Take home assignment

SCS 2111 Laboratory II

W.M.N.K.Madushanka

2013/CS/073

13000731

Q1

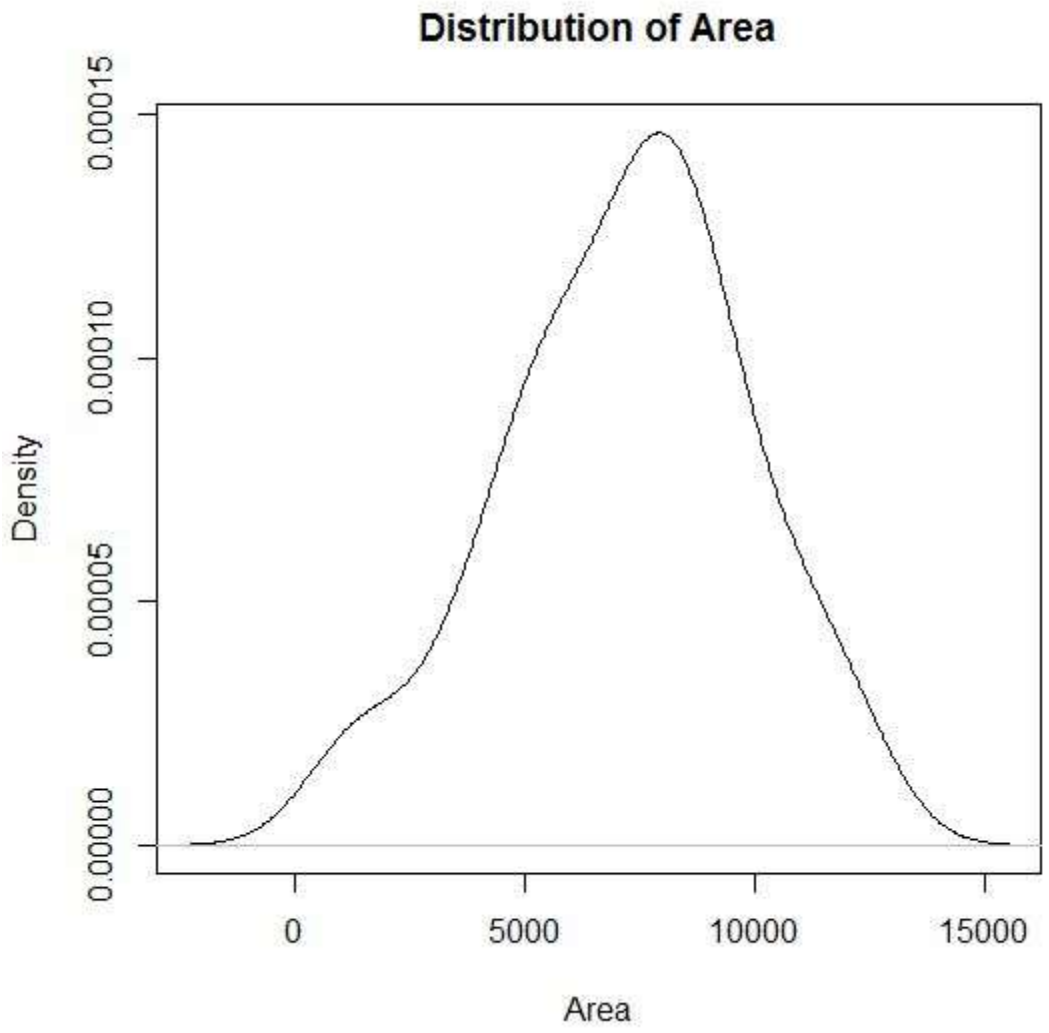
a)

```
> 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
```

b)

