Jordan Mayer (Lab 071-1:30pm, LEC 076-1:30pm)

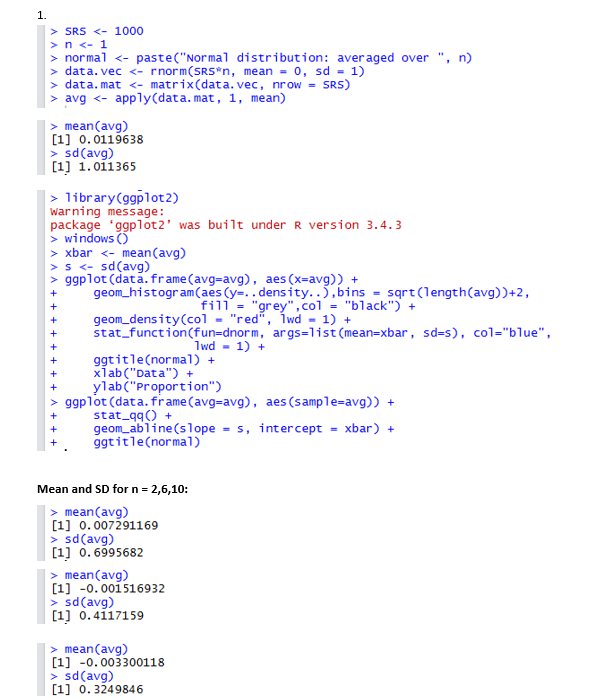
Harley Jo Rowland (Lab 071-1:30pm, LEC 076-1:30pm)

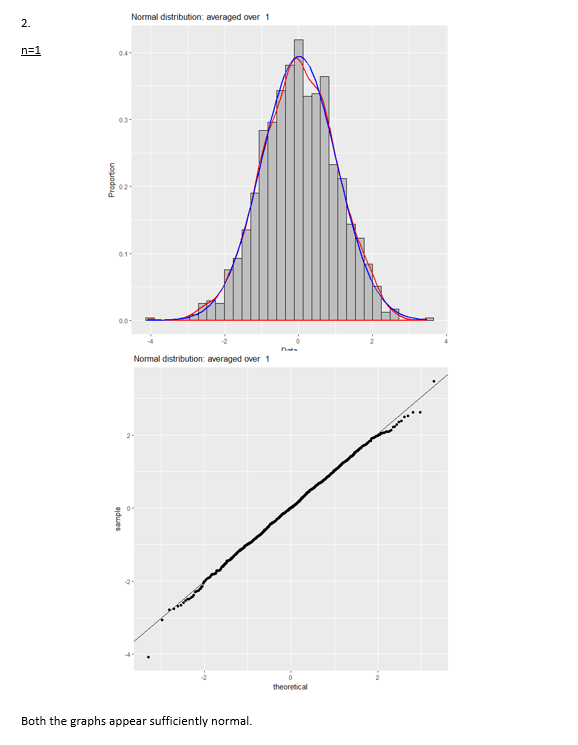
Qihang Xu(Lab Section 071-1:30pm, LEC 076-1:30pm)

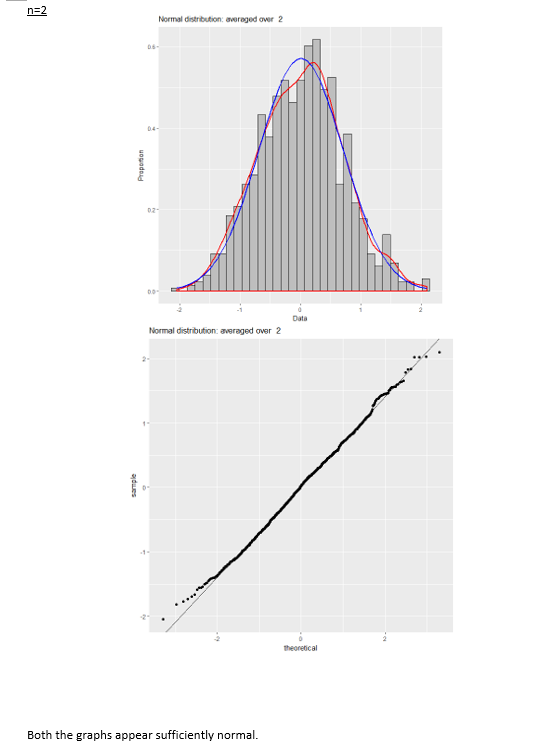
Ernest Lee (Lab Section 051-3.30pm, LEC 050-3.30pm)

