Instructor: Womble

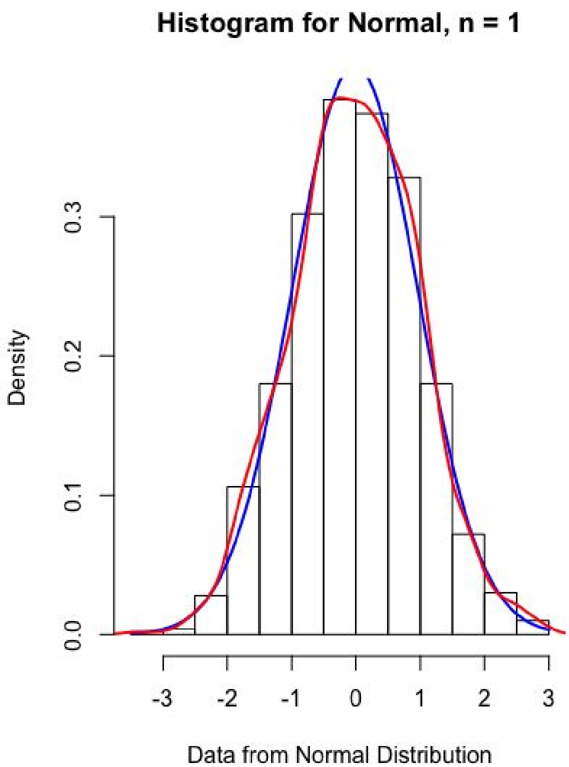
Team members: Johnny Saltanovitz, Zeqing Cai, Tian Qiu, Tyler Crumrin.

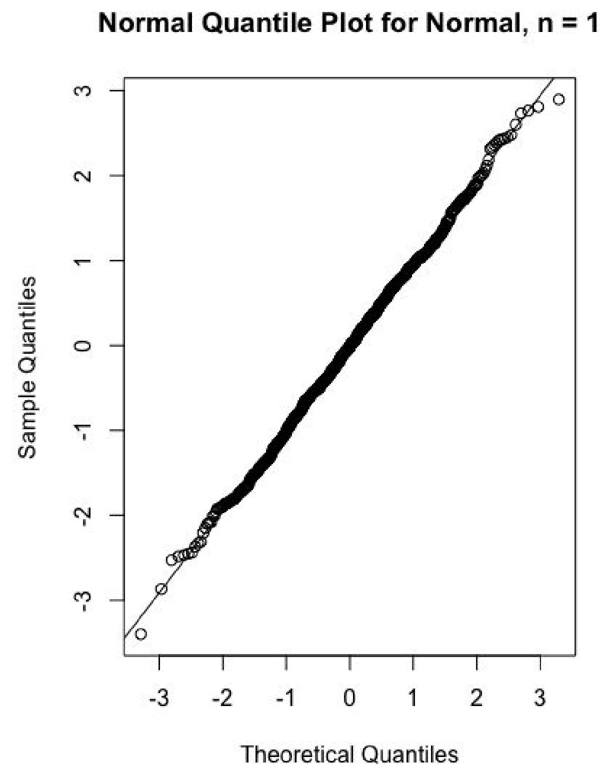
Part A)

1. Code

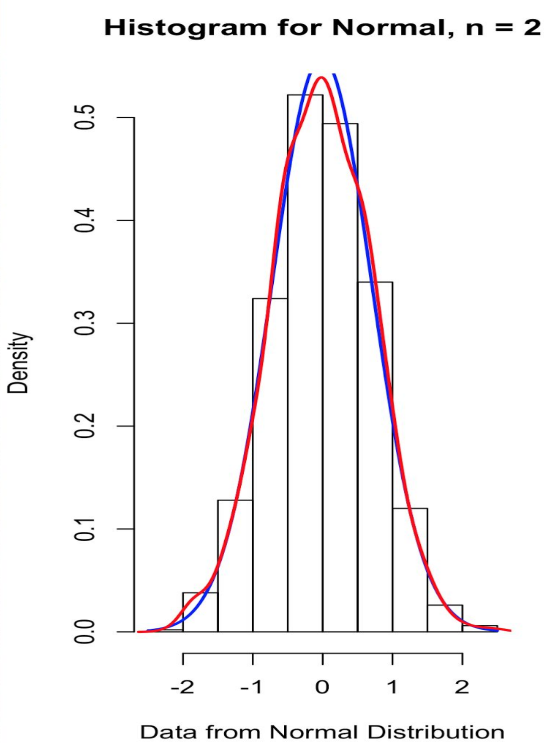
```
# A
SRS <- 1000
set <- c(1, 2, 6, 10)
for(i in set) {
	attach(mtcars)
	par(mfrow=c(1,2))
	n <- i
	data.vec <- rnorm(SRS*n,mean=0,sd=1)
	data.mat <- matrix(data.vec, ncol = n)
	avg <- apply(data.mat, 1, mean)

	m = mean(avg)
	std = sd(avg)
	# print(n, m, std)
	str = sprintf("n = %d, mean = %f, std = %f, sample std = %f", n,m, std, std*sqrt(n))
	print(str)

	# windows()#this is optional
	hist(avg, xlab="Data from Normal Distribution", freq = FALSE, main=sprintf("Histogram for Normal, n
= %d",n))
	curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
	lines(density(avg, adjust=1),col = "red", lwd=2)
	# windows()
	# quartz()
	qqnorm(avg,main=sprintf("Normal Quantile Plot for Normal, n = %d",n))
	qqline(avg)
	quartz()
}
```

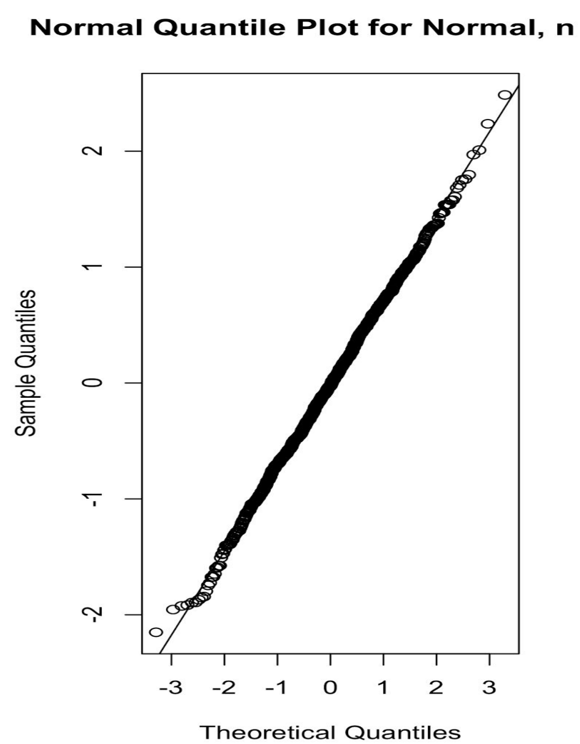
2. Graph



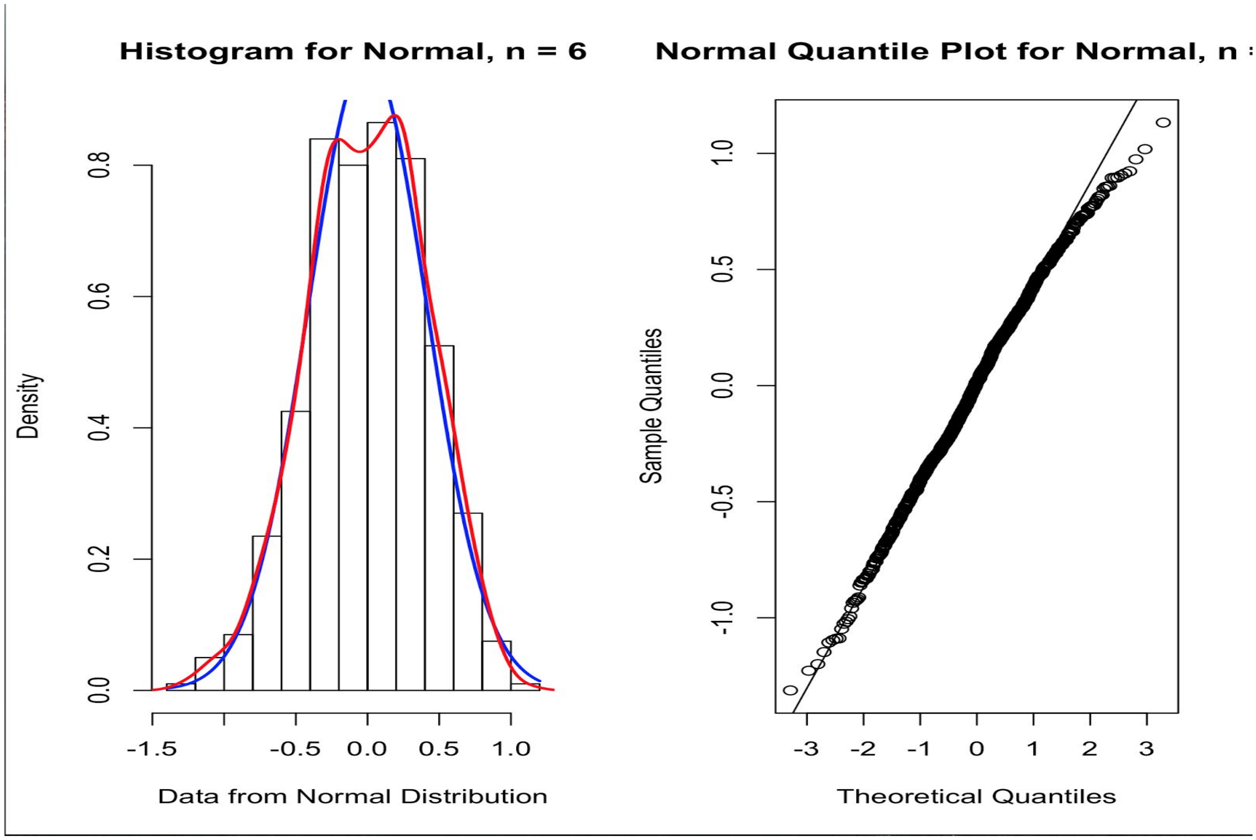


Normal

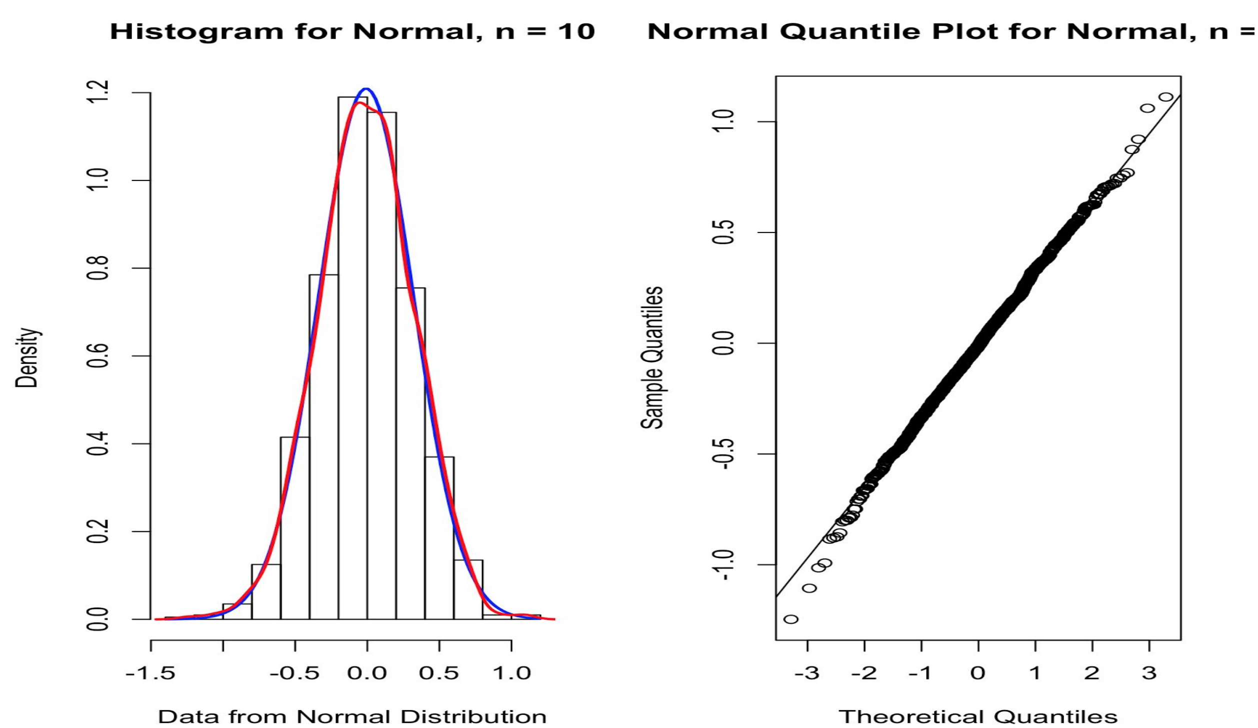




Normal



Normal



Normal

3.

n	experimental mean of your 1000 $\bar{x}$	theoretical mean (Equation 1)	experimental standard deviation of your 1000 $\bar{x}$	theoretical standard deviation (Equations 1)
1	0.010303	0	1.021060	1.021060
2	-0.010864	0	1.010475	0.714514
6	0.000369	0	1.012602	0.413393
10	-0.008641	0	1.042997	0.329825

4.

When n is larger, the plot is more normal and the mean, standard deviation become more closer to theoretical value.

