19BCE1412-EDA-LAB1

Aishwarya S S

31/01/2022

Reg. No: 19BCE1412

Name: Aishwarya S S

Title - Regression

Aim:

Understand the following operations/functions on 'mtcars' dataset and perform similar operations on 'CreditWorthiness.csv' dataset based on given instructions.

Clear the R environment

```
rm(list=ls())
```

Importing the dataset

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.0.2
```

```
data <- read_csv("CreditWorthiness.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   Cdur = col_double(),
##   Camt = col_double(),
##   InRate = col_double(),
##   age = col_double(),
##   NumCred = col_double(),
##   Ndepend = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
View(data)
```

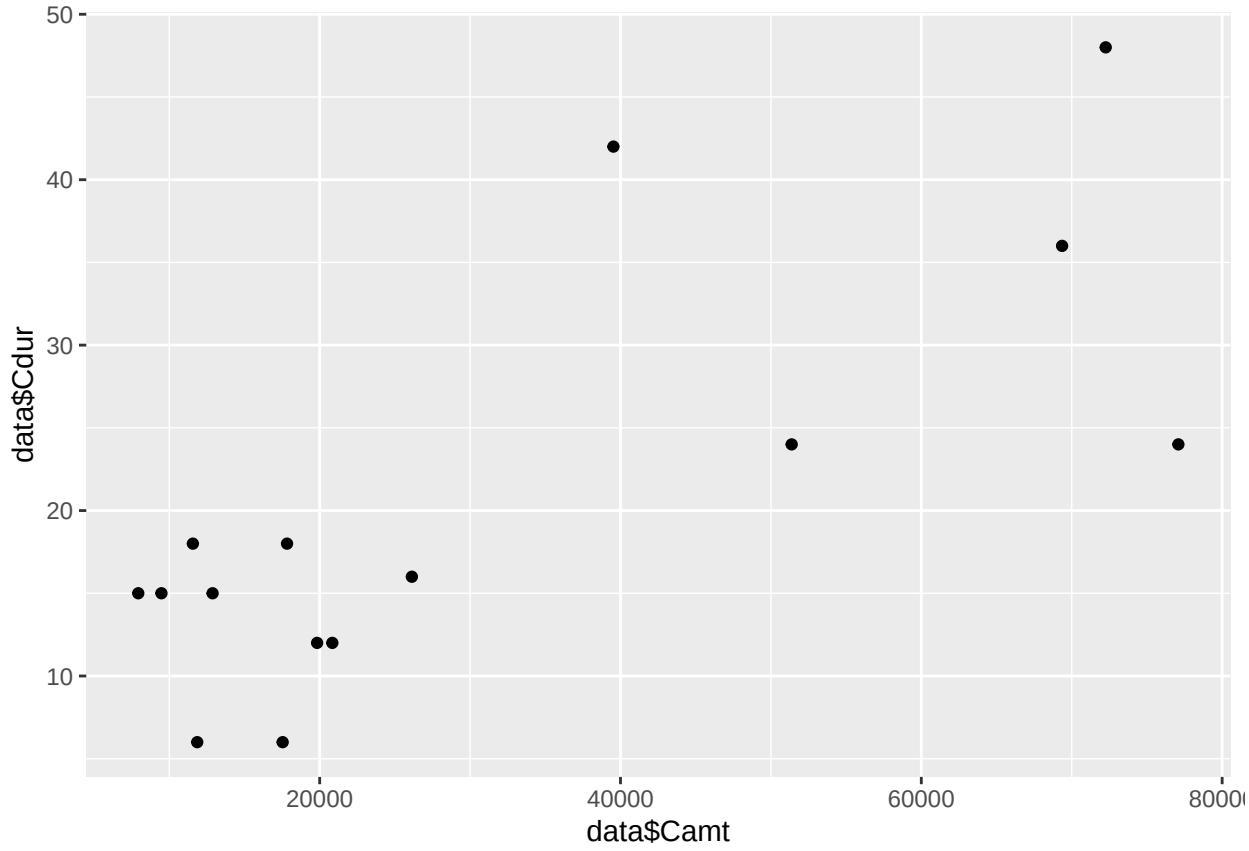
```
library(dplyr)      #dplyr is a library, which has functions related to data analysis
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data <- sample_n(data,15)

