Motor Trend

VR SREEGANESH

30/08/2020

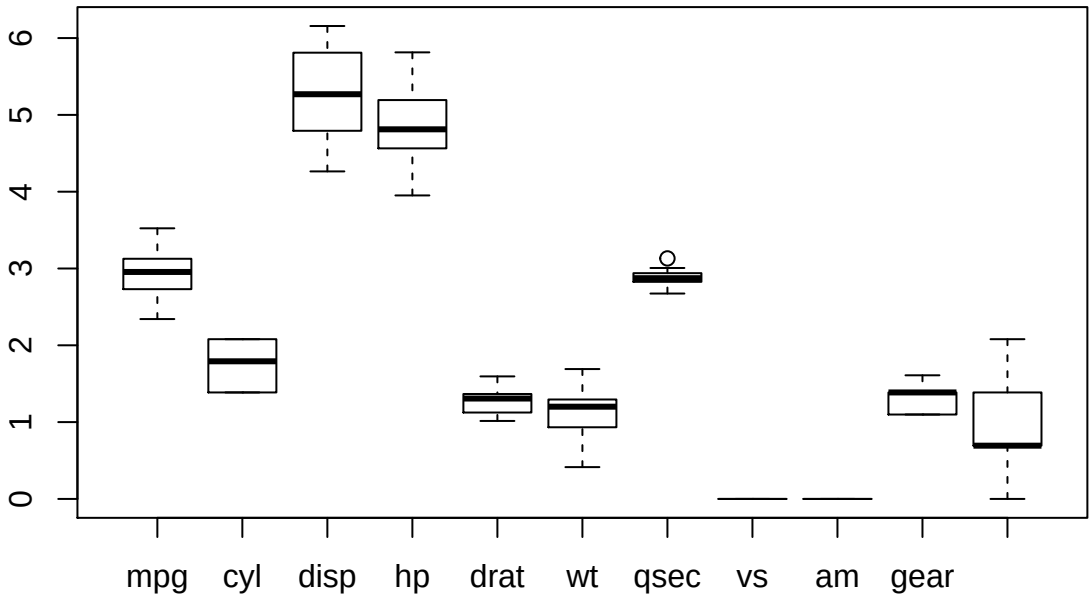
Introduction

This report was generated as part of the final year project work for the course, "Regression Models". The general interest is in exploring the relationship between the miles per gallon and the transmission parameter. We're specifically interested in two questions. 1. "Is an automatic or manual transmission better for miles per gallon" 2. "Quantify the MPG difference between automatic and manual transmission"

Exploratory Data Analysis

The dataset we deal with here is the mtcars dataset. The dataset has 11 columns corresponding to different parameters of the cars that are listen in the column. We can get an idea of the distribution of those values by observing the following box plots. Note that the boxplot presented here is the log of the values. This is done so because the "disp" and "hp" columns have a larger variance and showing them in regular values results in the other boxplots reducing themeselves into a single line.

```
data("mtcars")
boxplot(log(mtcars))
```



Regression Analysis

We first find the mean MPG for all cars dividing on the basis of transmission alone.

```
print(c("The mean MPG for all cars having manual transmission is ",summary(lm(mpg ~ factor(am) -1, data
```

```
## [1] "The mean MPG for all cars having manual transmission is "
## [2] "17.1473684210526"
```

```
print(c("The mean MPG for all cars having automatic transmission is ",summary(lm(mpg ~ factor(am) -1, da
```

```
## [1] "The mean MPG for all cars having automatic transmission is "
## [2] "24.3923076923077"
```

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

```
##
## 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
```

We see that the manual transmission gives a 7 mpg more than automatic. We also see that the P-value is less than 0.0003. We see that the R squared value for this test is around .35, suggesting that only a third of the variance in MPG can be attributed transmission type alone. This implies that we should perform an analysis on rest of the factors.

```
summary(aov(mpg~.,data = mtcars))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## cyl          1  817.7   817.7 116.425 5.03e-10 ***
## disp         1   37.6    37.6   5.353  0.03091 *
## hp           1    9.4     9.4   1.334  0.26103
## drat         1   16.5    16.5   2.345  0.14064
## wt           1   77.5    77.5  11.031  0.00324 **
## qsec         1    3.9     3.9   0.562  0.46166
## vs           1    0.1     0.1   0.018  0.89317
## am           1   14.5    14.5   2.061  0.16586
## gear         1    1.0     1.0   0.138  0.71365
## carb         1    0.4     0.4   0.058  0.81218
## Residuals   21  147.5     7.0
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above output, we consider those factors that has a p value less than 0.05. This includes the factors 'cyl', 'disp' and 'wt'. The fact that their p values are less than 0.05 indicates that they play a huge role in deciding the value of mpg. Now, we shall consider a linear model that includes those factors are regressors.

