

Motor_Trend_Regression_Model

Bamidele Tella

29/9/2020

Executive Summary

In this project, we take a look at a data set of a collection of cars, we aim to explore the relationship between a set of variables and miles per gallon (MPG) (outcome). At the completion of the project, we shall answer the following two questions: 1. "Is an automatic or manual transmission better for MPG?" 2. "what value(s) Quantifies the MPG difference between automatic and manual transmissions?"

Loading Required Library

First, the motor trend dataset was loaded into R, along with the required libraries for analysis.

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

Summary of Dataset

Next, using R's head function, the top rows of the dataset was viewed to see how it appears, then the properties of each column with the str function was viewed next. Also we changed the transmission type into a factor class of automatic(0) or manual(1).

```
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 ...
```

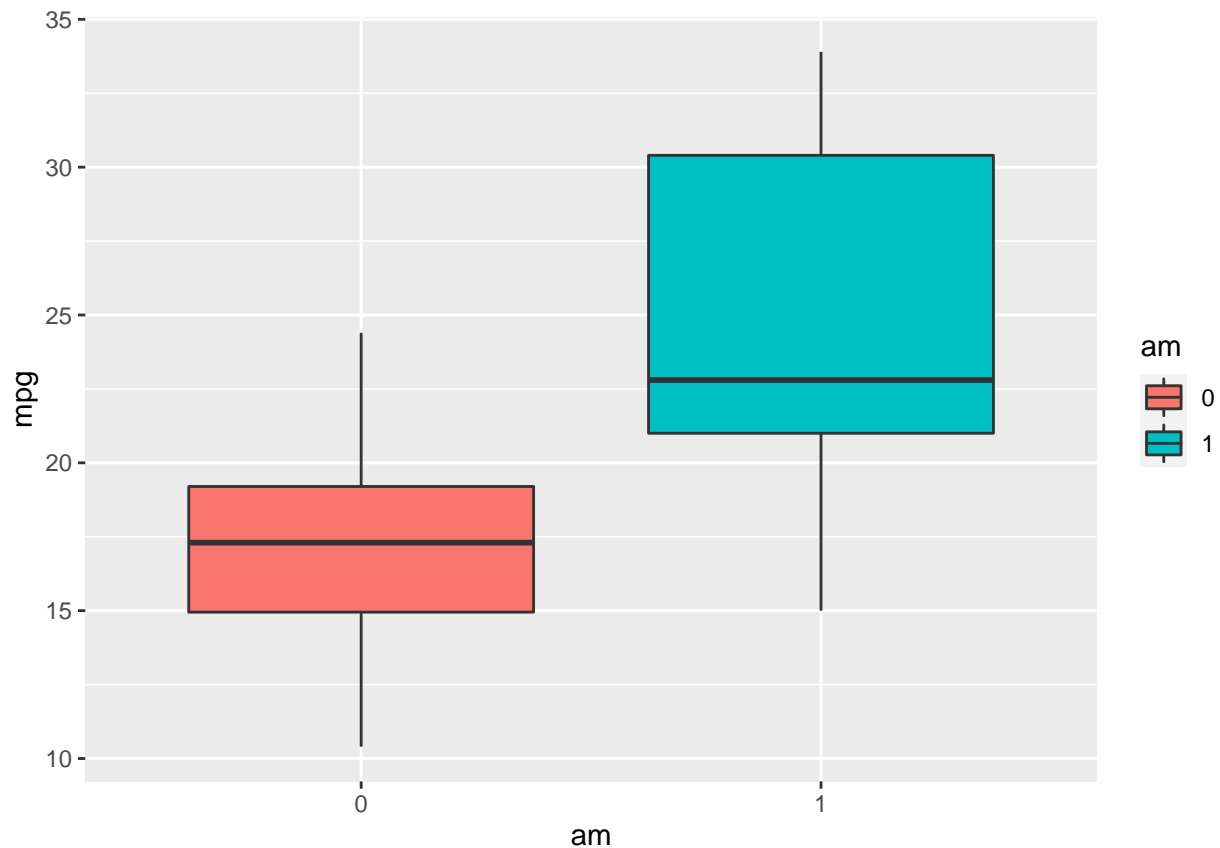
```
## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...
## $ disp: num 160 160 108 258 360 ...
## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...
## $ qsec: num 16.5 17 18.6 19.4 17 ...
## $ vs : num 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 ...
```

```
mtcars$am=as.factor(mtcars$am)
```

Exploratory Data Analysis

Next, a boxplot was visualized to understand the relationship between transmission type(am) and the miles per gallon(mpg).

```
gfit1=ggplot(mtcars,aes(x=am,y=mpg,fill=am))
gfit1=gfit1 + geom_boxplot()
gfit1
```



```

manual<-mean(mtcars[mtcars$am=="1",]$mpg)
auto<-mean(mtcars[mtcars$am=="0",]$mpg)
newTab<-data.frame(manual=manual,auto=auto)
rownames(newTab)<- "Mean"
newTab

```

```

##          manual      auto
## Mean 24.39231 17.14737

```

Is an automatic or manual transmission better for MPG?

From the boxplot and the mean analysis above, it can be seen that the cars with manual transmission type(1), offer more average miles per gallon than the cars with automatic transmission type(0).

