SCS

2111

LABORATORY II

TAKE HOME ASSIGNMENT-2

13000853

2013/CS/085

SCS 2111

- **Q1**

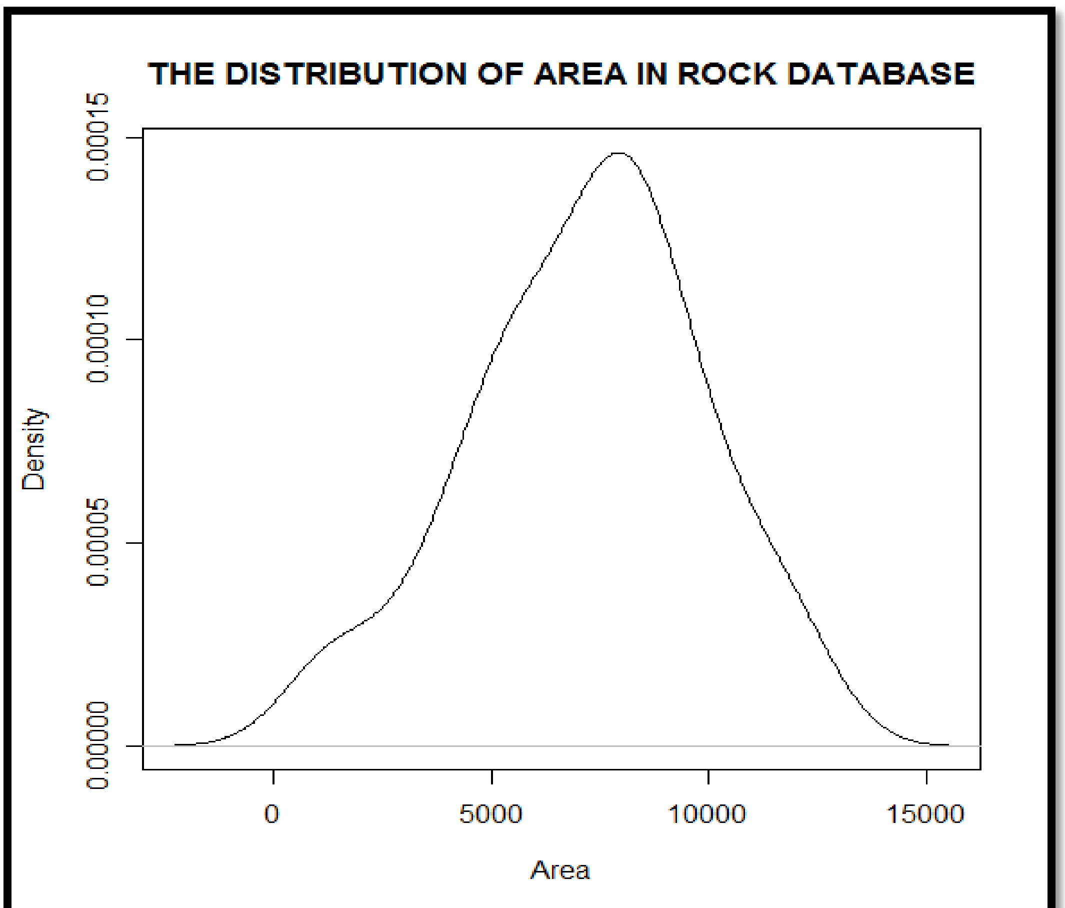
a) Load the data set in the package "dataset"?

```
> data(package="datasets")
> rock
    area     peri     shape   perm
1   4990 2791.900 0.0903296    6.3
2   7002 3892.600 0.1486220    6.3
3   7558 3930.660 0.1833120    6.3
4   7352 3869.320 0.1170630    6.3
5   7943 3948.540 0.1224170   17.1
6   7979 4010.150 0.1670450   17.1
7   9333 4345.750 0.1896510   17.1
8   8209 4344.750 0.1641270   17.1
9   8393 3682.040 0.2036540  119.0
10  6425 3098.650 0.1623940  119.0
11  9364 4480.050 0.1509440  119.0
12  8624 3986.240 0.1481410  119.0
13 10651 4036.540 0.2285950   82.4
14  8868 3518.040 0.2316230   82.4
15  9417 3999.370 0.1725670   82.4
16  8874 3629.070 0.1534810   82.4
17 10962 4608.660 0.2043140   58.6
18 10743 4787.620 0.2627270   58.6
19 11878 4864.220 0.2000710   58.6
20  9867 4479.410 0.1448100   58.6
21  7838 3428.740 0.1138520  142.0
22 11876 4353.140 0.2910290  142.0
23 12212 4697.650 0.2400770  142.0
24  8233 3518.440 0.1618650  142.0
25  6360 1977.390 0.2808870  740.0
26  4193 1379.350 0.1794550  740.0
27  7416 1916.240 0.1918020  740.0
28  5246 1585.420 0.1330830  740.0
29  6509 1851.210 0.2252140  890.0
30  4895 1239.660 0.3412730  890.0
31  6775 1728.140 0.3116460  890.0
32  7894 1461.060 0.2760160  890.0
33  5980 1426.760 0.1976530  950.0
34  5318  990.388 0.3266350  950.0
35  7392 1350.760 0.1541920  950.0
36  7894 1461.060 0.2760160  950.0
```

