Ishaan Taneja

Grading:
30 marks- Complete submission of Problem 2,3
70 marks- Problem 3

1. For each of the distributions: Beta(10,2) and Beta(10,10)

   (a) Generate 100 trials of $5, 50, 500$ samples respectively.

   (b) Using the data decide if the conclusion of the Central Limit Theorem applies in each of the three cases, $5, 50, 500$.

   Solution: 1

   ```
   > beta_plot=function(n,a,b){
   +   ssize=100 #Number of trials
   + mean_1=c()
   + for(i in 1:ssize){
   +   beta_1=rbeta(n,a,b)
   +   mean_1[i]=mean(beta_1)
   + }
   + par(mfrow=c(1,2))
   + hist(mean_1,main=paste("Histogram for n=",n),xlab=paste("Means"))
   + qqnorm(mean_1,pch=20)
   + qqline(mean_1,col="red")
   + }
   ```

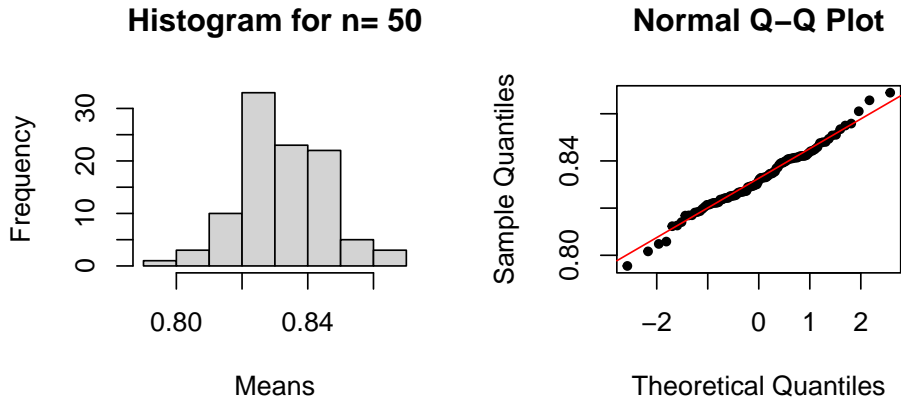   **For Beta(10,2):**

   ```
   > beta_plot(5,10,2)
   ```
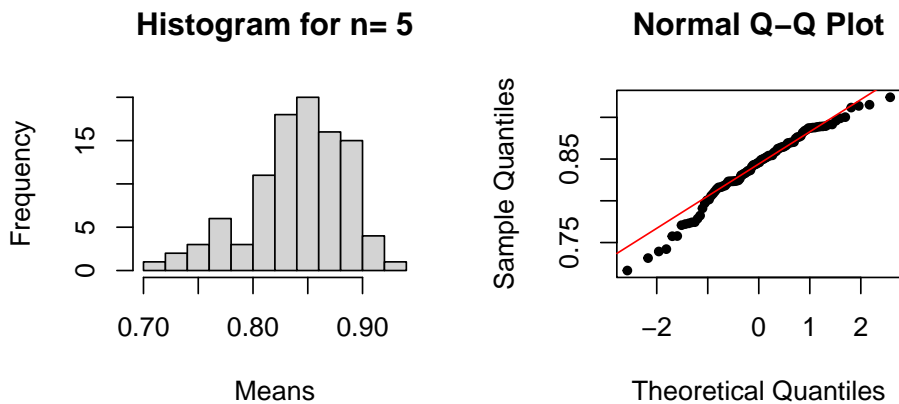

   It can be seen from histogram that data is left skewed. Also, from the QQ-Plot we can
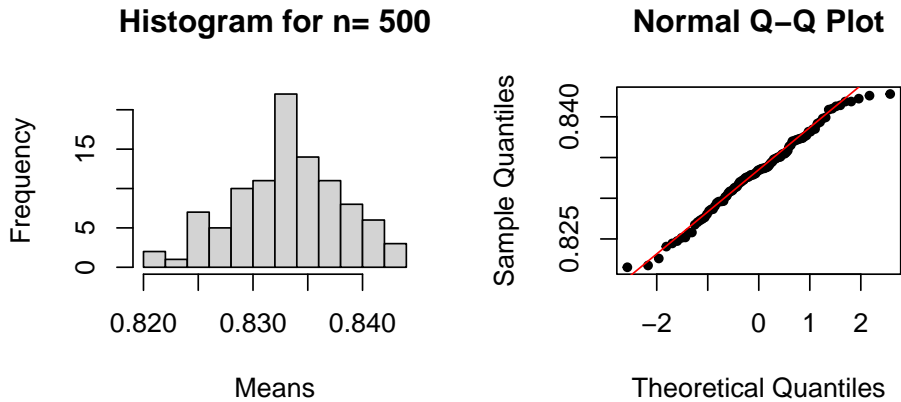
conclude that the data distribution is not normal.

