Motor Trend Project

Manuel Solans

December 3rd 2021

Executive Summary

The content of this document refers to the Assgiment of the Week 4 of the Linear Regression course. Through it you will find an analysis to determine wether automatic or manual transmissions in cars are better for gas consumptions of a car based on miles per galon.

Loading data

For this assigment I am going to use data from mtcars table. And you can see a summary below.

```
library(ggplot2)
library(corrplot)

## corrplot 0.92 loaded

library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

data("mtcars")
#Let's have a look into the content of 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  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

dim(mtcars)
```

```
## [1] 32 11

datacars<-mtcars
```

This mtcars dataset has 32 rows and 11 columns.

Question 1: "Is an automatic or manual transmission better for MPG"

Provide a basic summary of the transmission column data.

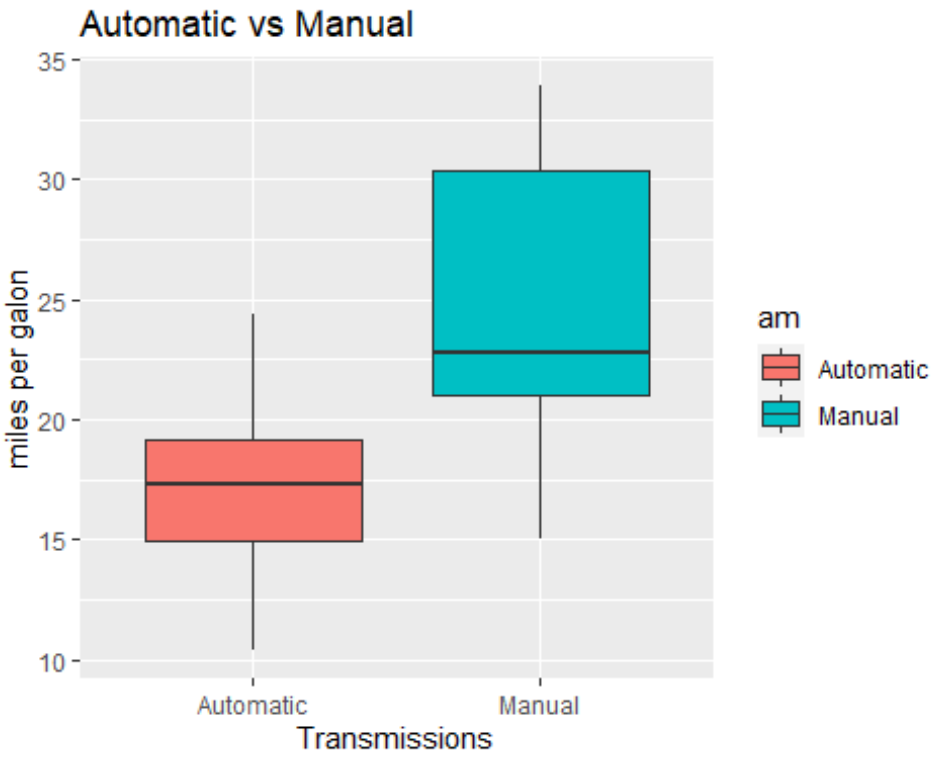
```
summary(mtcars$am)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  0.0000  0.4062  1.0000  1.0000

# We have to level both tranmissions within the data column
datacars$am <- as.factor(datacars$am)
levels(datacars$am) <-c("Automatic", "Manual")
```

As we have the data to start with, let's look into the content

```
aggregate(mpg ~ am, data = datacars, mean)

##          am      mpg
## 1 Automatic 17.14737
## 2    Manual 24.39231

autom.Data <- datacars[datacars$am == "Automatic",]
manual.Data <- datacars[datacars$am == "Manual",]
t.test(autom.Data$mpg, manual.Data$mpg)

##
##  Welch Two Sample t-test
##
## data:  autom.Data$mpg and manual.Data$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
```