Analysis for area

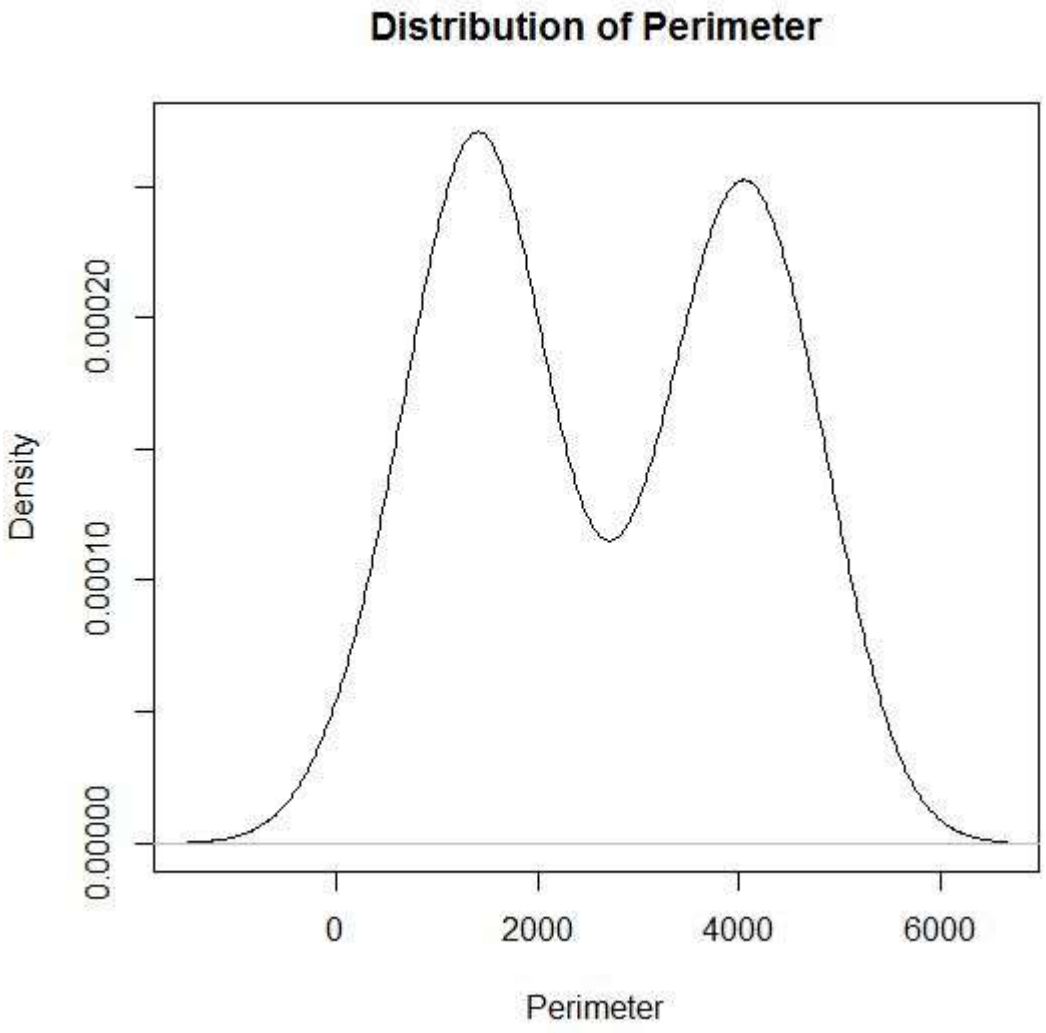
```
> summary(rock$area)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1016    5305    7487    7188    8870   12210
> plot(density(rock$area,na.rm=T),main="Distribution of Area",xlab="Area",ylab="Density")
```



This is a normal distribution symmetric graph. Graph has reached the peak value at about 7500. Also shows the median value around 7500.