Hence, conclusion of Central Limit Theorem doesn't applies in this case.

```
> beta_plot(50,10,2)
```

Histogram for n= 50

Normal Q–Q Plot

It can be seen from histogram that data tends to symmetric with increase in 'n'. Also, from the QQ-Plot we can conclude that the data distribution is approaching normal.
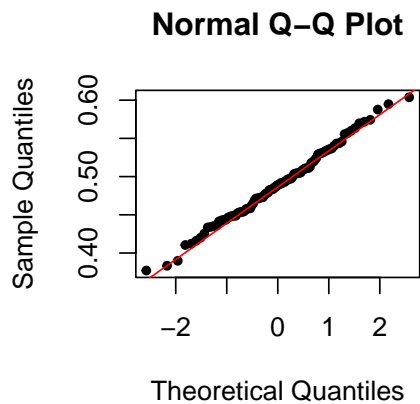
```
> beta_plot(500,10,2)
```

Histogram for n= 500

Normal Q–Q Plot

Again, it can be seen from histogram that data tends to symmetric with increase in 'n'. Also, from the QQ-Plot we can conclude that the data distribution is more normal than previous cases.

For Beta(10,10):

```
> beta_plot(5,10,10)
```

Histogram for n= 5

Normal Q–Q Plot

It can be seen from histogram that data is close to symmetric. Also, from the QQ-Plot we can conclude that the data distribution is close to normal, since most of the points are on the straight line.

```
> beta_plot(50,10,10)
```

Histogram for n= 50

Normal Q–Q Plot

It can be seen from histogram that data tends to symmetric with increase in 'n'. Also, from the QQ-Plot we can conclude that the data distribution is approaching normal.
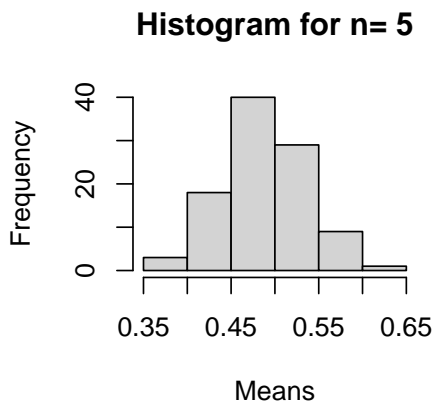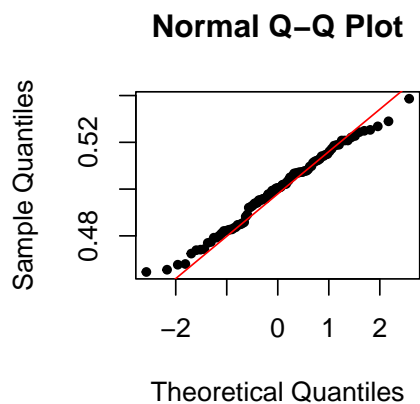
```
> beta_plot(500,10,10)
```