# install.packages("ggplot2")
library("ggplot2")
```

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
ggplot(data,aes(x=data$Camt,y=data$Cdur))+geom_point()
```



```r
# To plot - wt/gear
cor.test(data$Camt,data$Cdur)
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$Camt and data$Cdur
## t = 4.1857, df = 13, p-value = 0.001068
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.4010648 0.9148521
## sample estimates:
##       cor
## 0.7576649
```

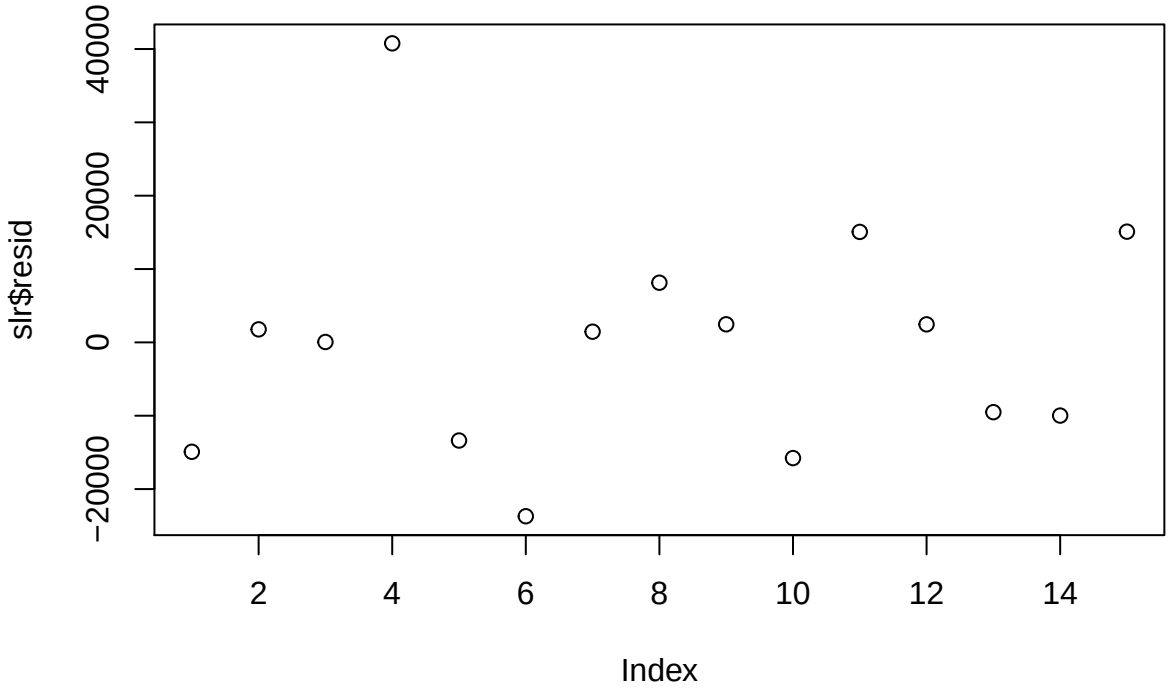
```r
# to find the correlation value
```

