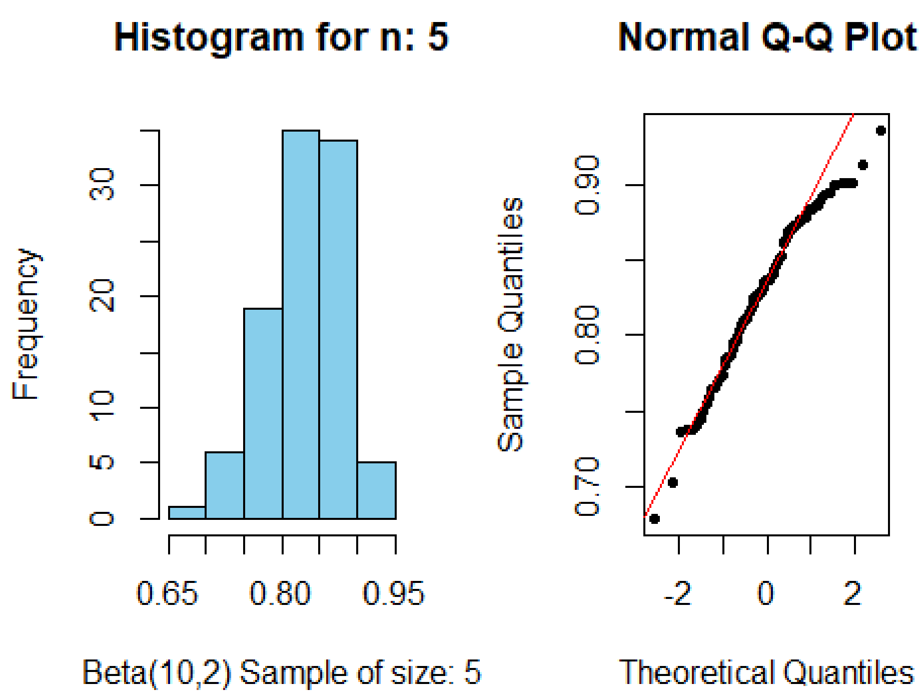
SuchetaHW8

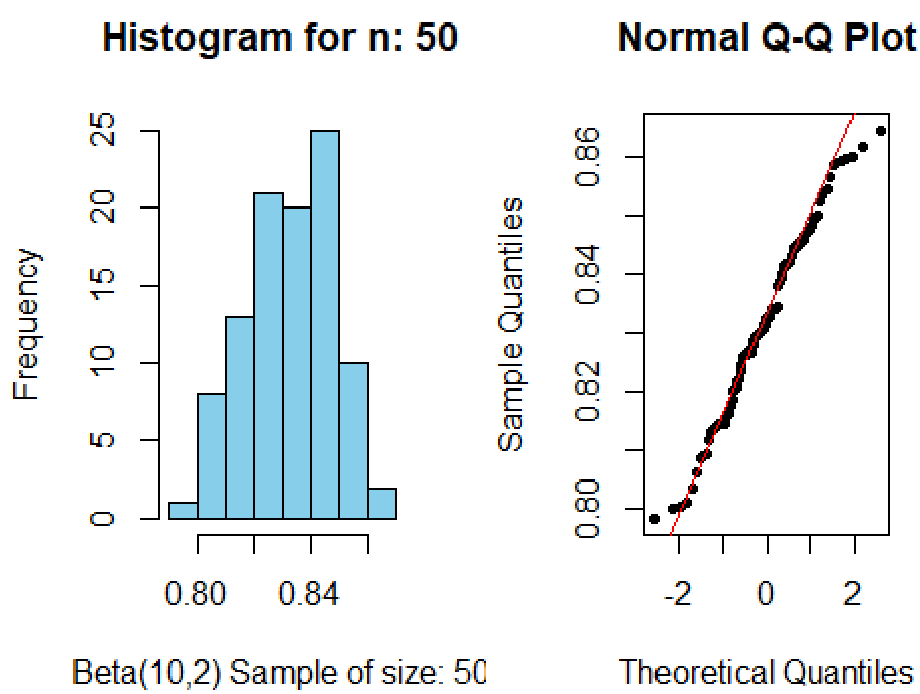
Sucheta

24/11/2021

```
#Solution 1

#100 trials of Beta(10,2)

#Sample size 5
n = 5  #Sample Size
sim.size = 100    #number of trials
mean1 = rep(NA,n)
for (i in 1:sim.size){   #To generate beta distribution
  beta1 = rbeta(n,10,2)
  mean1[i] = mean(beta1)
}
par(mfrow=c(1,2))
hist(mean1,col='skyblue',main = paste("Histogram for n:",n),
         xlab=paste("Beta(10,2) Sample of size:",n)) #Plot the histogram
qqnorm(mean1,pch=20)   #QQ plot to check if follow normal distribution
qqline(mean1,col='red')
```

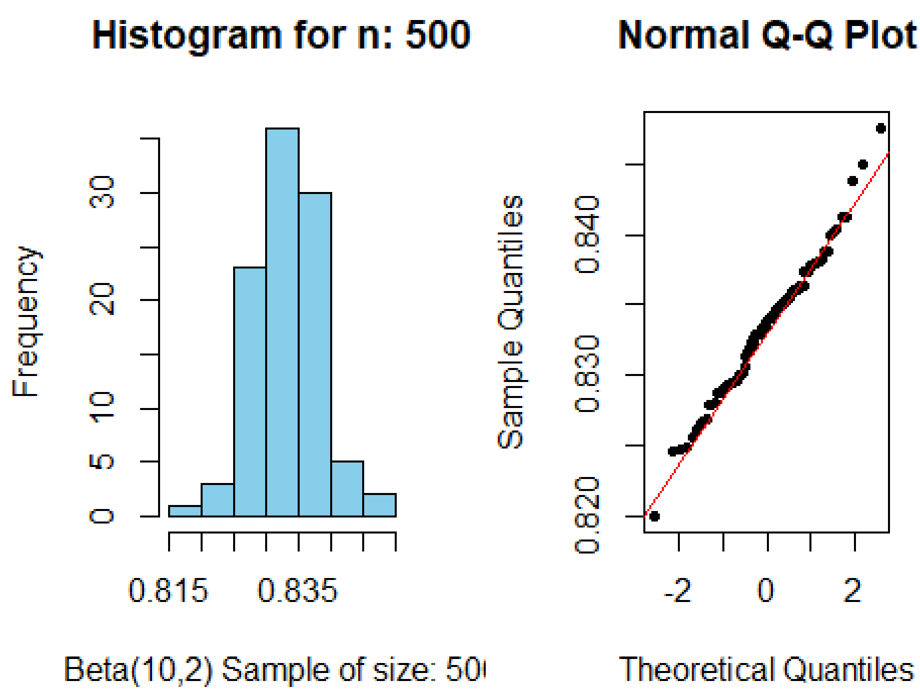


```
#The data is left skewed which can be seen from the histogram.Also, from
#the QQ plot we can infer that the data is away from the normal distribution.

#Sample size 50
n= 50  #Sample Size
sim.size = 100    #number of trials
mean2 = rep(NA,n)
for (i in 1:sim.size){  #To generate beta distribution
  beta2 = rbeta(n,10,2)
  mean2[i] = mean(beta2)
}

par(mfrow=c(1,2))
hist(mean2,col='skyblue',main = paste("Histogram for n:",n),
         xlab=paste("Beta(10,2) Sample of size:",n)) #Plot the histogram
qqnorm(mean2,pch=20)   #QQ plot to check if follow normal distribution
qqline(mean2,col='red')
```