Histogram for n= 500   **Normal Q–Q Plot**
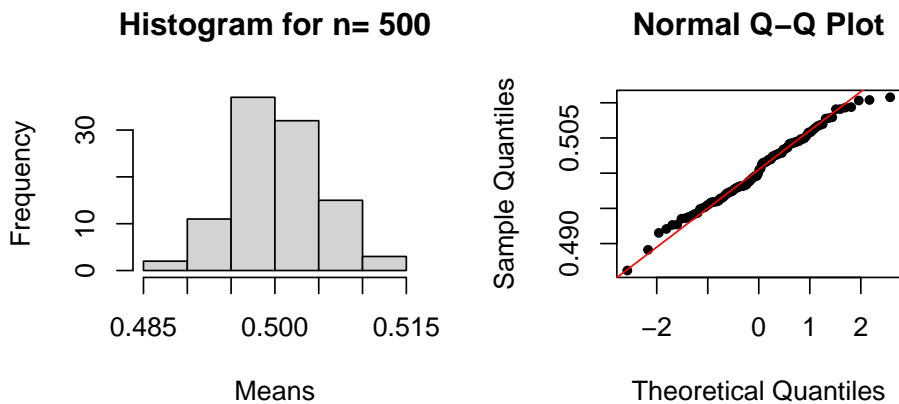
Again, it can be seen from histogram that data tends to symmetric with increase in 'n'. Also, from the QQ-Plot we can conclude that the data distribution is normal.
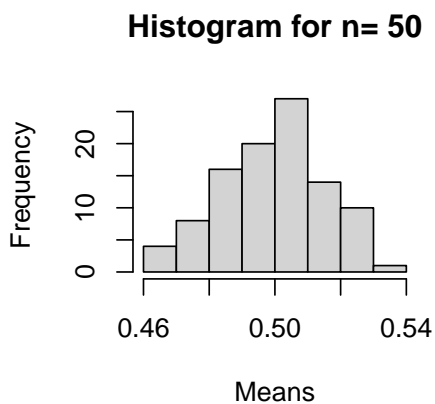
Hence, in case of Beta(5,10), as the sample size increases the histogram become more symmetric and the points of QQ plot stats coinciding with the normal line i.e. data is approaching normal distribution with increase in 'n'.
Whereas, in case of Beta(10,10), even a small sample size (n=5) is giving a good normal approximation because population distribution is symmetric.

2. Consider the Poisson(1) distribution.

   (a) Generate 100 trials of 500 samples respectively.

   (b) Find the 95%-confidence interval for the mean in each trial.

   (c) Compute the number of trials in which the true mean lies in the interval.

Solution: 2

(a)
```
> mean_p=c()
> for(i in 1:100){
+   x=rpois(500,1)  #Generating 100 trials of 500
+ mean_p[i]=mean(x)
+ }
```

(b)
```
> sd_p=sd(mean_p)
> #Computing 95% confidence interval
> confidence_interval=function(x){
+   z=qnorm((1-0.95)/2,lower.tail = FALSE)
+   sd_xbar=sd_p*sqrt(1/length(x))
+   c(mean(x)-z*sd_xbar, mean(x)+z*sd_xbar)
+ }
> c_i=sapply(mean_p,confidence_interval)
> c_i
           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
[1,] 0.9205189 0.8905189 0.9425189 0.9285189 0.8765189 0.9425189 0.9685189
[2,] 1.0954811 1.0654811 1.1174811 1.1034811 1.0514811 1.1174811 1.1434811
```