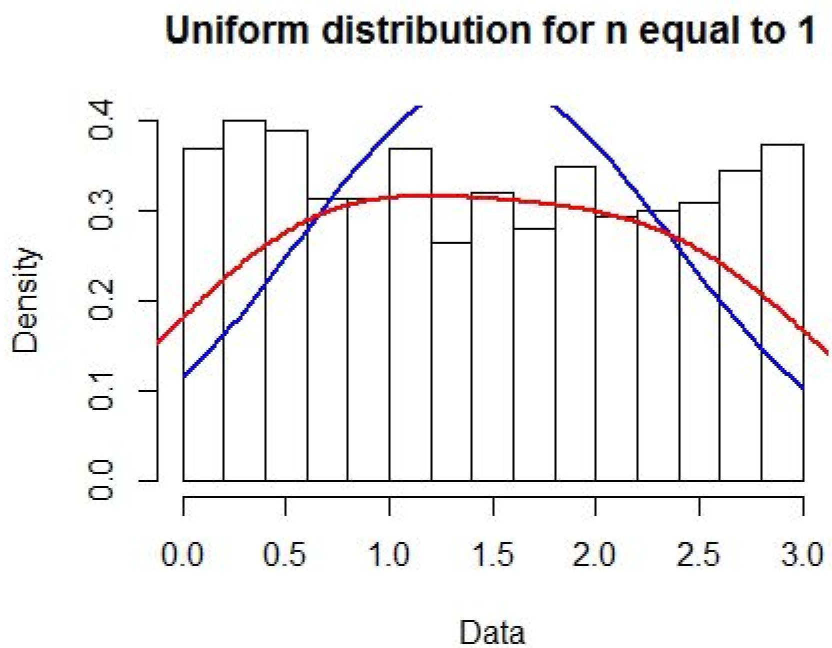
Part B)

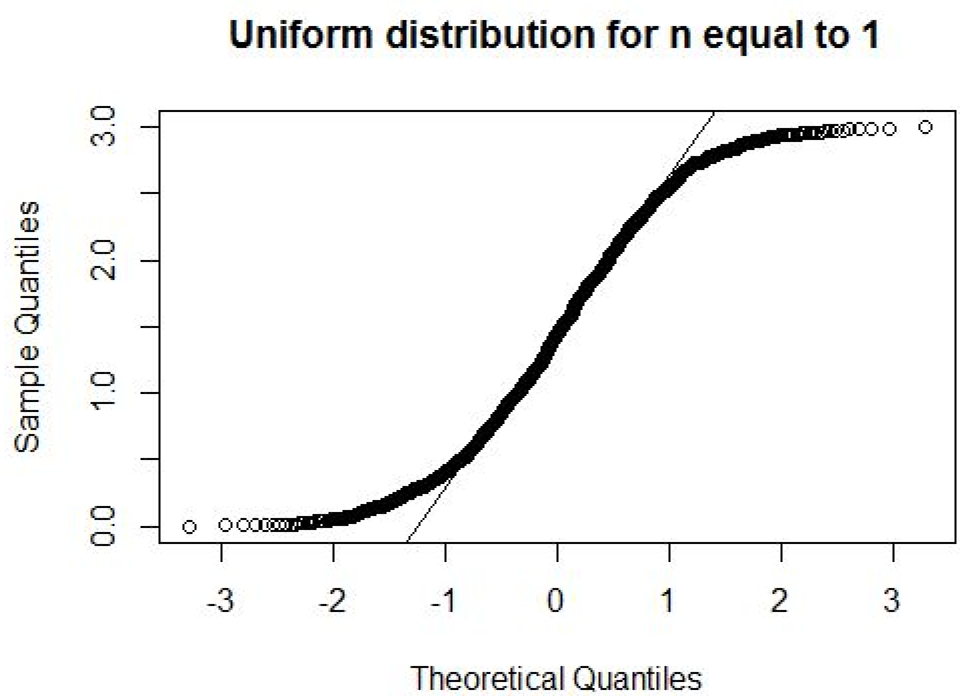
1. Code

This code is an example of what each run would look like. The only thing that would change is the n value. In this code the n value is 16 instead of 1,2 or 9.

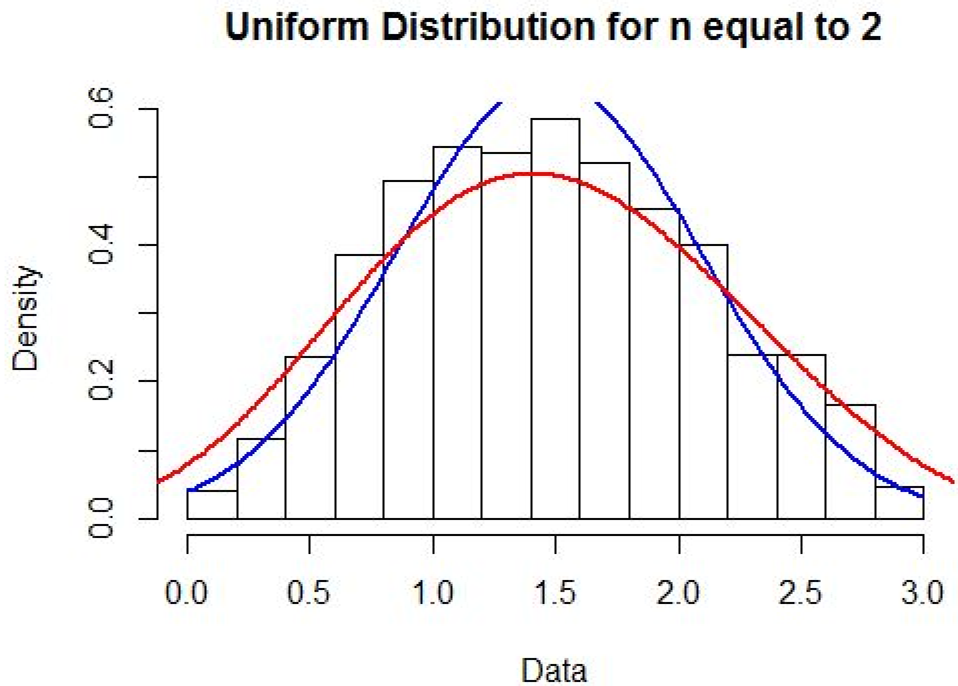
```
> SRS<-1000
> n<-16
> short<-runif(n*SRS,min=0,max=3)
> data.vec<-short
> data.mat<-matrix(data.vec,ncol=n)
> avg<-apply(data.mat,1,mean)
> m16=mean(avg)
> m16
[1] 1.499333
> s19=sd(avg)
> s16=sd(avg)
> s16
[1] 0.2191101
> title<-"Uniform Distribution for n equal to 16"
> windows()
> hist(avg,xlab="Data",freq=FALSE,main=title)
> curve(dnorm(x,mean=m16,sd=s16),col="blue",lwd=2,add=TRUE)
> lines(density(avg,adjust=3),col="red",lwd=2)
> qqnorm(avg,main=title)
> qqline(avg)
```

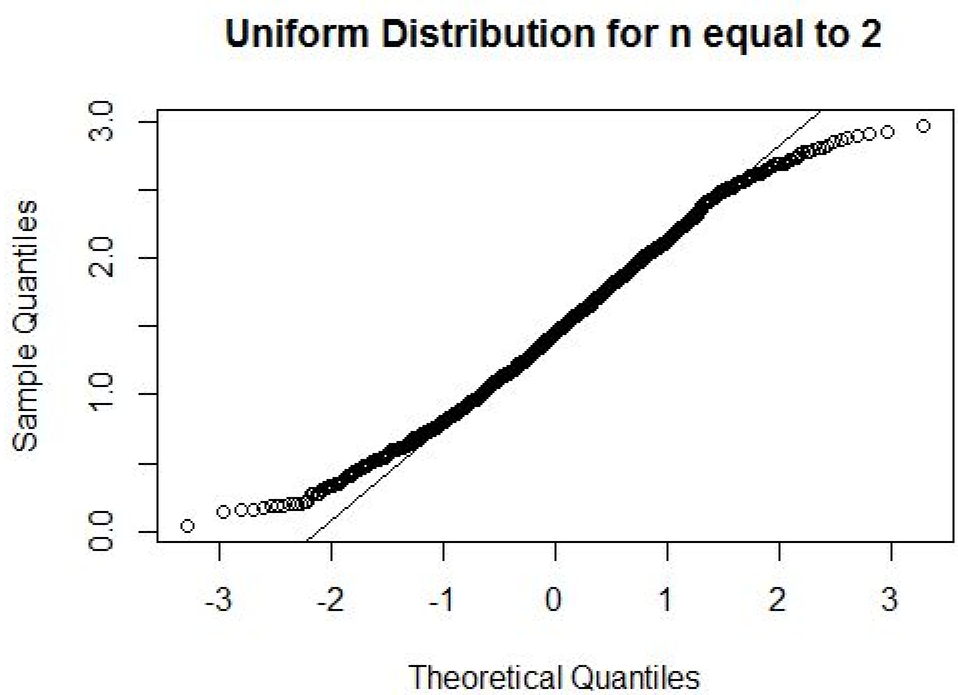
2. Graphs



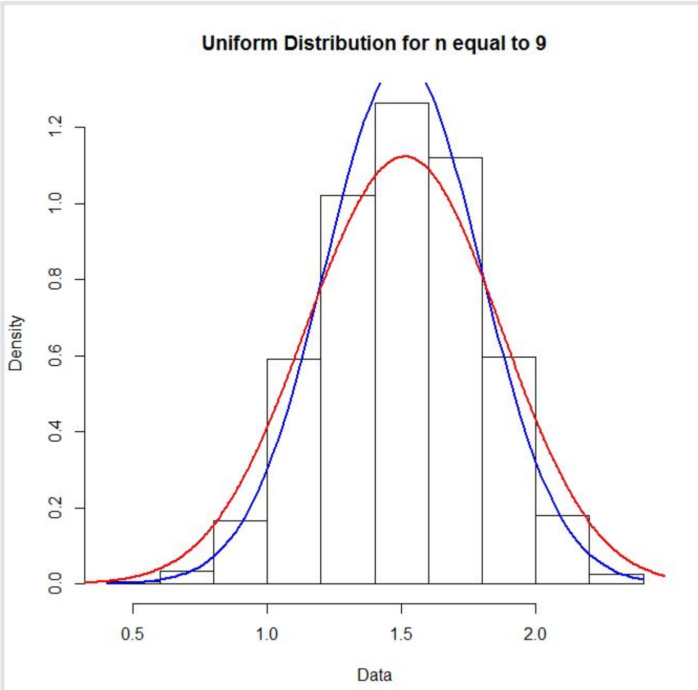


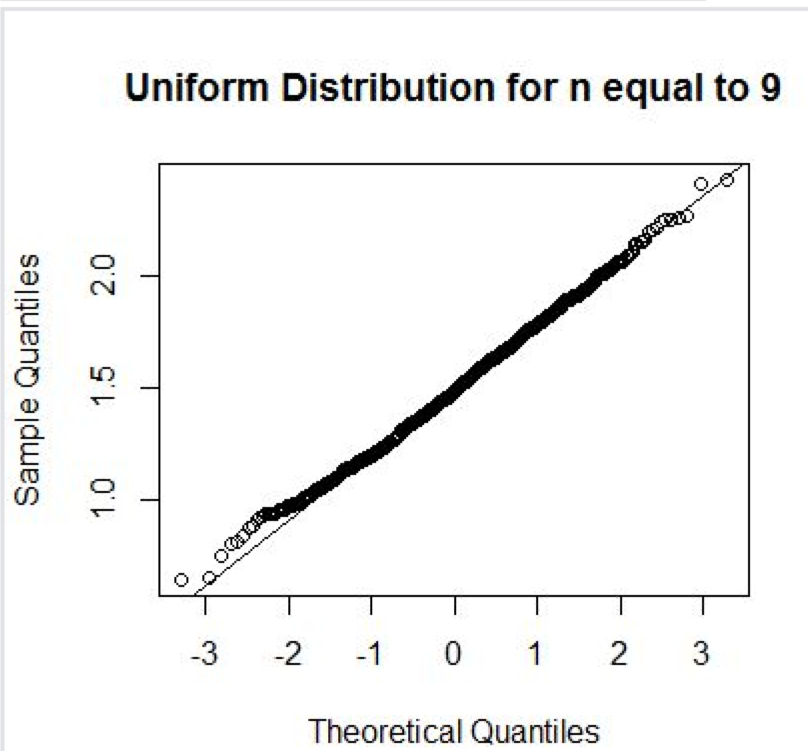
The situation is not normal. The red and blue lines are too far away. Additionally, the curve is not in a straight line for the quantile plot. The data is not in a straight line.



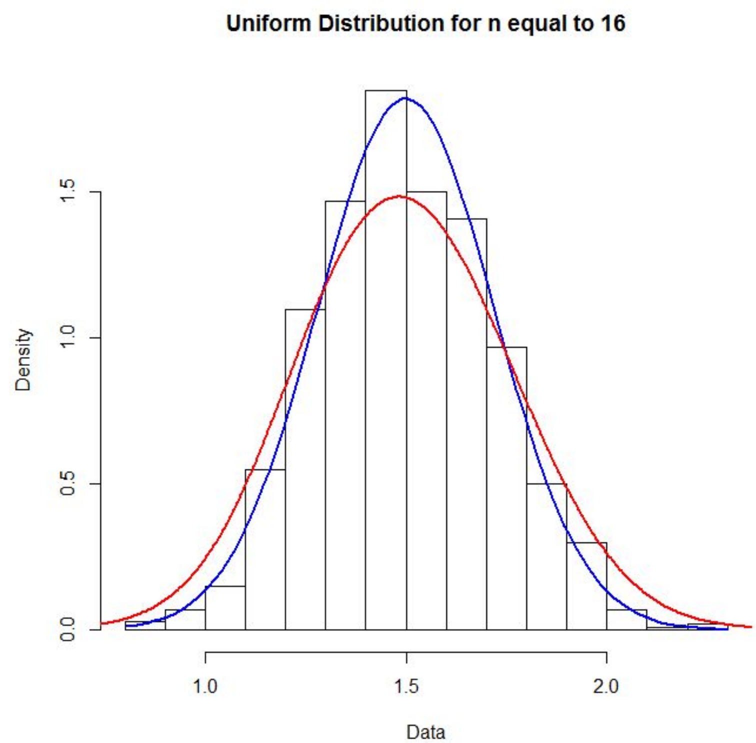


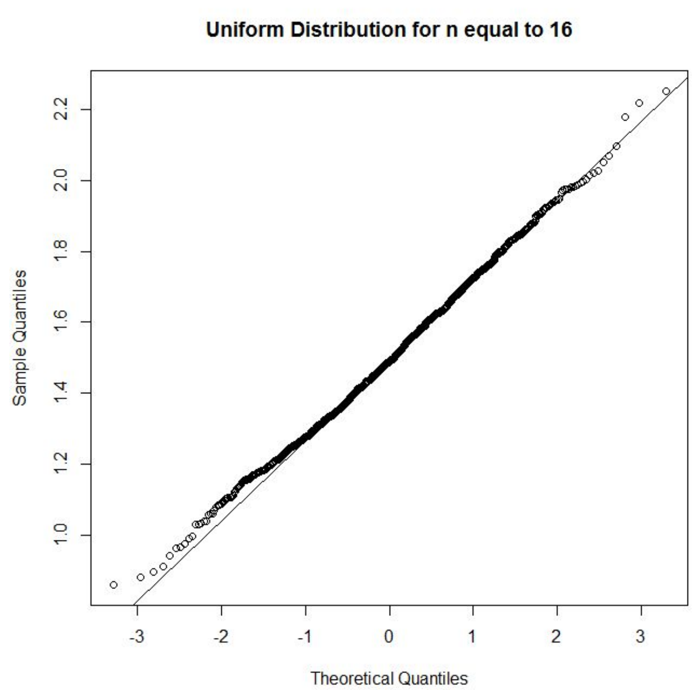
These graphs are still not normal but they are close. The data still does not fall on a straight line and the blue and red lines for the histogram do not match up.