b) Carryout a descriptive analysis for the above variables and comment on your findings?

<u>Descriptive analysis for Area</u>

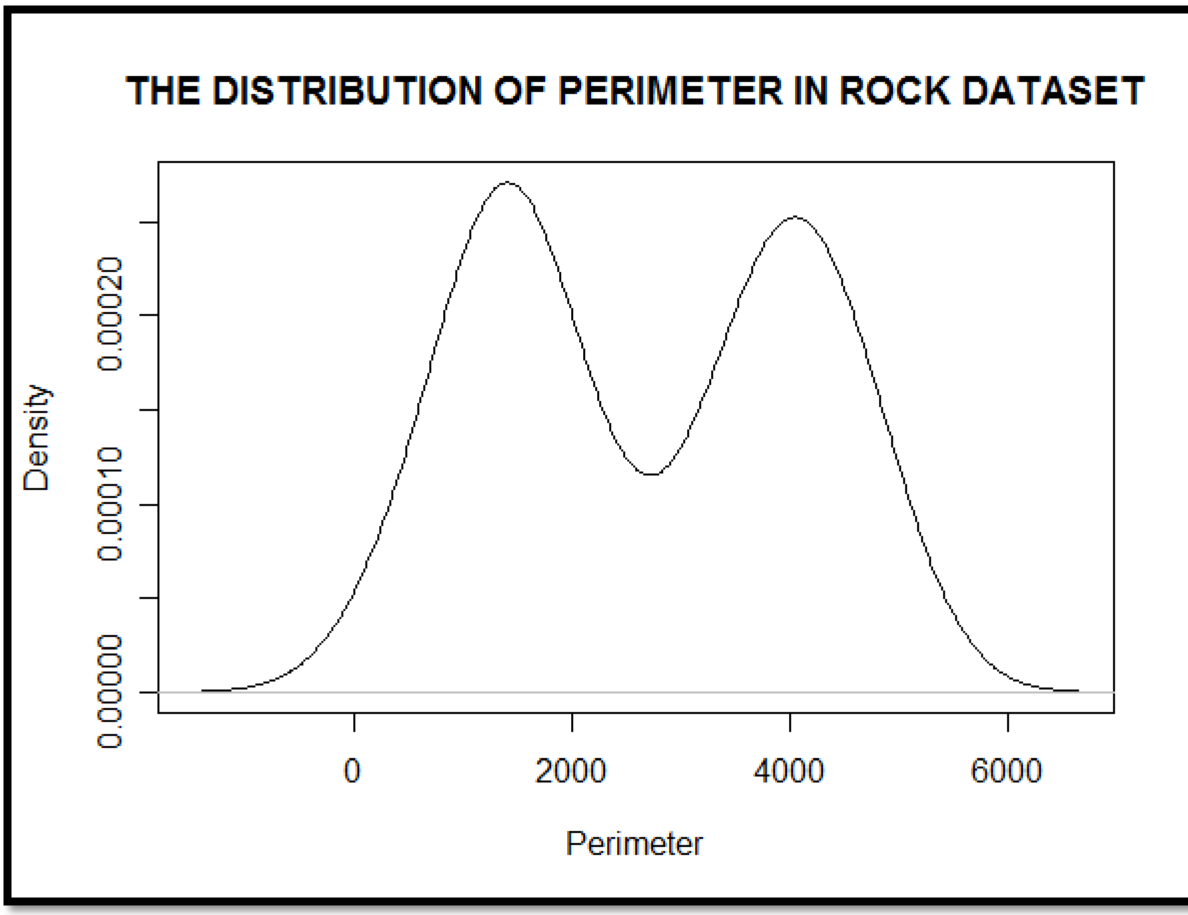
```
> summary(rock$area)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1016    5305    7487    7188    8870   12210
> plot(density(rock$area,na.rn=T),main="THE DISTRIBUTION OF AREA IN ROCK DATABASE",xlab="Area",ylab="Density")
Warning message:
In density.default(rock$area, na.rn = T) :
  non-matched further arguments are disregarded
> plot(density(rock$area,na.rm=T),main="THE DISTRIBUTION OF AREA IN ROCK DATABASE",xlab="Area",ylab="Density")
```



