LAB1_CSE3506

Sparsh Raj_19BPS1028

29/01/2022

R Markdown

```
rm(list=ls())
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.1.1
```

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

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

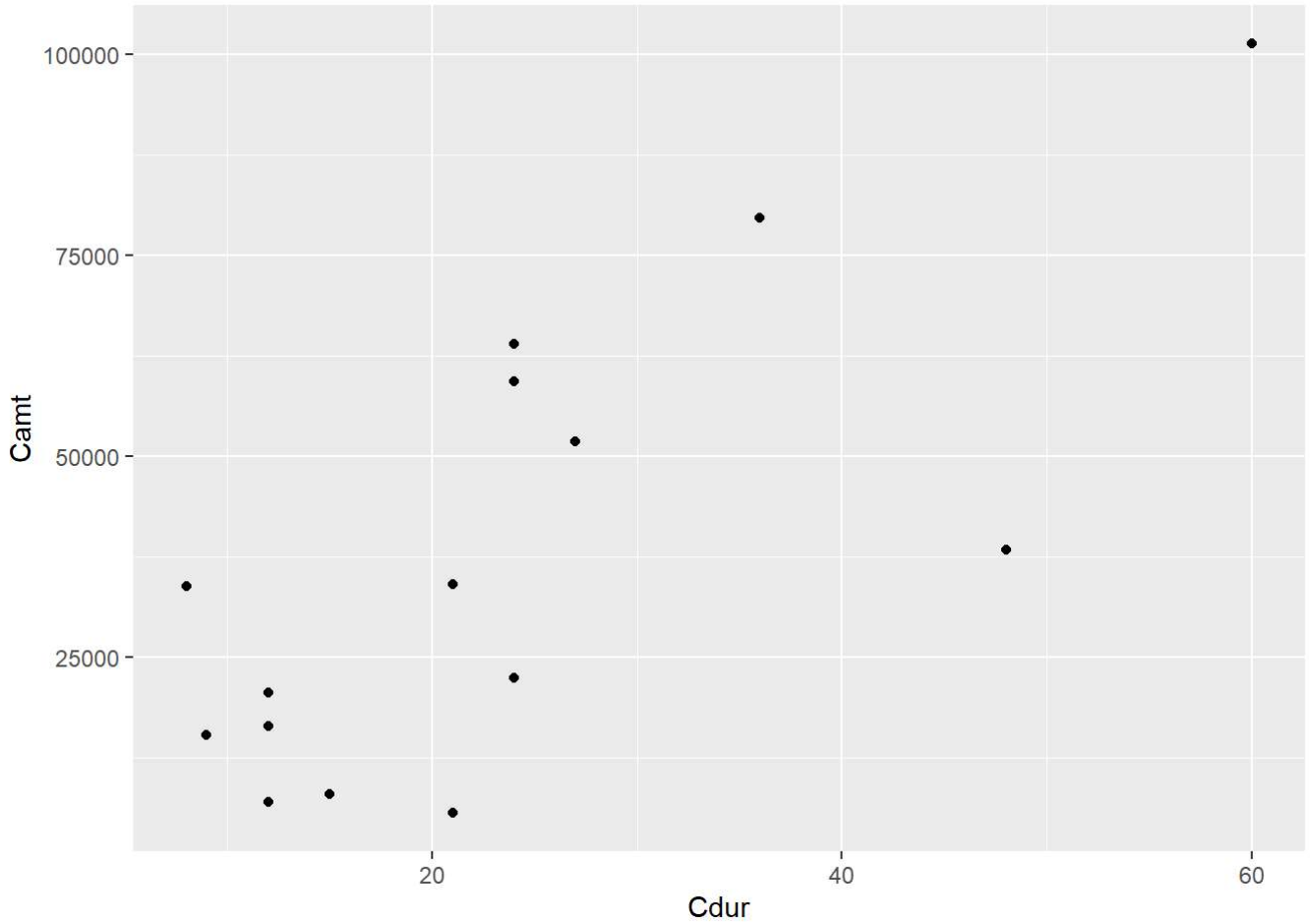
```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data=read.csv("CreditWorthiness.csv",header=TRUE)
data <- sample_n(data,15)
head(data)
```

```
##                  Cbal Cdur               Chist        Cpur   Camt
## 1             Rs. < 0   21 all settled till now   education  34020
## 2             Rs. < 0   12 all settled till now electronics   6970
## 3 no checking account   60 all settled till now electronics 101320
## 4     0 <= Rs. < 2000   24 all settled till now electronics  63910
## 5             Rs. < 0   15 all settled till now    Business   7940
## 6 no checking account   27 all settled till now  renovation  51780
##                  Sbal              Edur InRate
## 1          Rs. < 1000  less than 1 year      2
## 2          Rs. < 1000 more than 7 years      4
## 3 1000 <= Rs. < 5,000      4 to 7 years      2
## 4          Rs. < 1000  less than 1 year      1
## 5          Rs. < 1000      1 to 4 years      4
## 6  no savings account more than 7 years      4
##                                     MSG Oparties              Rdur
## 1                            single male   no one  less than a year
## 2                            single male   no one more than 3 years
## 3 divorced or separated or married female   no one more than 3 years
## 4                            single male   no one      1 to 2 years
## 5 divorced or separated or married female   no one more than 3 years
## 6                            single male   no one more than 3 years
##                            Prop age inPlans Htype NumCred
## 1 life insurance/building society  26    none   own       2
## 2                     real estate  57  stores   own       1
## 3                     real estate  21    none   own       1
## 4                 Other cars etc.  33    none   own       1
## 5 life insurance/building society  22    none   own       1
## 6 life insurance/building society  48    none   own       4
##                           JobType Ndepend telephone foreign creditScore
## 1 employee with official position       1        no      no         bad
## 2              resident unskilled       1        no      no         bad
## 3 employee with official position       1       yes      no        good
## 4 employee with official position       1        no      no        good
## 5              resident unskilled       1        no      no        good
## 6 employee with official position       2       yes      no        good
```