I believe that this data is normal. The blue and red lines closely match up and all of the data on the quantile plot is in a straight line. This is due to the n being higher.





I believe that this data is normal. The blue and red lines closely match up and all of the data on the quantile plot is in a straight line. This is the most normal plot as the n is the highest.

3.

n	experimental mean of your 1000 x	theoretical mean (Equation 1)	experimental standard deviation of your 1000 x	theoretical standard deviation (Equations 1)
1	1.46925	1.46925	0.89597999	0.89597999
2	1.470841	1.470841	0.6221789	0.439946919
9	1.504702	1.504702	0.2900263	0.096675433
16	1.499333	1.499333	0.2191101	0.054777525

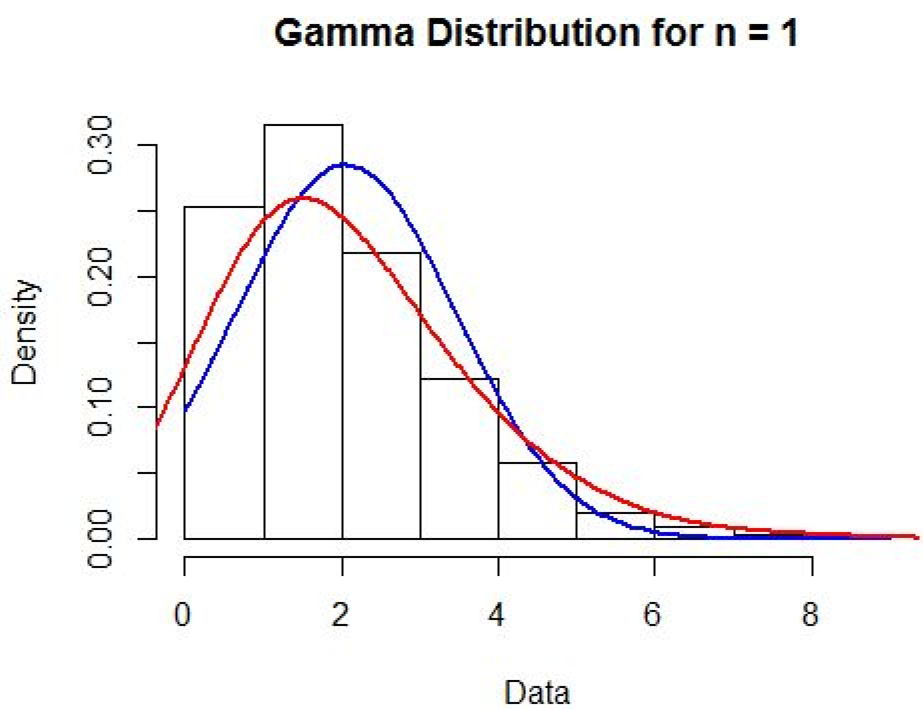
4.
As the n value increase the plot becomes more normal. The plot for this section was normal when n was equal to 9. The equations (1) are applicable to all n values. The graphs will simply become more accurate as the n value increases theoretically. As can be seen by the graphs and the chart, as the n increased the mean got closer to 1.5, and the standard deviation got smaller. These are characteristics of a normal plot.

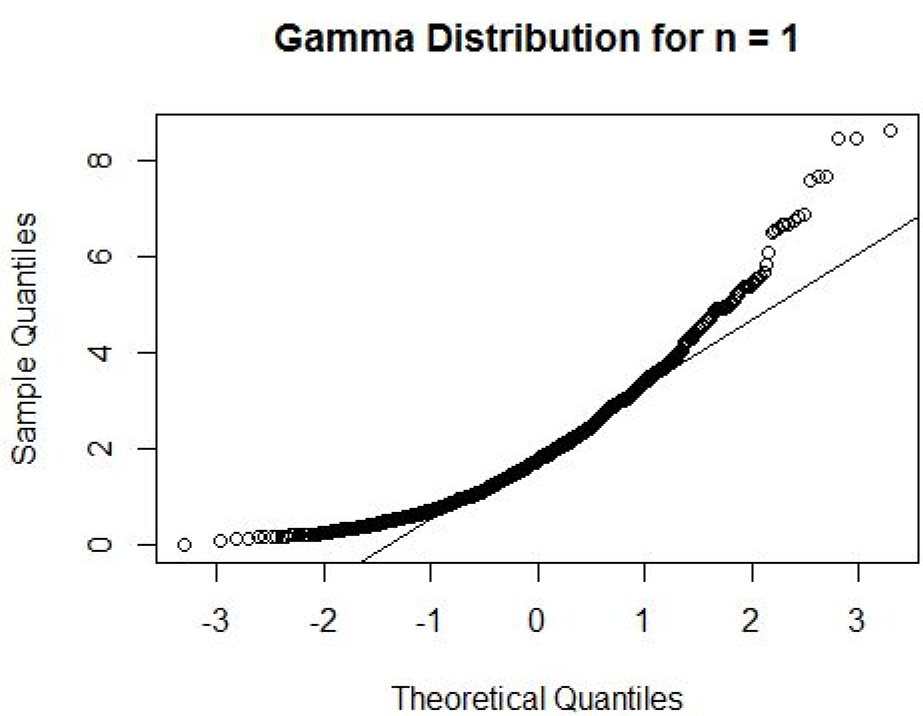
Part C)

1. Code: This is the code used to generate the gamma distribution histograms. The code shown generates the distribution with an n value of 1.

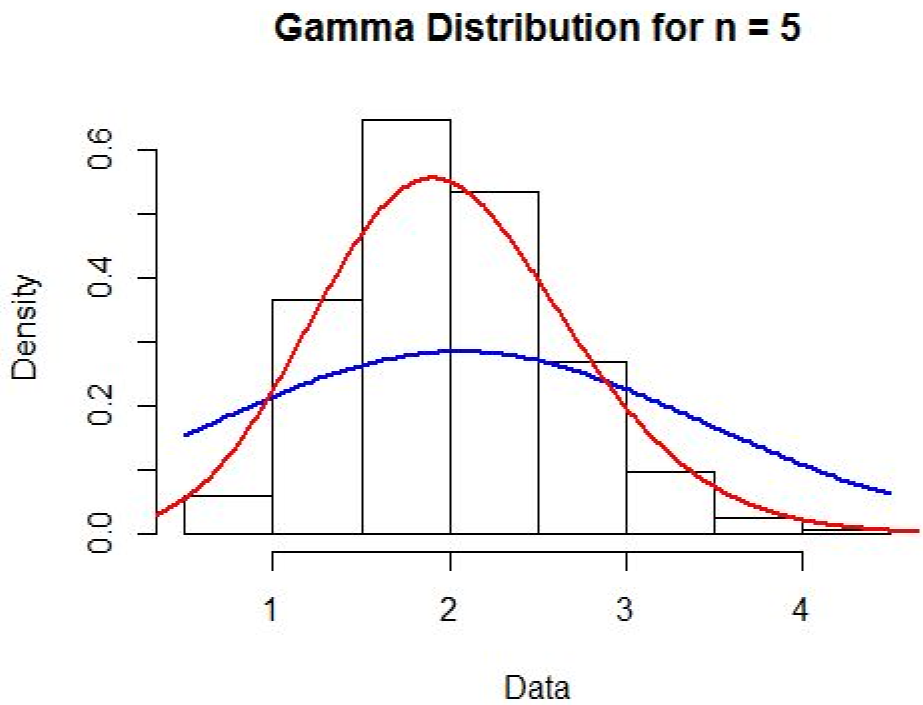
```
> SRS <- 1000
> n <- 1
>
> data.vec <- rgamma(SRS*n,2,rate=1)
> data.mat <- matrix(data.vec, ncol = n)
> avg <- apply(data.mat, 1, mean)
>
> m1 <- mean(avg)
> s1 <- sd(avg)
> title<-"Gamma Distribution for n = 1"
>
> hist(avg,xlab="Data",freq=FALSE,main=title)
> curve(dnorm(x,mean=m1,sd=s1),col="blue",lwd=2,add=TRUE)
> lines(density(avg,adjust=3),col="red",lwd=2)
>
> qqnorm(avg,main=title)
> qqline(avg)
>
> m1
[1] 2.053329
> s1
[1] 1.40293
```

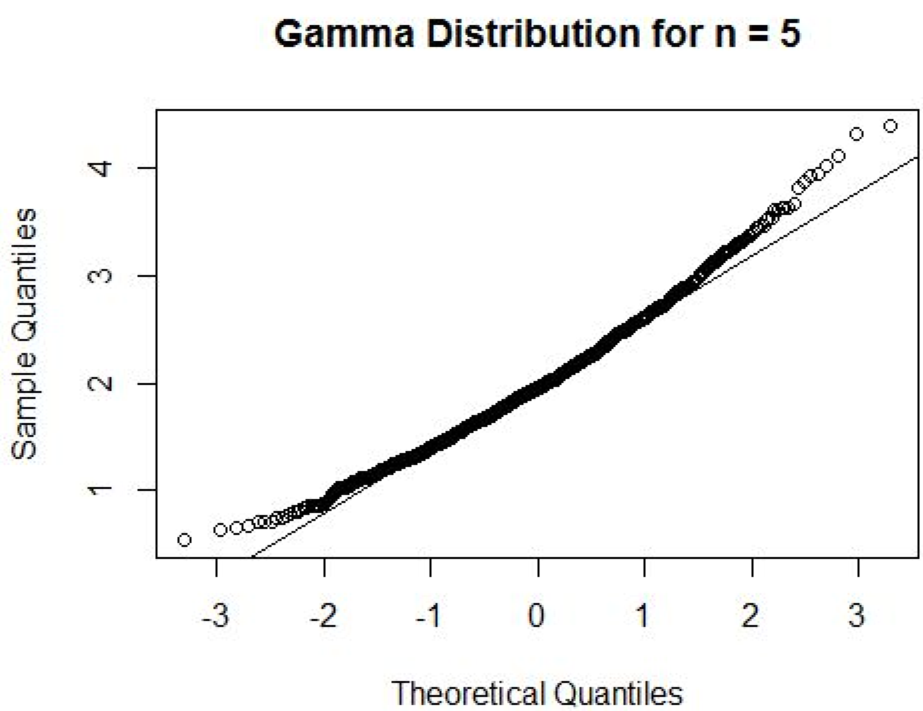
2) Graphs



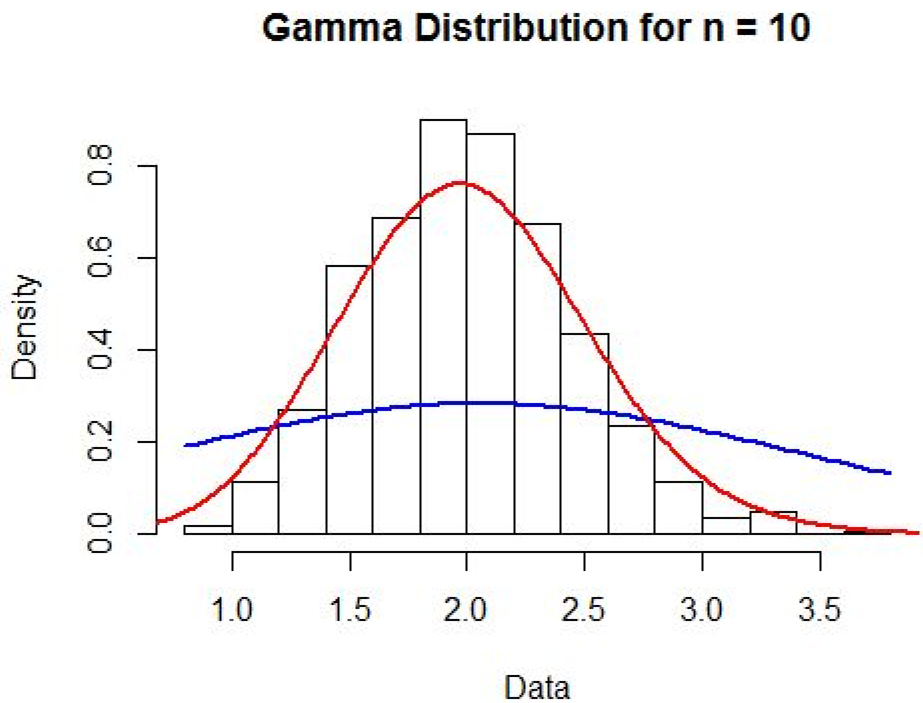


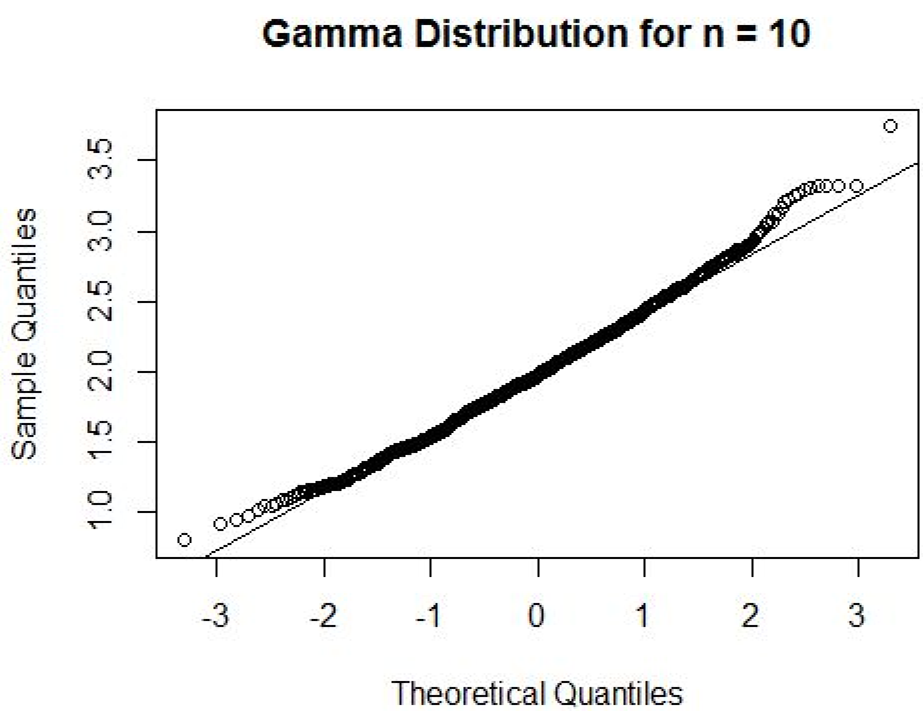
(n=1) This graph is NOT normal, and is in fact heavily skewed.