The median value of the area is around 7000. In here normal distribution when consider the minimum, maximum and the median value. The graph for the distribution of area in rock dataset is a normal distribution symmetric graph. It has a peak value near 7000.

Descriptive analysis for Perimeter

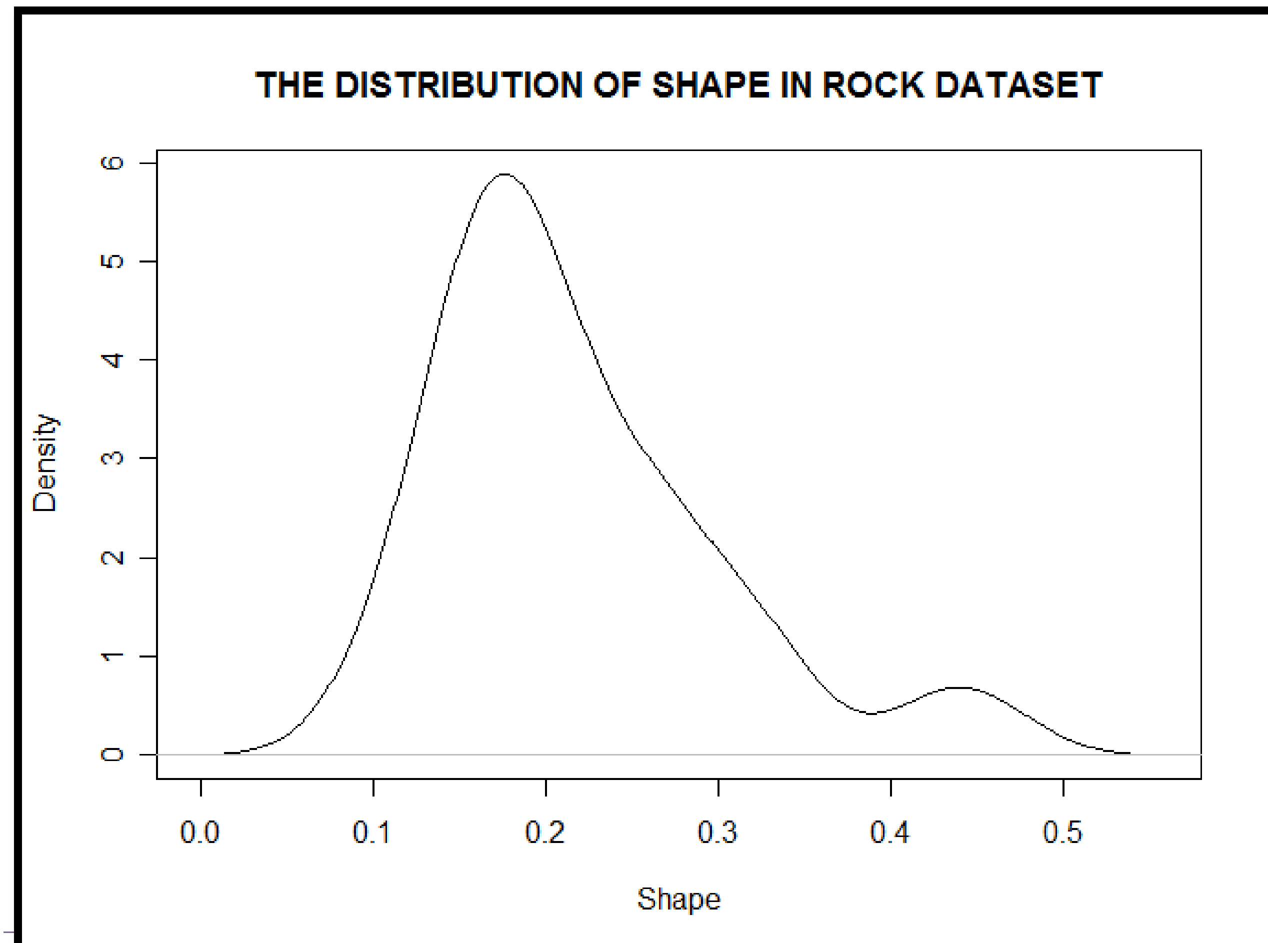
```
> summary(rock$peri)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  308.6  1415.0  2536.0  2682.0  3990.0  4864.0 
> plot(density(rock$peri,na.rm=T),main="THE DISTRIBUTION OF PERIMETER IN ROCK DATASET",xlab="Perimeter",ylab="Density")
> 
```



Distribution of perimeter has a median value of 2536 while mean is 2682. From the distribution graph, that most of the perimeter is distributed between 1000 and 4000 and the graph is a bimodal graph.