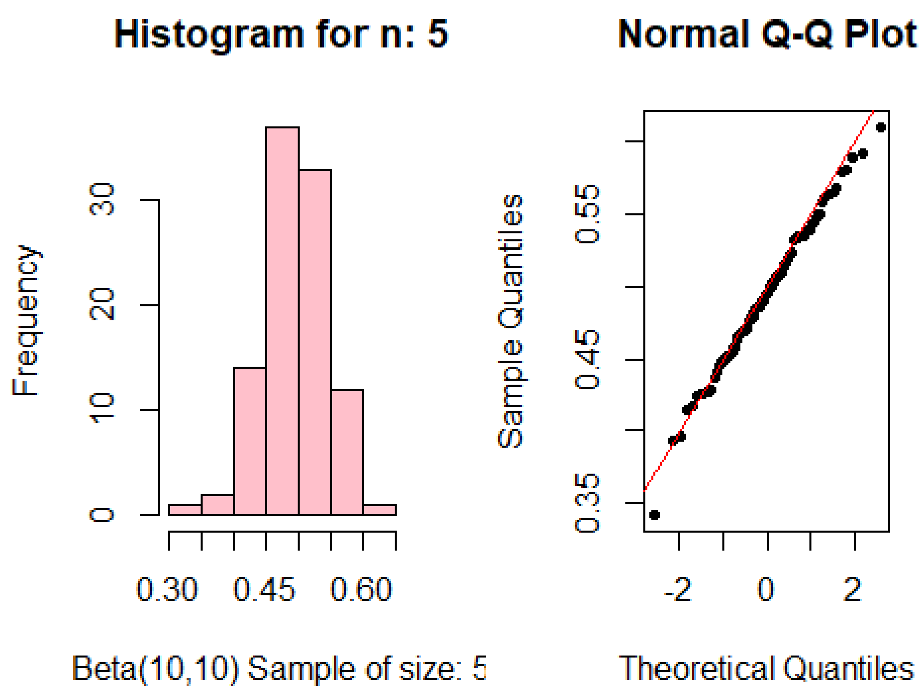

```
#The data is not skewed but it is not symmetric which can be seen from the
histogram.
#Also, from the QQ plot we can infer that the some of the boundary points
# of the data are away from the normal distribution.


#Sample size 500
n = 500  #Sample Size
```

```r
sim.size = 100    #number of trials
mean3 = rep(NA,n)
for (i in 1:sim.size){  #To generate beta distribution
  beta3 = rbeta(n,10,2)
  mean3[i] = mean(beta3)
}

par(mfrow=c(1,2))
hist(mean3,col='skyblue',main = paste("Histogram for n:",n),
         xlab=paste("Beta(10,2) Sample of size:",n)) #Plot the histogram
qqnorm(mean3,pch=20)   #QQ plot to check if follow normal distribution
qqline(mean3,col='red')
```



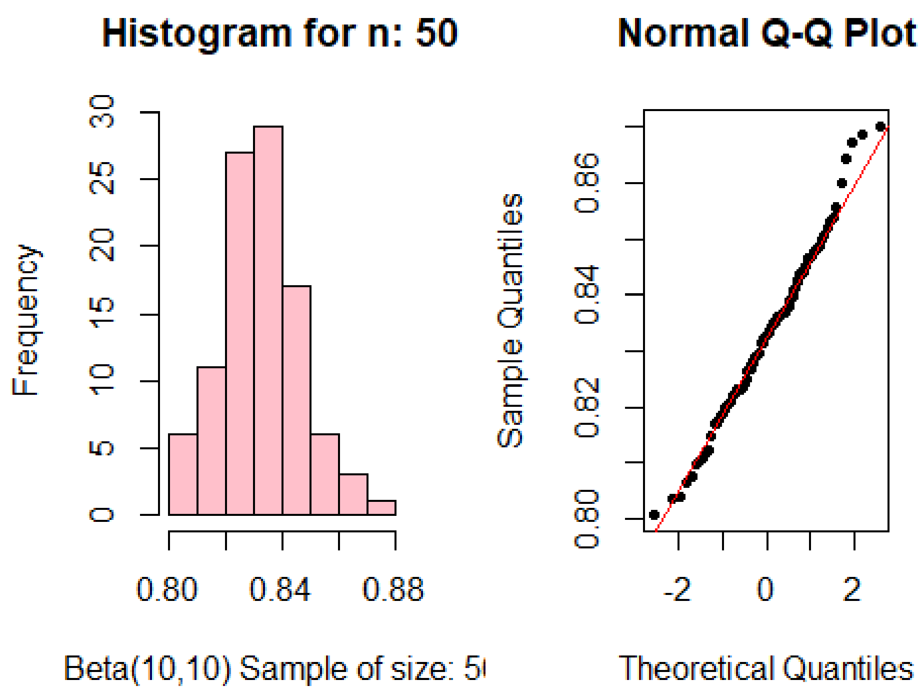

```r
#The data is not skewed and is also symmetric which can be seen from the
histogram.
#Also, from the QQ plot we can infer that the some of the boundary points of
#the data are still away from the normal distribution, but it is more
#"normal" than the previous samples.


#100 trials of Beta(10,10)

#Sample size 5
n = 5  #Sample Size
sim.size = 100    #number of trials
mean1 = rep(NA,n)
```

```
for (i in 1:sim.size){   #To generate beta distribution
  beta1 = rbeta(n,10,10)
  mean1[i] = mean(beta1)
}
par(mfrow=c(1,2))
hist(mean1,col='pink',main = paste("Histogram for n:",n),
         xlab=paste("Beta(10,10) Sample of size:",n)) #Plot the histogram
qqnorm(mean1,pch=20)   #QQ plot to check if follow normal distribution
qqline(mean1,col='red')
```

