Regression Models Course Project

Utkarsh Chauhan

9/1/2020

Executive Summary

When looking at the mtcars data set, we attempted to predict Miles per Gallon using different methods. Initially, regular inference testing indicated that the null hypothesis should be rejected proving manual transmissions did indeed have a significant positive impact on MPG. Next, we see average how many more miles per gallon a manual transmission car gets. The linear regression analysis estimated around a 7.24 miles per gallon increase.

It is also necessary to test how other variables effect MPG along with transmission. So, we tested a model with all the other variables and see difference between the two fitted models, and looked how the predictability increased.

Loading and Processing Data

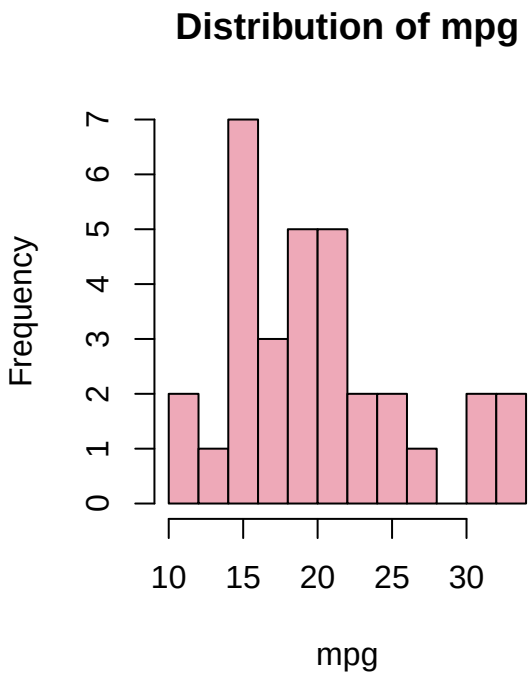
```
data("mtcars")

str(mtcars)
```

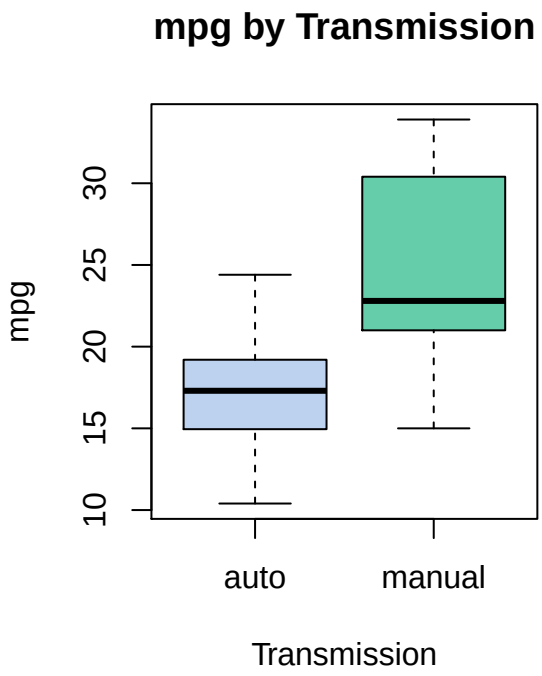
```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 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 ...
```

```
# converting categorical variables into factors
mtcars$cyl <- as.factor(mtcars$cyl)
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- as.factor(mtcars$am)
mtcars$gear <- as.factor(mtcars$gear)
mtcars$carb <- as.factor(mtcars$carb)
```

Exploratory Analysis

```
par(mfrow=c(1,2))
hist(mtcars$mpg,breaks = 15, col='pink2', main = 'Distribution of mpg', xlab = 'mpg')
plot(mtcars$am,mtcars$mpg,col=c('lightsteelblue2','aquamarine3'),
     xlab='Transmission',ylab='mpg',names=c('auto','manual'), main='mpg by Transmission')
```





Regression Models

Let's begin with a simple linear regression of MPG vs automatic/manual transmissions.