4

```
              [,8]       [,9]      [,10]      [,11]      [,12]      [,13]      [,14]
[1,] 0.8865189 0.9665189 0.9605189 0.9845189 0.8725189 1.020519 0.9705189
[2,] 1.0614811 1.1414811 1.1354811 1.1594811 1.0474811 1.195481 1.1454811
             [,15]      [,16]      [,17]      [,18]      [,19]      [,20]      [,21]
[1,] 0.8485189 0.9405189 0.8845189 0.9125189 0.9205189 0.9365189 0.8845189
[2,] 1.0234811 1.1154811 1.0594811 1.0874811 1.0954811 1.1114811 1.0594811
             [,22]      [,23]      [,24]      [,25]      [,26]      [,27]      [,28]
[1,] 0.9365189 0.9205189 0.8145189 0.8825189 0.8925189 0.8605189 0.8665189
[2,] 1.1114811 1.0954811 0.9894811 1.0574811 1.0674811 1.0354811 1.0414811
             [,29]      [,30]      [,31]      [,32]      [,33]      [,34]      [,35]
[1,] 0.8965189 0.9545189 1.044519 1.048519 0.9225189 0.8045189 0.9345189
[2,] 1.0714811 1.1294811 1.219481 1.223481 1.0974811 0.9794811 1.1094811
             [,36]      [,37]      [,38]      [,39]      [,40]      [,41]      [,42]
[1,] 0.9485189 0.8465189 0.9605189 0.9345189 0.8765189 0.9145189 0.9585189
[2,] 1.1234811 1.0214811 1.1354811 1.1094811 1.0514811 1.0894811 1.1334811
             [,43]      [,44]      [,45]      [,46]      [,47]      [,48]      [,49]
[1,] 0.9225189 0.9025189 1.014519 0.9045189 0.9045189 0.9405189 0.9065189
[2,] 1.0974811 1.0774811 1.189481 1.0794811 1.0794811 1.1154811 1.0814811
             [,50]      [,51]      [,52]      [,53]      [,54]      [,55]      [,56]
[1,] 0.9305189 0.9105189 0.9765189 0.8365189 0.9105189 0.8365189 0.9325189
[2,] 1.1054811 1.0854811 1.1514811 1.0114811 1.0854811 1.0114811 1.1074811
             [,57]      [,58]      [,59]      [,60]      [,61]      [,62]      [,63]
[1,] 0.9325189 0.9625189 0.8545189 0.9665189 0.9125189 0.9845189 0.9305189
[2,] 1.1074811 1.1374811 1.0294811 1.1414811 1.0874811 1.1594811 1.1054811
             [,64]      [,65]      [,66]      [,67]      [,68]      [,69]      [,70]
[1,] 0.9025189 0.9265189 0.9045189 0.8665189 0.9525189 0.9545189 0.9765189
[2,] 1.0774811 1.1014811 1.0794811 1.0414811 1.1274811 1.1294811 1.1514811
             [,71]      [,72]      [,73]      [,74]      [,75]      [,76]      [,77]
[1,] 0.8565189 0.9025189 0.9045189 0.9205189 0.9545189 0.8805189 0.9285189
[2,] 1.0314811 1.0774811 1.0794811 1.0954811 1.1294811 1.0554811 1.1034811
             [,78]      [,79]      [,80]      [,81]      [,82]      [,83]      [,84]
[1,] 0.9445189 0.9385189 0.9245189 0.8965189 0.9225189 0.9205189 0.9085189
[2,] 1.1194811 1.1134811 1.0994811 1.0714811 1.0974811 1.0954811 1.0834811
             [,85]      [,86]      [,87]      [,88]      [,89]      [,90]      [,91]
[1,] 0.9245189 0.9165189 0.9185189 0.9645189 0.9105189 0.9725189 0.8705189
[2,] 1.0994811 1.0914811 1.0934811 1.1394811 1.0854811 1.1474811 1.0454811
             [,92]      [,93]      [,94]      [,95]      [,96]      [,97]      [,98]
[1,] 0.8725189 0.8945189 0.9545189 0.9145189 1.008519 0.9585189 0.8965189
[2,] 1.0474811 1.0694811 1.1294811 1.0894811 1.183481 1.1334811 1.0714811
             [,99]     [,100]
[1,] 0.9105189 0.9325189
[2,] 1.0854811 1.1074811
```

(c)
```
> true_mean=1 #Lambda=1
> TRUEIN=(c_i[1,]-1)*(c_i[2,]-1)<0
> tab=table(TRUEIN)
> tab

TRUEIN
```

```
FALSE    TRUE
    7      93
> sprintf("The number of times true mean lies in the intervals =%d",tab[2][1])
[1] "The number of times true mean lies in the intervals =93"
```

3. The dataset `BangaloreRain.csv` in the course website at:
   https://www.isibang.ac.in/~athreya/Teaching/PaSwR/BngaloreRain.csv