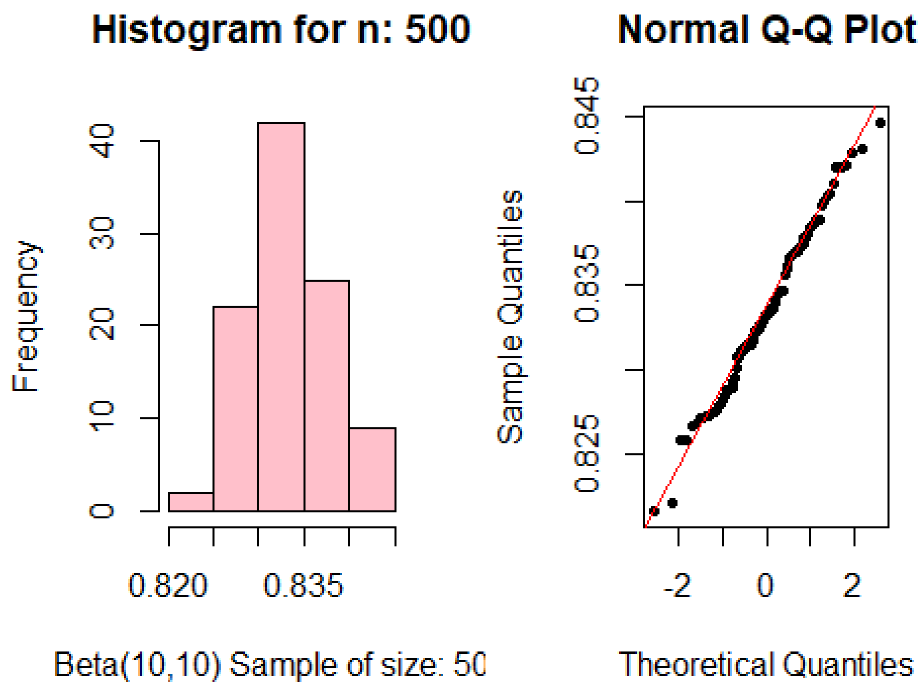



```
#The data is not symmetric which can be seen from the histogram. Also, from
#the QQ plot we can infer that the data is away from normal distribution
#since most of the points are not on the straight line

#Sample size 50
n= 50  #Sample Size
sim.size = 100    #number of trials
mean2 = rep(NA,n)
for (i in 1:sim.size){  #To generate beta distribution
  beta2 = rbeta(n,10,2)
  mean2[i] = mean(beta2)
}

par(mfrow=c(1,2))
hist(mean2,col='pink',main = paste("Histogram for n:",n),
         xlab=paste("Beta(10,10) Sample of size:",n)) #Plot the histogram
```

```
qqnorm(mean2,pch=20)   #QQ plot to check if follow normal distribution
qqline(mean2,col='red')
```




```
##The data is left skewed which can be seen from the histogram. Also, from
#the QQ plot we can infer that the many points of the
#data are away from normal distribution.


#Sample size 500
n = 500  #Sample Size
sim.size = 100    #number of trials
mean3 = rep(NA,n)
for (i in 1:sim.size){  #To generate beta distribution
  beta3 = rbeta(n,10,2)
  mean3[i] = mean(beta3)
}

par(mfrow=c(1,2))
hist(mean3,col='pink',main = paste("Histogram for n:",n),
         xlab=paste("Beta(10,10) Sample of size:",n)) #Plot the histogram
qqnorm(mean3,pch=20)   #QQ plot to check if follow normal distribution
qqline(mean3,col='red')
```

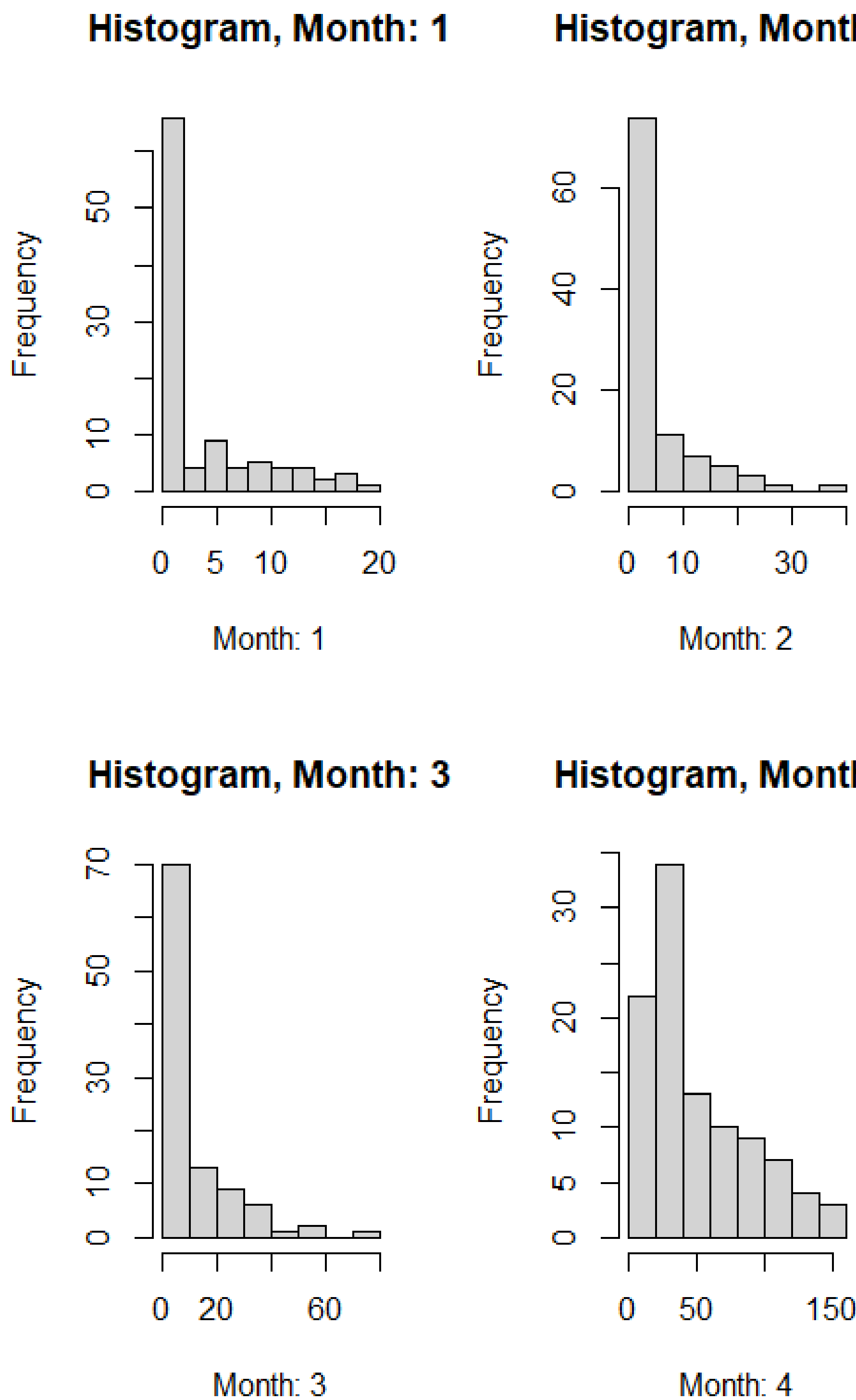



```
##The data is not skewed but it isn't symmetric which can be seen from the
histogram.
#Also, from the QQ plot we can infer that the most of the points of the
#data are close normal distribution since they lie on the straight line.


#According to the Central Limit theorem, as we increase the sample size,
#our distribution must tend towards a normal distribution. In both the cases
#we can see as sample size increases the histograms become less skewed and
more
#symmetric. The points of QQ plot also start coinciding with the the normal
line.


#Solution 2.a

pois_mean = rep(NA,100)
for (i in 1:100){   #Generate a Poisson(1) distribution
  pois_mean[i] = mean(rpois(500,1))
}

#Solution 2.b

sd_p = sd(pois_mean) #Standard Deviation of the sample
cifn = function(x, alpha=0.95){ #function to find Confidence Interval
  z = qnorm( (1-alpha)/2, lower.tail=FALSE)
```

