Regression models course assignment

Siva

February 27, 2016

Assignment Case Study - Regression Models Course Project - Motor Trend
You work 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) (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"

Case study with solution step by step :
First, we load the dataset mtcars and lets preview the dataset beforehand using header command.

```
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
```

Model Vs and am columns as categorical variables

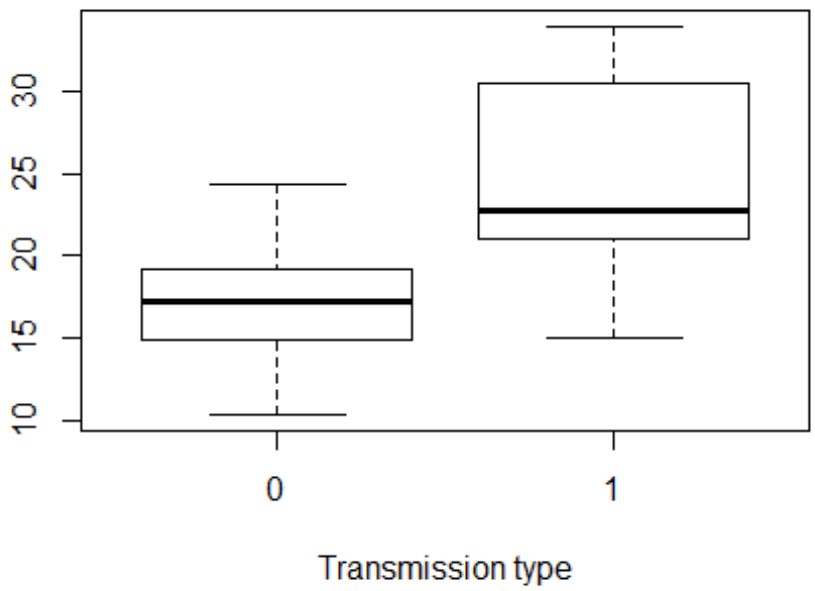
```
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- as.factor(mtcars$am)
```

View the summary of variable mpg

```
summary(mtcars$mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.40   15.42   19.20   20.09   22.80   33.90
```

Draw a quick boxplot to understand the relationship between variable mpg and variable am. The boxplot shows clear diffierence between both variables. In order to prove this signifcat difference we would need to perform addtional exploratory analysis. Next steps would be understand how these variables behave as liner regression model and mulitvariate regression models



Simple linear regression model
Try to fit using linear regression model

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

```
##
## Call:
## lm(formula = mpg ~ factor(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 ***
## factor(am)1    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
```

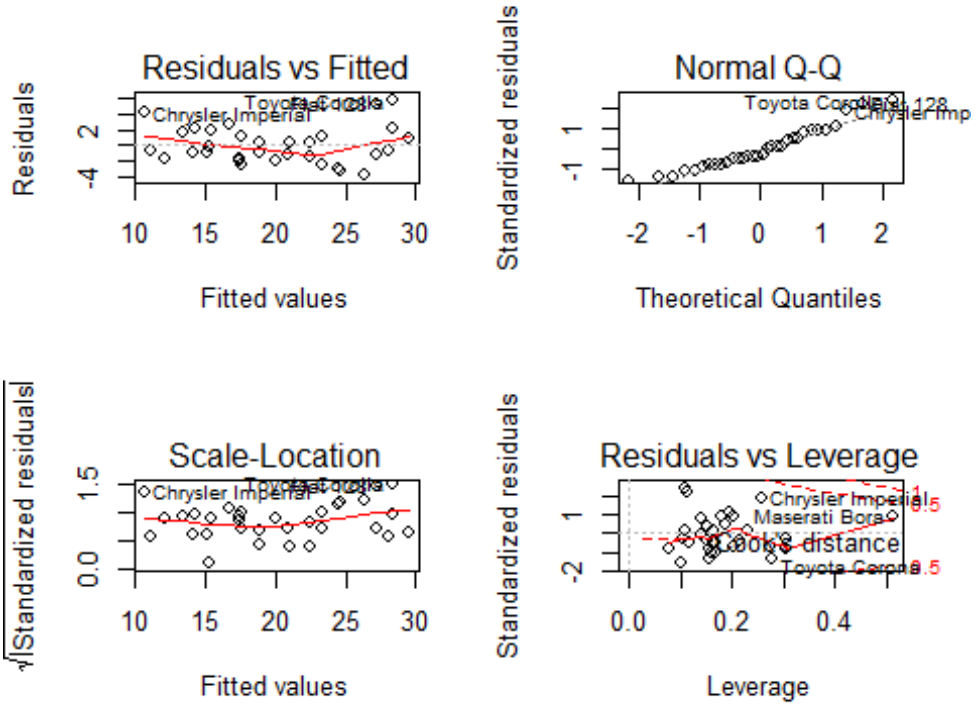
P Values is 0.000285 and R - square adjusted value is 33.85 % which leaves us with question of how the remainaing models behaves and hence not a best fit model.

Multivariable Regression Model

```
data(mtcars)
fit_multi <- lm(mpg ~ . ,data=mtcars)
summary(fit_multi)
```

```
##
## 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
```

```
##        mpg        cyl       disp         hp       drat         wt
##  1.0000000 -0.8521620 -0.8475514 -0.7761684  0.6811719 -0.8676594
##       qsec         vs         am       gear       carb
##  0.4186840  0.6640389  0.5998324  0.4802848 -0.5509251
```



The adjusted R- Square value is 80.66 % which is closer to our expectation and multivariate models fits best for this mtcars dataset. I have ploted residuals and check for non normality behavior. From the first graph, we can observe that the variance is almost the same across. we can see that the multiple R squared value is much higher at 0.85 in third graph using the final high impacting variables.

Conclusion :

The 'wt' and 'cyl'are confounding variables in the relationship between 'am and 'mpg' and that manual transmission cars on average have 1.55 miles per gallon more than automatic cars. 1. "Is an automatic or manual transmission better for MPG" : Answer Manual transsion 2. "Quantify the MPG difference between automatic and manual transmissions" : Manual cars have 1.55 miles per gallan mpg compared to an automatic transmission.