Analysis for Perimeter

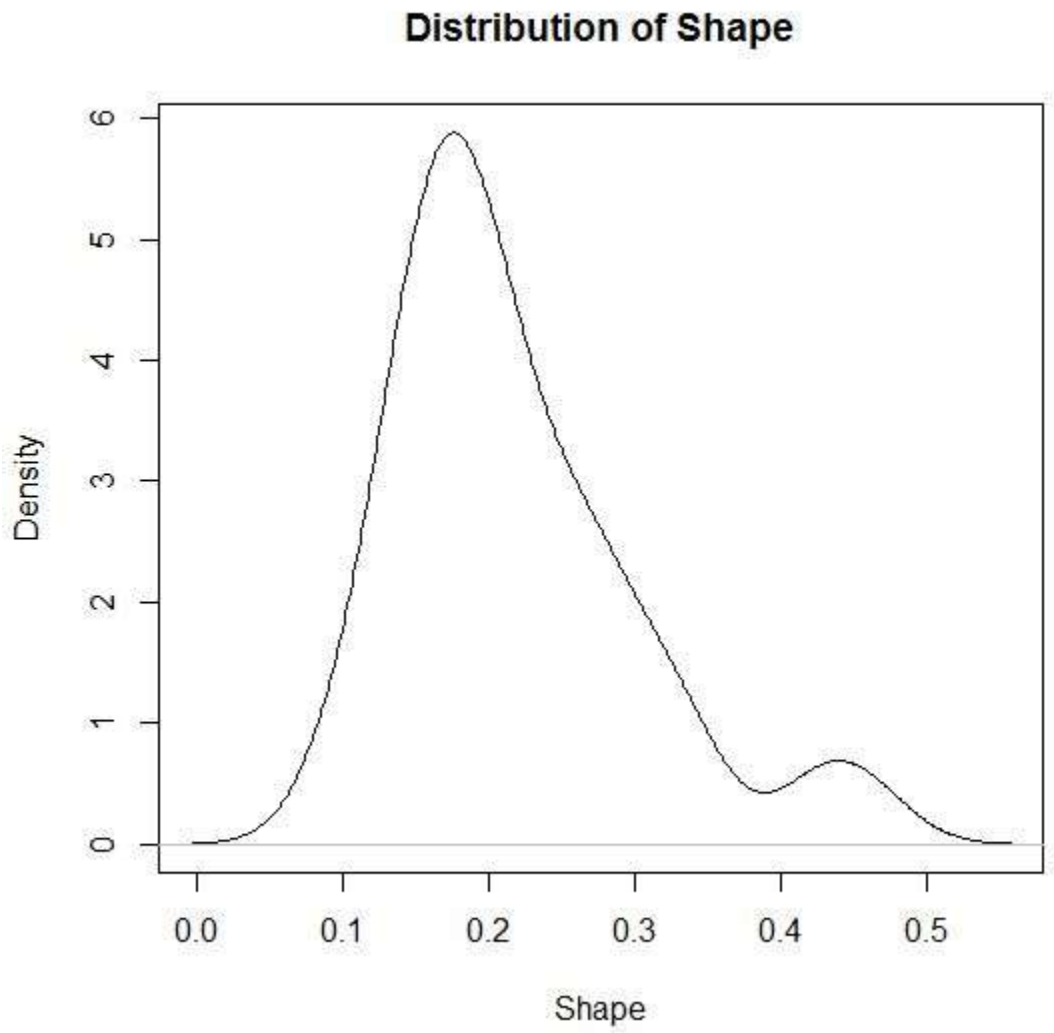
```
> summary(rock$peri)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  308.6  1415.0  2536.0  2682.0  3990.0  4864.0
> plot(density(rock$peri,na.rm=T),main="Distribution of Perimeter",xlab="Perimeter",ylab="Density")
```



The graph is a bimodal graph. Median is around 2500. Graph has two peaks and the density of one is greater than 0.00025 and other is very close to 0.00025.