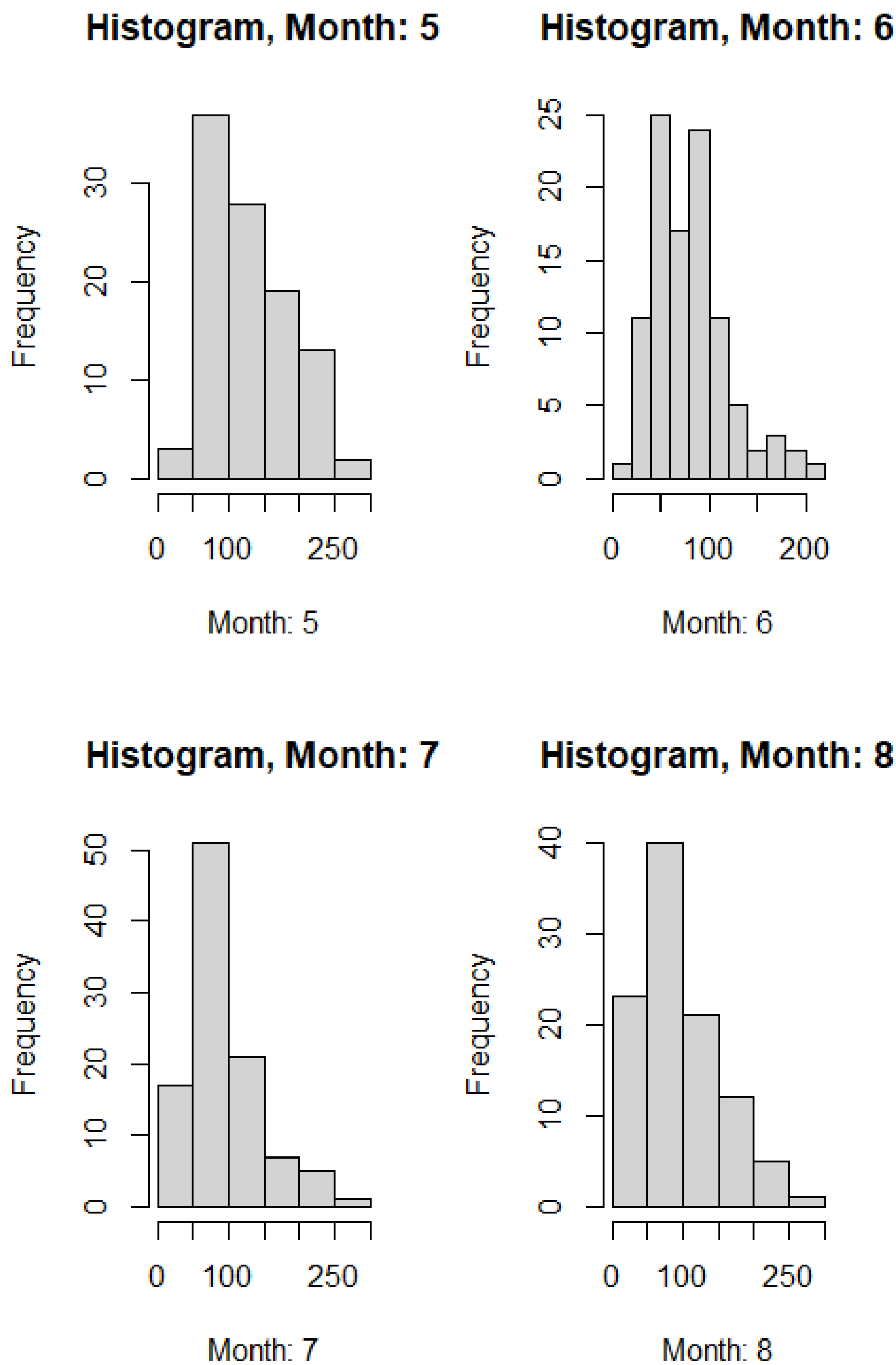
```
  sdx = sd_p*sqrt(1/length(x))
  c(mean(x) - z*sdx, mean(x) + z*sdx)
}

cidata = sapply(pois_mean,cifn) #apply the function to each element
print(cidata)
```