simple linear regression

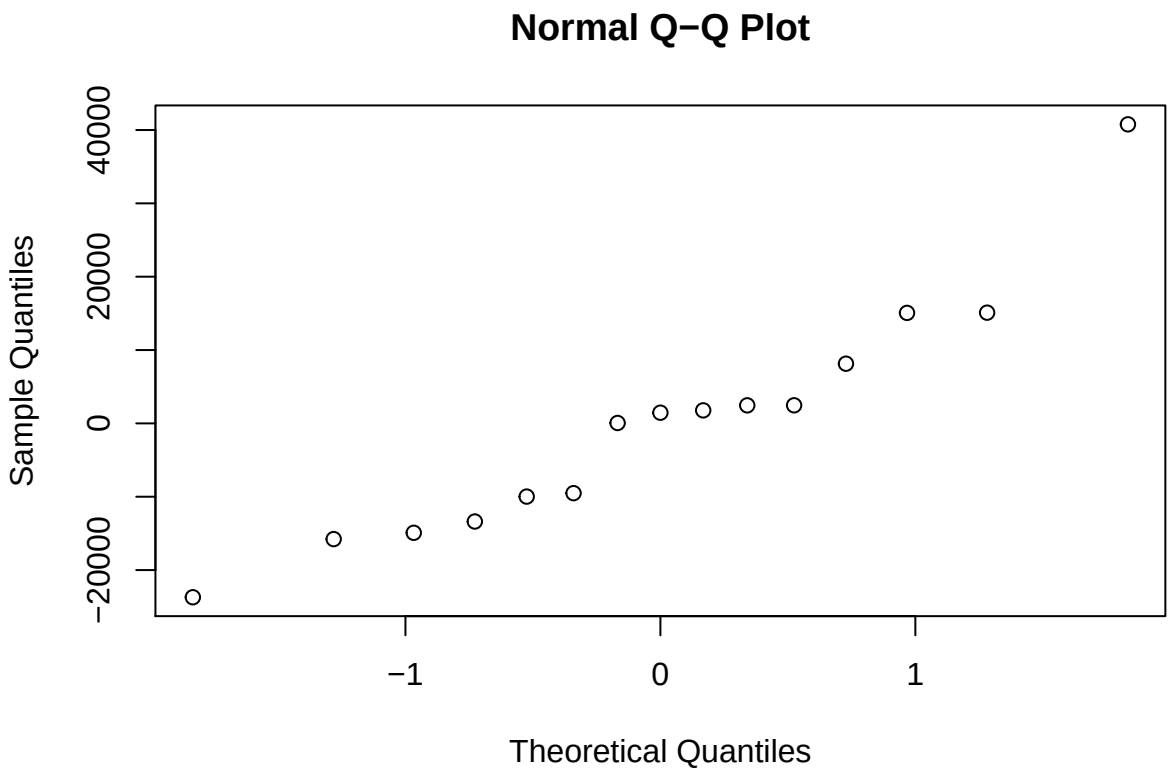
```r
slr = lm(Camt~Cdur, data)
summary(slr)
```

```
##
## Call:
## lm(formula = Camt ~ Cdur, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -23705 -11680   1456   5302  40772
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    429.3     8481.8   0.051  0.96040
## Cdur          1495.4      357.3   4.186  0.00107 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16650 on 13 degrees of freedom
## Multiple R-squared:  0.5741, Adjusted R-squared:  0.5413
## F-statistic: 17.52 on 1 and 13 DF,  p-value: 0.001068
```

```r
plot(slr$resid)  # Residual plot
```




```
qqnorm(slr$resid)   #Q-Q Plot
```




```
slr$residuals
```

```
##            1            2            3            4            5            6
## -14919.94597   1774.67829     52.65472  40771.67240 -13379.94597 -23705.09086
##            7            8            9           10           11           12
##   1456.18124   8138.43566   2466.18124 -15776.07318  15061.67240   2458.43566
##           13           14           15
##  -9516.07318  -9979.94597  15097.16356
```

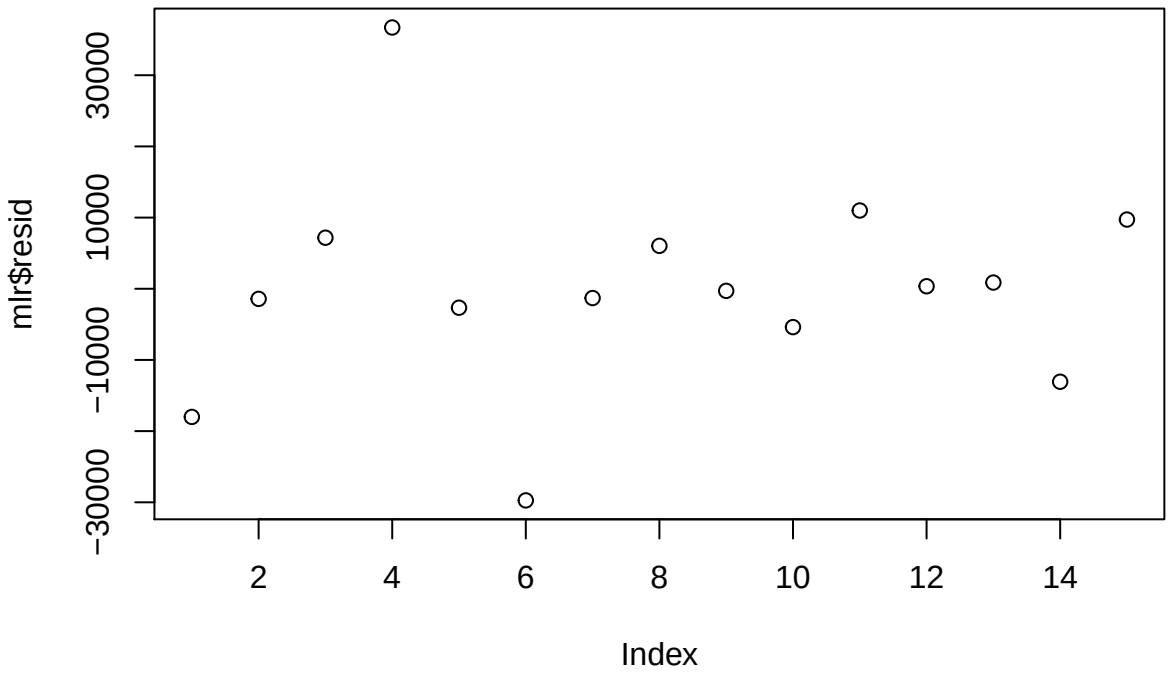
Multiple linear regression

```
mlr = lm(Camt~Cdur+NumCred, data)
summary(mlr)
```