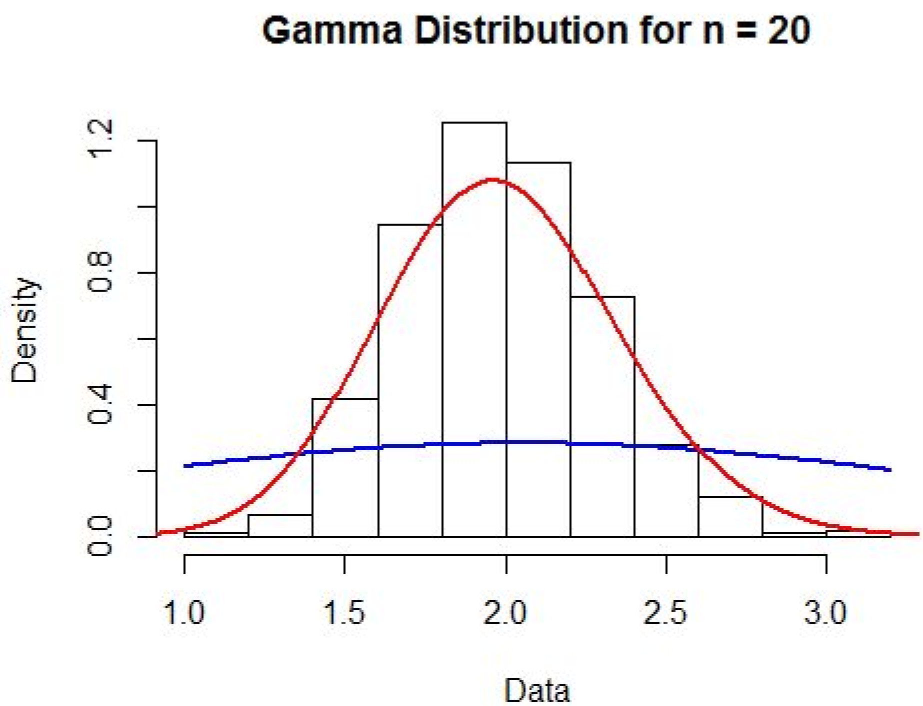


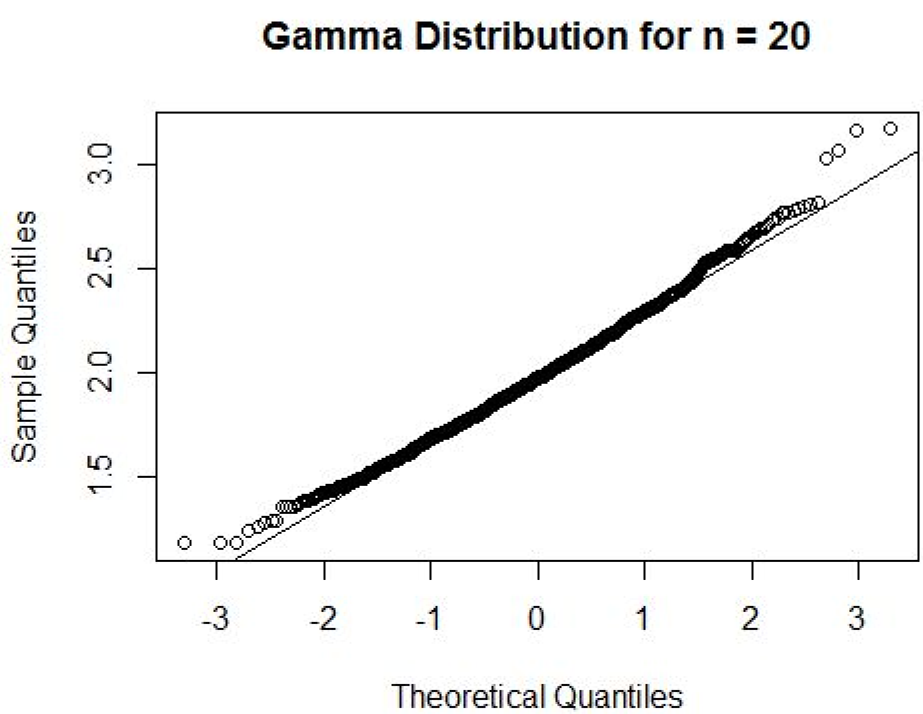
(n=5) This graph is NOT normal,it is slightly skewed.


Gamma Distribution for n = 10
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.5
2.0
2.5
3.0
3.5
Data

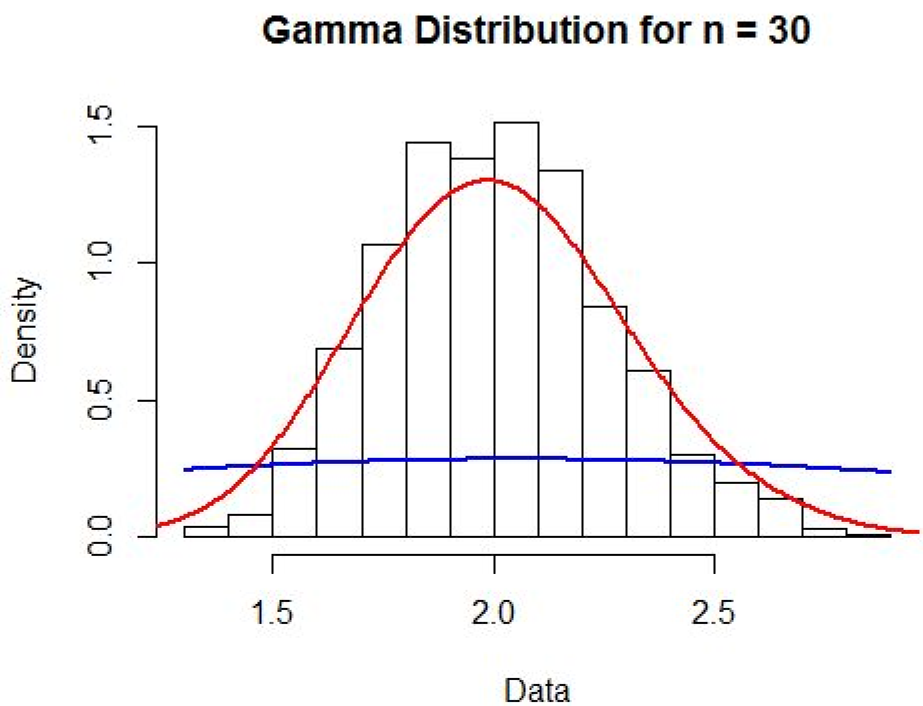

Gamma Distribution for n = 10
Sample Quantiles
1.0
1.5
2.0
2.5
3.0
3.5
-3
-2
-1
0
1
2
3
Theoretical Quantiles

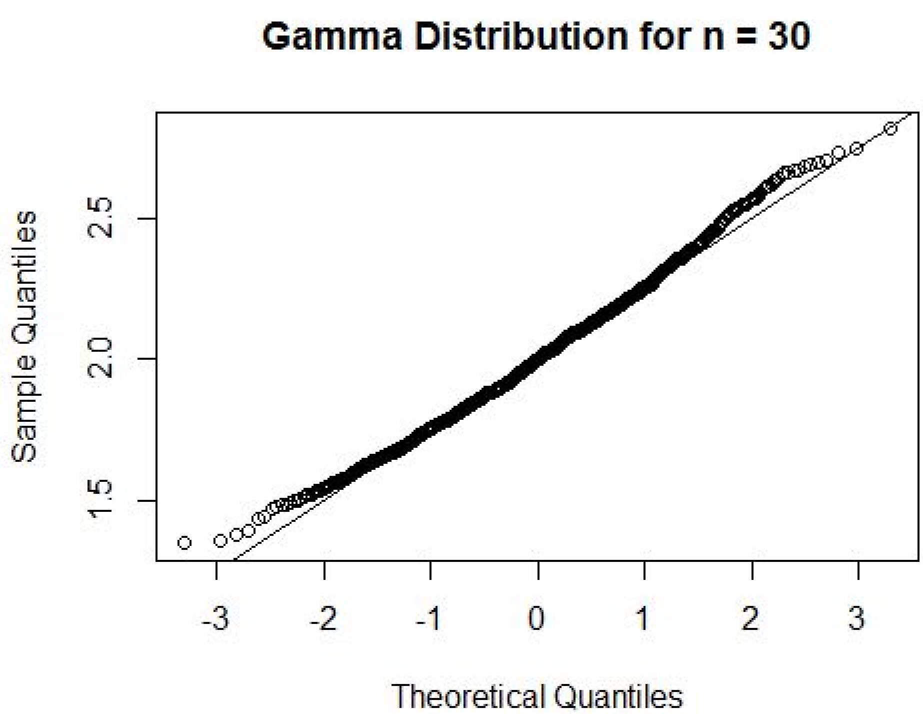
(n=10) This graph is NOT normal, it is slightly skewed with a large anomaly in the upper values.



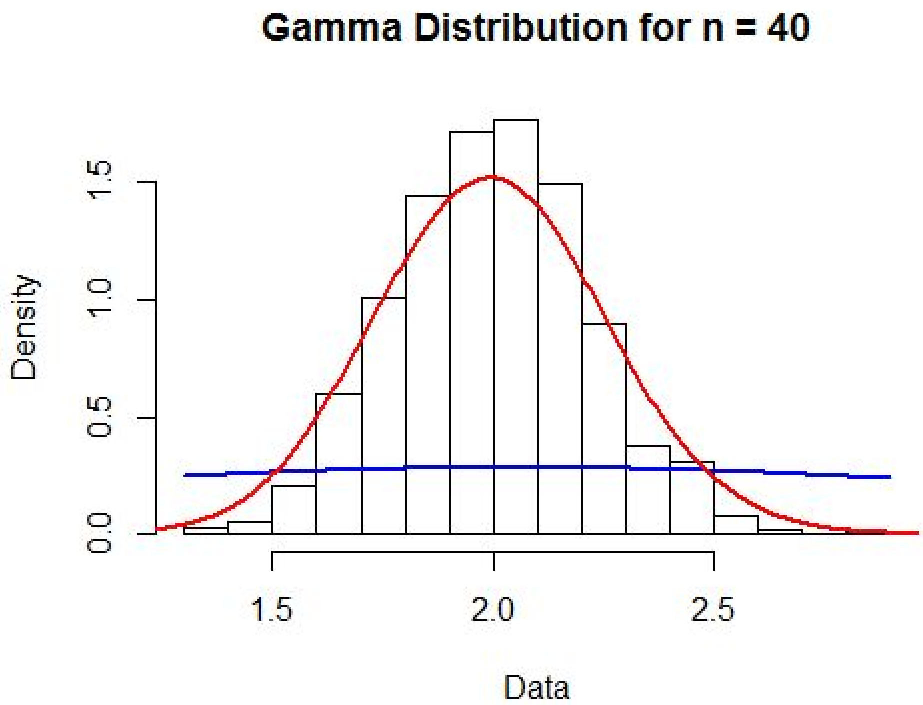


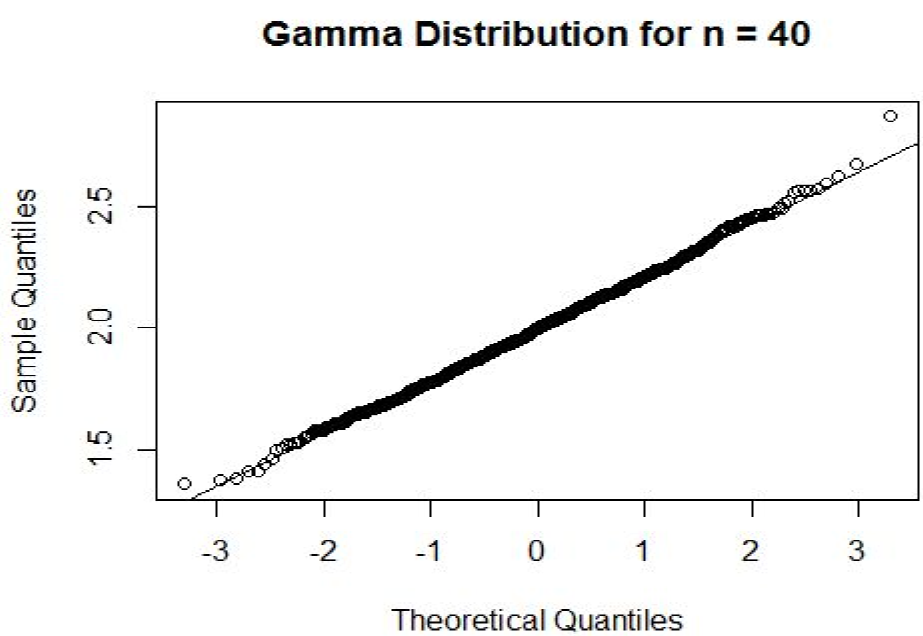
(n=20) This graph is NOT normal, it is very slightly skewed with an anomaly in the higher values.