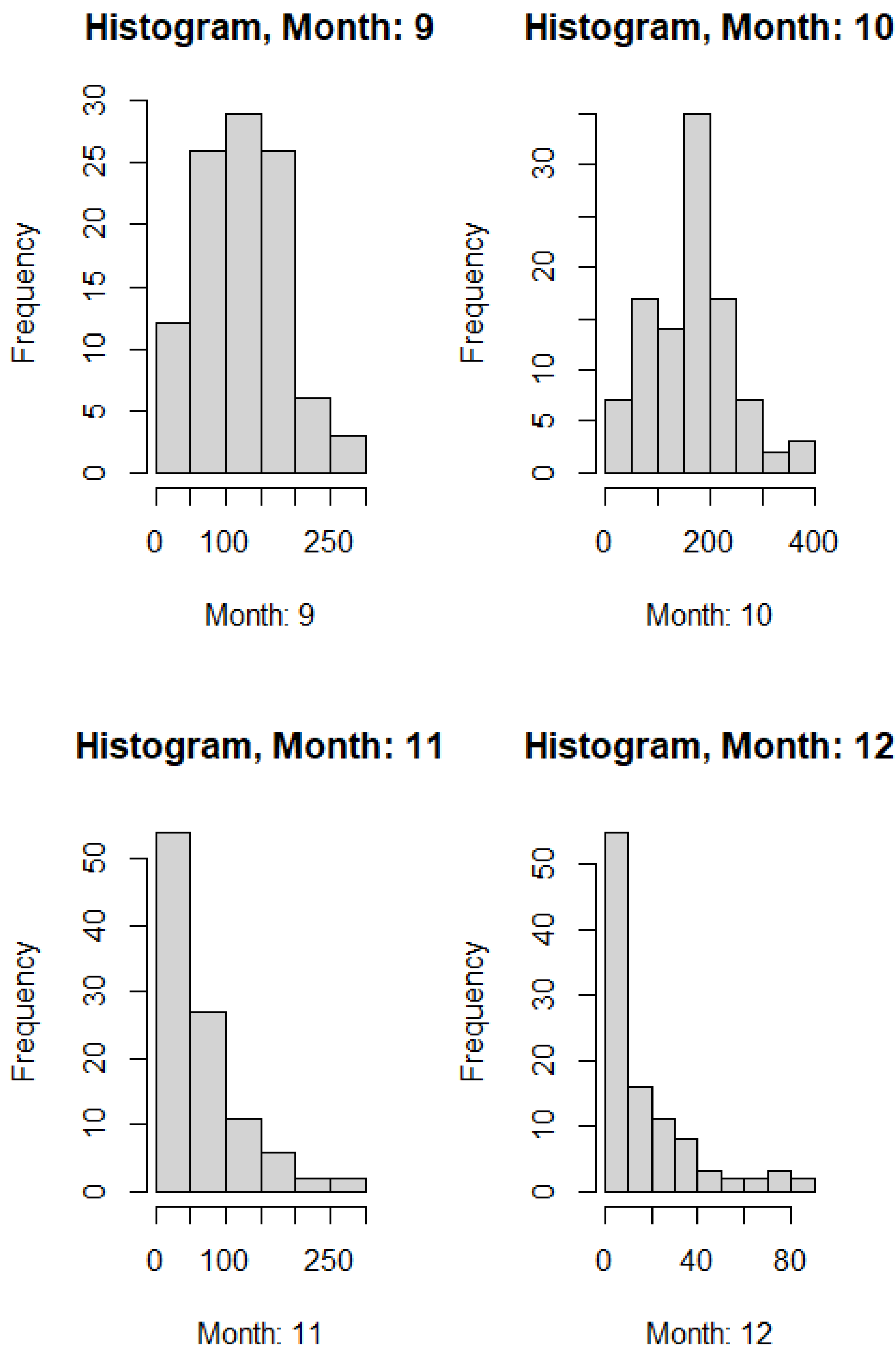
```
##           [,1]     [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
## [1,] 0.9822686 1.036269 0.9162686 0.8542686 0.9422686 0.9702686 0.9342686
## [2,] 1.1337314 1.187731 1.0677314 1.0057314 1.0937314 1.1217314 1.0857314
##           [,8]      [,9]     [,10]     [,11]     [,12]     [,13]     [,14]
## [1,] 0.8902686 0.9062686 0.8842686 0.9622686 0.9062686 0.9442686 0.9282686
## [2,] 1.0417314 1.0577314 1.0357314 1.1137314 1.0577314 1.0957314 1.0797314
##          [,15]     [,16]     [,17]     [,18]     [,19]     [,20]     [,21]
## [1,] 0.9582686 0.9482686 0.9242686 0.9242686 0.9282686 0.8962686 0.8582686
## [2,] 1.1097314 1.0997314 1.0757314 1.0757314 1.0797314 1.0477314 1.0097314
##          [,22]     [,23]     [,24]     [,25]     [,26]     [,27]     [,28]
## [1,] 0.9482686 0.9082686 0.9462686 0.9682686 0.9142686 0.8882686 0.9762686
## [2,] 1.0997314 1.0597314 1.0977314 1.1197314 1.0657314 1.0397314 1.1277314
##          [,29]     [,30]     [,31]     [,32]     [,33]     [,34]     [,35]
## [1,] 0.9462686 0.9982686 0.9162686 0.9282686 0.9582686 0.8462686 0.9042686
## [2,] 1.0977314 1.1497314 1.0677314 1.0797314 1.1097314 0.9977314 1.0557314
##          [,36]     [,37]     [,38]     [,39]     [,40]     [,41]     [,42]
## [1,] 0.9122686 0.9402686 0.9942686 0.9242686 0.9382686 0.9442686 0.8822686
## [2,] 1.0637314 1.0917314 1.1457314 1.0757314 1.0897314 1.0957314 1.0337314
##          [,43]     [,44]     [,45]     [,46]     [,47]     [,48]     [,49]
## [1,] 0.9662686 0.9142686 0.8862686 0.8962686 0.9862686 0.9522686 0.8822686
## [2,] 1.1177314 1.0657314 1.0377314 1.0477314 1.1377314 1.1037314 1.0337314
##          [,50]     [,51]     [,52]     [,53]     [,54]     [,55]     [,56]
## [1,] 0.8642686 0.9322686 0.9522686 0.9842686 0.9282686 0.9262686 0.8862686
## [2,] 1.0157314 1.0837314 1.1037314 1.1357314 1.0797314 1.0777314 1.0377314
##          [,57]     [,58]     [,59]     [,60]     [,61]     [,62]     [,63]
## [1,] 0.9062686 0.9422686 0.8802686 0.9762686 0.9162686 0.9282686 0.9402686
## [2,] 1.0577314 1.0937314 1.0317314 1.1277314 1.0677314 1.0797314 1.0917314
##          [,64]     [,65]     [,66]     [,67]     [,68]     [,69]     [,70]
## [1,] 0.9182686 0.9022686 0.8542686 0.8682686 0.8922686 0.8742686 0.9482686
## [2,] 1.0697314 1.0537314 1.0057314 1.0197314 1.0437314 1.0257314 1.0997314
##          [,71]     [,72]     [,73]     [,74]     [,75]     [,76]     [,77]
## [1,] 0.9542686 0.9322686 0.8802686 0.8782686 0.9362686 0.8342686 0.8842686
## [2,] 1.1057314 1.0837314 1.0317314 1.0297314 1.0877314 0.9857314 1.0357314
##          [,78]     [,79]     [,80]     [,81]     [,82]     [,83]     [,84]
## [1,] 0.9722686 0.9882686 0.9042686 0.9322686 0.8862686 0.8882686 0.9762686
## [2,] 1.1237314 1.1397314 1.0557314 1.0837314 1.0377314 1.0397314 1.1277314
##          [,85]     [,86]     [,87]     [,88]     [,89]     [,90]     [,91]
## [1,] 0.9702686 0.9162686 0.8762686 0.9642686 0.9262686 0.9282686 0.9062686
## [2,] 1.1217314 1.0677314 1.0277314 1.1157314 1.0777314 1.0797314 1.0577314
##          [,92]     [,93]     [,94]     [,95]     [,96]     [,97]     [,98]
## [1,] 0.9562686 0.9322686 0.8722686 0.9502686 0.9302686 0.9362686 0.9122686
## [2,] 1.1077314 1.0837314 1.0237314 1.1017314 1.0817314 1.0877314 1.0637314
##          [,99]   [,100]
```

```
## [1,] 0.9502686 1.004269
## [2,] 1.1017314 1.155731

#Solution 2.c
#The true mean= lambda = 1
TRUEIN = (cidata[1,]-1)*(cidata[2,]-1)<0 #Check whether true mean = 1 lies in
the confidence interval
tab = table(TRUEIN)
print("The number of times true mean lies in the intervals:")

## [1] "The number of times true mean lies in the intervals:"

print(tab[2][1])

## TRUE
##   96

#Solution 3.a

bang_rain =read.csv("D:\\Sucheta\\CMI\\PSWR\\Homework\\BangaloreRain.csv")
par(mfrow=c(1,2))
for (i in 2:13){  #plot the histogram
  hist(bang_rain[[i]],main=paste("Histogram, Month:",i-1), xlab =
paste("Month:",i-1))
}
```