Descriptive analysis for Shape

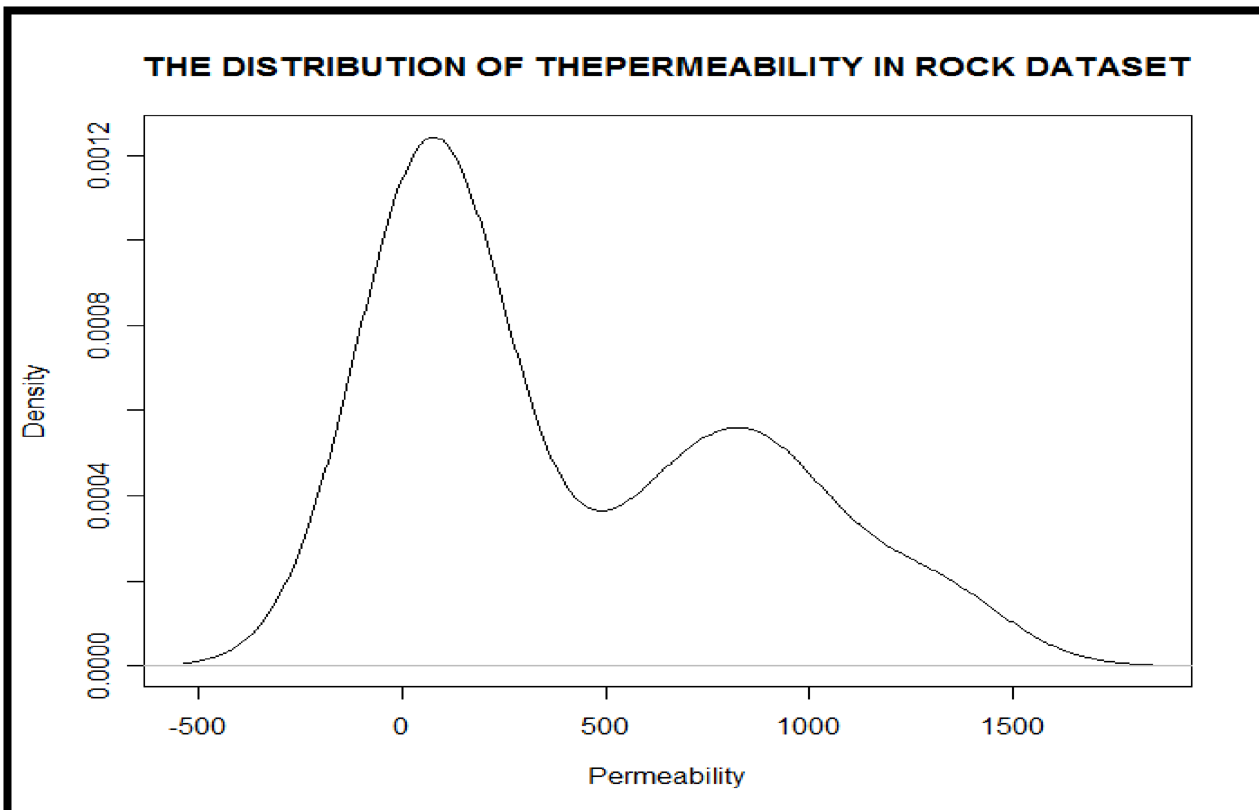
```
> summary(rock$shape)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.09033 0.16230 0.19890 0.21810 0.26270 0.46410
> plot(density(rock$shape,na.rm=T),main="THE DISTRIBUTION OF SHAPE IN ROCK DATASET",xlab="Shape",ylab="Density")
>
```



The shape distribution has a mean value of 0.21810 while 0.19890 value of median. The distribution graph has a continuous distribution with a positive skewness. The peak value is around 0.19.

Descriptive analysis for Permeability

```
> summary(rock$perm)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   6.30   76.45  130.50  415.40  777.50 1300.00 
> plot(density(rock$perm,na.rm=T),main="THE DISTRIBUTION OF THEPERMEABILITY IN ROCK DATASET",xlab="Permeability",ylab="Density")
> 
```



Permeability distribution has a median value around 130.5 and mean value of 415.4. The distribution graph has a positive skewness while graph has two peaks around 100 and 800.

c) Construct 95% confidence interval for the variable “area ” and interpret your results

```
> meanRock<-mean(rock$area)
> meanRock
[1] 7187.729
> lengthRock<-length(rock$area)
> lengthRock
[1] 48
> meanRock<-mean(rock$area)
> meanRock
[1] 7187.729
> sdRock<-sd(rock$area)
> sdRock
[1] 2683.849
> lengthRock<-length(rock$area)
> lengthRock
[1] 48
> errorCL<-qnorm(0.975)*sdRock/sqrt(lengthRock)
> errorCL
[1] 759.2513
> errorLeft<-meanRock-errorCL
> errorCL
[1] 759.2513
> errorLeft
[1] 6428.478
> errorRight<-meanRock+errorCL
> errorRight
[1] 7946.98
>
```

Since the standard deviation of area distribution is 2683.849, margin of error for the variable area at 95% confidence level is 759.2513 pixels. The confidence interval is between 6428.48 and 7946.98 pixels.