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

```
##
## 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 ***
## am1            7.245      1.764   4.106 0.000285 ***
```

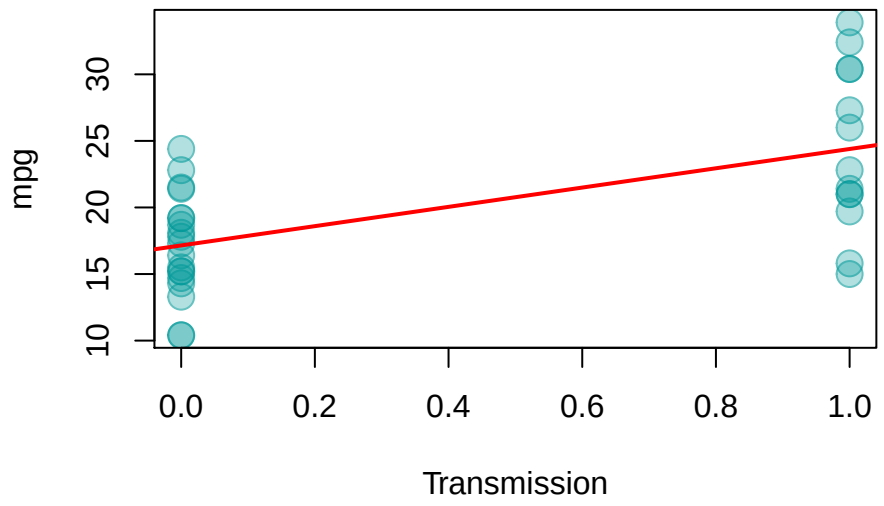
```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

```
confint(fit)
```

```
##                2.5 %   97.5 %
## (Intercept) 14.85062 19.44411
## am1          3.64151 10.84837
```

- Intercept = 17.15 and Slope = 7.24
  Implies with an *Automatic Transmission*, MPG avg starts at 17.15 and if the car had a *Manual Transmission*, you could estimate an increase in MPG of 7.24.
- Adjusted R square = 33.9%
  implies this much variation can be explained by this model.
- P-value = 0.0002
  implies this model is significant.
- confidence interval excludes zero.

```
plot(as.numeric(mtcars$am)-1,mtcars$mpg,xlab='Transmission',ylab='mpg',
     col=rgb(0,.6,.6,.6), bg=rgb(0,.6,.6,.3),cex=1.8, pch=21)
abline(fit, lwd=2, col='red')
```



Hence, we see there is a significant relationship between MPG and Transmission.

Now lets fit multivariable regression with all the variables.

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

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23.87913   20.06582   1.190   0.2525
## cyl6        -2.64870    3.04089  -0.871   0.3975
## cyl8        -0.33616    7.15954  -0.047   0.9632
## disp         0.03555    0.03190   1.114   0.2827
## hp          -0.07051    0.03943  -1.788   0.0939 .
## drat         1.18283    2.48348   0.476   0.6407
## wt          -4.52978    2.53875  -1.784   0.0946 .
## qsec         0.36784    0.93540   0.393   0.6997
## vs1          1.93085    2.87126   0.672   0.5115
## am1          1.21212    3.21355   0.377   0.7113
## gear4        1.11435    3.79952   0.293   0.7733
## gear5        2.52840    3.73636   0.677   0.5089
## carb2       -0.97935    2.31797  -0.423   0.6787
## carb3        2.99964    4.29355   0.699   0.4955
## carb4        1.09142    4.44962   0.245   0.8096
## carb6        4.47757    6.38406   0.701   0.4938
## carb8        7.25041    8.36057   0.867   0.3995
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF,  p-value: 0.000124
```

- MPG avg starts at 23.88 with *Automatic Transmission* and if the car had a *Manual Transmission*, you could estimate an increase in MPG of 1.21.
- Adjusted R square = 80% implies this much variation can be explained by this model.
- P-values shows large values implying non-significant result.

```r
anova(fit,fitall)
```

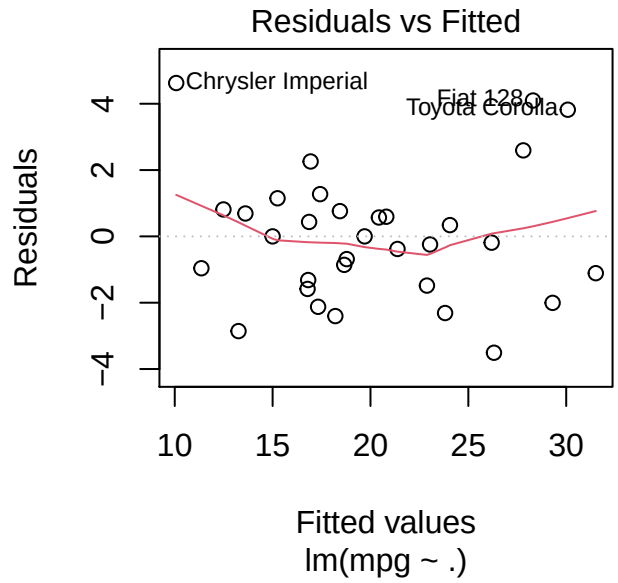
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df   RSS Df Sum of Sq      F   Pr(>F)
## 1     30 720.9
## 2     15 120.4 15    600.49 4.9874 0.001759 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