###GGPlot and Correlation calculation

```
library("ggplot2")
ggplot(data,aes(x=Cdur,y=Camt))+geom_point()
```




```
cor.test(data$Camt,data$age)
```

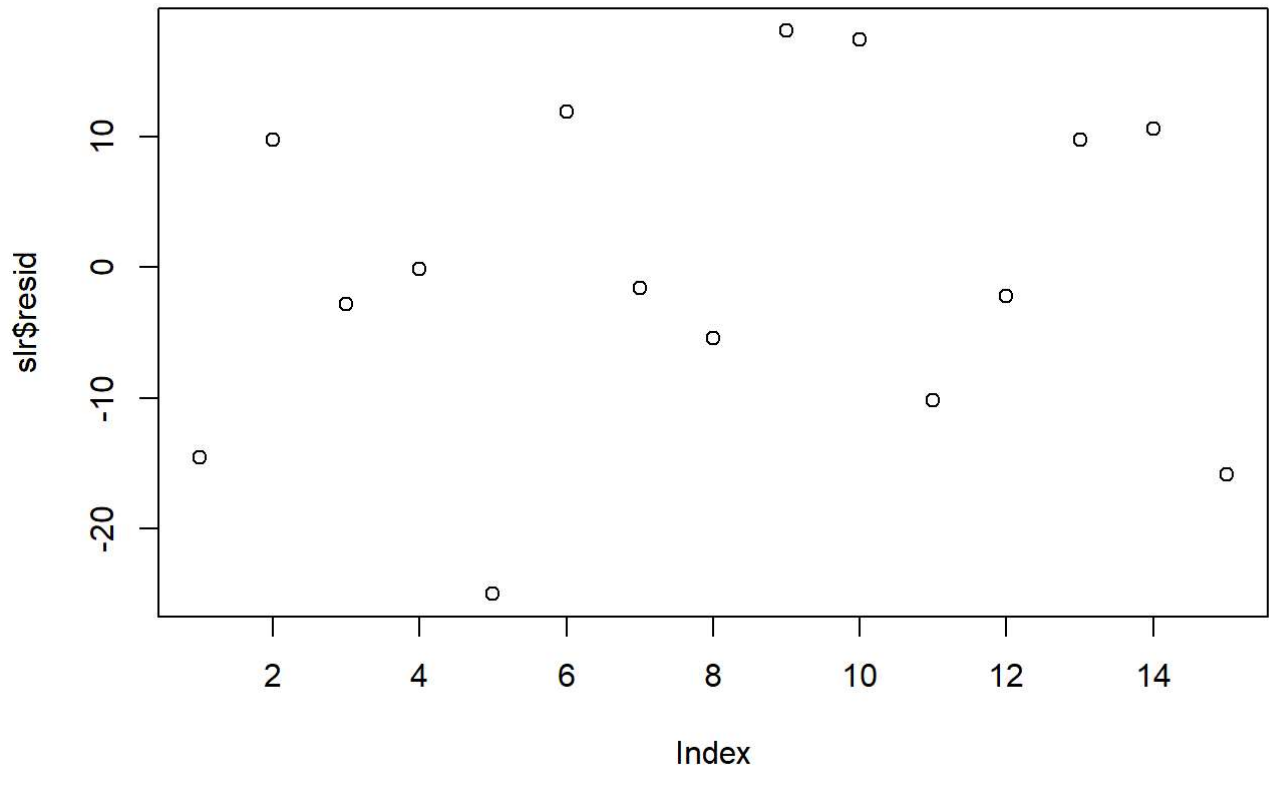
```
##
##  Pearson's product-moment correlation
##
## data:  data$Camt and data$age
## t = -2.002, df = 13, p-value = 0.0666
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.79900901  0.03570498
## sample estimates:
##       cor
## -0.4854367
```