   (a) Decide if any month's 100 year rainfail is Normally distributed.
   (b) Calculate the yearly total rain fall for each of the 100 years.
   (c) Plot the histogram and Decide if the distribution is Normal.
   (d) Find a 95% confidence interval for the expected annual rainfall in Bangalore.
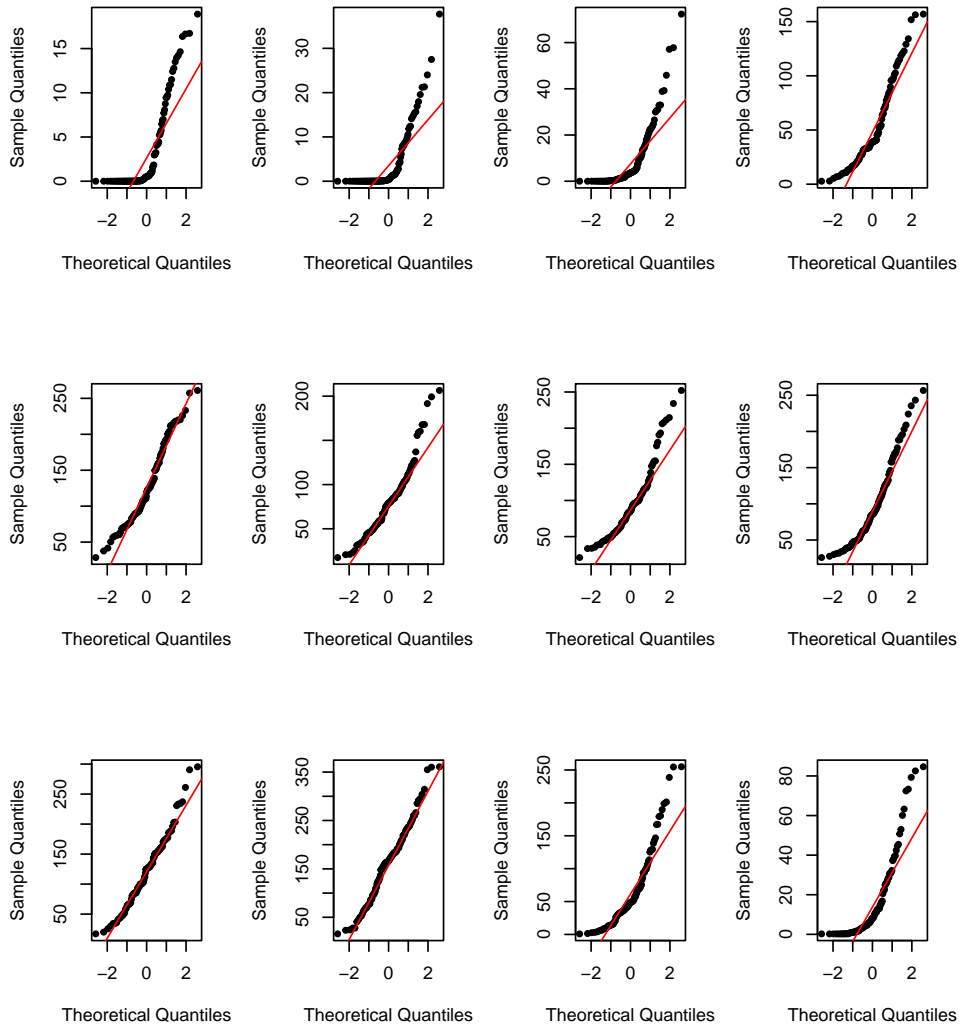
   Solution: 3

   (a)
   ```
   > df=read.csv("C:\\Users\\bt240\\Downloads\\BangaloreRain.csv",sep="\t",
   +    header=TRUE,stringsAsFactors = FALSE,quote = "")
   > par(mfrow=c(3,4))
   > for(i in 2:13){
   +    hist(x=df[[i]],main=colnames(df[i]),xlab = paste("Month:",i-1))
   + }
   ```
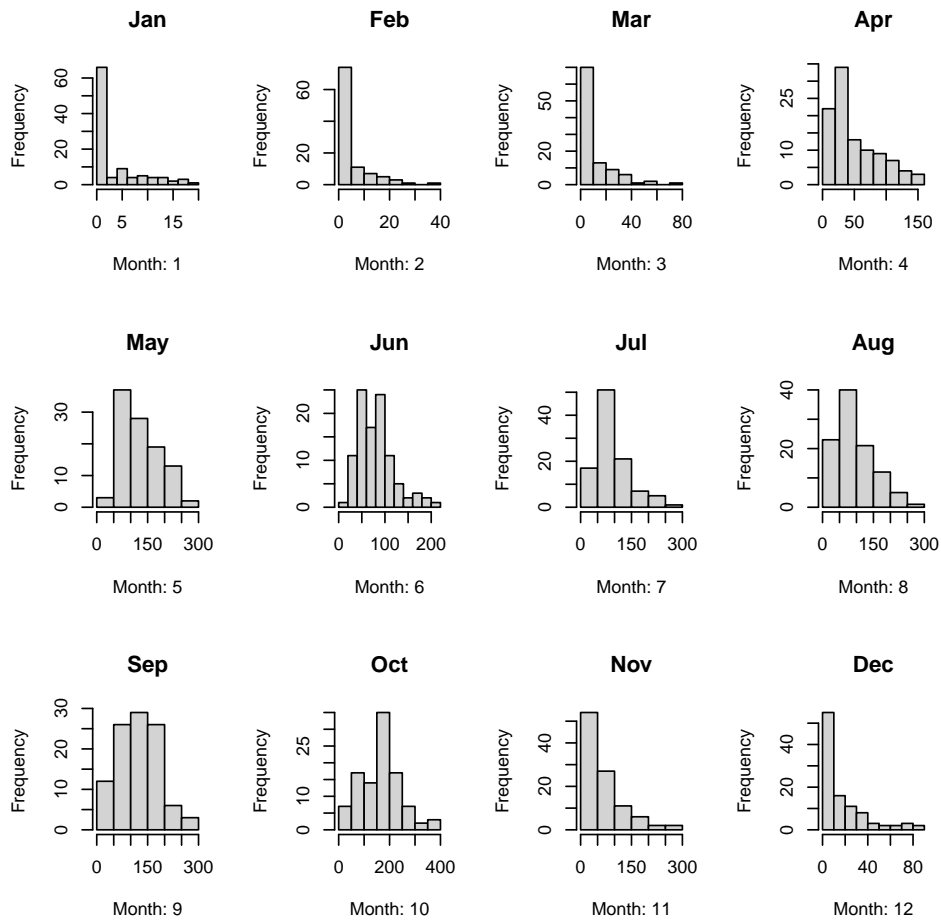
```
> par(mfrow=c(3,4))
> for(i in 2:13){
+   qqnorm(df[[i]],pch=20,main=paste("Month:",i-1))
+   qqline(df[[i]],col='red')
+ }
```