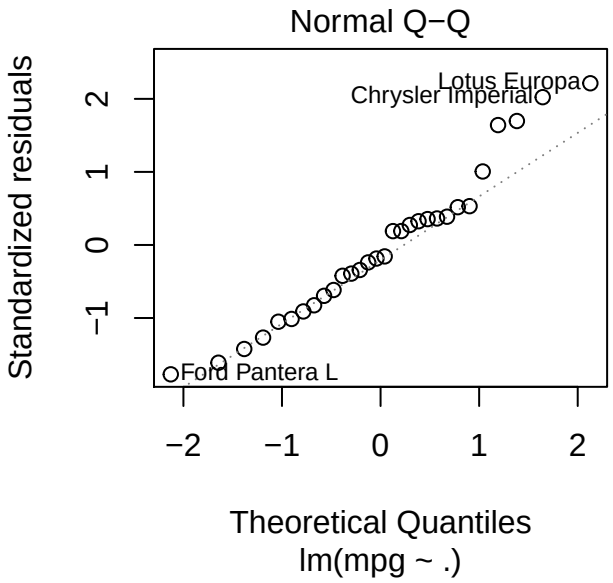
- P-value = 0.0017 implies second model is better than the first.

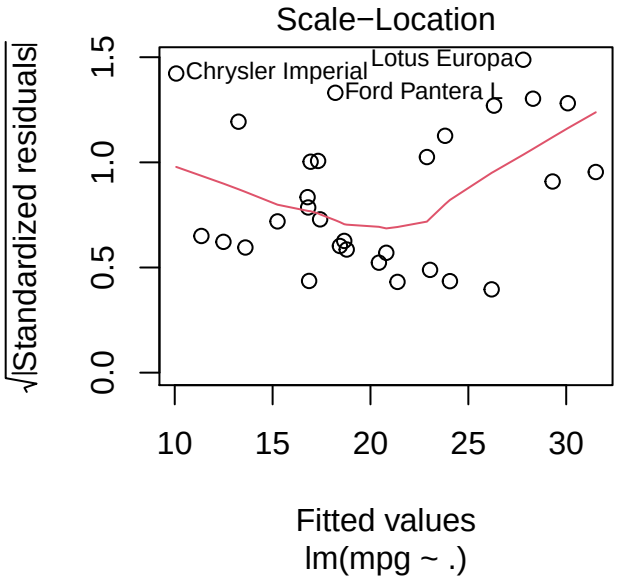
This shows not all of the of the remaining variables are significant and model should be fitted again with only the important ones.

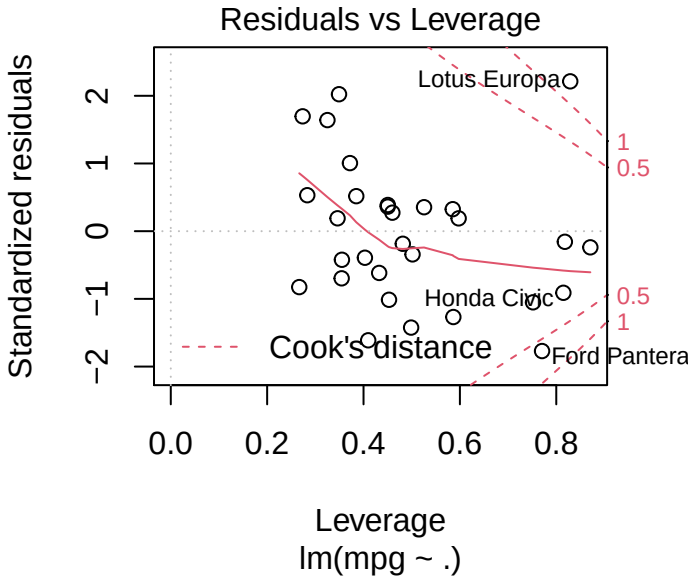
```
plot(fitall)
```

```
## Warning: not plotting observations with leverage one:
##   30, 31
```









Conclusion

From the above analysis, we can say that the manual transmission is better than automatic transmission for MPG. The factors which influence MPG can be multiple. By using all possible regression method, we can pick up the relatively right variables into our model.