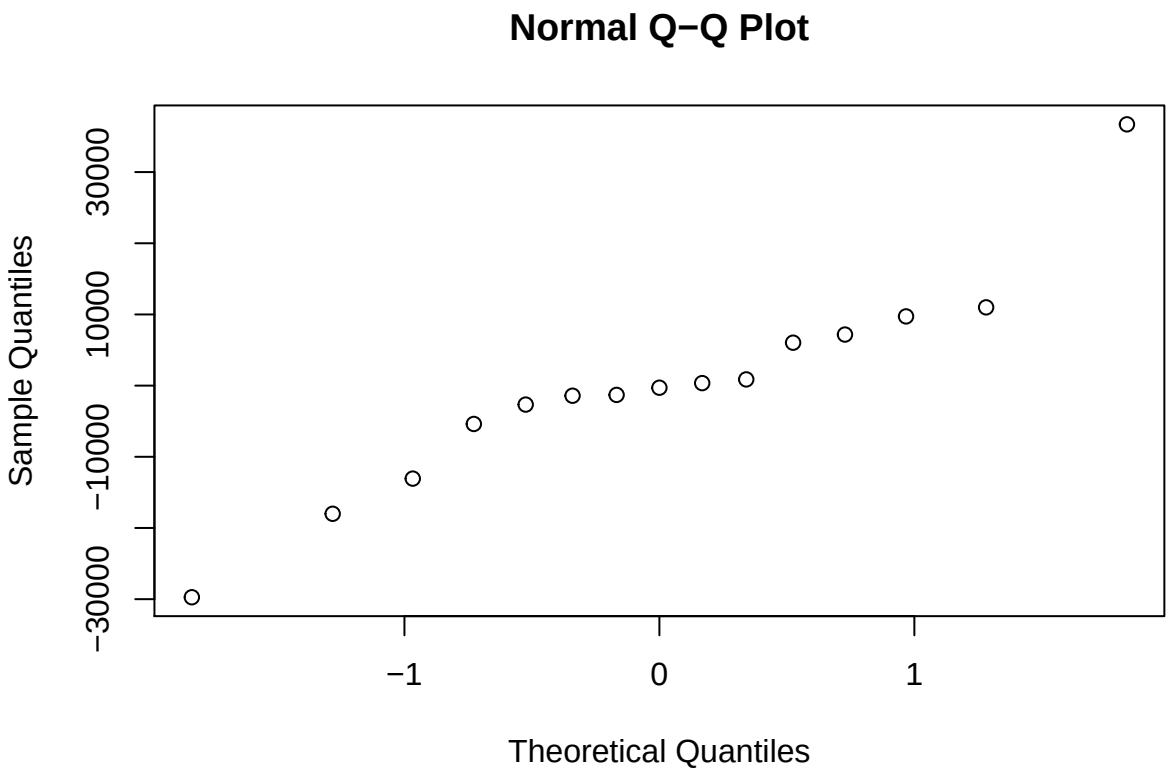
```
##
## Call:
## lm(formula = Camt ~ Cdur + NumCred, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -29730  -4025   -292   6605  36706
##
## Coefficients:
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15683.0    13345.3   1.175 0.262719
## Cdur          1604.3      351.4   4.565 0.000649 ***
## NumCred     -13801.7     9562.4  -1.443 0.174519
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16000 on 12 degrees of freedom
## Multiple R-squared:  0.6371, Adjusted R-squared:  0.5766
## F-statistic: 10.53 on 2 and 12 DF,  p-value: 0.002286
```

```
plot(mlr$resid)  # Residual plot
```




```
qqnorm(mlr$resid)   #Q-Q Plot
```


Normal Q−Q Plot
Sample Quantiles
−30000
−10000
10000
30000
−1
0
1
Theoretical Quantiles