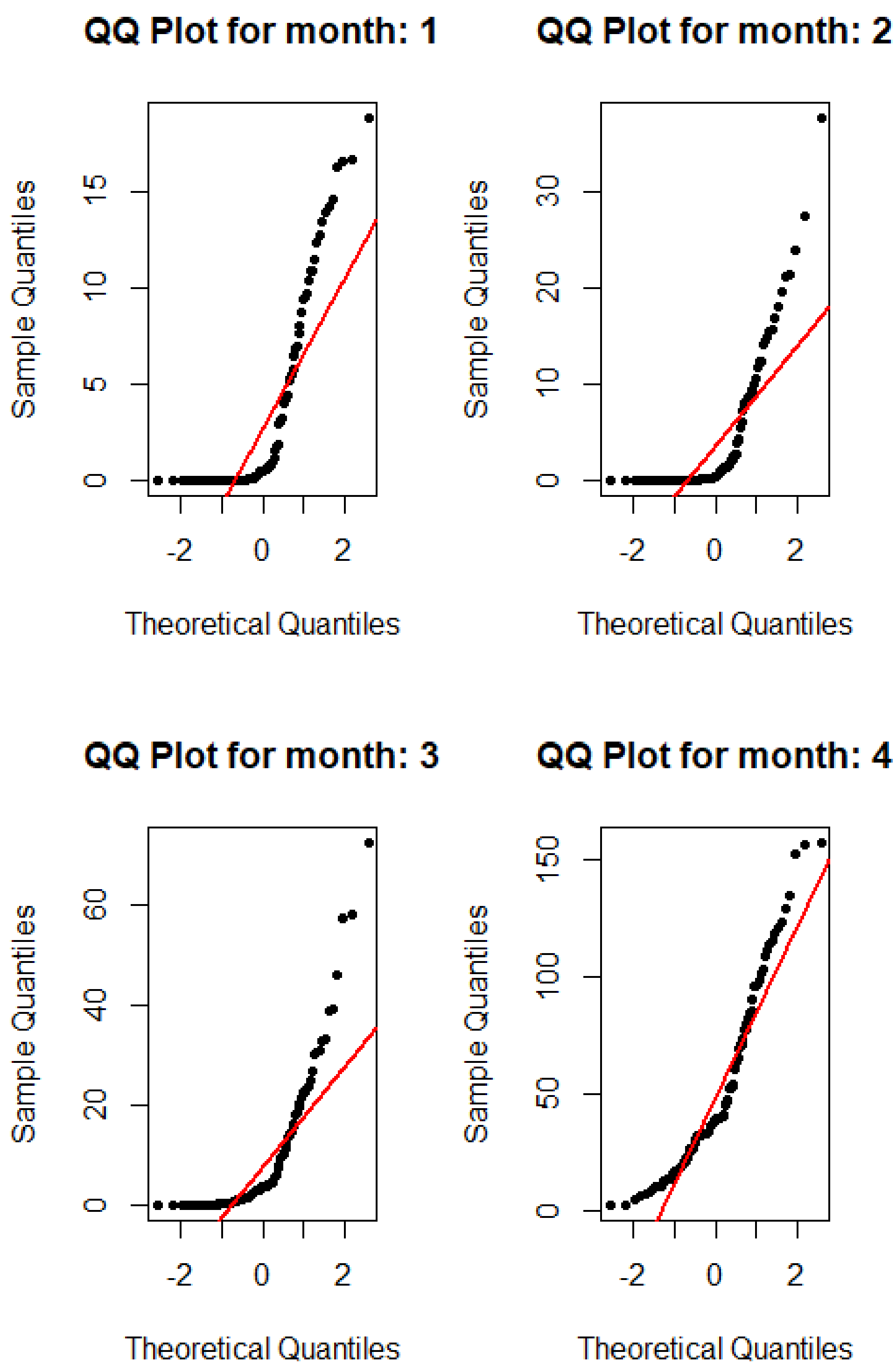
Histogram, Month: 1
Frequency
0
10
30
50
0
5
10
20
Month: 1
Histogram, Month: 2
Frequency
0
20
40
60
0
10
30
Month: 2
Histogram, Month: 3
Frequency
0
10
30
50
70
0
20
60
Month: 3
Histogram, Month: 4
Frequency
0
5
10
20
30
0
50
150
Month: 4

