Linear Regression Project

WendyD

9/15/2020

```
library(car)
```

```
## Loading required package: carData
```

```
data("mtcars")
```

Executive Summary

This project involve analyzing the correlation between mpg (miles per gallon) variable with different variables within mtcars dataset. After conducting analysis, we concluded that having manual transmission is positively correlated to mpg value. We also weighed the influence of other variables on mpg, and saw that the best model to estimate mpg is to include transmission type, weight(1000lbs) and 1/4 mile time

Objective

To answer these questions:

- "Is an automatic or manual transmission better for MPG"
- "Quantify the MPG difference between automatic and manual transmissions"

Data Processing/Explanatory Analysis

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
```

```
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

Since our analysis mainly concern with the dependence of MPG on tranmission types, we first transformed "am" variable into factor variable

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <-c("Automatic", "Manual")
```

Look over the mean MPG based on the transmission type

```
tapply(mtcars$mpg, mtcars$am, mean)
```

```
## Automatic    Manual
##  17.14737  24.39231
```

With the first look, the average MPG value for cars with automatic transmission are higher than automatic cars's

Please refer to the appendix for more figures

Regression Analysis

Conducting t test to test the hypothesis, and to see if there is a significant difference in the average mpg between automatic and manual cars

```
t.test(mtcars$mpg ~ mtcars$am, data = mtcars)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$am
## 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
```

The p value is $0.001374 < 0.05$. Thus we can reject the null hypothesis and conclude that: the average mpg for automatic car is significantly lower than that of manual car. However there are lots of factor that conrtibute to this difference

We further explore the relationship between mpg and other factors of the dataset via regression analysis

Quantifying the MPG difference between automatic and manual transmissions

```
fit<- lm(mtcars$mpg ~ ., data = mtcars)
summary(fit)$coefficients
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
```

```
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## amManual     2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

```
summary(fit)
```

```
##
## Call:
## lm(formula = mtcars$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
## amManual     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
```

Just by looking , we see that wt has the most influence on MPG (Slope coefficients is -3.71530). We also see that having manual tranmission increase mpg on the average of 2.52023 mile/US gallon .

We then use the stepwise selction function to select the input variables for our model.

```
stepfit<- step(fit, trace =0)
summary(stepfit)
```

```
##
## Call:
## lm(formula = mtcars$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 ***
## amManual      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
```

From the result of step regression, we chose weight (wt), qsec, and transmission types (am) as explanatory variable for our model