From the histogram and QQ Plots, we can see that the 100 years rainfall in the months of of September and October are fairly normal.

(b)
```
> Yearly_rain=rowSums(df[,2:13])
> Yearly_rain
 [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
```

(c) > hist(Yearly_rain)

Histogram of Yearly_rain
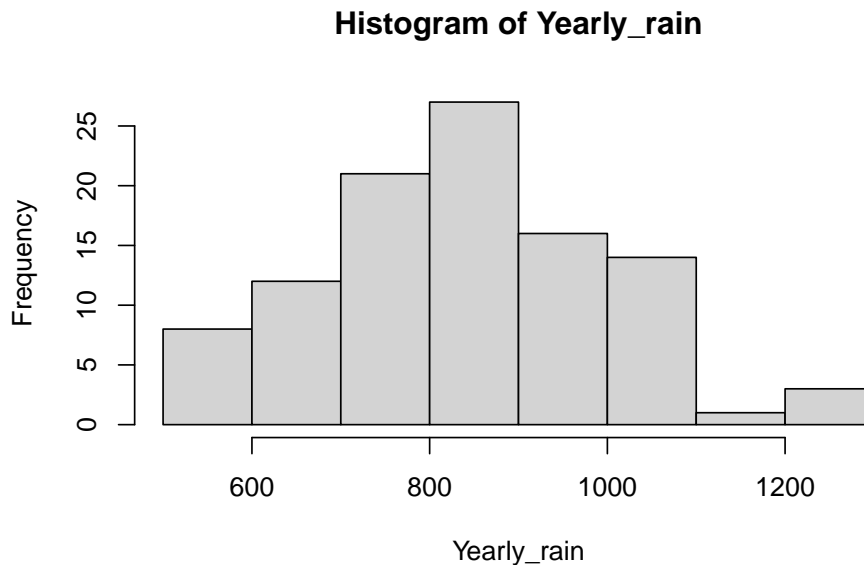
From the above histogram, the distribution of data of annual rainfall is nearly normal.

(d)
```
> z=qnorm((1-0.95)/2,lower.tail = FALSE)
> std_m=sd(Yearly_rain)/sqrt(length(Yearly_rain))
> c_i=c(mean(Yearly_rain)-z*std_m,mean(Yearly_rain)+z*std_m)
> c_i
```

```
[1] 807.9882 869.3301
```

95% confidence interval of the expected annual rainfall = (807.9882 , 869.3301)

4. Two types of coin are produced at a factory: a fair coin and a biased one that comes up heads 55% of the time. We have one of these coins but do not know whether it is a fair or biased coin. In order to ascertain which type of coin we have, we shall perform the following statistical test. We shall toss the coin 1000 times. If the coin comes up heads 525 or more times we shall conclude that it is a biased coin. Otherwise, we shall conclude that it is fair. If the coin is actually fair, what is the probability that we shall reach a false conclusion? What would it be if the coin were biased?

Solution: 4

n=1000

Let X be a random variable representing the number of heads in 1000 tosses of a fair coin.

