Regression_Model

Muneeb Shahid

1/18/2021

Executive Summary

In this analysis, we are going to study the mtcars data set and answer following questions:

1) Is an automatic or manual transmission better for MPG ?
2) Quantify the MPG difference between automatic and manual transmissions ?

From our analysis we can show that manual transmission has an MPG 2.93 greater than an automatic transmission hence, manual transmission have more miles per gallon than automatic transmissions. However, we can not be sure about exact quantitative difference as the number of observations in the dataset is very low.

Reading Dataset

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

Viewing some of the enteries in dataset

```
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
```

Data Conversion

Looking at the above table, our variable of concern is am. It indicates transmission type of a car, 0 for automatic and 1 for manual. Converting this variable into factor.Also, we need to convert other categorical value into factor as well.

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

Exploratory Analysis (Getting mpg average with respect to different transmission types)

```r
aggregate(mpg~am_factor, data = mtcars, mean)
```

```
##   am_factor      mpg
## 1         0 17.14737
## 2         1 24.39231
```

We can see from the above table that there is a signifcant difference between the averages of mpg, suggesting that mpg gets affected by transmission type.

Linear Model Analysis

Overall impact of am variable

```r
summary(lm(mpg~am, mtcars))
```

```
##
## 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 ***
## am             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
```

There are two conclusions that can be drawn from above results:

1. The p-value is very small, it supports the results from exploratory data analysis that manual transmission yields around 7 more mpg on average than automatic transmission.
2. Adjusted R-square value suggests that only 33% approx variation can be explained by our model i.e. using only transmission variable. There are other variables in the dataset as well, we should look whether those variables improves this value or not.

Multivariable model analysis

There are 9 other variables so, trying all of the combinations would take a lot of time and effort. So we will use step function to select the best model for us.

```
# excluding factor variables

summary(step(lm(mpg ~ ., mtcars[c(1:11)]), direction = "both", trace=FALSE))
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars[c(1:11)])
##
## 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
```

The best model includes am variable along with the qsec and wt variable. Combining these variables have increased the Adjusted R-square value to 0.83 approx which is way better than the one obtained from simple model with am variable only. Residual standard error has also decreased to 2.5 from 4.9. The p-values for all of these variables are also significant. Correlation of these variables can be seen in Appendix - 1

Comparing impact of manual (1) vs automatic (0) transmission

```
summary(lm(mpg~wt + qsec + am_factor, mtcars[c(1:12)]))
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am_factor, data = mtcars[c(1:12)])
##
## 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_factor1    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
```

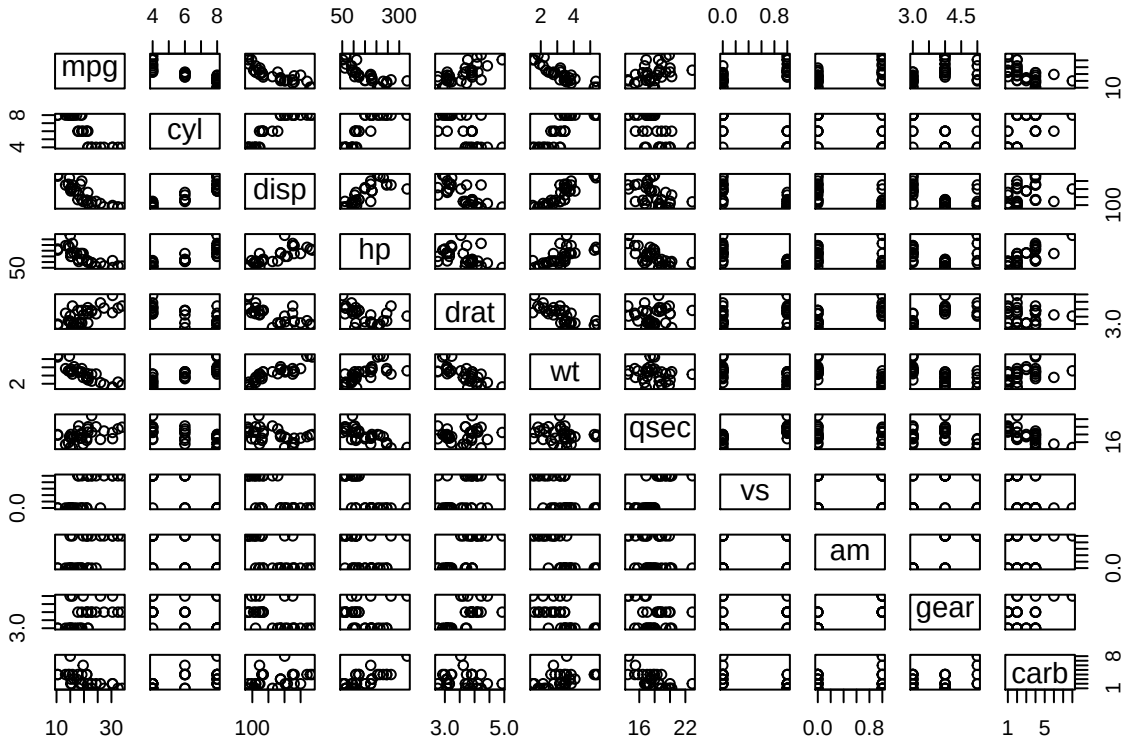
Residual Analysis

Following observations can be made from the Residual plots in Appendix - 2:

