Regression Models

Anushri

27/08/2019

Summary

This data analysis is for Motor Trend, a magazine about the automobile industry. 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). In this analysis of mtcar dataset we need to analyse data from ataomatic and manual transmission. In accoradance of this exploratory data analysis is performed. To answer the question that whether a manual or automatic transmission is better, some regression models are constructed. From this analysis one can conclude that manual transmisson gives better performance than automatic.

Exploratory Data Analysis

```
library(datasets)
data(mtcars)
head(mtcars)
```

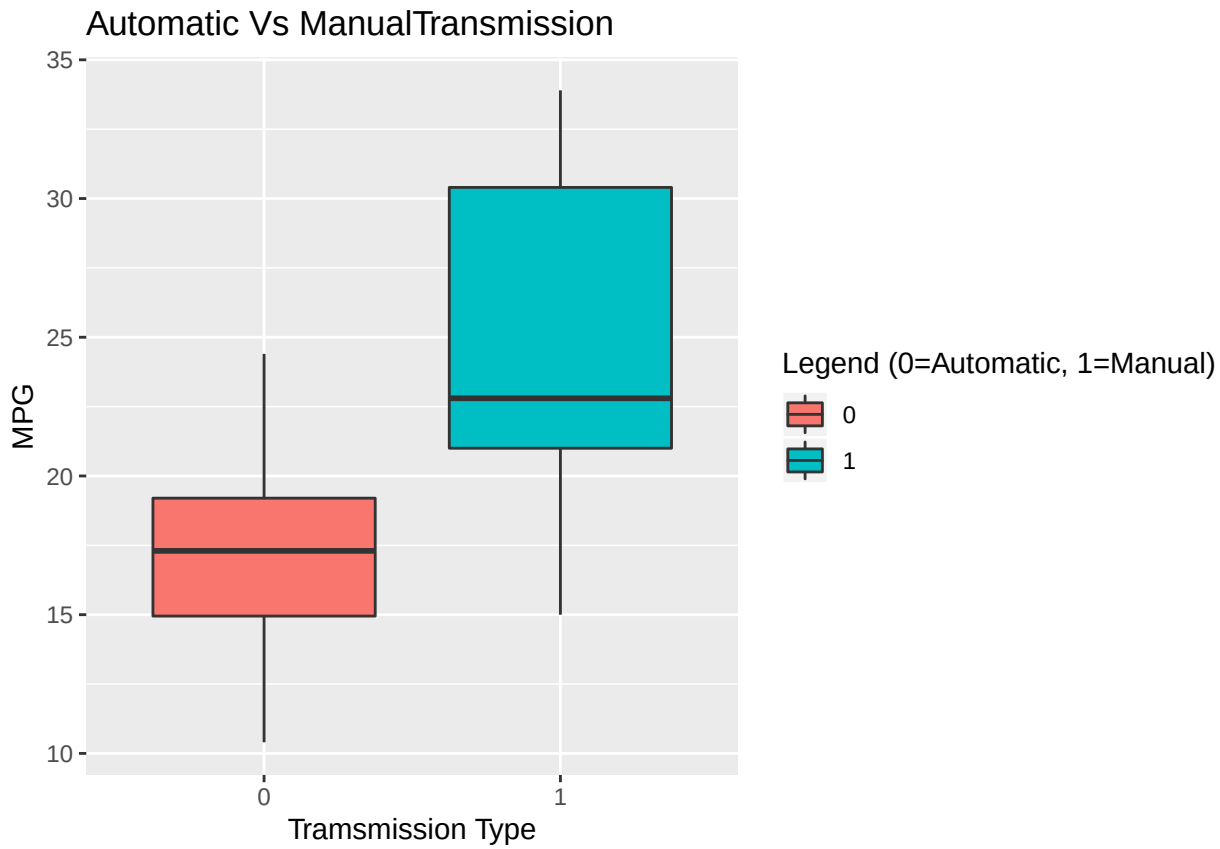
```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 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
```

```
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 ...
```

Let's see the graphical interpretation

```
library(ggplot2)
mtcars$am=as.factor(mtcars$am)
g<-ggplot(mtcars,aes(x=am, y=mpg)) + geom_boxplot(aes(fill=am))
g<-g + ggtitle("Automatic Vs ManualTransmission")+xlab("Tramsmission Type")+ylab("MPG")
g<- g+ labs(fill= "Legend (0=Automatic, 1=Manual)")
g
```



The boxplot clearly indicates that manual transmissions provide better MPG than automatic transmissions.