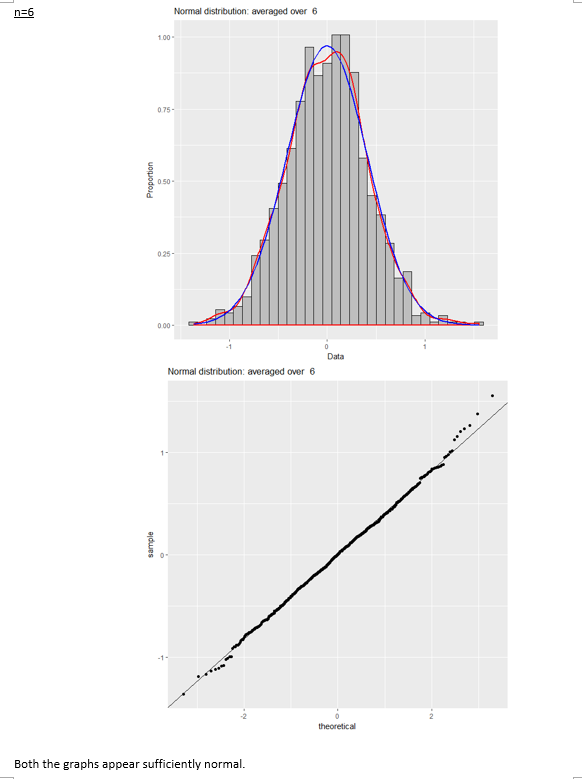
STAT 350 Lab 4

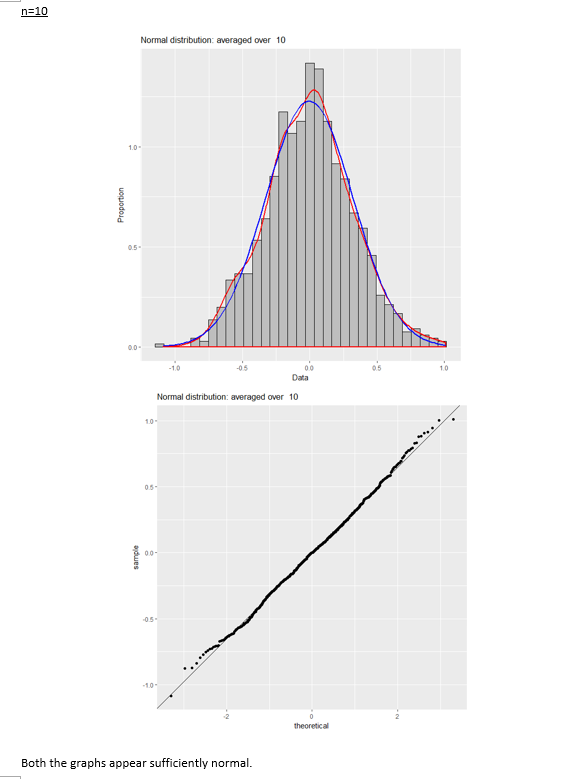
**B. (20 points) Standard Normal Distribution.**

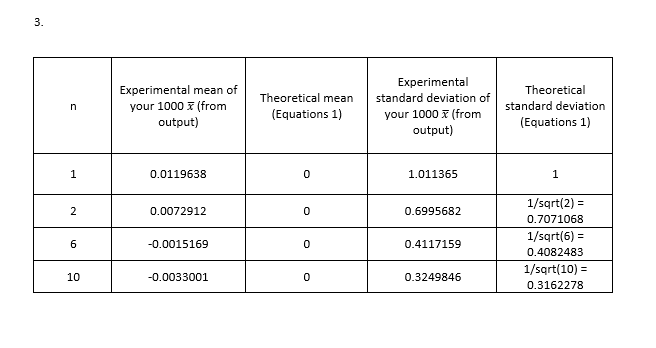




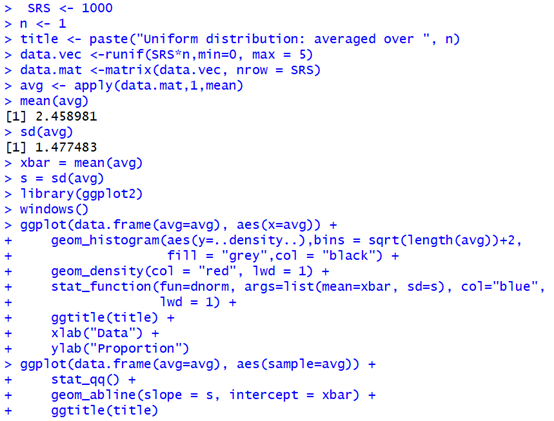


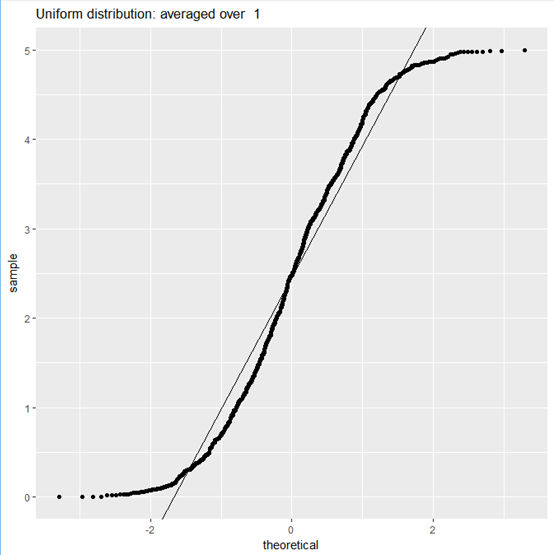
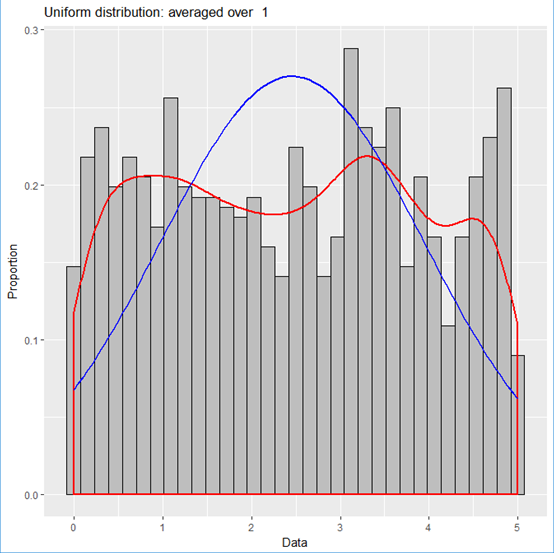




