RegressionModels

Rohit Jain

October 9, 2018

Relationship between a set of variables and miles per gallon (MPG)

Data Preprocessing and Exploratory Data Analysis:

Under the below block, Data is being preprocessed to use in the regression model for the relationship between MPG and tranmission and being analysed for any initaial trends i.e; EDA.

Data Preprocessing:

```
dt <- mtcars
str(dt)
```

```
## '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 ...
```

```
lb <- factor(dt$am)
levels(lb)[levels(lb)== 0] <- "automatic"
levels(lb)[levels(lb)== 1] <- "manual"
dt$trans <- lb
```

EDA

- Figure 1 in the Appendix is to give the maximun, minumum and the average millage for the cars on the basis of transmission which shows that the maunal transmmsion has better measures then the automatic transmission.
- Figure 2 in the Appendix is the histogram for the milles per gallon divided on the basis of the transmission and it gives that the range of millage is [10,25) and [15,35] for the automatic and manual transmission respectively.
- Figure 3 in the Appendix gives the correlation of diffrent variabls in the mtcars data set.
- t-test:

```
t.test(dt$mpg~dt$trans,conf.level=0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  dt$mpg by dt$trans
## 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 in group automatic    mean in group manual
##                17.14737                24.39231
```

- t-test concultion: The p-value is 0.001374, we may reject the null hypothesis and conclude, that automatic transmission cars have lower mpg compared with manual transmission cars - but this assumption is based on all other characteristics of automatic transmission cars and manual transmission cars are same (e.g: both have same weight distribution). This needs to be further explored in a multiple linear regression analysis.

Regression Models:

Model1: Simple multiple regression

```
MultilinearMod <- lm(mpg ~ ., data=mtcars)  # build linear regression model on full data
summary(MultilinearMod)
```

```
##
## 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
## am           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
```

- form the above model it is clear that the variable which are siginficant for the mpg prediction are am, qsec, wt so now we will use a stepwise algorithm, to choose the best model.

Model2: Step Model

```
stepfit = step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
summary(stepfit)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## am            2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

- we found from the above step model that the model contain 3 vaibles

1. am
2. wt
3. qsec

Residue square for the above model is 0.85.

Model3: Best fit model