Analysis for Shape

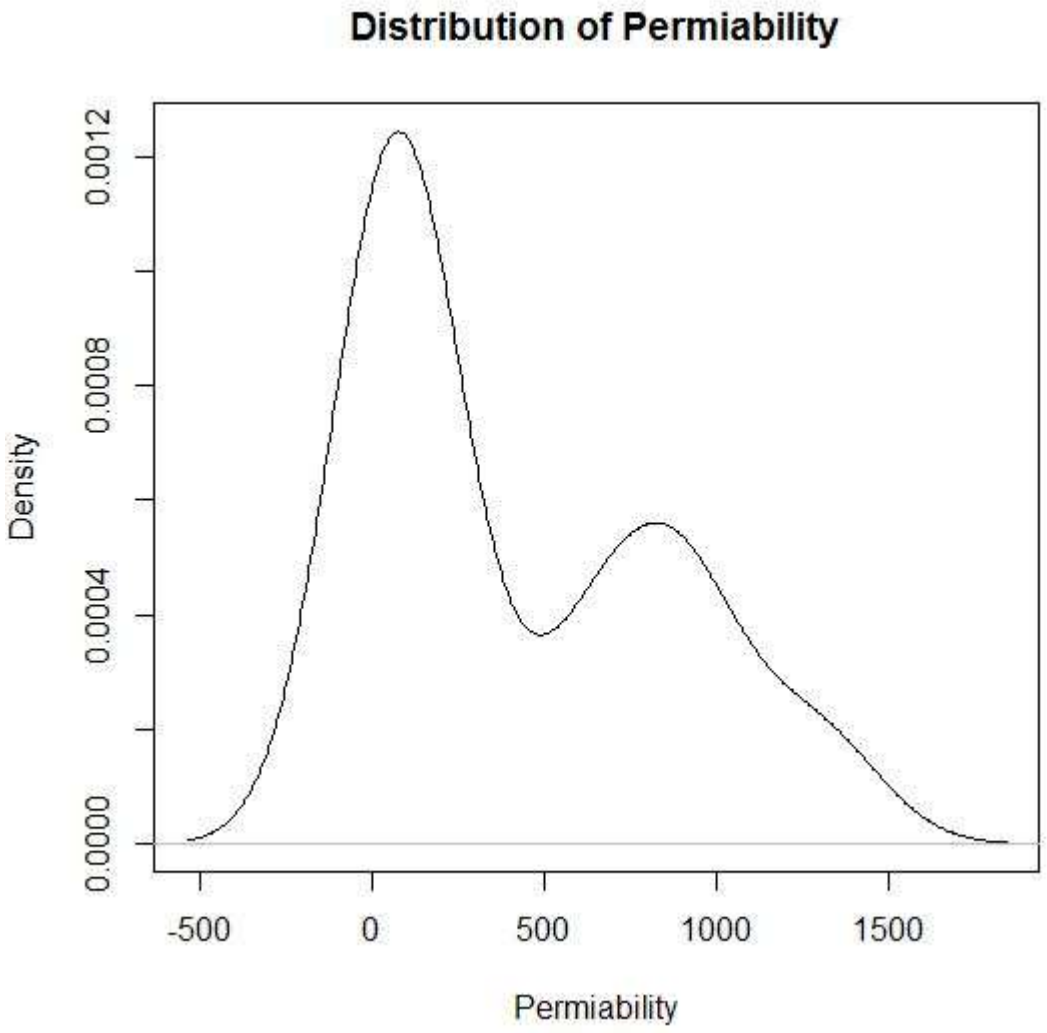
```
> summary(rock$shape)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.09033 0.16230 0.19890 0.21810 0.26270 0.46410
> plot(density(rock$shape,na.rm=T),main="Distribution of Shape",xlab="Shape",ylab="Density")
```



Peak value is very close to 0.2 in shape. Graph has positive skewness and has a continuous distribution. Mean value is0.21810 while median is 0.19890.

Analysis of 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="Distribution of Permiability",xlab="Permiability",ylab="Density")
```



The distribution graph has a positive skewness while graph has two peaks around 100 and 800.

c)

```
> M<-mean(rock$area)
> M
[1] 7187.729
> SD<-sd(rock$area)
> SD
[1] 2683.849
> L<-length(rock$area)
> L
[1] 48
> error1<-qnorm(0.975)*SD/sqrt(L)
> error1
[1] 759.2513
> error2<-M-error1
> error2
[1] 6428.478
> error3<-M+error1
> error3
[1] 7946.98
```

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

d)