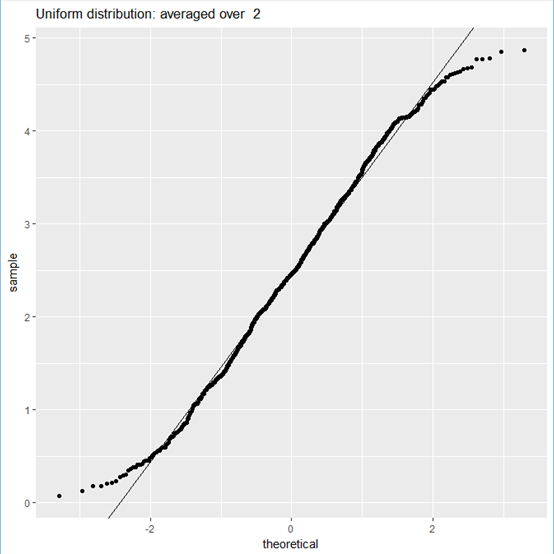
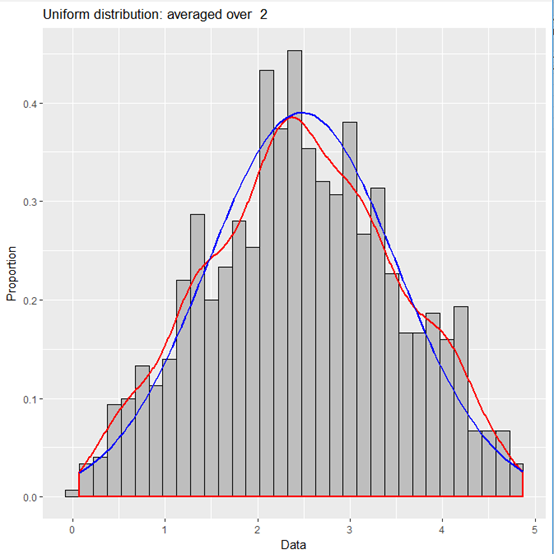


**C. (20 points) Uniform distribution**

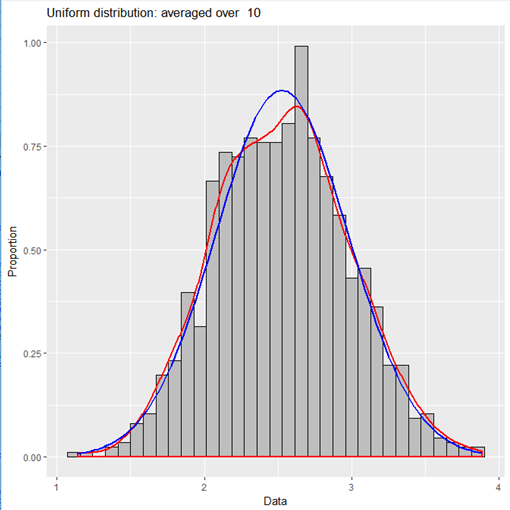




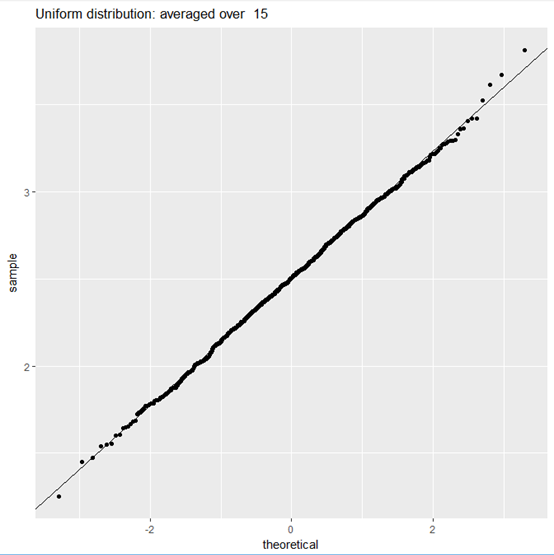
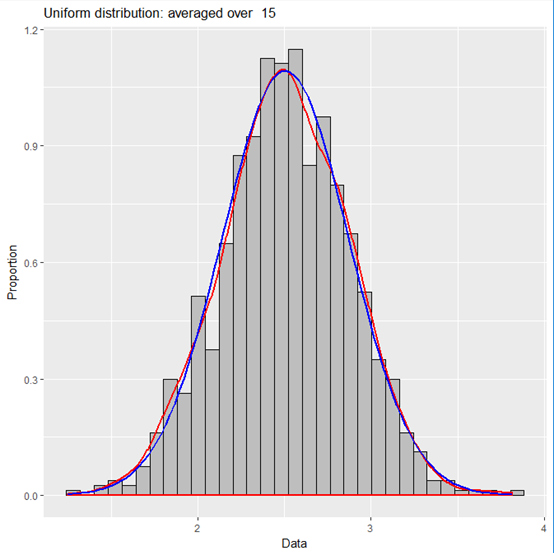
Both still look like uniform distribution.



Both of the graphs are somewhere in between uniform and normal.



Graphs are close to uniform distribution.



Graphs are reasonably normal.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n | Experimental mean of your 1000 𝑥̅ (from output) | Theoretical mean (Equations 1) | Experimental standard deviation of your 1000 𝑥̅ (from output) | Theoretical standard deviation (Equations 1) |
| 1 | 2.458981 | (0+5)/2=2.5 | 1.477483 | Sqrt((5-0)^2/12)/1=1.443375673 |
| 2 | 2.480496 | 2.5 | 1.022201 | 1.443375673/sqrt(2) =  1.020620726 |
| 10 | 2.526079 | 2.5 | 0.4509768 | 1.443375673/sqrt(10)=0.4564354646 |
| 15 | 2.504869 | 2.5 | 0.3652246 | 1.443375673/sqrt(15)=0.3726779962 |

**D. (20 points) Gamma distribution**

**1. Code**

# set working directory to Lab 04

setwd("C:/Users/jordan/Google Drive/Courses Spring 2018/STAT 350/STAT 350 Labs/Lab 04")

# set up ggplot2 for plotting

library(ggplot2)

