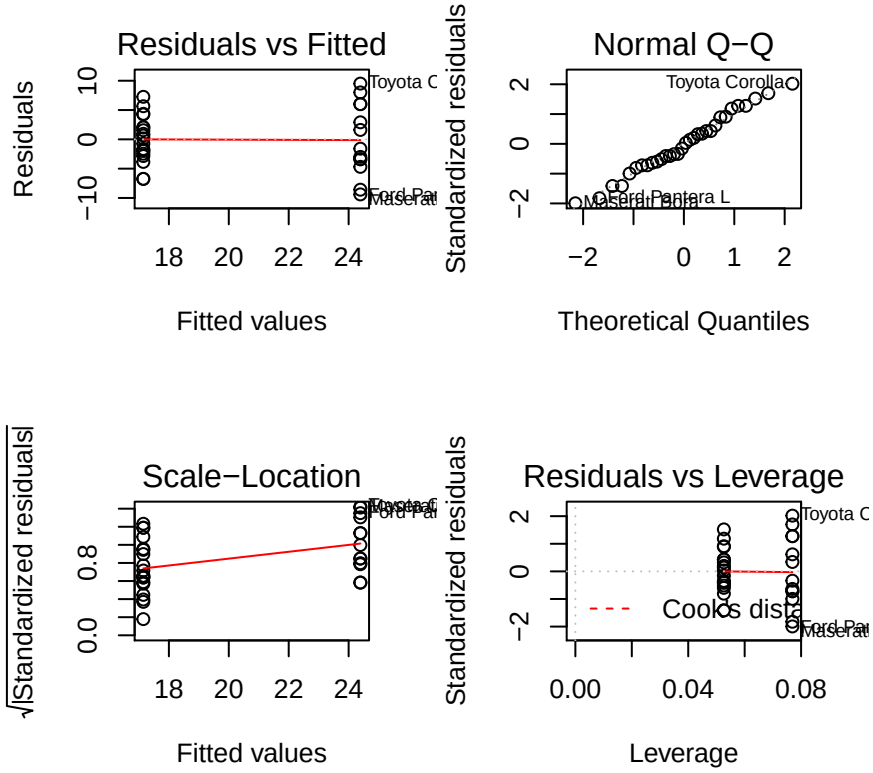
4- Anova for comparing the two models

```
fit2 <- step(lm(mpg~., mtcars),direction = "both" ,trace = 0)
anova <- anova(fit1,fit2)
```

5- Model Diagnostic plots

A- SLR model

```
par(mfrow=c(2,2))
plot(fit1)
```



B- ML model

```
par(mfrow=c(2,2))
plot(fit2)
```