###Linear Regression

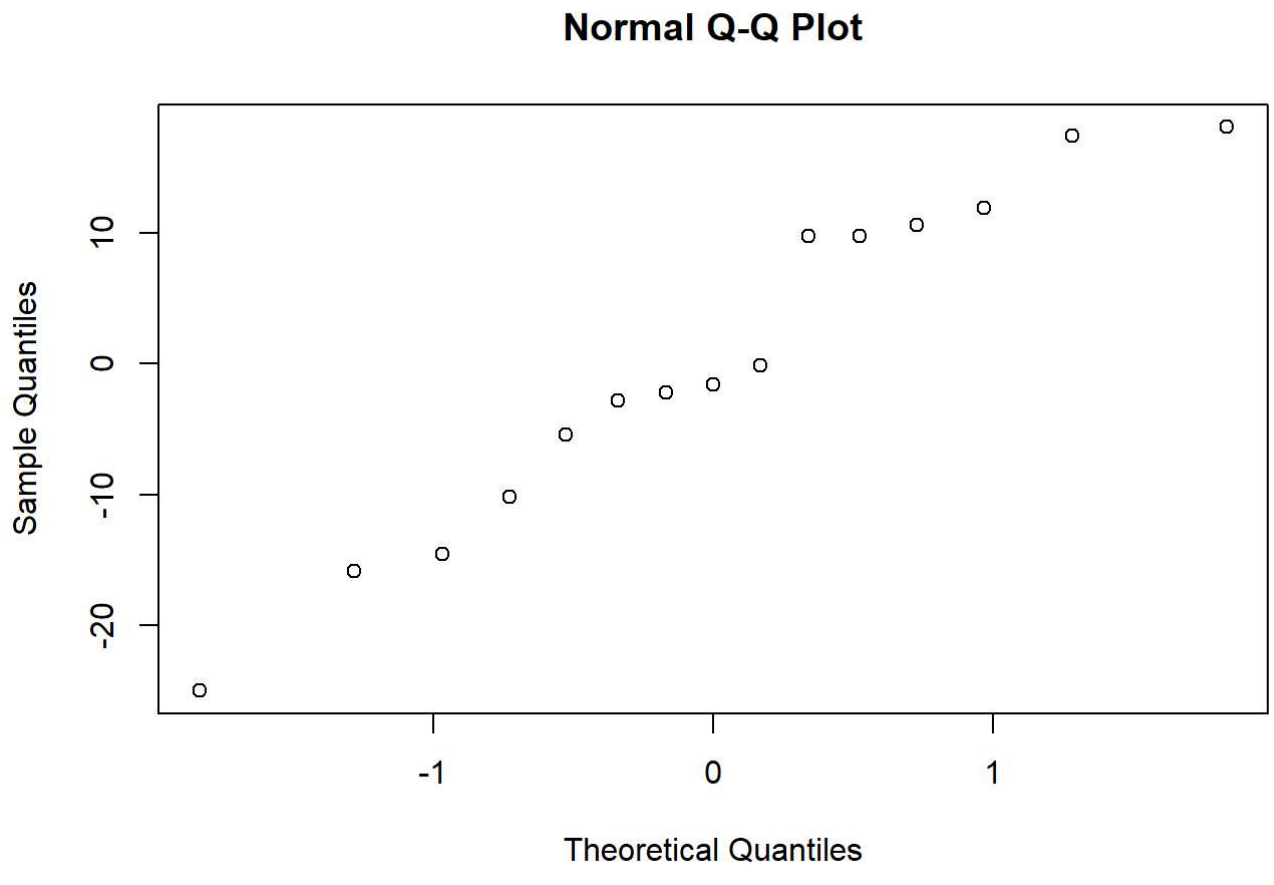
```
slr = lm(age~Camt, data)
summary(slr)
```

```
##
## Call:
## lm(formula = age ~ Camt, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -25.000  -7.800  -1.556  10.187  18.110
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.9746479  5.7582635   8.505 1.14e-06 ***
## Camt        -0.0002486  0.0001242  -2.002   0.0666 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.33 on 13 degrees of freedom
## Multiple R-squared:  0.2356, Adjusted R-squared:  0.1769
## F-statistic: 4.008 on 1 and 13 DF,  p-value: 0.0666
```

```
plot(slr$resid)
```




```
qqnorm(slr$resid)
```