(n=30) This graph is NOT normal, it is very slightly skewed with a large anomaly in the higher values.





(n=40) This plot is almost entirely straight so I would call this plot normal. There are still a couple of outliers but I feel enough of the points are in a line straight enough to justify calling this distribution normal.

3) Table

n	Experimental Mean	Theoretical Mean	Experimental Standard Deviation	Theoretical Standard Deviation
1	2.053329	2.053329	1.40293	1.40293
5	2.006196	2.006196	0.6145314	1.374134
10	1.999161	1.999161	0.4421559	1.39822
20	1.988048	1.988048	0.3106505	1.389271
30	2.007985	2.007985	0.2539808	1.39111
40	1.996513	1.996513	0.2183277	1.380826

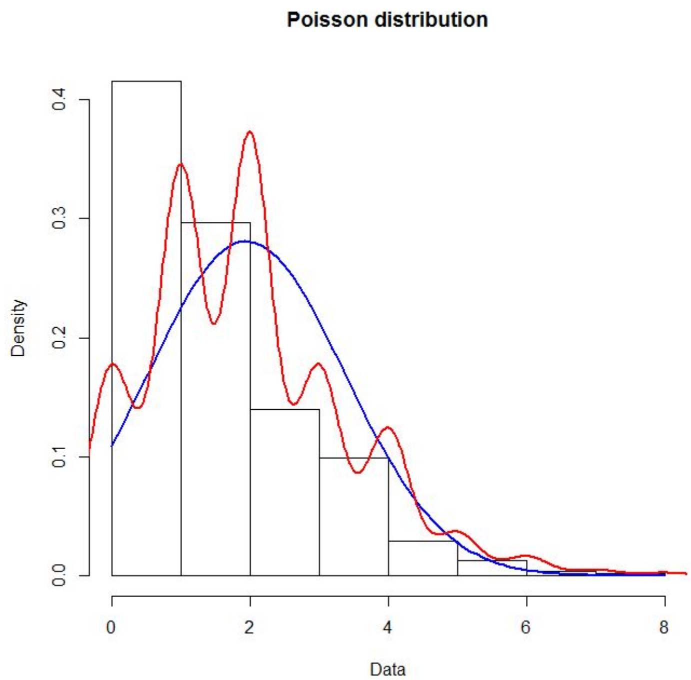
4) Concluding Remarks

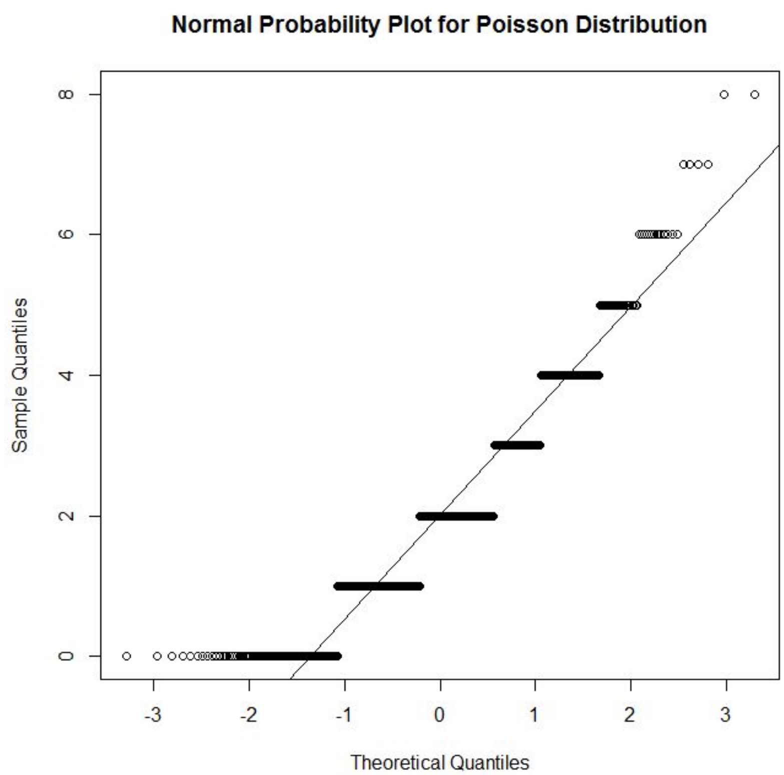
For the Gamma Distribution graphs as the value of n increased the peak of the red curve, the curve representing the distribution, began to approach the x coordinate of the peak of the blue curve, the curve representing actual normal., which means that the mean of the distributions got closer and closer to their predicted normal mean (2) as the number of columns went up to about 40 or higher. The equations comparing the sample mean and standard deviation to the theoretical population mean and standard deviation seem to hold true no matter the value of n. The population standard deviation value seemed to hover quite closely around an average value of 1.389 and the population mean always came out to be a value very close to 2.000.

Part D)

```
SRS<-1000
n<-1
data.vec<-rpois(SRS*n,2)
data.mat<-matrix(data.vec,ncol=n)
avg<-apply(data.mat,1,mean)
windows()
hist(avg,xlab="Data",freq=FALSE,main="Poisson distribution")
curve(dnorm(x, mean=mean(avg), sd=sd(avg)), col="blue", lwd=2, add=TRUE)
lines(density(avg, adjust=1),col = "red", lwd=2)
windows()
qqnorm(avg,main="Normal Probability Plot for Poisson Distribution")
qqline(avg)
mean(data.vec)
mean(avg)
sd(data.vec)
sd(avg)

```

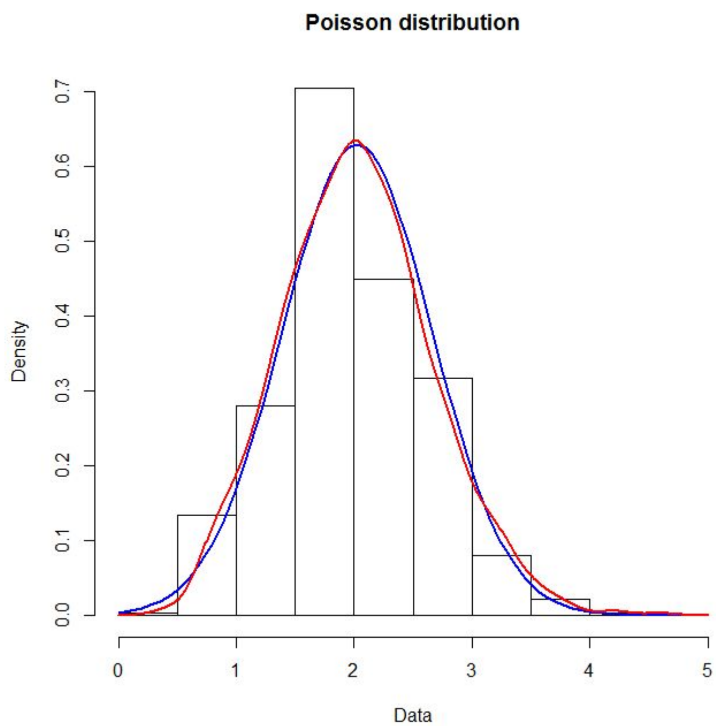
n=1

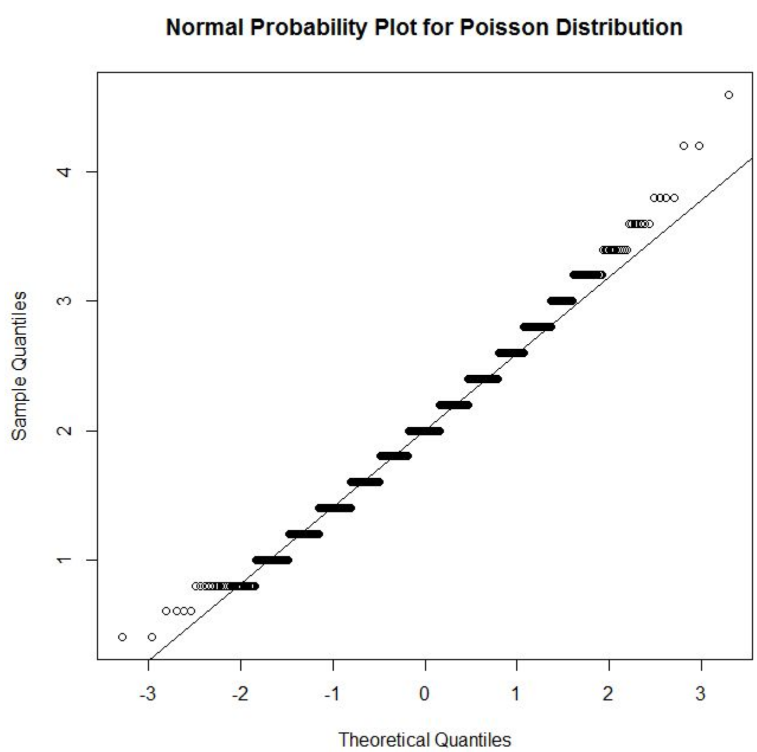




Not normal

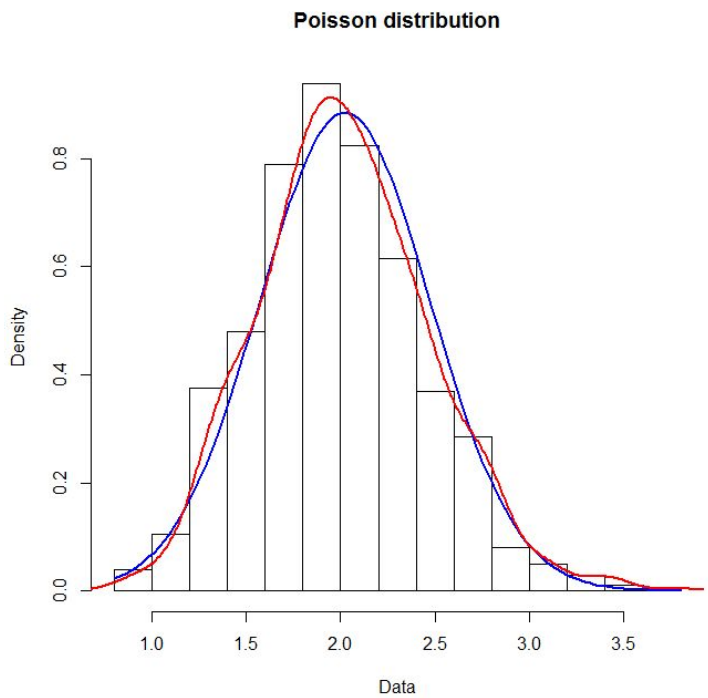
n=5


Poisson distribution
Density
Data
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
1
2
3
4
5

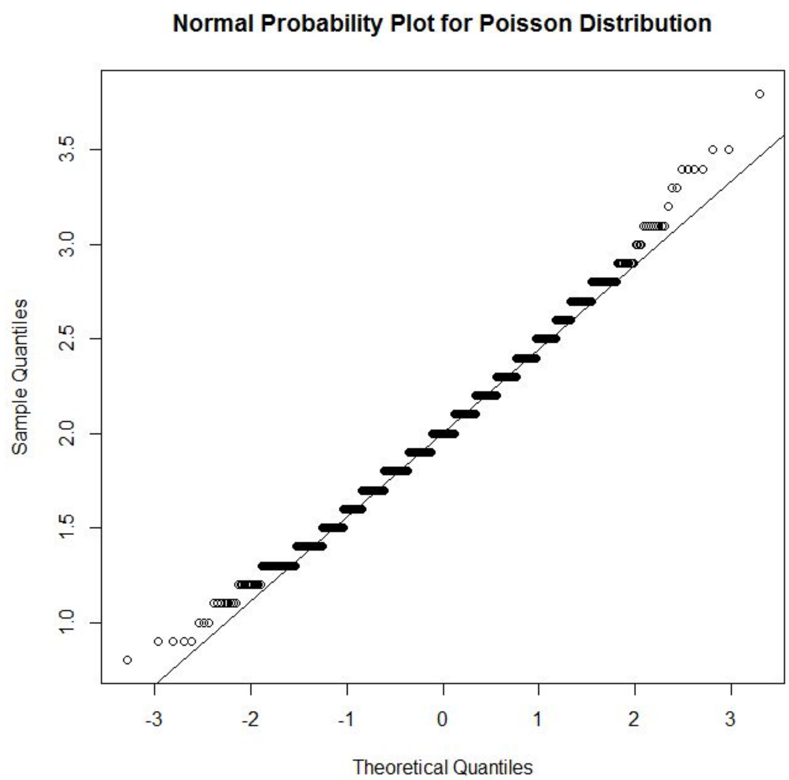

Normal Probability Plot for Poisson Distribution
Sample Quantiles
Theoretical Quantiles
1
2
3
4
-3
-2
-1
0
1
2
3

Normal

n=10



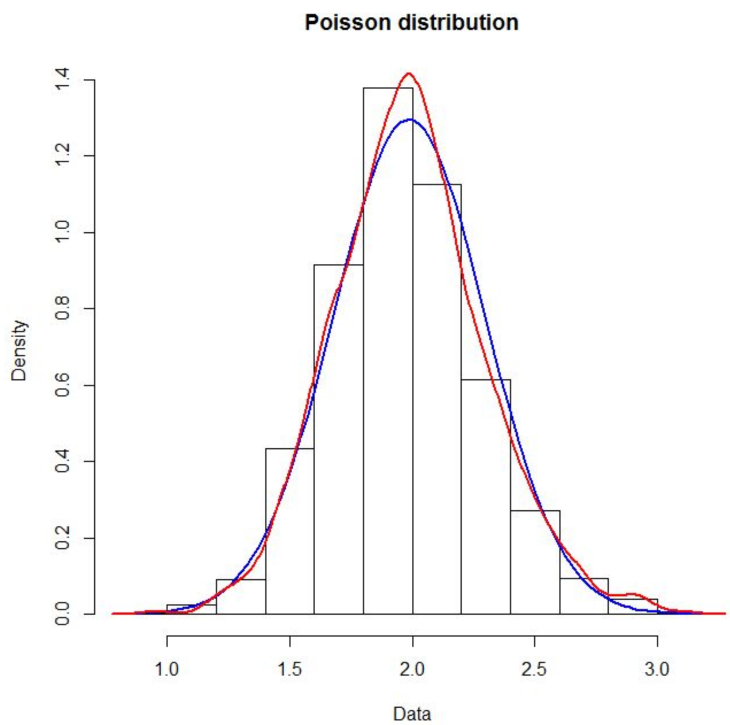
Poisson distribution
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.5
2.0
2.5
3.0
3.5
Data