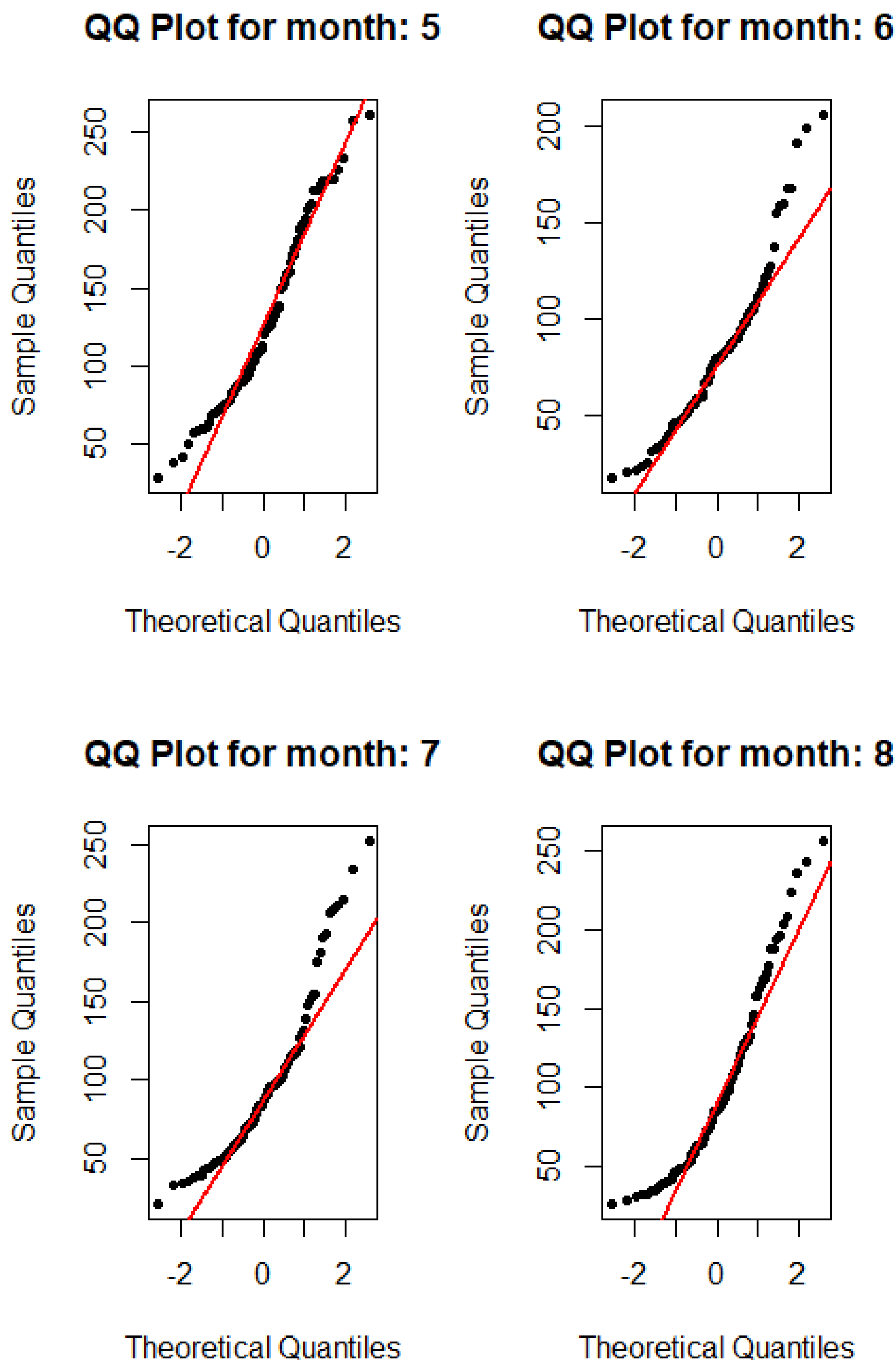

Histogram, Month: 5
Frequency
0
10
20
30
0
100
250
Month: 5
Histogram, Month: 6
Frequency
0
5
10
15
20
25
0
100
200
Month: 6
Histogram, Month: 7
Frequency
0
10
20
30
40
50
0
100
250
Month: 7
Histogram, Month: 8
Frequency
0
10
20
30
40
0
100
250
Month: 8

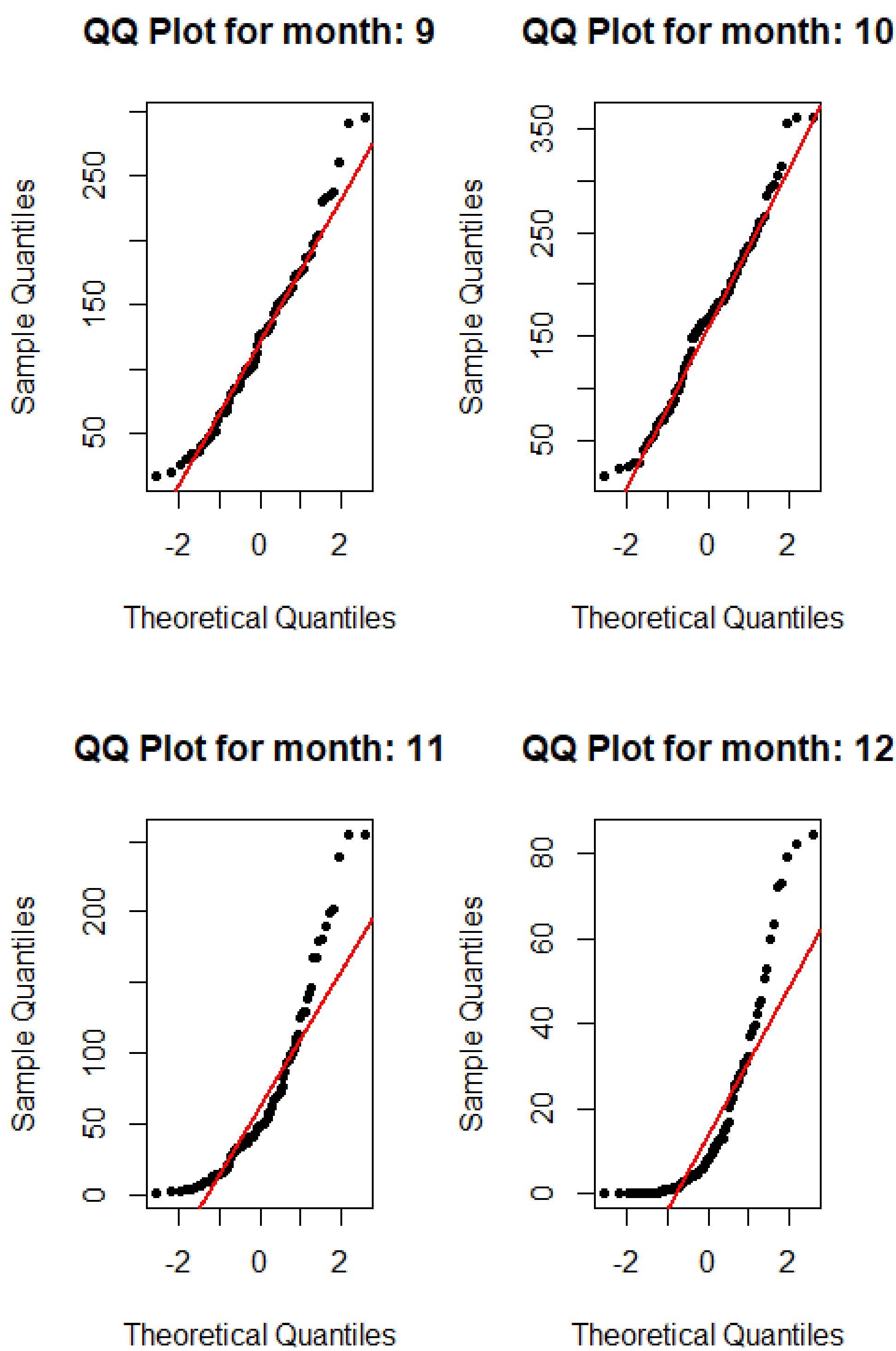



```
par(mfrow=c(1,2))
for (i in 2:13){ #plot QQ plot for each month
  qqnorm(bang_rain[[i]],pch=20, main=paste("QQ Plot for month:",i-1))
```

```
  qqline(bang_rain[[i]],col='red',lwd=2)
}
```