###

# PART D: Gamma distribution with parameters alpha=3, beta=2

#

# For a gamma distribution with alpha = 3 and beta = 2,

# generate 1000 random samples of size n = 1, 5, 10, 20, 40, 60, 80, ...

#

# Then for each sample:

# find sample average and

# create histogram (with two colored curves) and

# normal probability plot of sample mean

# for each graph pair, indicate whether they appear sufficiently normal.

# NOTE: do not need to explain judgment of normality.

#

# Do this until n is large enough that distribution of sample mean

# appears normal.

###

samples <- 1000 # number of samples to collect

sizes <- c(1, 5, 10, 20, 40, 60, 80, 100) # sample sizes

# create graph pair for each sample size

for (n in sizes) {

title <- paste("Gamma Distribution: Averaged Over", n) # title for graphs

# generate data and calculate means

gamma.vec <- rgamma(samples\*n, 3, rate=2) # random gamma data

gamma.mat <- matrix(gamma.vec, nrow=samples) # separate data into rows

gamma.means <- apply(gamma.mat, 1, mean) # calculate means

# create histogram

hist <- ggplot(data.frame(gamma.means=gamma.means),aes(gamma.means))+

geom\_histogram(aes(y=..density..),bins=sqrt(samples)+2,

fill="grey",col="black")+

geom\_density(col="red",lwd=1)+

stat\_function(fun=dnorm,args=list(mean=mean(gamma.means),

sd=sd(gamma.means)),

col="blue",lwd=1)+

ggtitle(title)+

xlab("Data")+

ylab("Proportion")

ggsave(hist, filename=paste("gammaHist",n,".png",sep=""))

# create normal probability plot

qq <- ggplot(data.frame(gamma.means=gamma.means),aes(sample=gamma.means))+

stat\_qq()+

geom\_abline(slope=sd(gamma.means),intercept=mean(gamma.means))+

ggtitle(title)+

xlab("Theoretical")+

ylab("Sample")

ggsave(qq, filename=paste("gammaQQ",n,".png",sep=""))

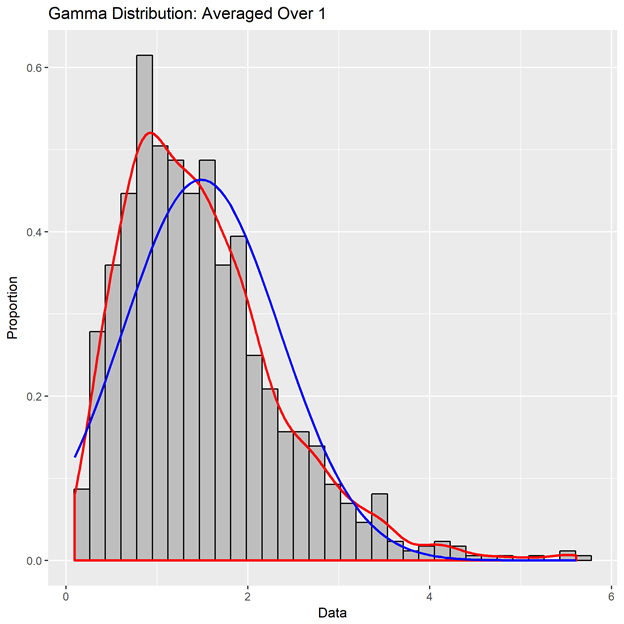
print(paste("n = ", n))

print(paste("mean = ", mean(gamma.means)))

print(paste("sd = ", sd(gamma.means)))

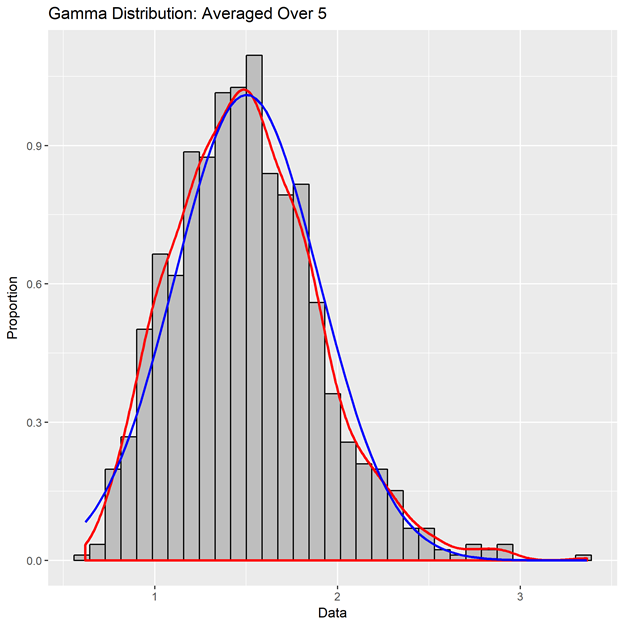
}

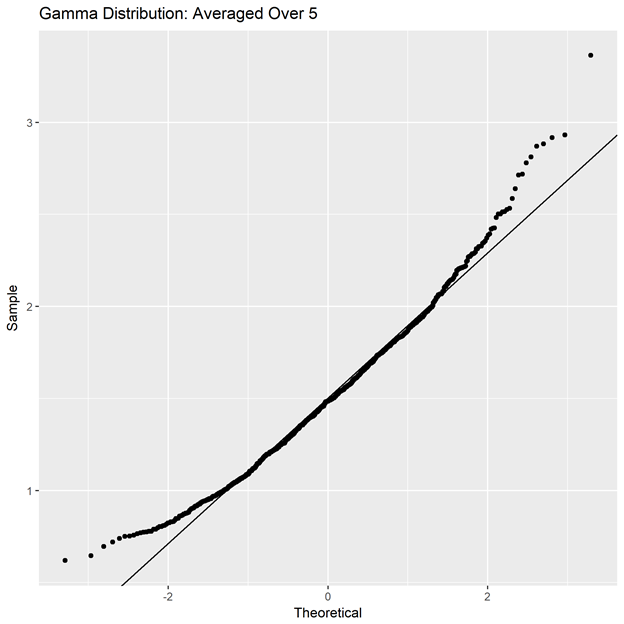
**2. Histogram/normal probability plots**