Model Selection

The anova function was then used to determine the best model fit for the regression model. First, I generated a model of the miles per gallon(mpg) with all variables , then I generated a model with specific variables starting with transmission type(am),then I added the number of cylinder(cyl) as a confounding variable to the transmission type,then I added weight(wt), gross horsepower(hp), and displacement(displacement) respectively in different model fitting.

```

fit1=lm(mpg~.,data = 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
```

```
fit2=lm(mpg~am,data = mtcars)
fit3=lm(mpg~am+cyl,data = mtcars)
fit4=lm(mpg~am+cyl+wt,data = mtcars)
fit5=lm(mpg~am+cyl+hp+wt,data = mtcars)
fit6=lm(mpg~am+cyl+hp+wt+disp,data = mtcars)
anova(fit2,fit3,fit4,fit5,fit6,fit1)
```

```
## Analysis of Variance Table
##
```

```
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl + wt
## Model 4: mpg ~ am + cyl + hp + wt
## Model 5: mpg ~ am + cyl + hp + wt + disp
## Model 6: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      30 720.90
## 2      29 271.36  1    449.53 64.0039 8.231e-08 ***
## 3      28 191.05  1     80.32 11.4351 0.002817 **
## 4      27 170.00  1     21.05  2.9970 0.098089 .
## 5      26 163.12  1      6.88  0.9793 0.333646
## 6      21 147.49  5     15.63  0.4449 0.812059
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the variance analysis done, I deduced that model 5 and model 6 do not offer significant difference to the change in miles per gallon hence we can assume that model 4 is the best model to quantify the effects of different variables on mile per gallon(mpg).

Summary of Selected Model

After selection of the best model fit, I use the summary function to know the corresponding effect of each variable on miles per gallon(mpg).

```
fit5=lm(mpg~am+cyl+hp+wt, data = mtcars)
summary(fit5)
```

```
##
## 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
```

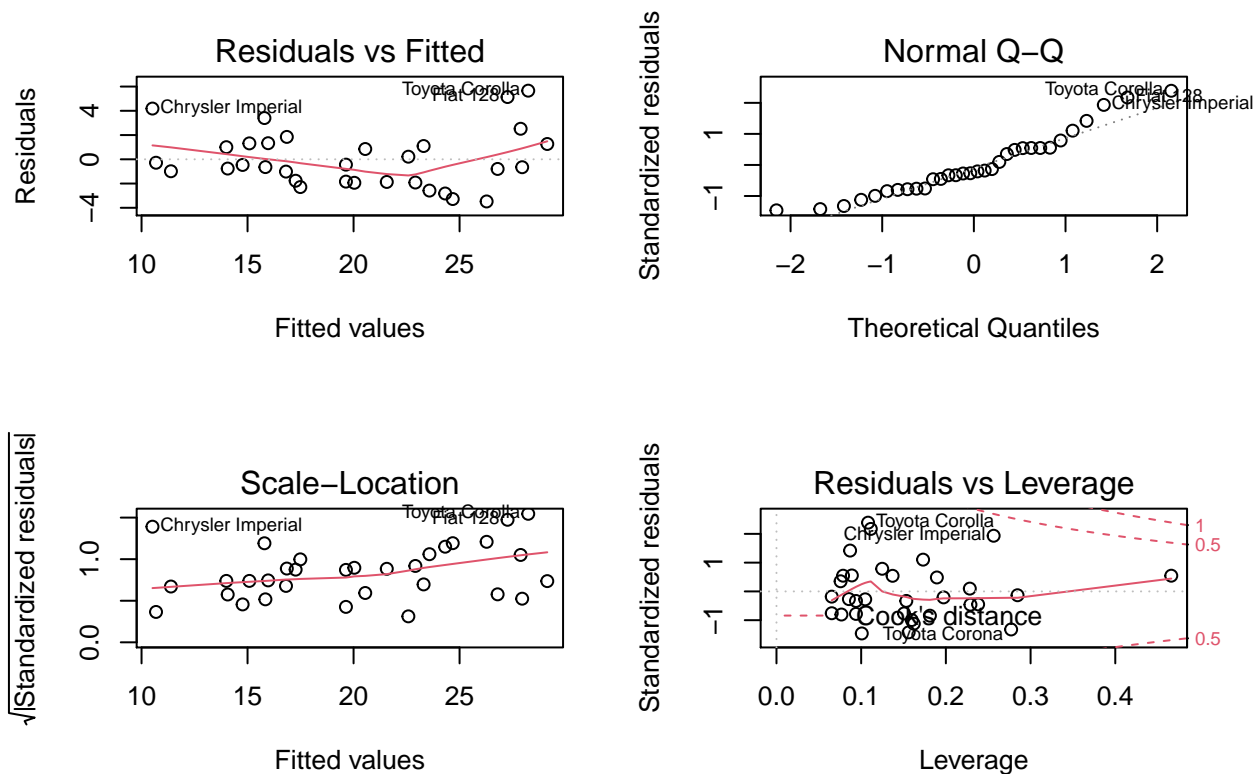
Quantify the MPG difference between automatic and manual transmissions

From the summary of the linear model fit, it can be estimated that cars with automatic transmission offer an average of 36.15 miles per gallon while cars with manual as transmission type offer 1.47 more miles per gallon than that of automatic cars which implies that manual cars offer 37.61 miles per gallon. We can also see from the value R-squared, that the model accounts for 85% of the factors that affect the miles per gallon.

Appendix

Here we take a visual look at the relationships of the properties of the model and also feature a pair plot.

```
par(mfrow=c(2,2))
plot(fit5)
```



```
ggpairs(mtcars, columns = c(1,2,4,6,9), aes(colour=am))
```