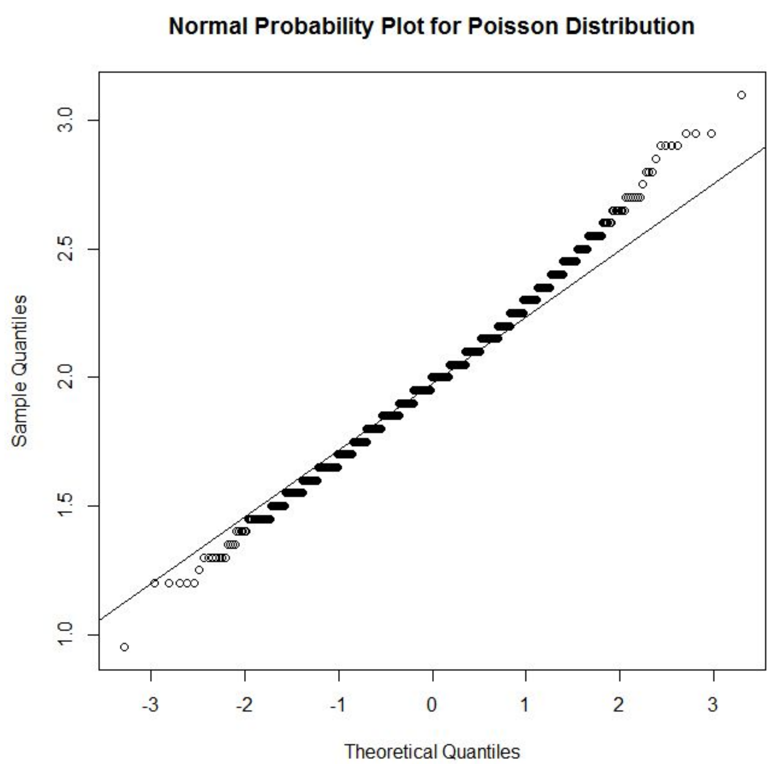


Normal Probability Plot for Poisson Distribution
Sample Quantiles
1.0
1.5
2.0
2.5
3.0
3.5
-3
-2
-1
0
1
2
3
Theoretical Quantiles

Normal

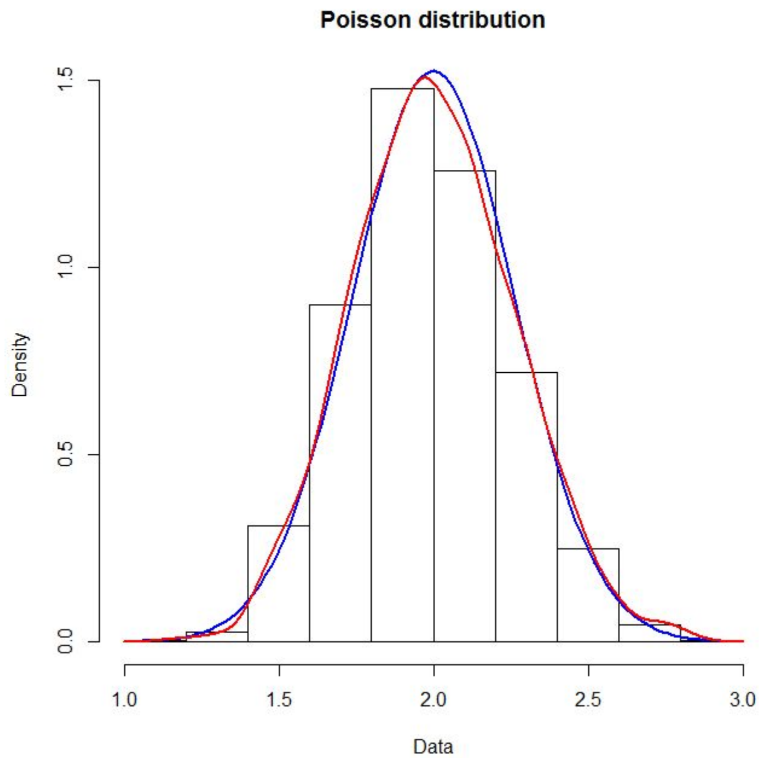
n=20

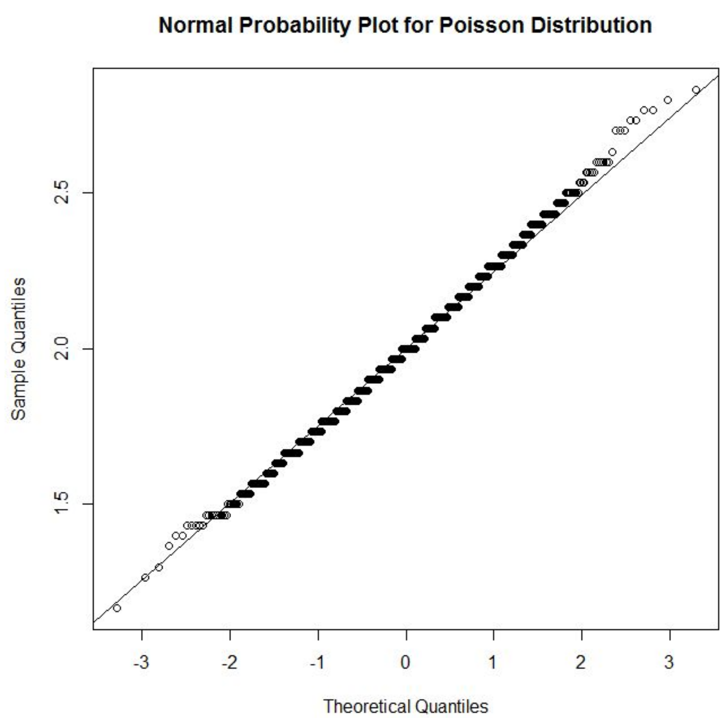

Poisson distribution
Density
Data
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.0
1.5
2.0
2.5
3.0


Normal Probability Plot for Poisson Distribution
Sample Quantiles
Theoretical Quantiles
1.0
1.5
2.0
2.5
3.0
-3
-2
-1
0
1
2
3

Normal

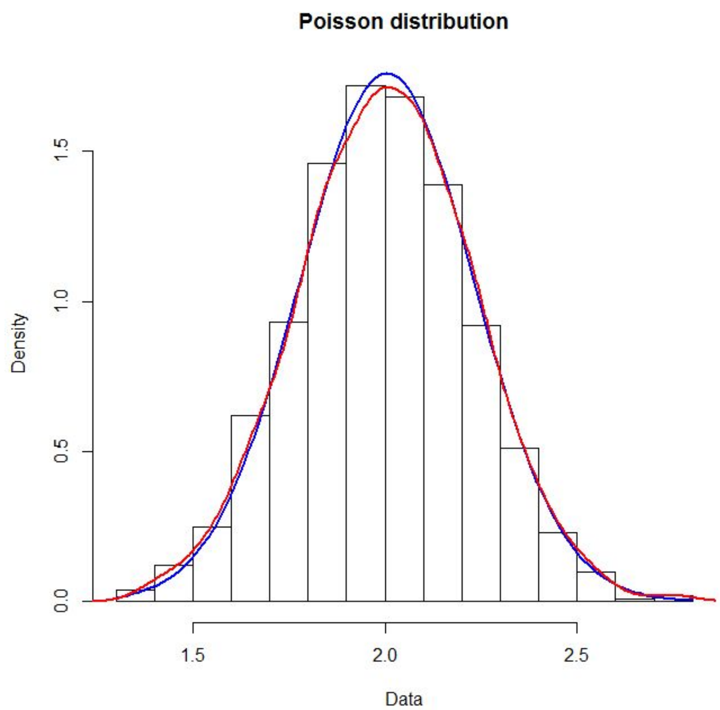
n=30

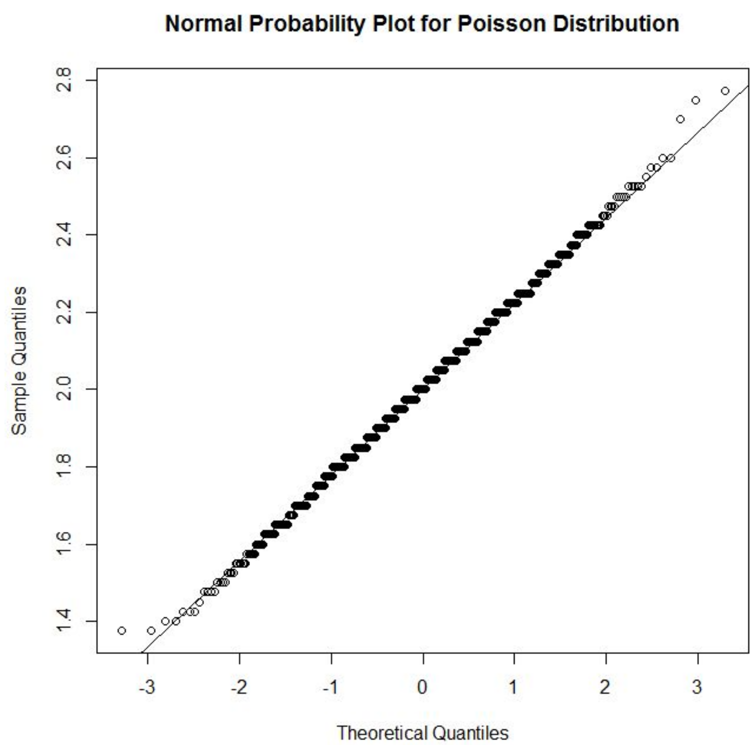

Poisson distribution
Density
Data
0.0
0.5
1.0
1.5
1.0
1.5
2.0
2.5
3.0


Normal Probability Plot for Poisson Distribution
Sample Quantiles
Theoretical Quantiles
1.5
2.0
2.5
-3
-2
-1
0
1
2
3

Normal

n=40


Poisson distribution
Density
0.0
0.5
1.0
1.5
1.5
2.0
2.5
Data


Normal Probability Plot for Poisson Distribution
Sample Quantiles
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
-3
-2
-1
0
1
2
3
Theoretical Quantiles

Normal

n	experimental mean of your 1000 x	theoretical mean (Equation 1)	experimental standard deviation of your 1000 x	theoretical standard deviation (Equations 1)
1	2.06	2.06	1.403699	1.403699
5	2.021	2.021	1.452095	0.640632
10	1.9981	1.9981	1.409076	0.4613884
20	2.00535	2.00535	1.423806	0.3156795
30	2.010467	2.010467	1.421088	0.2619949
40	1.99995	1.99995	1.416633	0.231484

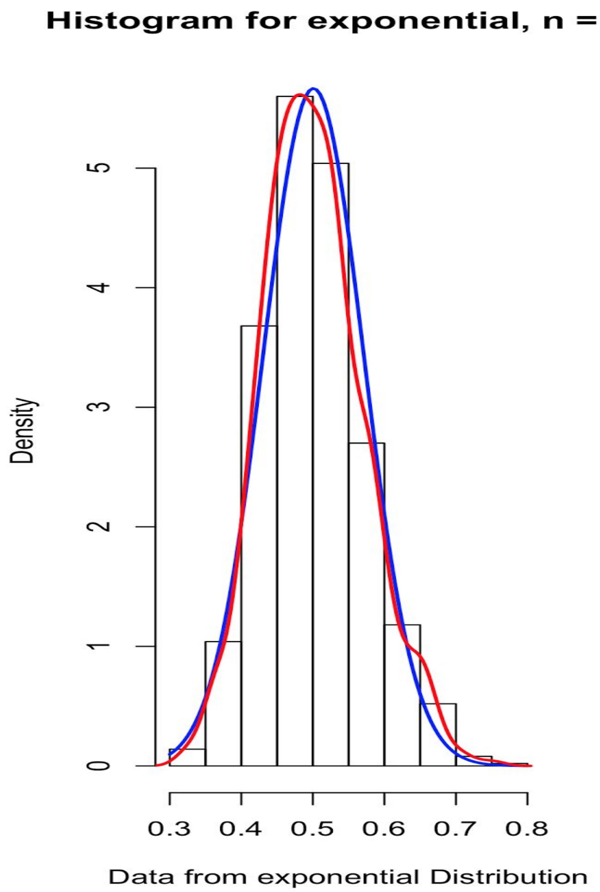
As the n value increase the plot becomes more normal. The plot for this section was normal when n was equal to 5. The equations (1) are applicable to all n values. The graphs will simply become more accurate as the n value increases theoretically.

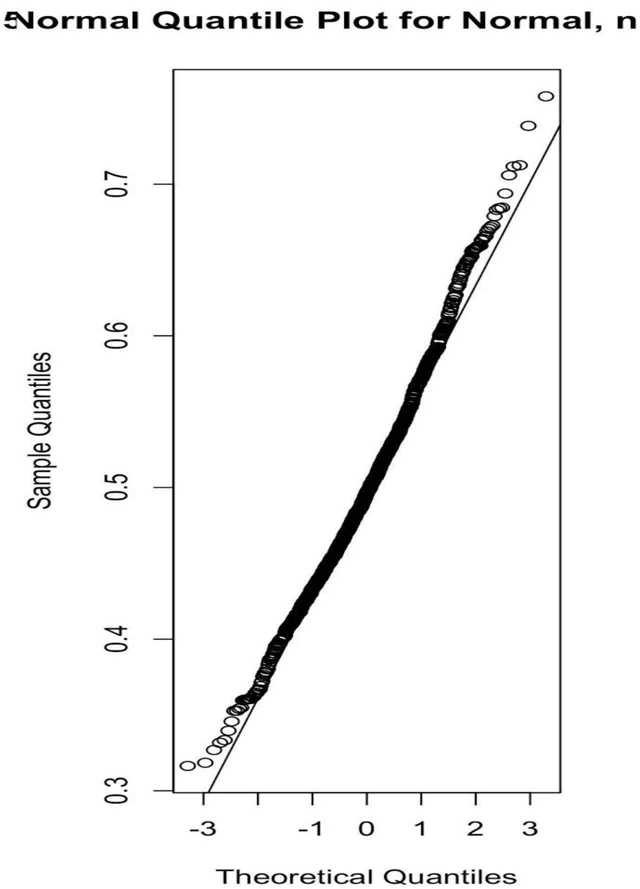
Part E.