****

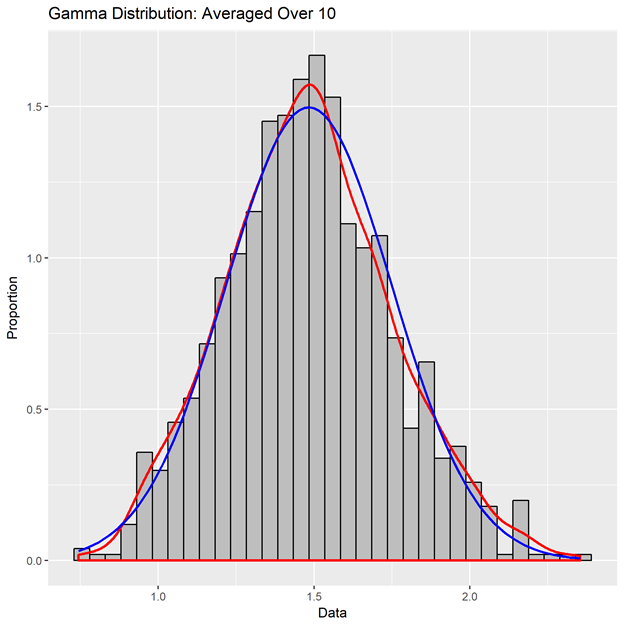
****

Not sufficiently normal



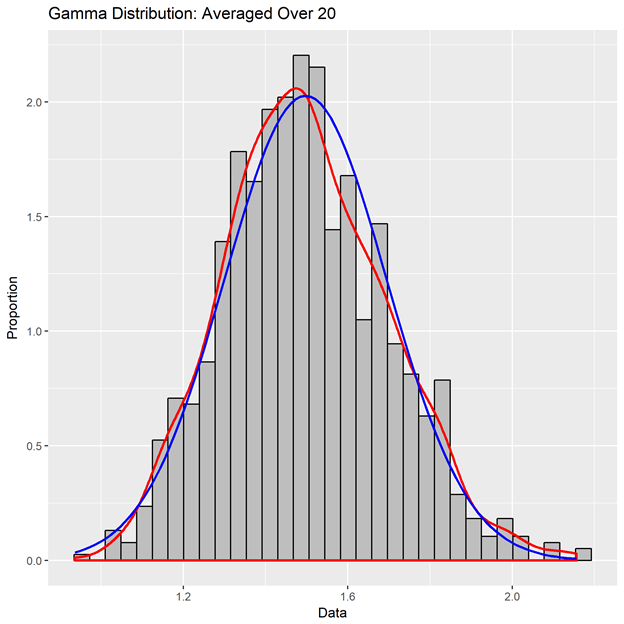


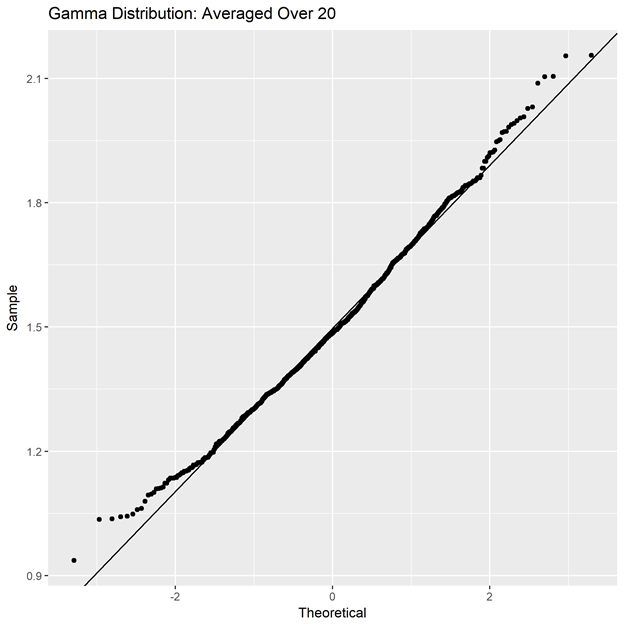
Not sufficiently normal

****

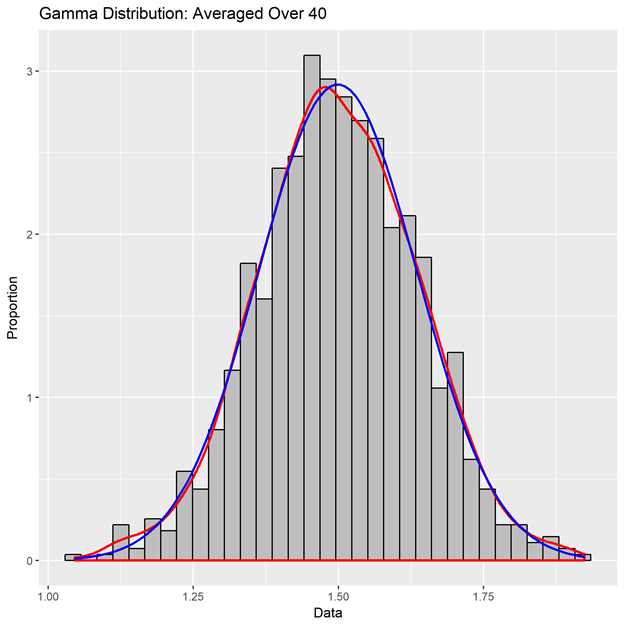
****

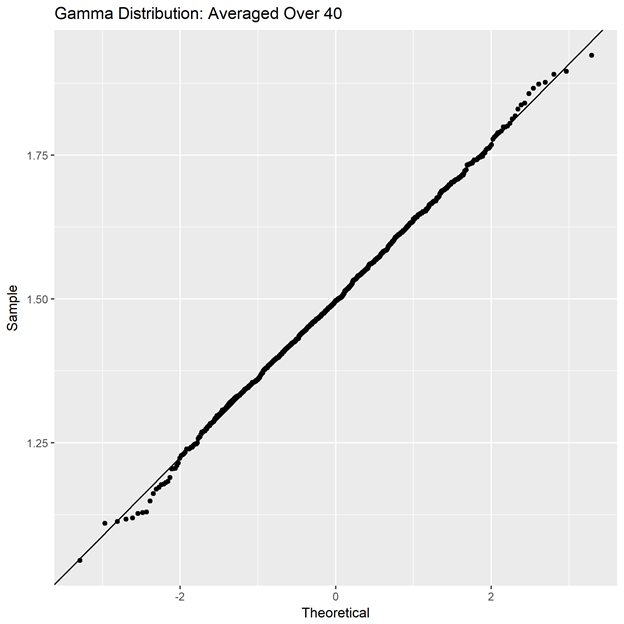
Not sufficiently normal

****

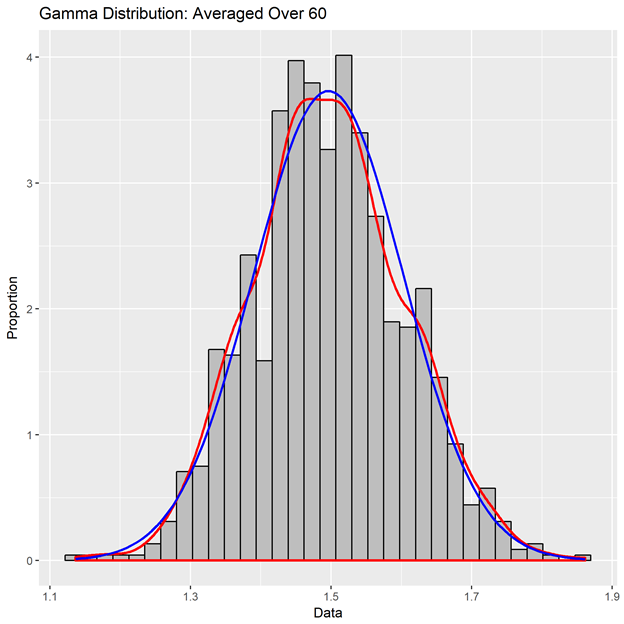
****

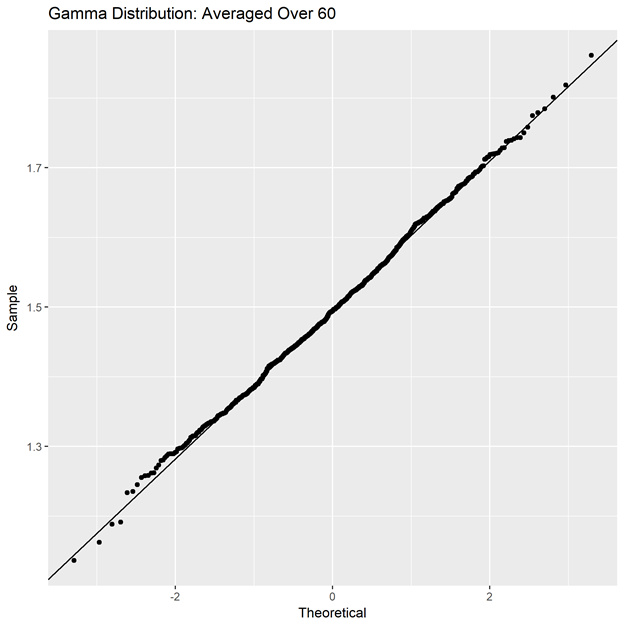
Not sufficiently normal

****

****

Not sufficiently normal, but pretty close