1. Residuals vs. Fitted plot confirms the independence due to randomness.
2. Most of The points of the Normal Q-Q plot are close to the line indicating that the distribution of residuals is normal.
3. The Scale-Location plot randomness supports the constant variance assumption.
4. Residuals and Leverage show that most of the points are within 0.5 band hence, no outliers.

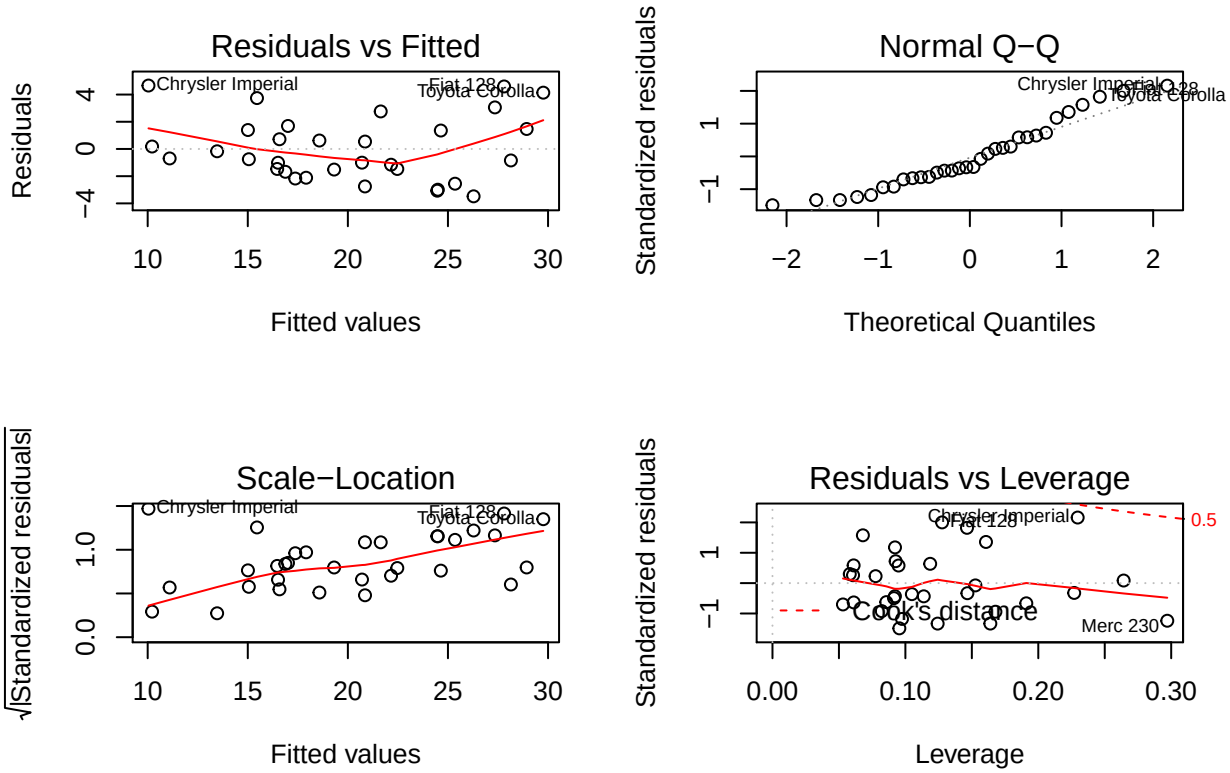
Appendix - 1: Correlation plot

```
pairs(mpg ~ ., data = mtcars[c(1:11)])
```



Appendix - 2: Residual Plots

```
par(mfrow = c(2, 2))
plot(lm(mpg ~ am + wt + qsec, mtcars[c(1:11)]))
```



Conclusion

Is an automatic or manual transmission better for MPG?

Based on the simple model and as well as multivariable model we can say that manual transmission have more miles per gallon than automatic transmissions.

Quantify the MPG difference between automatic and manual transmissions?

Using wt and qsec along with am, With a $p < 0.05$ confidence cars with manual transmission have 2.93 times more miles per gallon than automatic transmissions. However, we can not be certain as the dataset has only 32 observations and this may not have been enough to answer this question more clearly.