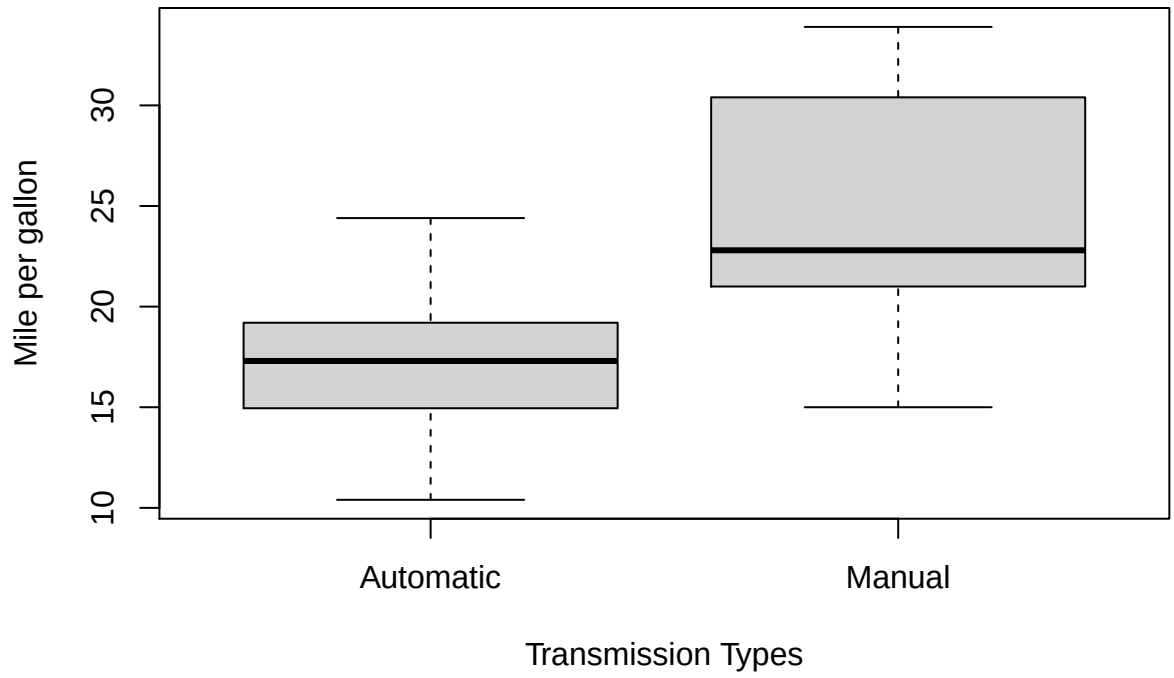
Conclusion

With multiple R squared value at 0.8497, our model explains nearly 84.5% variance of the observed value from the mean. The expected change in mpg is 2.9358 miles per gallon for manual cars in comparision to automatic cars. Thus we can say that manual car is better for mpg than automatic car

Appendix

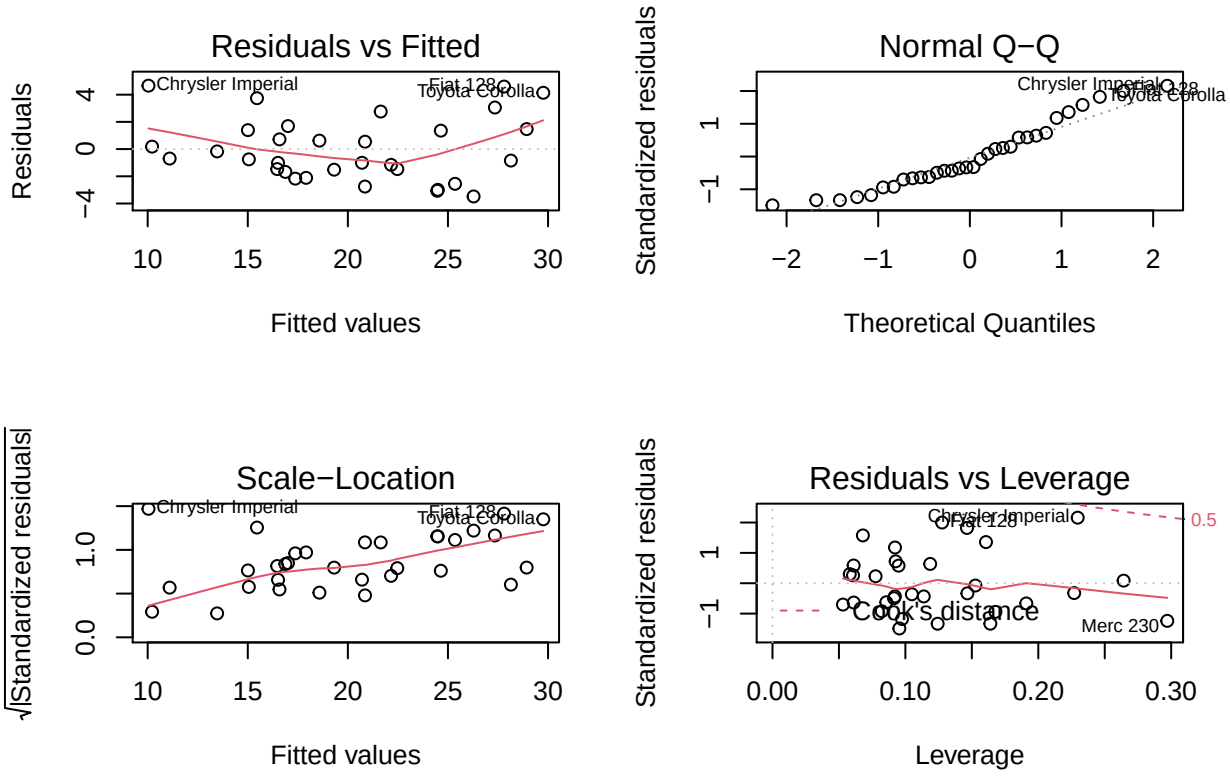
Plotting the mpg value for automatic and transmission cars

```
boxplot(mtcars$mpg ~ mtcars$am, data = mtcars, xlab="Transmission Types", ylab="Mile per gallon")
```



Residual Plot

```
par(mfrow = c(2,2))
plot(stepfit)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Cook's distance
0.5
Merc 230
Leverage

“‘