****

****

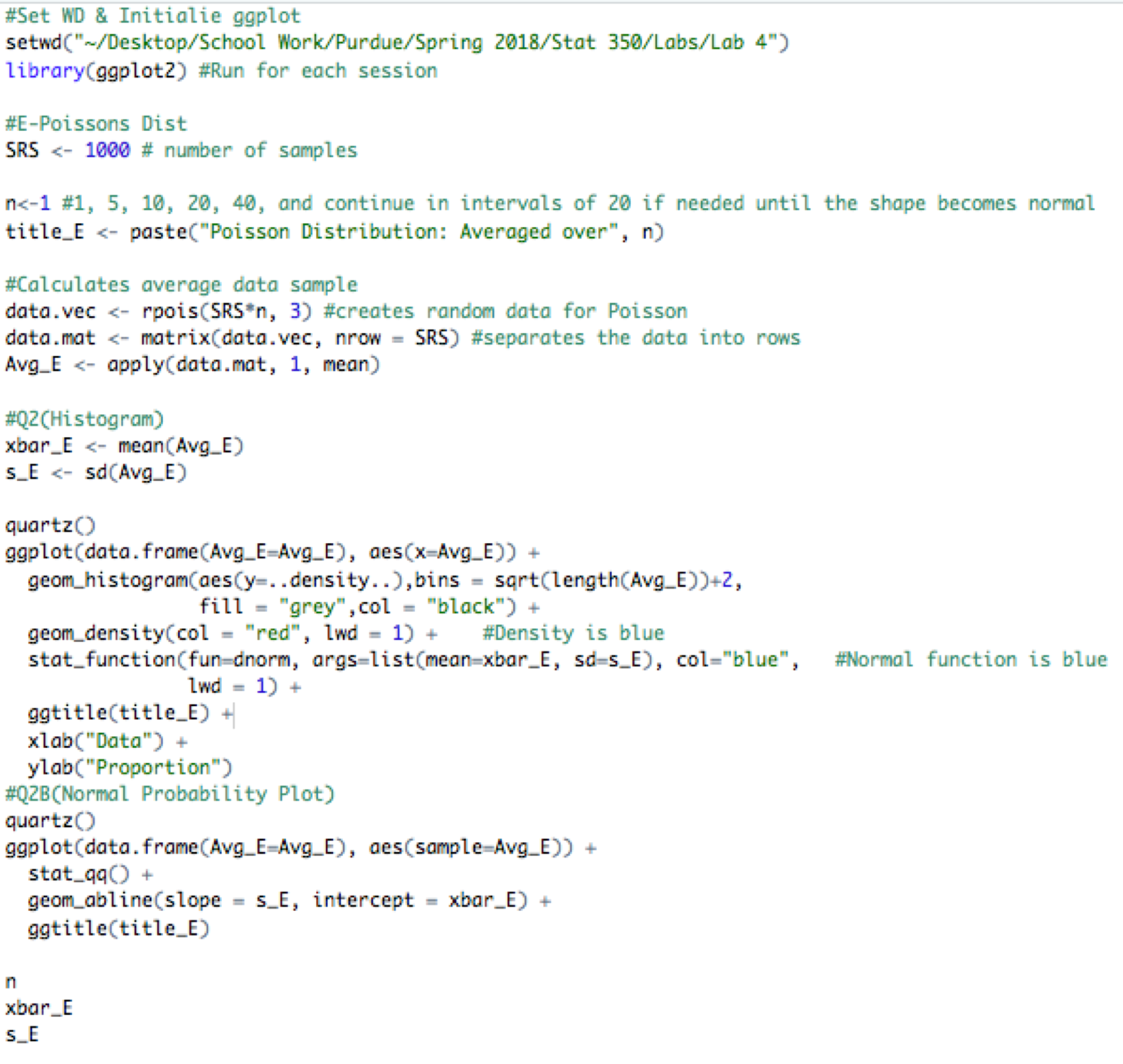
Approximately normal

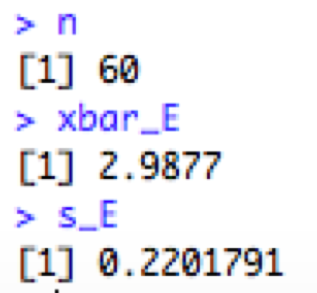
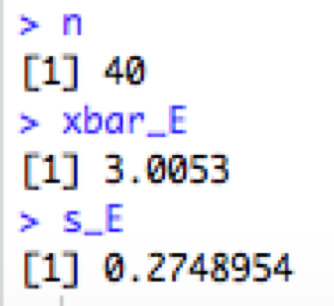
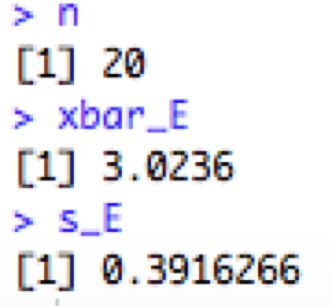
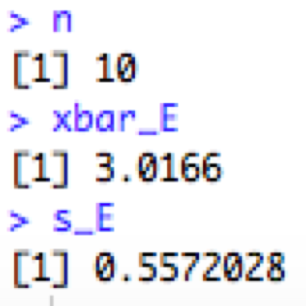
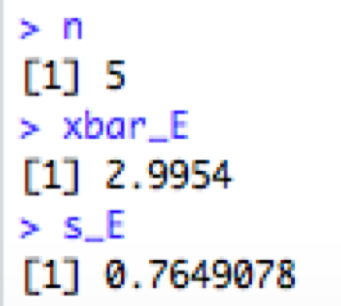
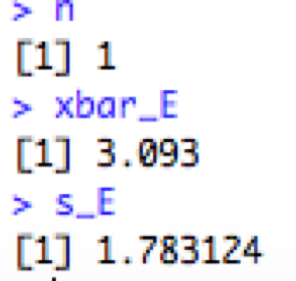
**3. Summary table**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***n*** | **experimental mean of 1000**  **(from output)** | **theoretical mean**  **(given)** | **experimental standard deviation of 1000**  **(from output)** | **theoretical standard deviation**  **(given and equation)** |
| 1 | 1.4934 | 1.5 | 0.863 | sqrt(3)/2  = 0.8660 |
| 5 | 1.4946 | 1.5 | 0.383 | 0.8660/sqrt(5)  = 0.3873 |
| 10 | 1.5076 | 1.5 | 0.2721 | 0.8660/sqrt(10)  = 0.2739 |
| 20 | 1.4951 | 1.5 | 0.1899 | 0.8660/sqrt(20)  = 0.1936 |
| 40 | 1.5027 | 1.5 | 0.1354 | 0.8660/sqrt(40)  = 0.1369 |
| 60 | 1.5018 | 1.5 | 0.1169 | 0.8660/sqrt(60)  = 0.1118 |

