Regression Model Project

Expedito Pinto de Paula Junior

10/16/2020

1. Summary

Motor Trend Magazine about the automobile industry are looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

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

2. Exploratory Data Analysis

2.1 Load Libraries and Data Set

```
library(datasets)
library(ggplot2)
data("mtcars")
```

2.2 Basic Exploratory Data Analysis

```
# Sample 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
```

```
# Variables
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
```

```
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

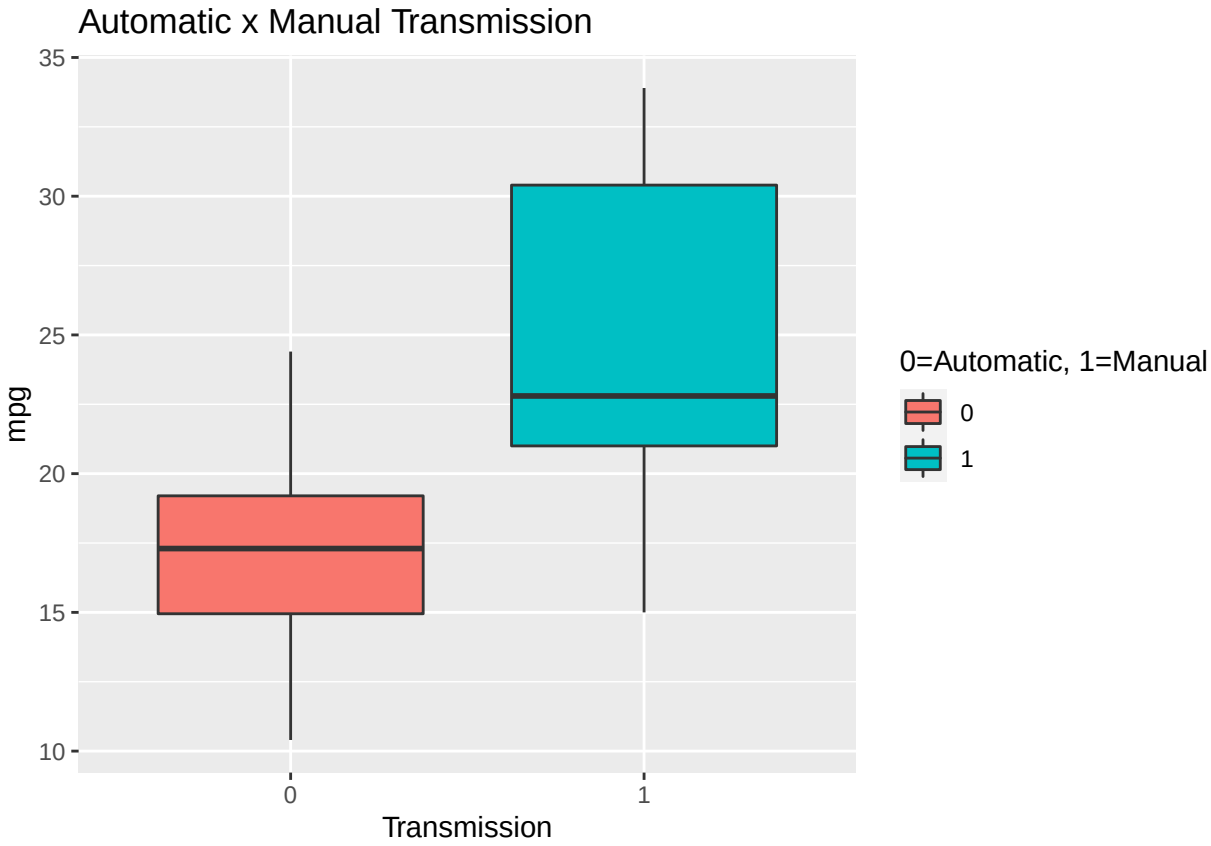
```
# summary
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

2.3 Automatic x Manual Transmission

```
mtcars$am <- as.factor(mtcars$am)

ggplot(aes(x=am, y=mpg), data=mtcars) +
     geom_boxplot(aes(fill=am)) +
     labs(title = "Automatic x Manual Transmission") +
     xlab("Transmission") +
     ylab("mpg") +
     labs(fill="0=Automatic, 1=Manual")
```



The plot above indicates that manual transmission provide better mileage per galon than the automatics.

```
AutomaticT <- mtcars[mtcars$am == "0",]
ManualT <- mtcars[mtcars$am == "1",]
t.test(AutomaticT$mpg, ManualT$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  AutomaticT$mpg and ManualT$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 hypothesis test performed rejects the null hypothesis that the transmission type is in fact significantly correlateed to mileage per gallon (mpg).

3.Regression Models

3.1 Linear Regression Model