d) A resercher claims that the area of pores space is greater than 7000 pixels.Formulate suitable hypothese to test the researcher's claim.Assuming the area is normally distributed test the validity of the resercher's claim and interpret your results

```
> meanRock<-mean(rock$area)
> sdRock<-sd(rock$area)
> lengthRock<-lenght(rock$area)
Error: could not find function "lenght"
> lengthRock<-length(rock$area)
> zValue<-(meanRock-7000)/(sdRock/sqrt(lengthRock))
> zValue
[1] 0.4846122
> pValue<-pnorm(zValue)
> pValue
[1] 0.6860243
>
```

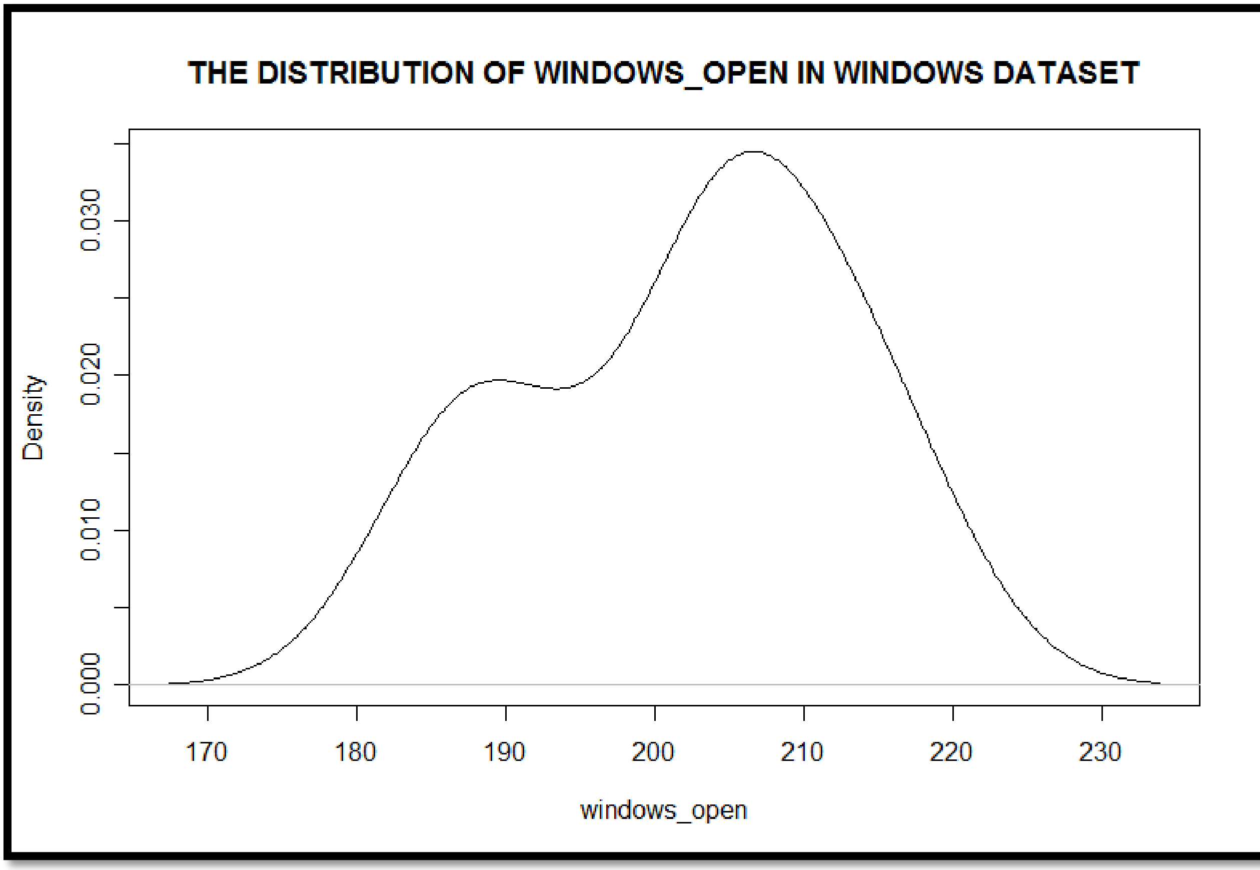
- **Q2**

a) Carryout the discriptive analysis for the two samples and comment on your findings.

```
> Windows_open<-c(202.0,204.5,207.0,215.5,190.8,215.6,208.8,187.8,204.1,185.7)