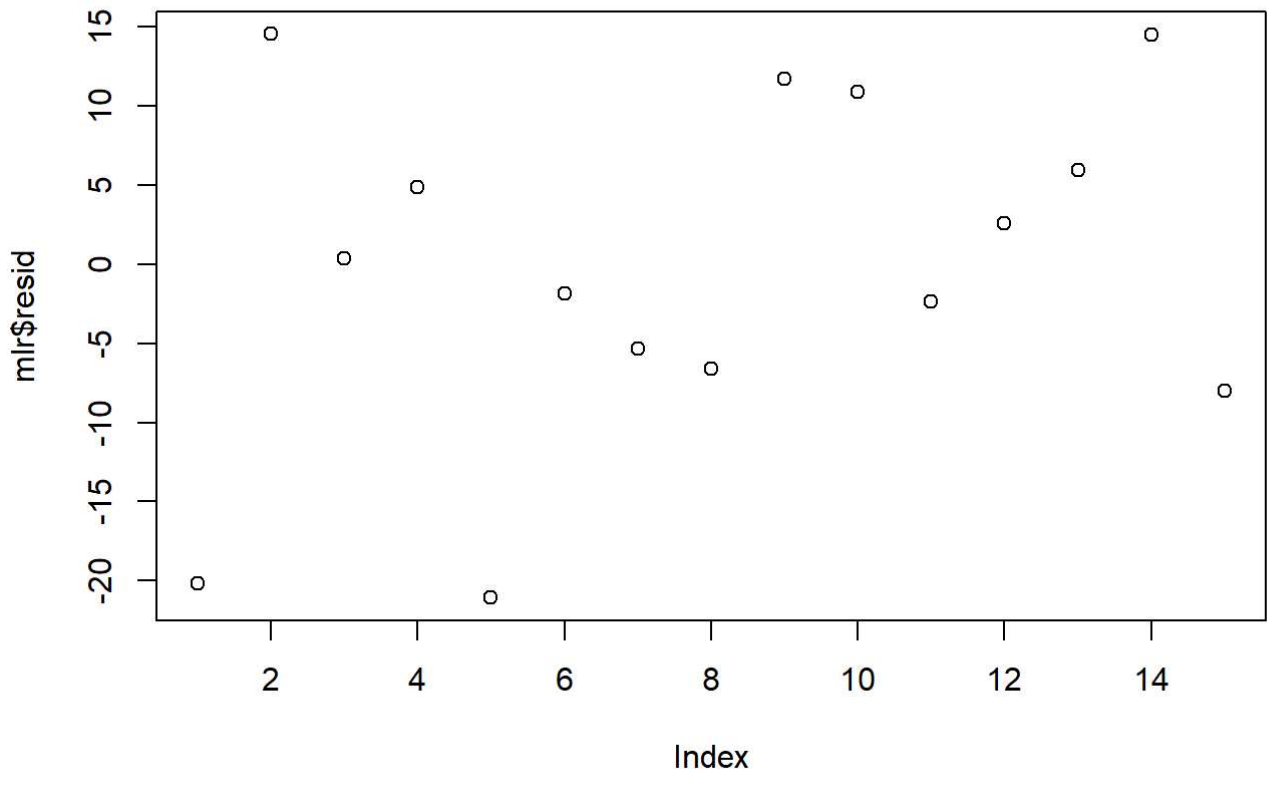


###Multiple Regression

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

```
##
## Call:
## lm(formula = age ~ Cdur + Camt + InRate + NumCred, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.0722  -5.9769   0.3615   8.4120  14.5518
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 44.8159018 13.8625841   3.233  0.00898 **
## Cdur         0.3612352  0.4424100   0.817  0.43322
## Camt        -0.0004740  0.0002512  -1.887  0.08851 .
## InRate      -2.7859707  3.4824562  -0.800  0.44230
## NumCred      7.7451103  4.5019598   1.720  0.11610
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.2 on 10 degrees of freedom
## Multiple R-squared:  0.4238, Adjusted R-squared:  0.1933
## F-statistic: 1.839 on 4 and 10 DF,  p-value: 0.1981
```

```
plot(mlr$resid)
```




```
qqnorm(mlr$resid)
```