1. Code

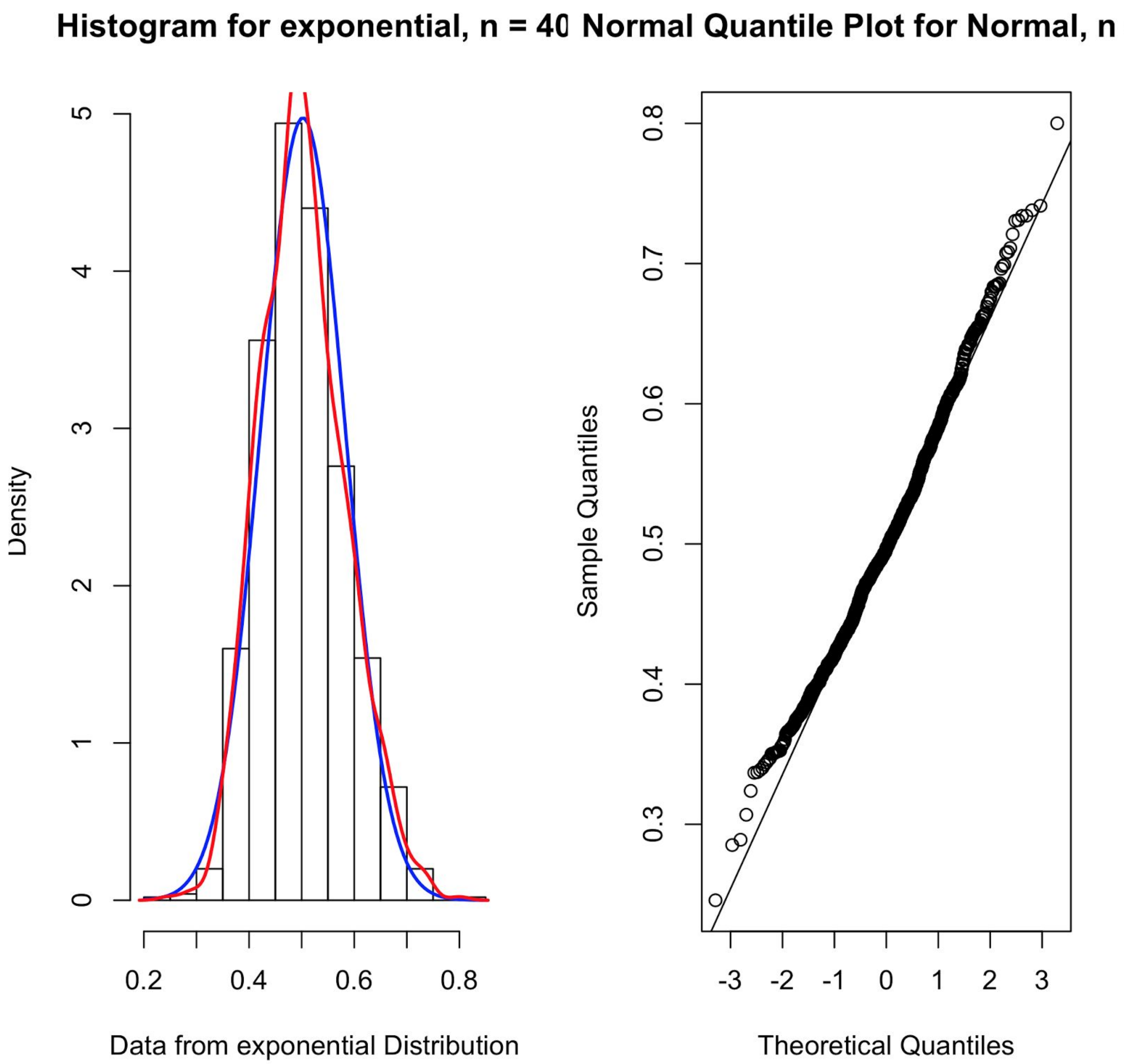
```
# E
SRS <- 1000
set <- c(1, 5, 10, 20, 30, 40 ,50)
for(i in set) {
  attach(mtcars)                # two plots in line
  par(mfrow=c(1,2))
  n <- i
  data.vec <- rexp(SRS*n,1/2) #creates the random data
  data.mat <- matrix(data.vec, ncol = n)
  avg <- apply(data.mat, 1, mean)

  m = mean(avg)
  std = sd(avg)
  # print(n, m, std)
  str = sprintf("n = %d, mean = %f, std = %f, sample std = %f", n,m, std, std*sqrt(n))
  print(str)

  # windows()#this is optional
  hist(avg, xlab="Data from exponential Distribution", freq = FALSE, main=sprintf("Histogram for exponential,
n = %d",n))

  curve(dnorm(x, mean=m, sd=std), col="blue", lwd=2, add=TRUE)
  lines(density(avg, adjust=1),col = "red", lwd=2)
  # windows()
  # quartz()
  qqnorm(avg,main=sprintf("Normal Quantile Plot for Normal, n = %d",n))
  qqline(avg)
  quartz()
}
```

2.Normal

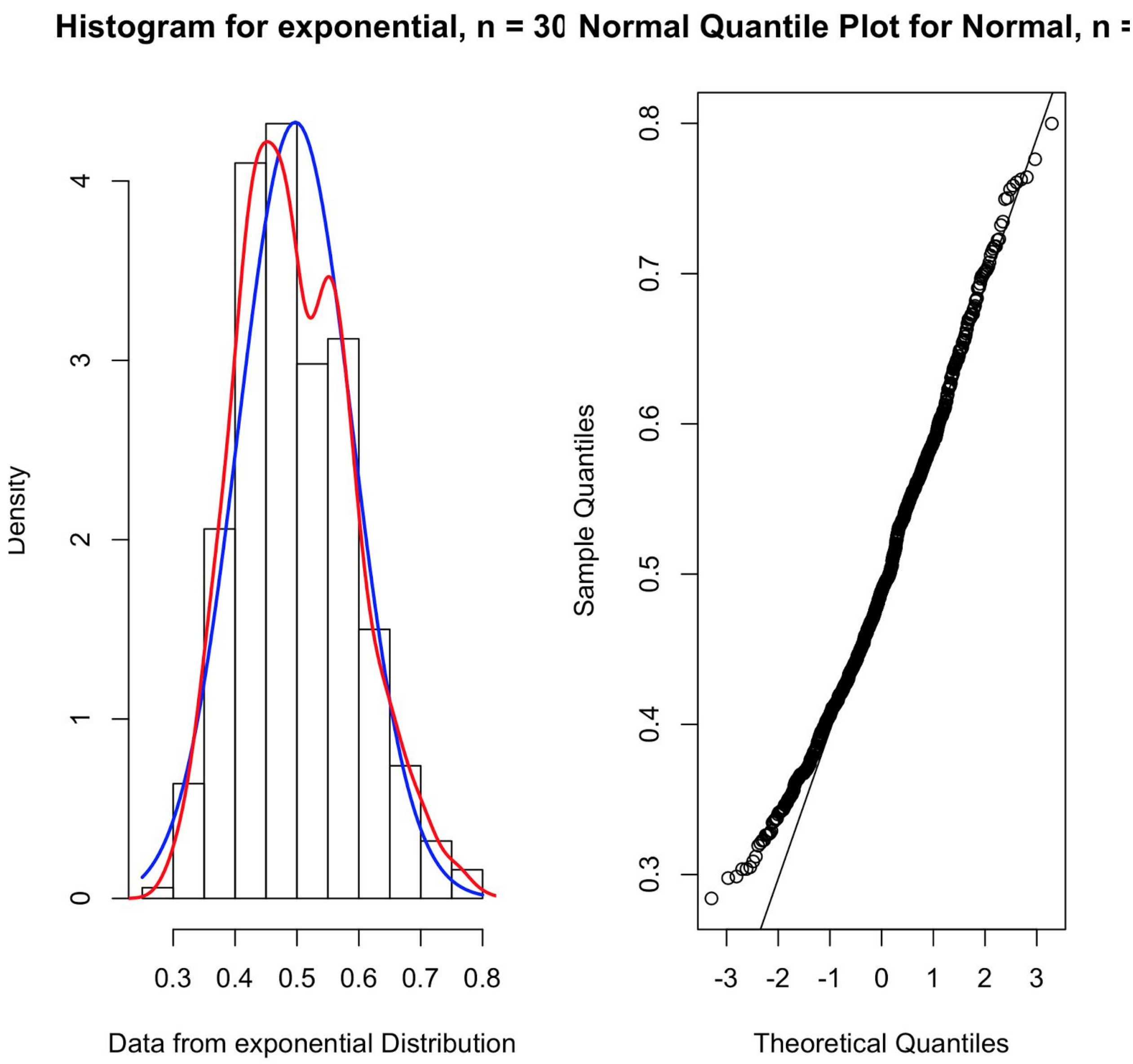

Histogram for exponential, n = 5
Density
Data from exponential Distribution
0.3 0.4 0.5 0.6 0.7 0.8
0 1 2 3 4 5


Normal Quantile Plot for Normal, n
Sample Quantiles
Theoretical Quantiles
-3 -1 0 1 2 3
0.3 0.4 0.5 0.6 0.7

Normal

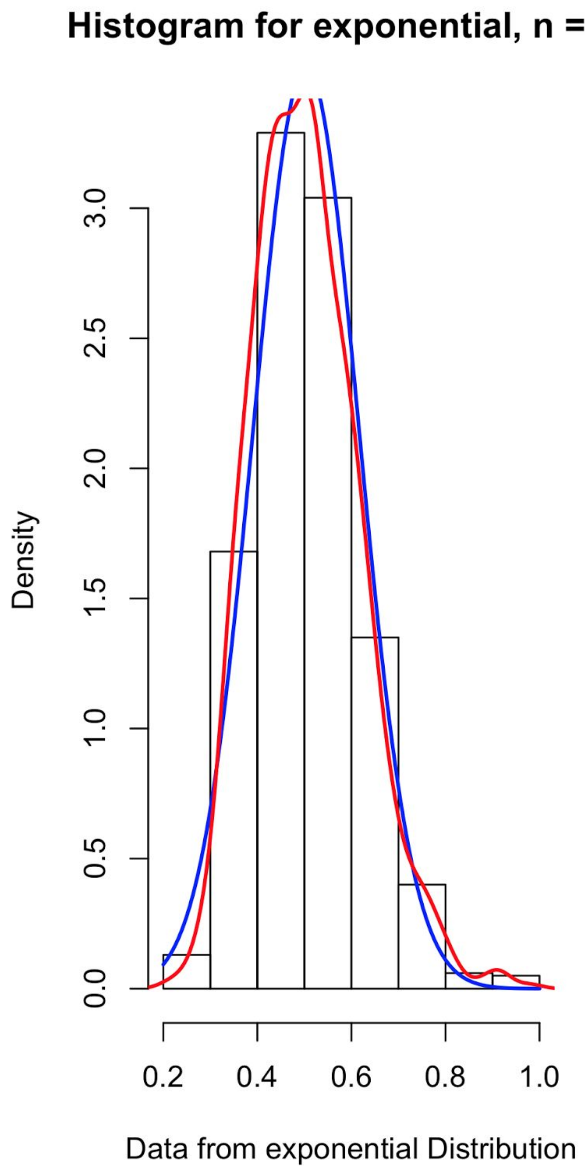


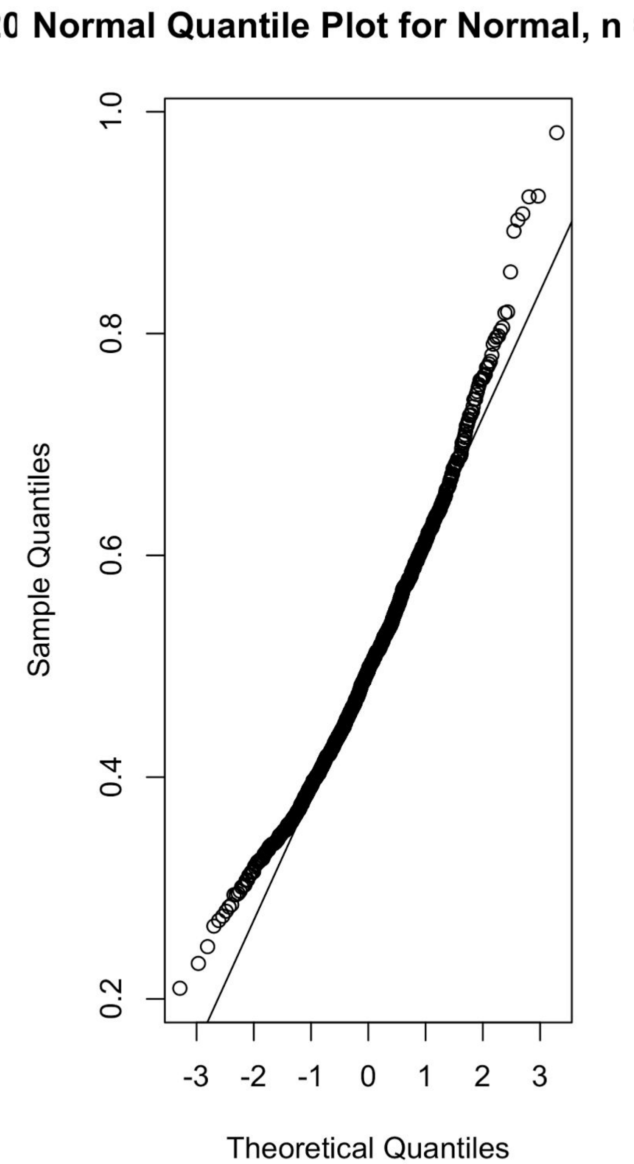
Normal



Histogram for exponential, n = 30
Density
Data from exponential Distribution
0 1 2 3 4
0.3 0.4 0.5 0.6 0.7 0.8
Normal Quantile Plot for Normal, n =
Sample Quantiles
Theoretical Quantiles
0.3 0.4 0.5 0.6 0.7 0.8
-3 -2 -1 0 1 2 3