```
lmModel <- lm(mpg ~ am, data=mtcars)
summary(lmModel)
```

```
##
## 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 ***
## am1            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
```

The linear regression shows how much transmission type affect mpg performance. Specifically manual transmissions provide 7.25 MPG better performance than automatics. But, based in R-squares value, transmission types only explain 36% of the MPG performance, based on this a simple linear regression is not enought to answer the Motor Trend's questions.

3.2 Multivariable Regression Model

```
mModel <- lm(mpg ~ am + cyl + hp + wt, data = mtcars)
anova(lmModel, mModel)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl + hp + wt
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.9
## 2     27 170.0  3     550.9 29.166 1.274e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(mModel)
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + hp + wt, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4765 -1.8471 -0.5544  1.2758  5.6608
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 36.14654    3.10478  11.642 4.94e-12 ***
## am1          1.47805    1.44115   1.026   0.3142
## cyl         -0.74516    0.58279  -1.279   0.2119
## hp          -0.02495    0.01365  -1.828   0.0786 .
```

```
## wt          -2.60648    0.91984  -2.834   0.0086 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.509 on 27 degrees of freedom
## Multiple R-squared:  0.849,  Adjusted R-squared:  0.8267
## F-statistic: 37.96 on 4 and 27 DF,  p-value: 1.025e-10
```

The multivariable regression model (MRM) includes other factors (variables) that most likely affect a vehicle's including number of cylinders, engine horsepower, vehicle weight and others. The MRM shows that:

1. Each 1.48 MPG increase from manual transmission over automatic.
2. The multivariable model explain 85% of the MPG performance.

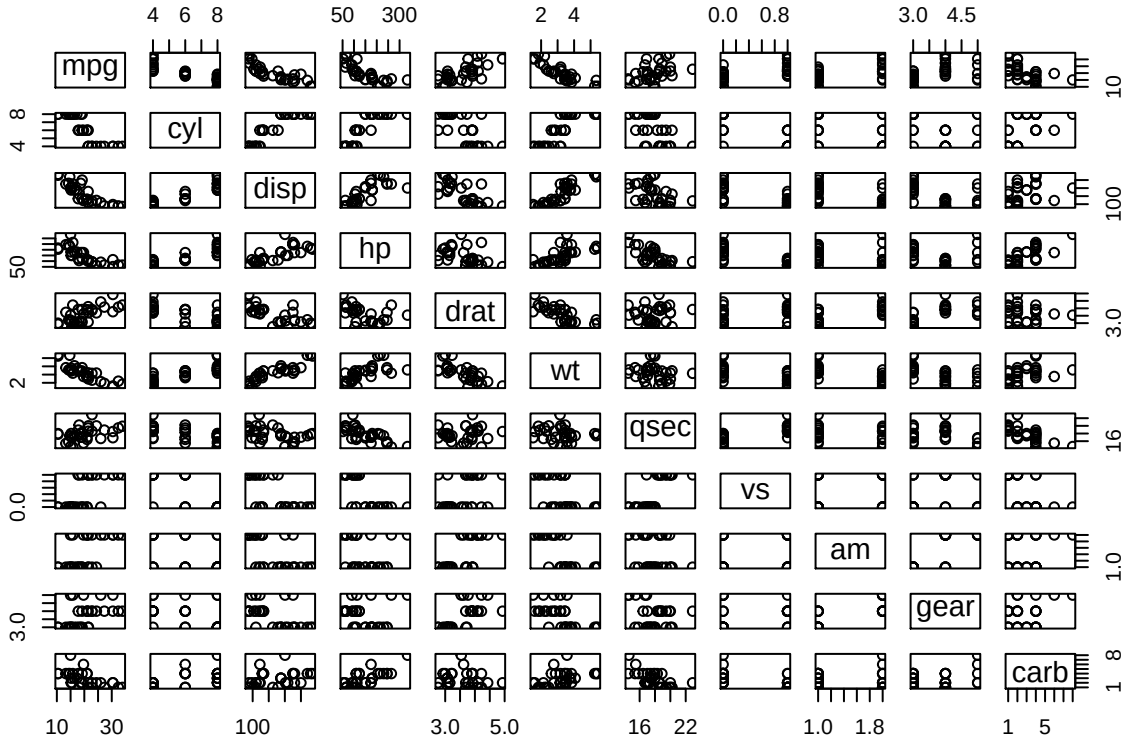
4. Conclusions

1. **"Is an automatic or manual transmission better for MPG?"** A: Based in MPG, manual transmission perform better than automatic.
2. **"Quantify the MPG difference between automatic and manual transmissions?"** A: In the first model, manual transmission perform better than automatic by 7.25 MPG, however this factor explain only 36% of the relation. When measure MPG adding cylinders, horsepower and weight variable, the manual transmission provide an additional 1.48 MPG of performance over automatic accounting 85% of the relation explanation.

5. Appendix

As appendix is showed a pairs plot based in mpg and residual from multivariable regression model.

```
pairs(mpg ~ ., data = mtcars)
```



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