```
finalfit <- lm(mpg ~ trans:wt + trans:qsec, data = dt)
summary(finalfit)
```

```
##
## Call:
## lm(formula = mpg ~ trans:wt + trans:qsec, data = dt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          13.9692     5.7756   2.419  0.02259 *
## transautomatic:wt    -3.1759     0.6362  -4.992 3.11e-05 ***
## transmanual:wt       -6.0992     0.9685  -6.297 9.70e-07 ***
## transautomatic:qsec   0.8338     0.2602   3.205  0.00346 **
## transmanual:qsec      1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13
```

Summary

Interpreting the results, we can see this model has a 89.5% total variance with an adjusted variance of 0.879. By adding the coefficients, we have the following conclusions:

- When the weight increased by 1000 lbs, the mpg decreased by -3.176 for automatic transmission cars, and -6.09 for manual transmission cars.
- With increasing car weight we should choose manual transmission cars.
- When the acceleration speed dropped, and 1/4 mile time increased (by 1 sec), the mpg factor increased by 0.834 miles for automatic transmission cars, and 1.446 miles for manual transmission cars
- With lower acceleration speed, but same weight, manual transmission cars are better for mpg
- Figure 4 in the Appendix give us the Residue relations.

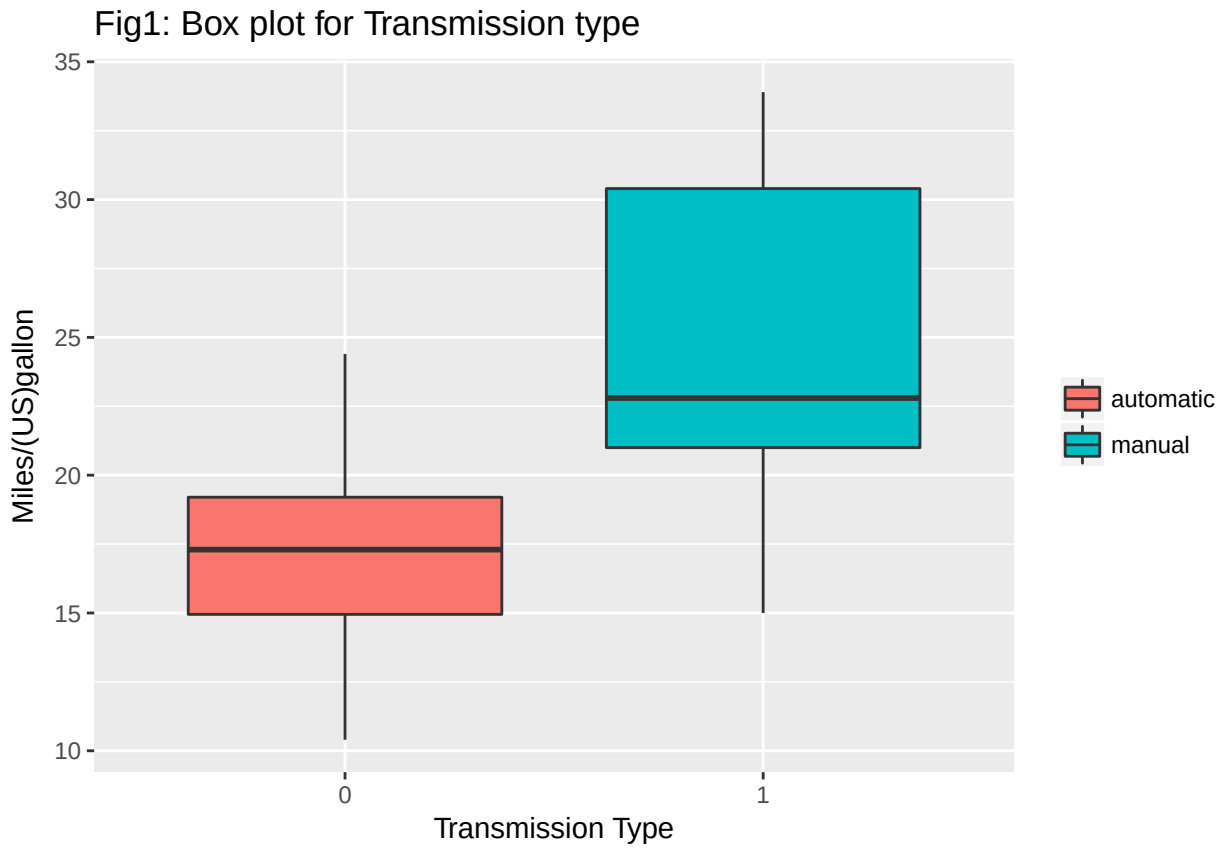
Main conclusion

The mpg is largely determined by the interplay between weight, acceleration and transmission. therefore we cannot perfectly answer which transmission mode is best as mpg depende upon weight, acceleration and transmission significantly.

Appendix

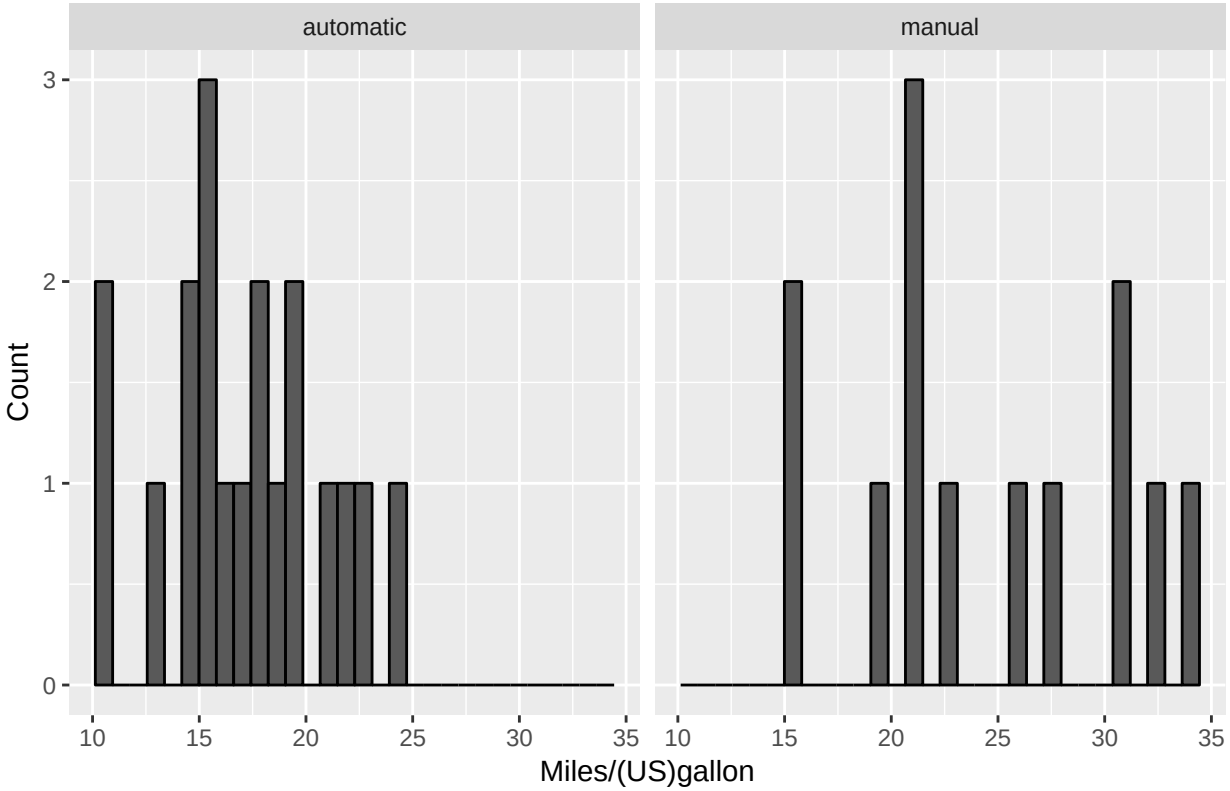