```
> M<-mean(rock$area)
> SD<-sd(rock$area)
> L<-length(rock$area)
> x<-(M-7000)/(SD/sqrt(L))
> x
[1] 0.4846122
> y<-pnorm(x)
> y
[1] 0.6860243
```

Hypothesis will not be rejected Since pnorm() is greater than 0.05.

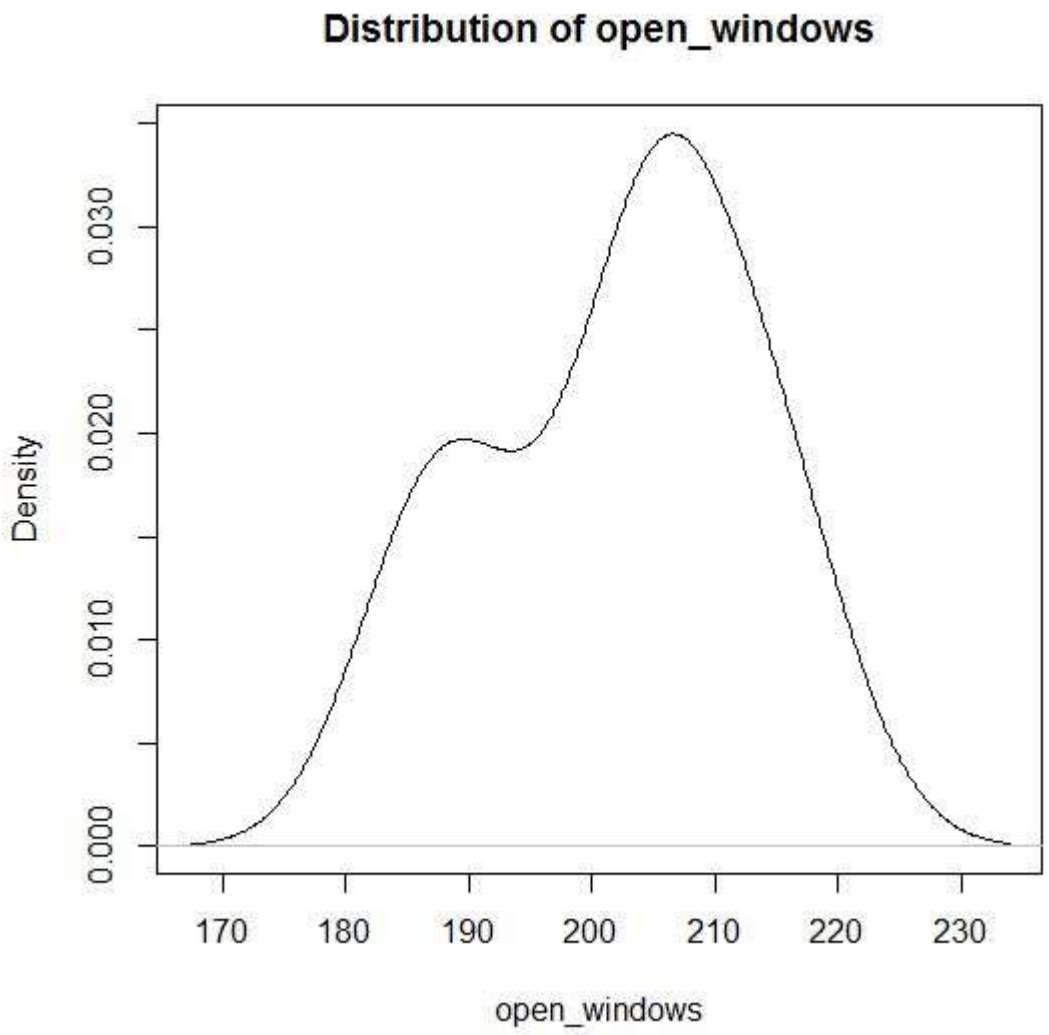
Q2

a)