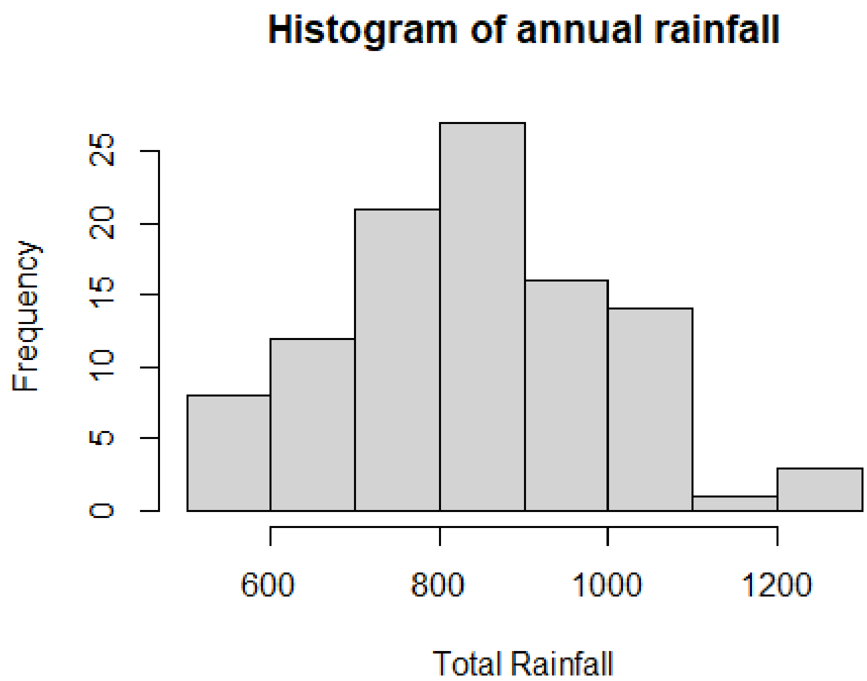



QQ Plot for month: 5
QQ Plot for month: 6
QQ Plot for month: 7
QQ Plot for month: 8
Sample Quantiles
Theoretical Quantiles
50
100
150
200
250
-2
0
2




```
#On looking at the histogram for months Jan to Aug and Nov to Dec,
#the data is right skewed, hence it is not normally distributed.
#For months September and October, The histogram is not skewed. Though the
histogram
```

```
#is less symmetric and some points do not lie on the normal line,
#rainfall for Sept and Oct are more normally distributed than the others.


#Solution 3.b

total = rep(0,length(bang_rain[,1]))
for (i in 1:length(bang_rain[,1])){   #Find the total rainfall for each year
  total[i] = sum(bang_rain[i,][2:13])
}
print("The yearly total rainfall for each year is:")

## [1] "The yearly total rainfall for each year is:"

print(total)

##   [1]  839.196  885.426 1207.198  772.655  707.271  908.493  783.518
528.732
##   [9]  948.122 1004.432  730.049  856.961  578.097  693.868  826.053
1239.555
##  [17]  862.039  683.766  952.050  592.117  814.067  822.288  588.502
683.451
##  [25]  768.307  674.042  580.463  922.159  776.576  974.938  637.935
936.440
##  [33] 1089.707  662.077  900.988  821.416  807.853  658.934  876.928
985.239
##  [41]  793.846  733.979 1067.055  877.299  555.482  983.474  728.105
892.919
##  [49]  848.229  738.061  783.783  728.705 1003.849  838.300  852.992
914.995
##  [57]  715.581 1015.276  862.460  805.080  841.178  987.809  865.976
1089.027
##  [65]  556.088 1182.929  696.461  725.312 1020.702  863.189  825.656
850.930
##  [73]  810.674  788.415 1002.065  629.694 1038.282  749.111 1039.553
719.099
##  [81]  838.170  593.735  831.284  742.650  604.686  906.760  729.308
1041.755
##  [89]  864.628  698.485  996.754  819.437  933.128  602.847  941.275
1081.408
##  [97] 1091.201 1218.563  922.834 1000.767  731.834  748.194

#solution 3.c

par(mfrow=c(1,1))
hist(total,main='Histogram of annual rainfall',xlab='Total Rainfall')
```




```
#The distribution of the yearly total rainfall is not completely symmetric
#so it is still near a Normal Distribution

#Solution 3.d

sd = sd(total)  #Standard deviation of annual rainfall
cifn = function(x, alpha=0.95){    #function to find Confidence Interval
  z = qnorm( (1-alpha)/2, lower.tail=FALSE)
  root = sqrt(1/length(x))
  sdx = sd*root
  c(mean(x) - z*sdx, mean(x) + z*sdx)
}

ci_total = cifn(total)

print(paste("The lower bound of confidence interval is:",ci_total[1]))

## [1] "The lower bound of confidence interval is: 807.988182135121"

print(paste("The lower bound of confidence interval is:",ci_total[2]))

## [1] "The lower bound of confidence interval is: 869.33007276684"
```

4) Let X be the number of heads that come up in a fair coin.
Let Y be the no. of heads that come in a biased coin.

$n = 1000$

$X \sim \text{Binomial}(1000, 0.5)$

$Y \sim \text{Binomial}(1000, 0.55)$

$B|F$ is the event when a fair coin comes but the result is biased coin.

$\therefore$ $B|F$ occurs if the conclusion is that the coin is biased, which happens if we get heads more than 525 times i.e. a fair coin shows at least 525 heads

$$\Rightarrow P(B|F) = P(X \geq 525)$$

$$= P\left(\frac{X - E(X)}{\sqrt{Var(X)}} \geq \frac{525 - E(X)}{\sqrt{Var(X)}}\right)$$

$E(X) = np = 1000 \times 0.5 = 500$

$Var(X) = npq = 1000 \times (0.5) \times 0.5 = 250$

$$\therefore P(B|F) = P\left(\frac{X - 500}{\sqrt{250}} \geq \frac{525 - 500}{\sqrt{250}}\right)$$

$$= P\left(\frac{X - 500}{\sqrt{250}} \geq 1.58114\right)$$

Let $Z = \frac{X - 500}{\sqrt{250}}$, $Z \sim \text{Normal}(0, 1)$

$$\therefore P(B|F) = P(Z \geq 1.58114)$$

$$= 1 - P(Z < 1.58114)$$

$$= 1 - 0.943077 = 0.057$$

F|B is the event that the coin was a biased one but our conclusion was that it is fair. This event will occur if the biased coin shows less than 525 heads.

i.e. $P(F|B) = P(Y < 525)$

$$= P\left(\frac{Y - E(Y)}{\sqrt{Var(Y)}} < \frac{525 - E(Y)}{\sqrt{Var(Y)}}\right)$$

$E(Y) = np_Y = 1000 \times 0.55 = 550$

$Var(Y) = np_Y(1-p_Y) = 1000 \times 0.55 \times 0.45 = 247.5$

$$P(F|B) = P\left(\frac{Y-550}{\sqrt{247.5}} < \frac{525-550}{\sqrt{247.5}}\right)$$

$$= P\left(\frac{Y-550}{\sqrt{247.5}} < -1.58910\right)$$

let $\tilde{Z} = \dfrac{Y-550}{\sqrt{247.5}}$, $\tilde{Z} \sim \text{Normal}(0, 1)$

$\therefore P(F|B) = P(Z < -1.58910)$

$= 0.056$

$P(B|F) \approx 0.057$, $P(F|B) = 0.056$