> Windows_closed<-c(193.5,192.2,199.4,177.6,205.4,200.6,181.8,169.2,172.2,192.8)
> Windows<-data.frame(Windows_open,Windows_closed)
> Windows
   Windows_open Windows_closed
1         202.0          193.5
2         204.5          192.2
3         207.0          199.4
4         215.5          177.6
5         190.8          205.4
6         215.6          200.6
7         208.8          181.8
8         187.8          169.2
9         204.1          172.2
10        185.7          192.8
> 
```

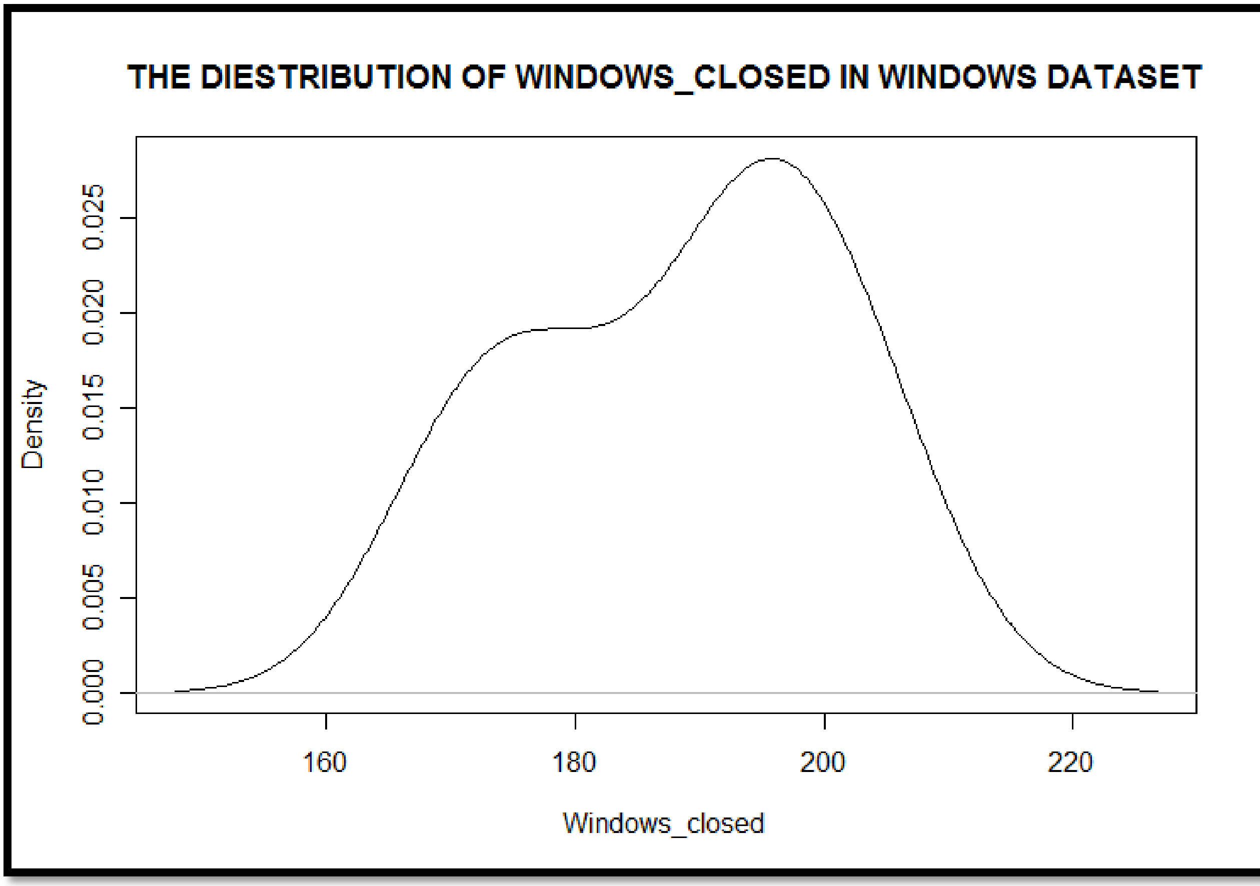
Descriptive analysis for Windows open varible

```
> summary(Windows_open)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  185.7   193.6   204.3   202.2   208.4   215.6 
> plot(density(Windows_open,na.rm=T),main="THE DISTRIBUTION OF WINDOWS DATASET",xlab="windows_open",ylab="Density")
> 
```