X ∼ Binomial(1000,0.5)

Let Y be a random variable representing the number of heads in 1000 tosses of a biased coin.

8

$Y \sim$ Binomial(1000,0.55)

Let us define the events:

B: The biased coin is tossed.

F: The fair coin is tossed.

E: Number of heads are 525 or more in 1000 tosses of a coin.

$$P(E|F) = P(X > 525)$$
$$= P\left(\frac{X - E(X)}{S.D(X)} \geq \frac{525 - E(X)}{S.D(X)}\right)$$
$$= P\left(\frac{X - (1000 \times 0.5)}{\sqrt{1000 \times 0.5 \times (1 - 0.5)}} \geq \frac{525 - (1000 \times 0.5)}{\sqrt{1000 \times 0.5 \times (1 - 0.5)}}\right)$$
$$= P\left(Z_1 \geq \frac{525 - 500}{\sqrt{250}}\right) \qquad (\text{where, } Z_1 \sim \text{Normal(0,1)})$$
$$= P(Z_1 \geq 1.581139)$$
$$= 1 - P(Z_1 < 1.581139)$$
$$\therefore P(E|F) = 0.0569$$

Hence, the probability that we reach a false conclusion with a fair coin is 0.0569.

$$P(E^c|B) = P(Y < 525)$$
$$= P\left(\frac{Y - E(Y)}{S.D(Y)} < \frac{525 - E(Y)}{S.D(Y)}\right)$$
$$= P\left(\frac{Y - (1000 \times 0.55)}{\sqrt{1000 \times 0.55 \times (1 - 0.55)}} < \frac{525 - (1000 \times 0.55)}{\sqrt{1000 \times 0.55 \times (1 - 0.55)}}\right)$$
$$= P\left(Z_2 < \frac{525 - 550}{\sqrt{247.5}}\right) \qquad (\text{where, } Z_2 \sim \text{Normal(0,1)})$$
$$= P(Z_2 < -1.589104)$$
$$\therefore P(E^c|B) = 0.0560$$

Hence, the probability that we reach a false conclusion with a biased coin is 0.0560.

5. The length of time (in appropriate units) that a certain type of component functions before failing is a random variable with probability density function

$$f(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Once the component fails it is immediately replaced with another one of the same type. Using the central limit theorem approximation, can you find, how many components would one need to have on hand to be approximately 90% certain that the stock would last at least 35 units of time ?

Solution: 5

$$E(X) = \int_0^1 x \times 2x \, dx$$

$$= \left. \frac{2x^3}{3} \right|_0^1$$

$$= \frac{2}{3}(1^3 - 0^3)$$

$$\therefore E(X) = \frac{2}{3}$$

And,

$$E(X^2) = \int_0^1 x^2 \times 2x \, dx$$

$$= \left. \frac{2x^4}{4} \right|_0^1$$

$$= \frac{1}{2}(1^4 - 0^4)$$

$$\therefore E(X^2) = \frac{1}{2}$$

Hence,

$$V(X) = E(X^2) - (E(X))^2$$

$$= \frac{1}{2} - \left(\frac{2}{3}\right)^2$$

$$= \frac{1}{2} - \frac{4}{9}$$

$$\therefore V(X) = \frac{1}{18}$$

Let, Y be the sum of independent X_i's i.e. $Y = \sum_{i=1}^{n} X_i$

Then, $E(Y) = nE(X) = \dfrac{2n}{3}$

and, $V(Y) = nV(X) = \dfrac{n}{18}$

i.e. $\sum_{i=1}^{n} X_i = Y \sim \text{Normal}(2n/3, \, n/18)$

Therefore,

$$P(Y \geq 35) = P\left(\frac{Y - E(Y)}{S.D(Y)} \geq \frac{35 - E(Y)}{S.D(Y)}\right)$$

$$= P\left(Z \geq \frac{35 - 2n/3}{\sqrt{n/18}}\right) = 0.90 \qquad \text{(given)}$$

where, Z $\sim$ Normal(0,1)

Also, we know that, $P(Z \geq -1.28) = 0.90$

Hence,

$$\frac{35 - 2n/3}{\sqrt{n/18}} = -1.28$$

$$35 - \frac{2n}{3} = -1.28 \times \sqrt{\frac{n}{18}}$$

10

On solving, we get $n \approx 56$

Hence, One needs to have hand on approximately 56 components to be approximately 90% certain that the stock would last at least 35 units of time.