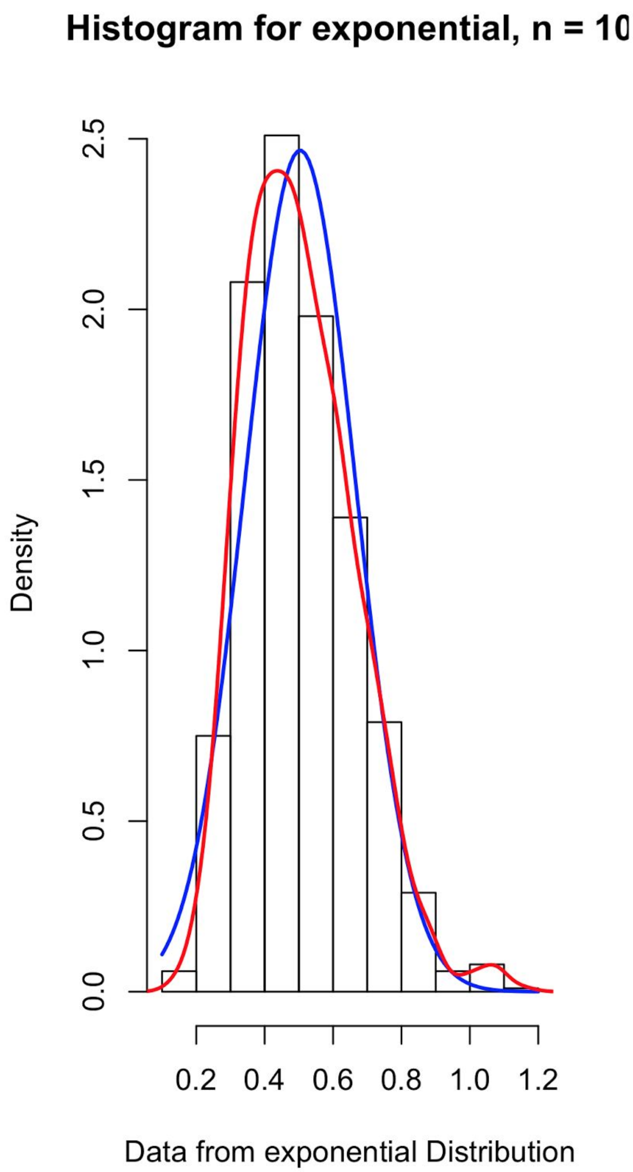
Normal

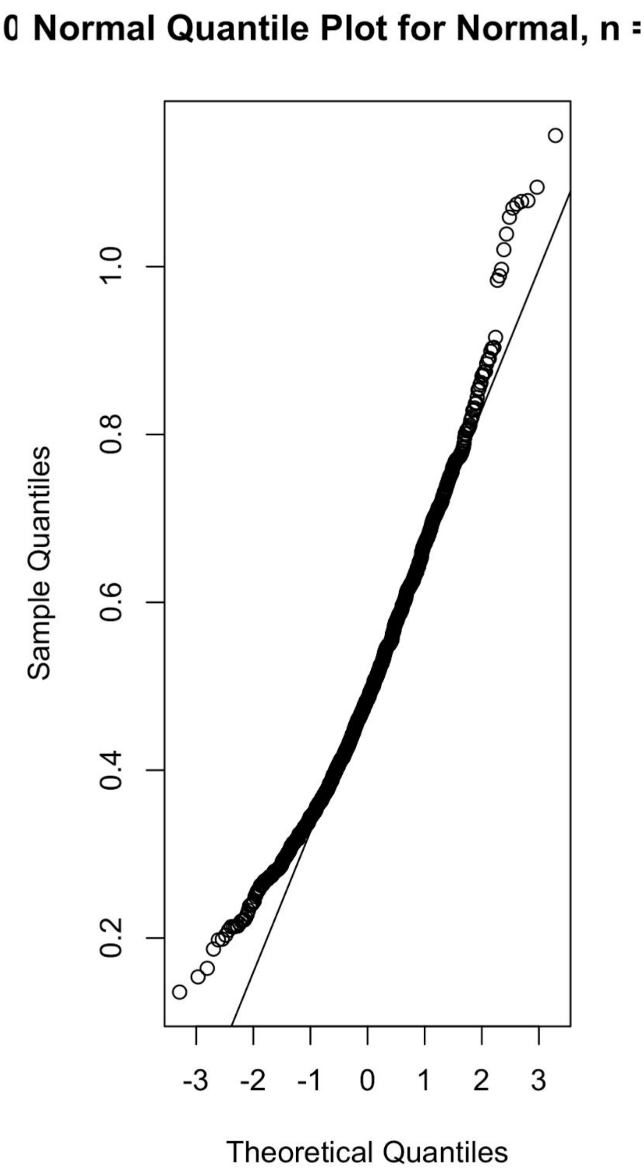

Histogram for exponential, n = 20
Density
Data from exponential Distribution
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.2
0.4
0.6
0.8
1.0


Normal Quantile Plot for Normal, n =
Sample Quantiles
Theoretical Quantiles
0.2
0.4
0.6
0.8
1.0
-3
-2
-1
0
1
2
3

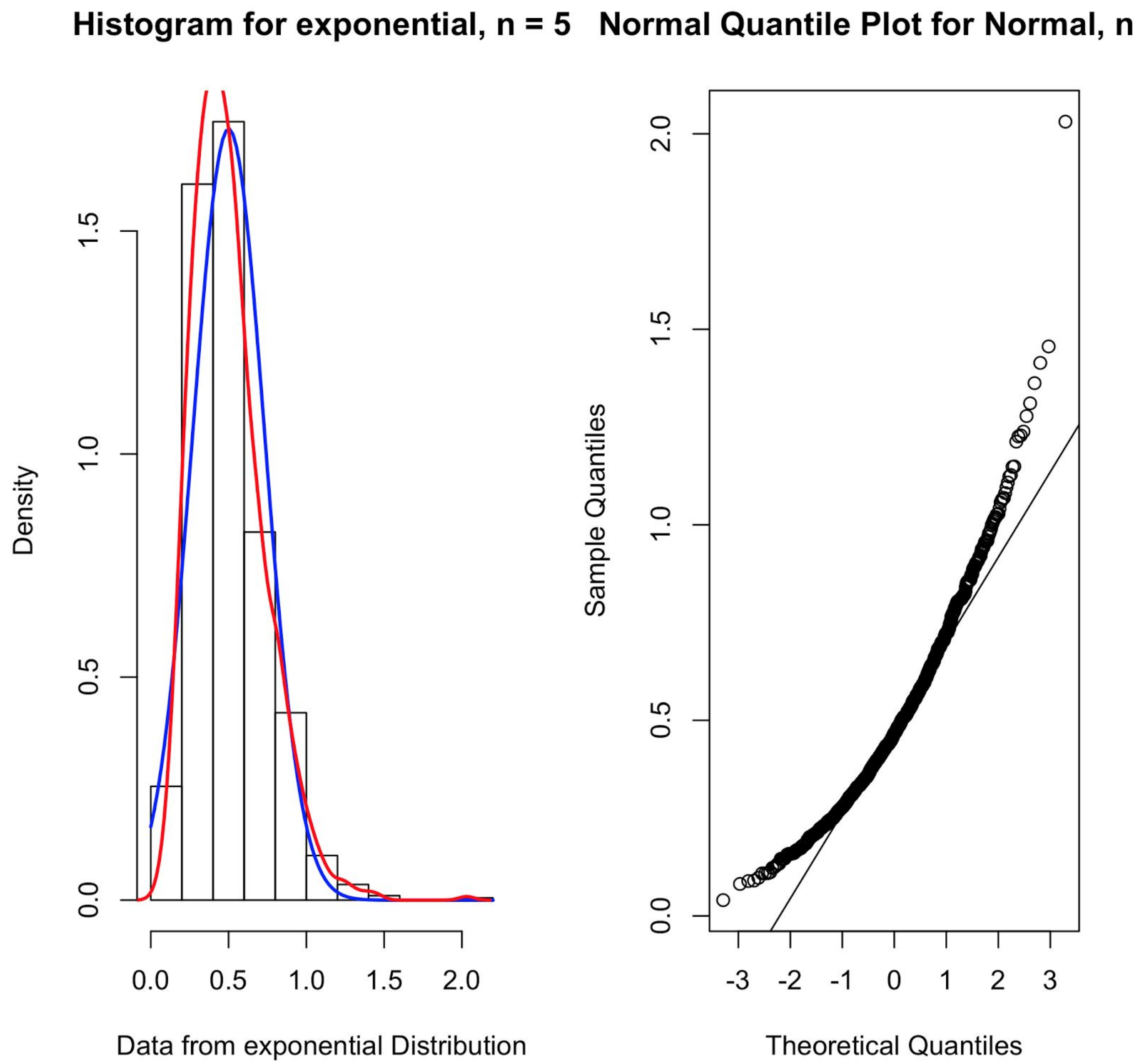
Normal

Normal

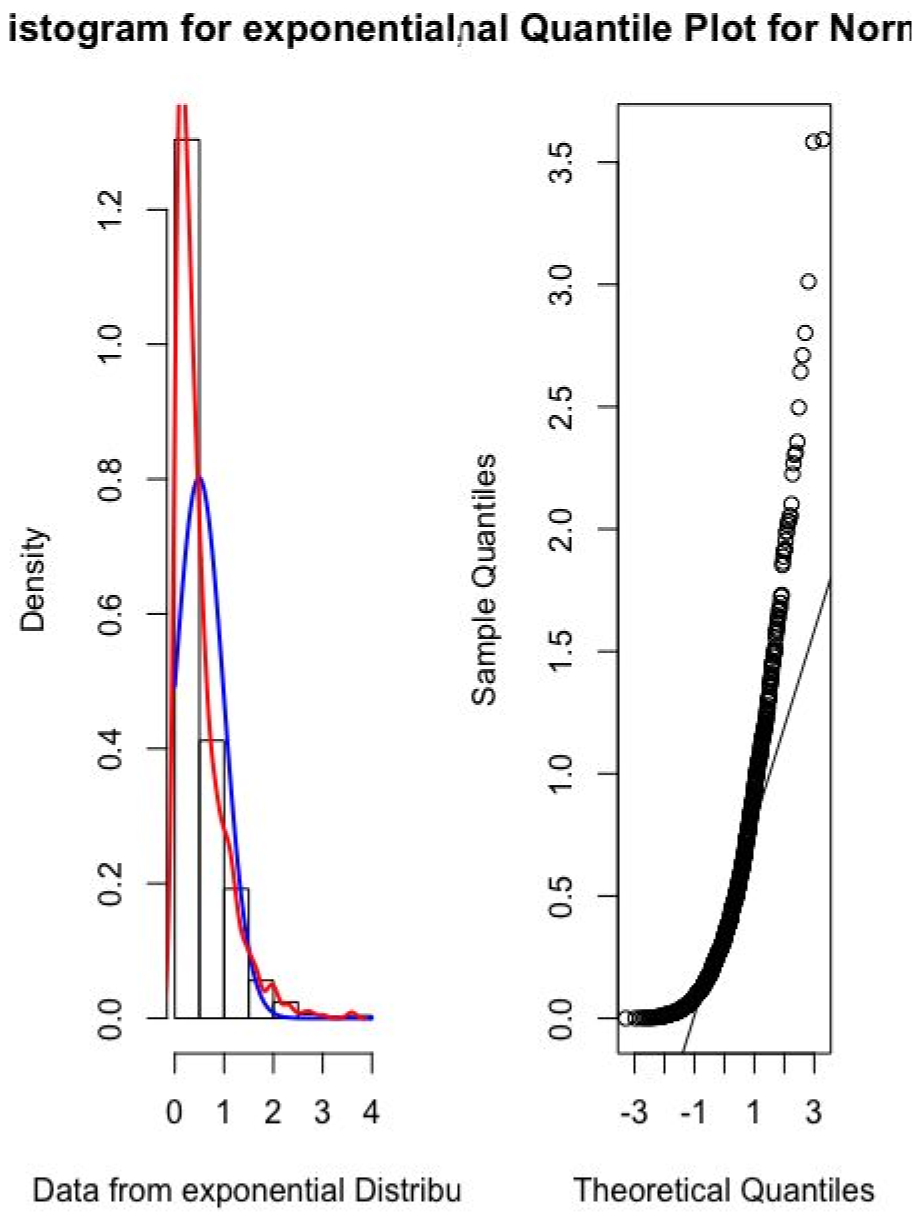




Not Normal


Histogram for exponential, n = 5
Normal Quantile Plot for Normal, n
Density
Data from exponential Distribution
Sample Quantiles
Theoretical Quantiles
0.0
0.5
1.0
1.5
2.0
-3
-2
-1
0
1
2
3

Not Normal!


istogram for exponential
nal Quantile Plot for Norn
Density
Data from exponential Distribu
Sample Quantiles
Theoretical Quantiles
0.0
0.2
0.4
0.6
0.8
1.0
1.2
0
1
2
3
4
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-3
-1
1
3

n	experimental mean of your 1000 x	theoretical mean (Equation 1)	experimental standard deviation of your 1000 x	theoretical standard deviation (Equations 1)
1	0.490893	0.5	0.497336	0.497336
5	0.500633	0.5	0.515984	0.230755
10	0.503915	0.5	0.511547	0.161766
20	0.503671	0.5	0.503348	0.112552
30	0.497481	0.5	0.504851	0.092173
40	0.502456	0.5	0.507165	0.080190
50	0.500830	0.5	0.497795	0.070399

For the Poisson Distribution graphs as the value of n increased the peak of the red curve, the curve representing the distribution, began to approach the x coordinate of the peak of the blue curve, the curve representing actual normal., which means that the mean of the distributions got closer and closer to their predicted normal mean (0.5) as the number of columns went up to about 50 or higher.