```r
summary(lm(mpg ~ cyl + disp + wt + am, data = mtcars))
```

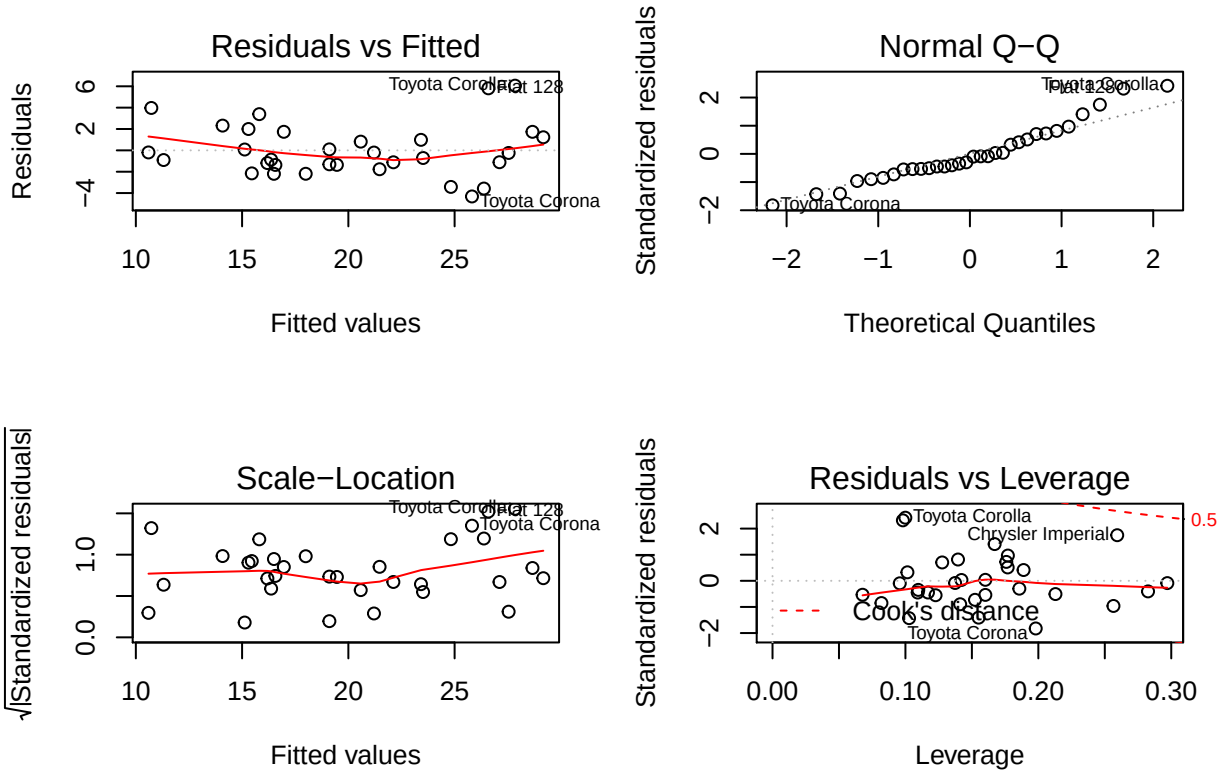
```
##
## Call:
## lm(formula = mpg ~ cyl + disp + wt + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.318 -1.362 -0.479  1.354  6.059
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.898313   3.601540  11.356 8.68e-12 ***
## cyl         -1.784173   0.618192  -2.886  0.00758 **
## disp         0.007404   0.012081   0.613  0.54509
## wt          -3.583425   1.186504  -3.020  0.00547 **
## am           0.129066   1.321512   0.098  0.92292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.642 on 27 degrees of freedom
## Multiple R-squared:  0.8327, Adjusted R-squared:  0.8079
## F-statistic: 33.59 on 4 and 27 DF,  p-value: 4.038e-10
```

From above, we see that the R squared value is approximately 0.83. This indicates that 83% of the variance can be explained by the regressors included in the above model. From the P values, we see that the factors 'cyl' and 'wt' have P values less than 0.05. This implies that the 'cyl' and 'wt' are confounding variables in the relation between transmission and the mpg.

Residual Plot and Analysis

The following plots show the different plots related to residuals.

```r
par(mfrow = c(2,2))
plot(lm(mpg ~ cyl + disp + wt + factor(am), data = mtcars))
```



We see that the residuals are homoscedastic and normally distributed bar some outliers.

Summary

Thus, we've answered the questions that was asked as part of the objectives for course project. We've seen that the manual vehicles provided an additional average of 7 miles per gallon for a device. Thus, the manual transmission is better when we consider the parameter, miles per gallon.