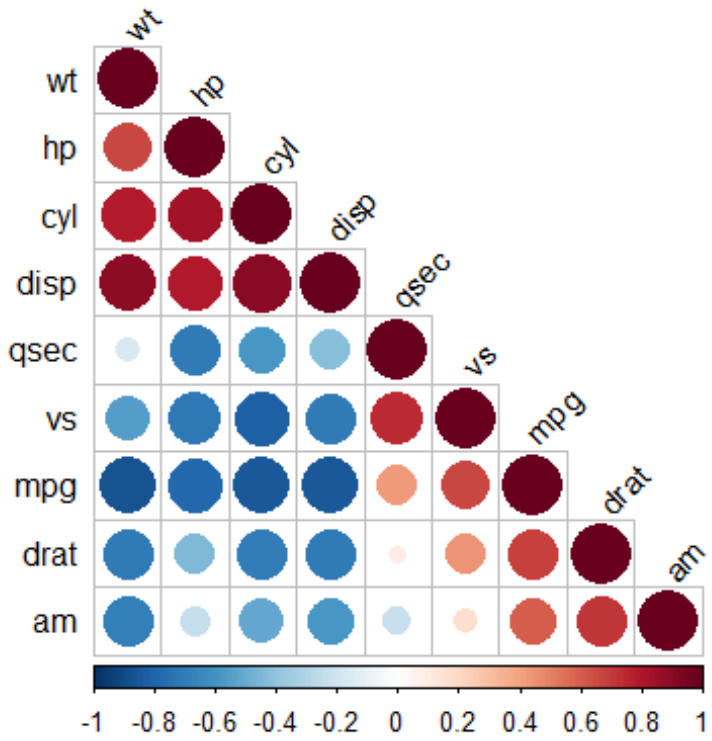
Exploratory Data Analysis

```
plot1 <- ggplot(datacars, aes(x=am, y=mpg,fill=am)) +  geom_boxplot() +
ggtitle("Automatic vs Manual") + xlab("Transmissions") + ylab("miles per
galon")
plot1
```



What we can observe in this chart is that the mpg (miles per galon) mean in mtcars depends on wether the transmission is Automatic or Manual. And the manual tranmission implies 7.245 mpg more.

```
source("http://www.sthda.com/upload/rquery_cormat.r")
mydata <- mtcars[, c(1,2,3,4,5,6,7,8,9)]
rquery.cormat(mydata)
```




```
## $r
##         wt    hp   cyl  disp  qsec   vs  mpg drat am
## wt       1
## hp    0.66     1
## cyl   0.78  0.83     1
## disp  0.89  0.79   0.9     1
## qsec -0.17 -0.71 -0.59 -0.43     1
## vs   -0.55 -0.72 -0.81 -0.71  0.74    1
## mpg  -0.87 -0.78 -0.85 -0.85  0.42 0.66    1
## drat -0.71 -0.45  -0.7 -0.71 0.091 0.44 0.68    1
## am   -0.69 -0.24 -0.52 -0.59 -0.23 0.17  0.6 0.71  1
##
## $p
##           wt      hp     cyl    disp  qsec      vs     mpg    drat am
## wt         0
## hp   4.1e-05       0
## cyl  1.2e-07 3.5e-09       0
## disp 1.2e-11 7.1e-08 1.8e-12       0
## qsec    0.34 5.8e-06 0.00037   0.013     0
## vs   0.00098 2.9e-06 1.8e-08 5.2e-06 1e-06       0
## mpg  1.3e-10 1.8e-07 6.1e-10 9.4e-10 0.017 3.4e-05       0
## drat 4.8e-06    0.01 8.2e-06 5.3e-06  0.62   0.012 1.8e-05       0
## am   1.1e-05    0.18  0.0022 0.00037  0.21    0.36 0.00029 4.7e-06  0
##
## $sym
##      wt hp cyl disp qsec vs mpg drat am
## wt   1
```

```
## hp   ,  1
## cyl  ,  +  1
## disp +  ,  +   1
## qsec    ,  .   .    1
## vs   .  ,  +   ,    ,    1
## mpg  +  ,  +   +    .    ,  1
## drat ,  .  ,   ,         .  ,   1
## am   ,     .   .           .   ,    1
## attr(,"legend")
## [1] 0 ' ' 0.3 '.' 0.6 ',' 0.8 '+' 0.9 '*' 0.95 'B' 1
```

Comments

The result of rquery.cormat function is a list containing the following components :

r : The table of correlation coefficients p : Table of p-values corresponding to the significance levels of the correlations sym : A representation of the correlation matrix in which coefficients are replaced by symbols according to the strength of the dependence.

In the generated graph, negative correlations are in blue and positive ones in red color.

Correlations: cyl (0.78), disp (0.89) and hp (0.66) show highest correlation. # Let's apply linear regression

Applying linear regression with transmission vs miles per galon

```
lm1 <- lm( mpg ~ am, data = datacars)
summary( lm1 )

##
## Call:
## lm(formula = mpg ~ am, data = datacars)
##
## 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
```

Conclusions

Manual cars deliver 7.245 miles per galon more than Automatic cars. The results of this model explains ~36% of the variance (R-Squared value). Other factors could influence as well as weight, acceleration, etc.