```
> open_windows<-c(202.0,204.5,207.0,215.5,190.8,215.6,208.8,187.8,204.1,185.7)
> closed_windows<-c(193.5,192.2,199.4,177.6,205.4,200.6,181.8,169.2,172.2,192.8)
> windows<-data.frame(open_windows,closed_windows)
> windows
   open_windows closed_windows
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
```

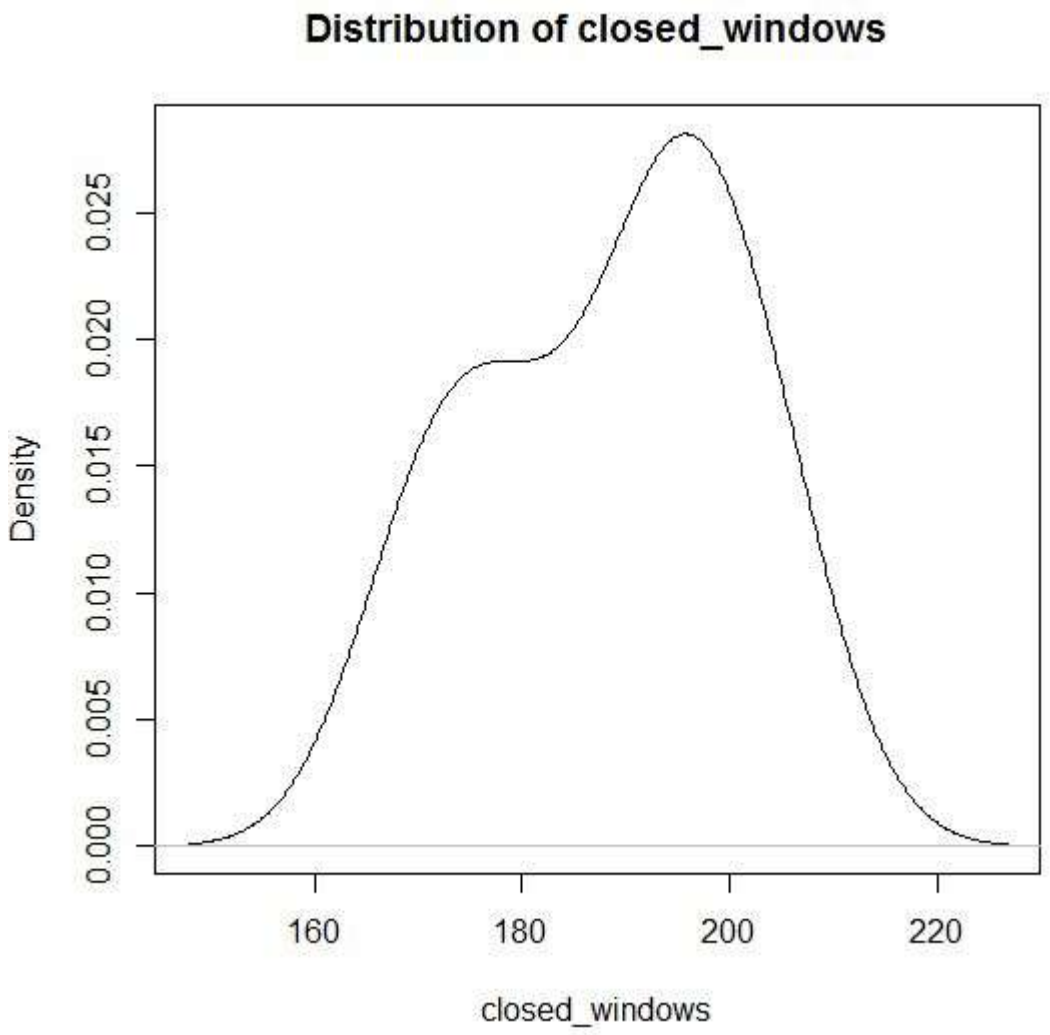
Analysis of open_windows variable

```
> summary(open_windows)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  185.7   193.6   204.3   202.2   208.4   215.6
> plot(density(open_windows,na.rm=T),main="Distribution of open_windows",xlab="open_windows",ylab="Density")
```



Analysis of closed_windows variable

```
> summary(closed_windows)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  169.2   178.6   192.5   188.5   197.9   205.4
> plot(density(closed_windows,na.rm=T),main="Distribution of closed_windows",xlab="closed_windows",ylab="Density")
```



open_windows distribution variable has median value of 204.3 while for closed_windows variable distribution is 192.5. According to graphs open_windows have two mode values at values 185 and 205 nearly and closed_windows also have two mode values at 175 and 195 nearly. Both graphs are bimodal.

b)

```
> O_mean<-mean(open_windows)
> O_sd<-sd(open_windows)
> O_sd
[1] 10.75772
>
> O_length<-length(open_windows)
> O_error<-qt(0.975,df=O_length-1)*O_sd/sqrt(O_length)
> O_error
[1] 7.695606
>
> O_error_L<-O_mean-O_error
> O_error_L
[1] 194.4844
>
> O_error_R<-O_mean+O_error
> O_error_R
[1] 209.8756
> C_mean<-mean(closed_windows)
> C_sd<-sd(closed_windows)
> C_sd
[1] 12.51613
>
> C_length<-length(closed_windows)
> C_error<-qt(0.975,df=C_length-1)*C_sd/sqrt(C_length)
> C_error
[1] 8.953498
>
> C_error_L<-C_mean-C_error
> C_error_L
[1] 179.5165
>
> C_error_R<-C_mean+C_error
> C_error_R
[1] 197.4235
```

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

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

c)

```
> t.test(open_windows,closed_windows,alternative="greater")

        Welch Two Sample t-test

data:  open_windows and closed_windows
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 = 0.01732 <0.05. Therefore we have to reject null hypothesis. Since we reject the hypothesis, alternative is true. That means sales on window open days is higher than in closed days. Therefore baker's belief is true.