Supporting graphs and figures:

- Figure 1:



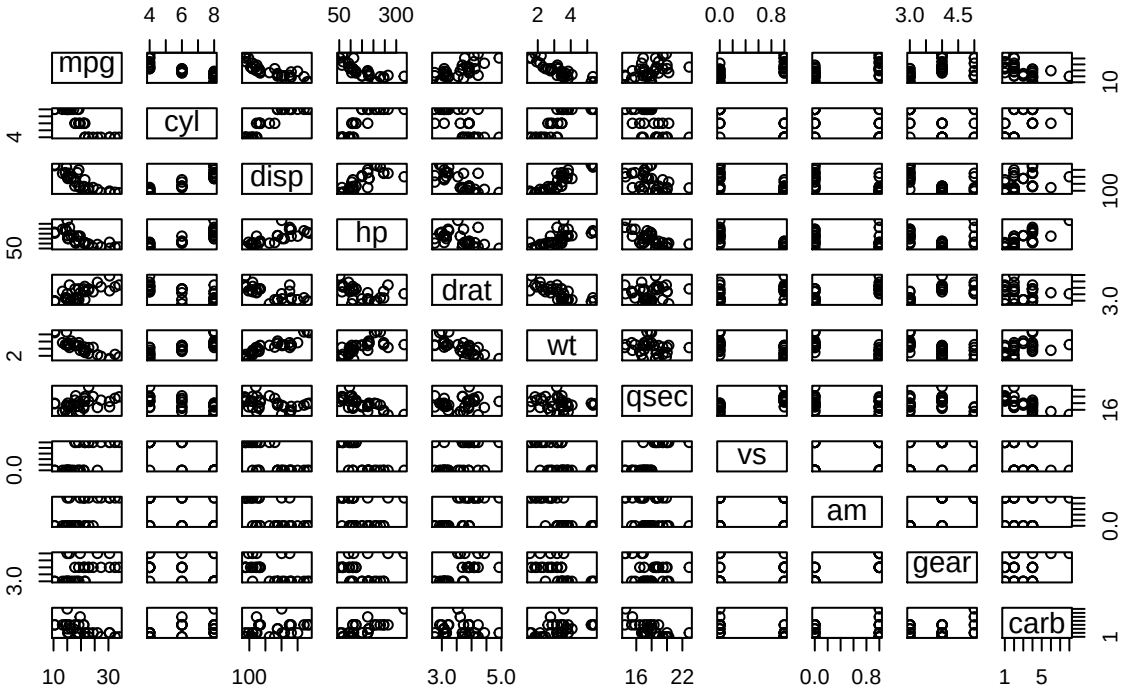
- Figure 2:

Fig2: histogram of miles per gallon for Transmission type



- Figure 3:

Fig3: Correlation diagram



- Figure 4:

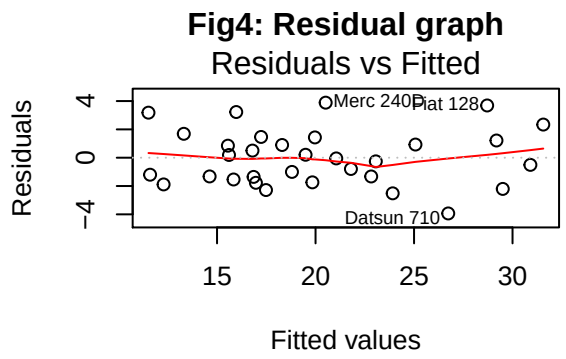


Fig4: Residual graph
Residuals vs Fitted
Residuals
Fitted values
Merc 240D
Fiat 128
Datsun 710

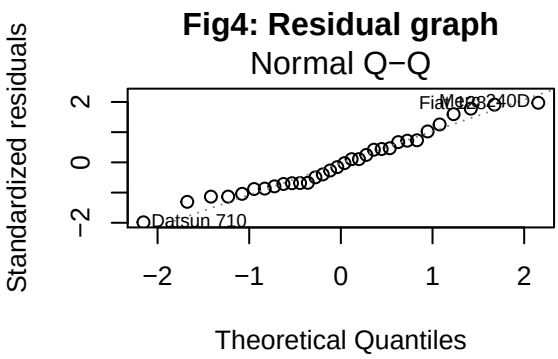


Fig4: Residual graph
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Datsun 710

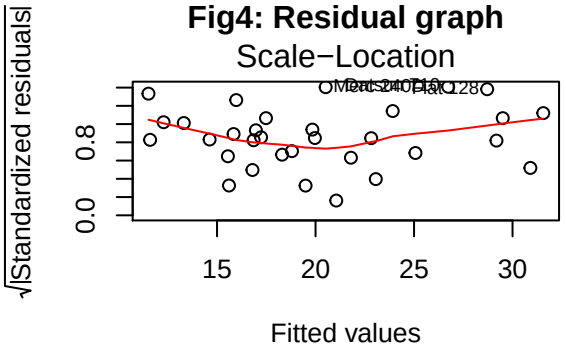


Fig4: Residual graph
Scale−Location
√|Standardized residuals|
Fitted values
Fiat 128

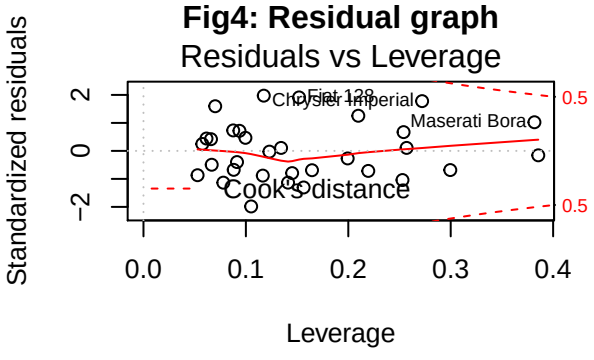


Fig4: Residual graph
Residuals vs Leverage
Standardized residuals
Leverage
Cook's distance
Fiat 128
Chrysler Imperial
Maserati Bora
0.5
0.5