Regression

Kaaltho

26/12/2020

Executive Summary

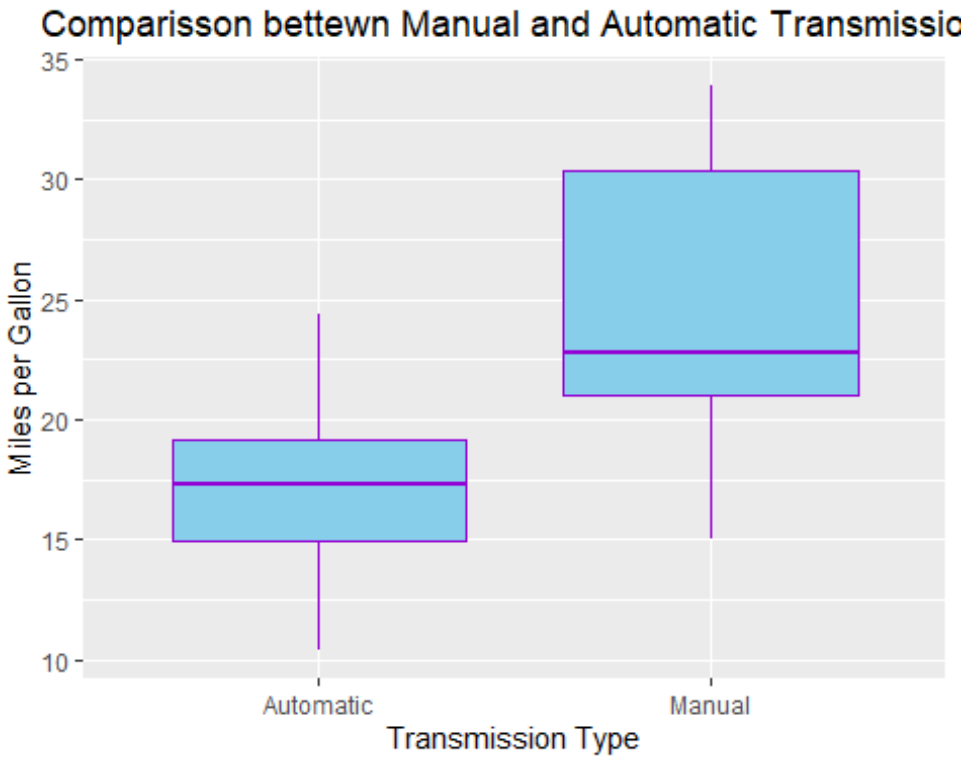
One of the features that are important in cars is miles per gallon (MPG) consume, every vehicle has his own specifications, one of the most frequent question ask by buyers about MPG consume is: What is better a manual transmission or automatic transmission. This analysis provide an aswer to the following questions:

1. “Is an automatic or manual transmission better for MPG”
2. “Quantify the MPG difference between automatic and manual transmissions”

Exploratory data analysis

The data used to this analysis is from Motor Trend US magazine from 1974, which give some metric for the fuel consumption and performance of 32 different automobiles

```
#Oppening libraries
library(ggplot2)
library(cowplot)
#Loading the data:
data(mtcars)
#changing the type of variables for convenience in order to do
calculations.
mtcars2 <- mtcars
cols <- c("cyl", "vs", "gear", "carb")
mtcars2[cols] <- lapply(mtcars2[cols], factor)
#new column name transmissions for better plotting understanding
mtcars2$transmission <- factor(mtcars$am, labels=c("Automatic","Manual"))
#mean fo automatic and manual transmission
mean.trans <- aggregate(data=mtcars2, mpg~transmission, mean)
#Visualization of the data
plot1 <- ggplot(mtcars2, aes(x=transmission, y=mpg)) +
  geom_boxplot(colour="darkviolet", fill="skyblue")+
  guides(fill=FALSE) +
  labs(x="Transmission Type", y="Miles per Gallon", title=("Comparisson
bettewn Manual and Automatic Transmission"))+
  theme(plot.title = element_text(hjust = 0.5))
plot1
```



It is possible to make initial assumptions, for instance, in the figure there is a clear difference between Automatic transmission and Manual transmission it seem that Automatic transmission is more efficient when it comes to fuel consumption with a mean of 17.15 Miles per galon , while Manual transmission show a consumption of 24.39 Miles per galon.

Applying Regresion models

Considering the previous results it is possible to say that there is a difference of 7.25 MPG in relation to the type of transmission, being Automatic transmission the one that consumes less fuel. This assumption could be analyzed through a simple linear regression.

```
#linear regression
linear.regression <- lm(mpg ~ factor(am), data=mtcars2)
summary(linear.regression)
```

From the calculation p-value is 0.000285, the hypothesis is not rejected, seeing the R-squared value can be said that a third of the variance (36%) could be attributed to just transmission type. To further analysis the next step is to do an Analysis of Variance.

```
#variance analysis
mt.variance <- aov(mpg ~ ., data = mtcars2)
summary(mt.variance)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          2  824.8   412.4  51.377 1.94e-07 ***
```

```
## disp         1   57.6    57.6   7.181   0.0171 *
## hp           1   18.5    18.5   2.305   0.1497
## drat         1   11.9    11.9   1.484   0.2419
## wt           1   55.8    55.8   6.950   0.0187 *
## qsec         1    1.5     1.5   0.190   0.6692
## vs           1    0.3     0.3   0.038   0.8488
## am           1   16.6    16.6   2.064   0.1714
## gear         2    5.0     2.5   0.313   0.7361
## carb         5   13.6     2.7   0.339   0.8814
## Residuals   15  120.4     8.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Some of the variables that could affect the variance in greater way are cyl, disp and wt. This variables are going to be used in the next analysis to see how those affect MPG.

```
#Multivariable analysis
mt.multi <- lm(mpg ~ am + cyl + disp + wt, data = mtcars2)
summary(mt.multi)

##
## Call:
## lm(formula = mpg ~ am + cyl + disp + wt, data = mtcars2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5029 -1.2829 -0.4825  1.4954  5.7889
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.816067   2.914272  11.604 8.79e-12 ***
## am           0.141212   1.326751   0.106  0.91605
## cyl6        -4.304782   1.492355  -2.885  0.00777 **
## cyl8        -6.318406   2.647658  -2.386  0.02458 *
## disp         0.001632   0.013757   0.119  0.90647
## wt          -3.249176   1.249098  -2.601  0.01513 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.652 on 26 degrees of freedom
## Multiple R-squared:  0.8376, Adjusted R-squared:  0.8064
## F-statistic: 26.82 on 5 and 26 DF,  p-value: 1.73e-09
```

This model show that 83% of the variability can be attribute to the cyl, disp, hp and wt variables, furthermore p-value for cy16 and cy18 are below 0.5, denoting that those variables could have a fair share of influence over the variability of Miles per Gallon so we can say that there is a correlation between those variables and transmission type and Miles per Gallon.

MPG difference between automatic and manual transmission

```
modelfit <- lm(mpg~., data=mtcars2)
stepmodel <- step(modelfit)
summary(stepmodel)$coeff
```

Finally the diferrence between automatic and manual transmissions is 1.81 MPG.

Appendix

Residual Analysis

The "Residuals vs Fitted" plot shows that the residuals are homoscedastic. Also can be seen that they are normally distributed.

```
par(mfrow = c(2,2))
plot(mt.multi)
```