Simple Linear Regression Model

```
fit<- lm(mpg~am,mtcars)
summary(fit)
```

```
##
## 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
```

In a simple linear regression model transmission type, specifically manual transmissions, provide 7.25MPG better performance than automatic transmission. The R-squared value is 0.3598, i.e.trasmission types only explain 36% of the MPG performance, and thus this simple linear regression is not a very good model to answer Motor Trend's questions.

Multivariable Regression Model

```
fit1<- lm(mpg~.,mtcars)
summary(fit1)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am1          2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

This multivariable regression model gives R-squared value as 0.849,i.e.transmission types explains 85% of the MPG performance.

Residual Plot

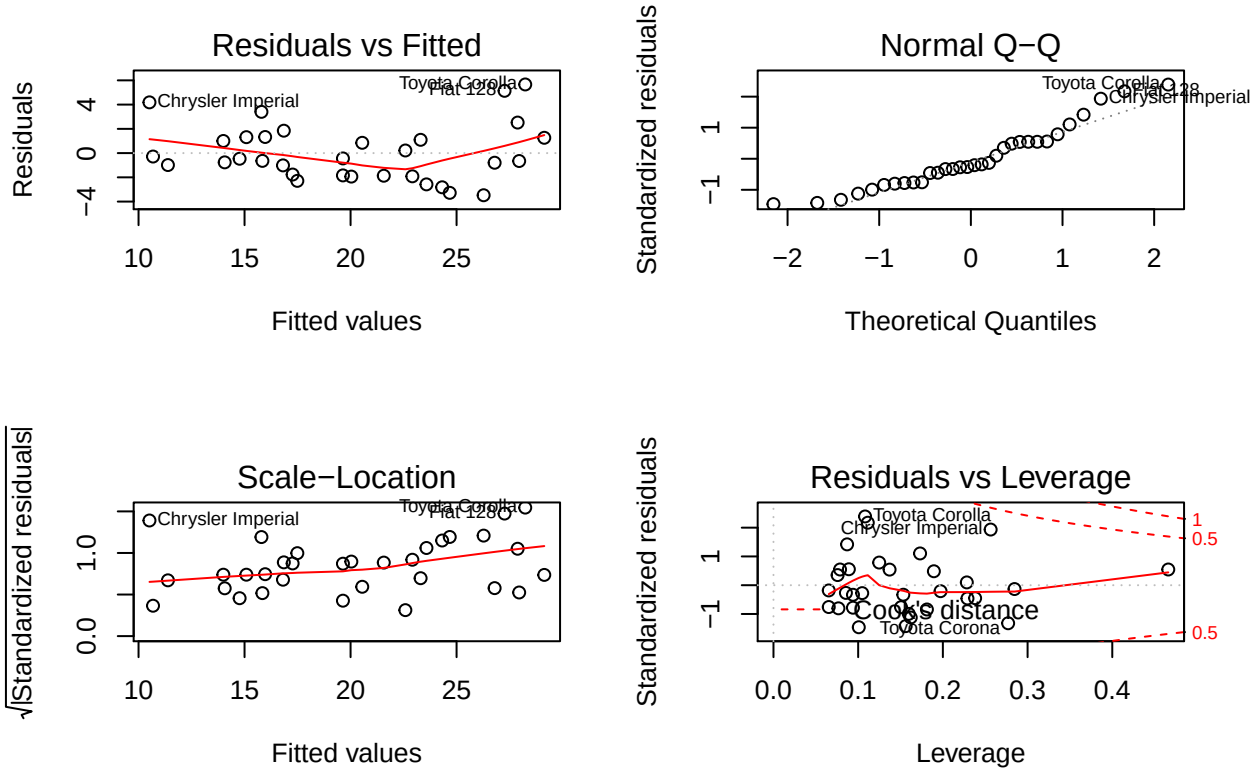
In residual plot we will consider only am,cyl,wt,hp as our predictors

```
fit3 <- lm(mpg~am + cyl + hp + wt, data = mtcars)
summary(fit3)
```

```
##
## 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
```

```
par(mfrow = c(2, 2))
plot(fit3)
```



We can thus conclude that 'wt' and 'cyl'are confounding variables in the relationship between 'am and 'mpg' and that manual transmission cars on average have 1.45 miles per gallon more than automatic cars.

Appendix

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