Descriptive analysis for Windows closed variable

```
> summary(Windows_closed)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  169.2   178.6   192.5   188.5   197.9   205.4
> plot(density(Windows_closed,na.rm=T),main="THE DIESTRIBUTION OF WINDOWS_CLOSED IN WINDOWS DATASET",xlab="Windows_closed",ylab="Density")
> 
```



Interpretation

Median value of Windows_open variable distribution is 204.3 while for Windows_closed variable distribution is 192.5. According to the distribution graph of Windows_open, it has two mode at values 185 and 205 nearly. Similarly for the distribution graph of Windows_closed, two mode values are at 175 and 195 nearly. But the peak values are highest for the Windows_open variable while both graphs are bimodal. Similarly mean and The median values are also highest for the Windows_open variable. Therefore it shows that most sales happen during the days where windows are open.

b) Construct 95% confidence interval for this incident and interpret your results.

```
> meanOpen<-mean(Windows_open)
> sdOpen<-sd(Windows_open)
> sdOpen
[1] 10.75772
> lengthOpen<-length(Windows_open)
> errorOpen<-qt(0.975,df=lengthOpen-1)*sdOpen/sqrt(lengthOpen)
> errorOpen
[1] 7.695606
> errorOpenLeft<-meanOpen-errorOpen
> errorOpenLeft
[1] 194.4844
> errorOpenright<-meanOpen+errorOpen
> errorOpenRight
Error: object "errorOpenRight" not found
> errorOpenRight<-meanOpen+errorOpen
> errorOpenRight
[1] 209.8756
> meanClosed<-mean(Windows_closed)
> sdClosed<-sd(Windows_closed)
> sdClosed
[1] 12.51613
> lengthClosed<-length(Windows_closed)
> errorClosed<-qt(0.975,df=lengthClosed-1)*sdClosed/sqrt(lengthClosed)
> errorClosed
[1] 8.953498
> errorClosedLeft<-meanClosed-errorClosed
> errorClosedLeft
[1] 179.5165
> errorClosedRight<-meanClosed+errorClosed
> errorClosedRight
[1] 197.4235
>
```

Since the standard deviation of Windows_open distribution is 10.76, margin of error for the variable Windows_open at 95% confidence level is 7.696. The confidence interval is distributed between 194.48 and 209.87.

Since the standard deviation of Windows_closed distribution is 12.52, margin of error for the variable Windows_closed at 95% confidence level is 8.953. The confidence interval is distributed between 179.52 and 197.42

c) Investigate the baker's belief by formulating a suitable hypothesis and interpret your results.

```
> t.test(Windows_open,Windows_closed,alternative="greater")

        Welch Two Sample t-test

data:  Windows_open and Windows_closed
t = 2.6269, df = 17.603, p-value = 0.008659
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 4.648803      Inf
sample estimates:
mean of x mean of y
   202.18    188.47

>
```

Here the p-value of the test is 0.01732, which is <0.05. Therefore we reject the null hypothesis. Since we reject the hypothesis, alternative is true. That means sales on window open days is higher than in closed days. So baker's belief is true.