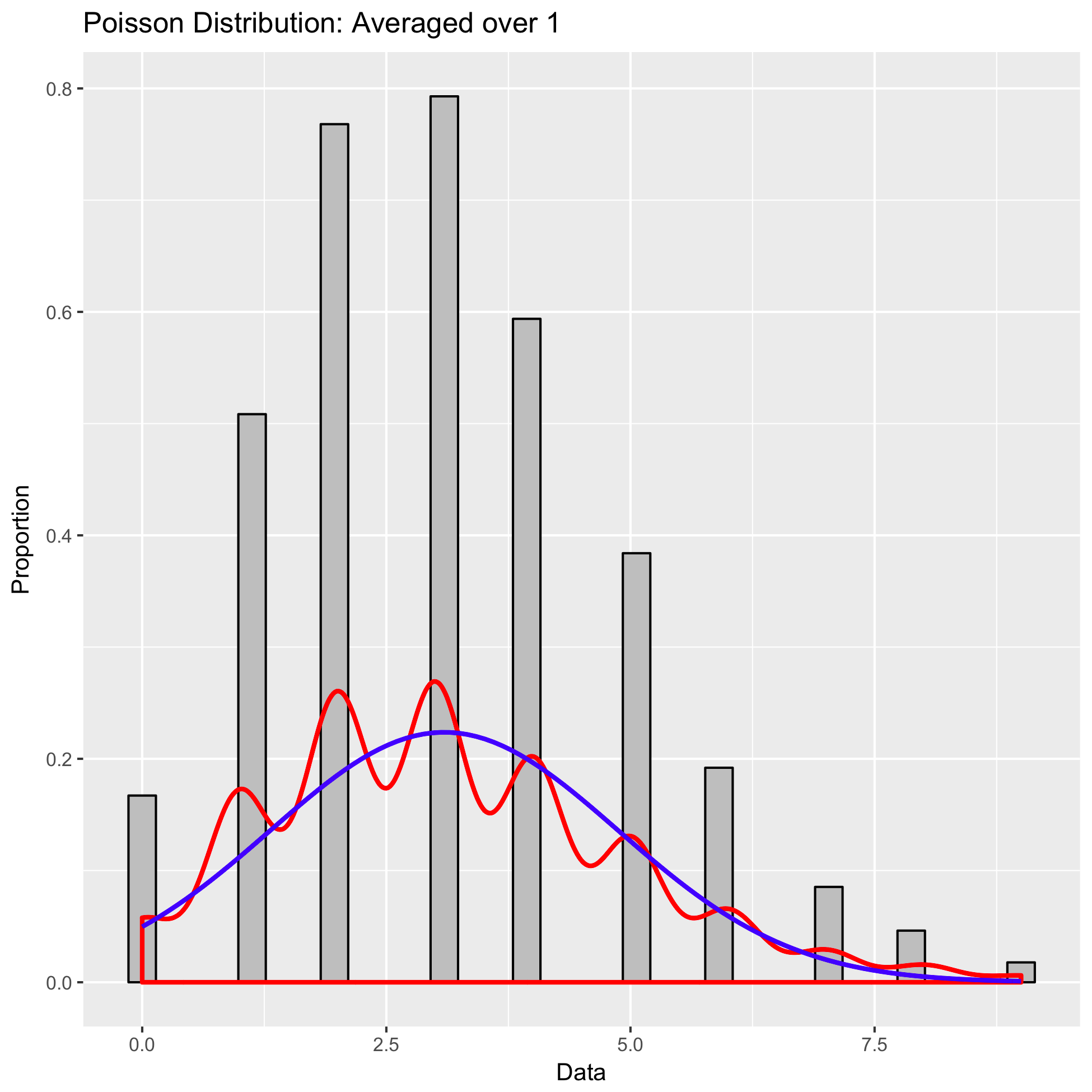
**E. (20 points) Poisson distribution**

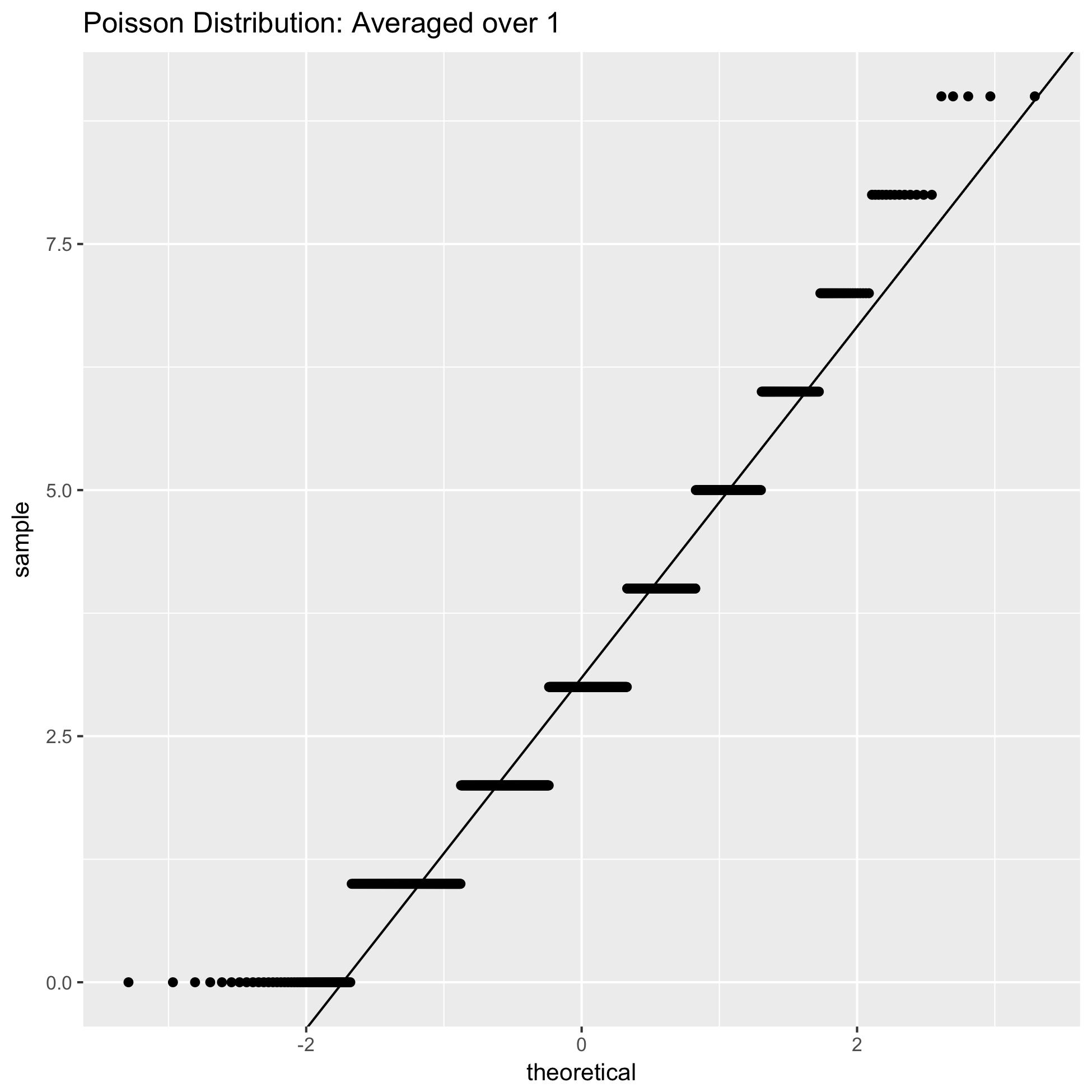
**1.**





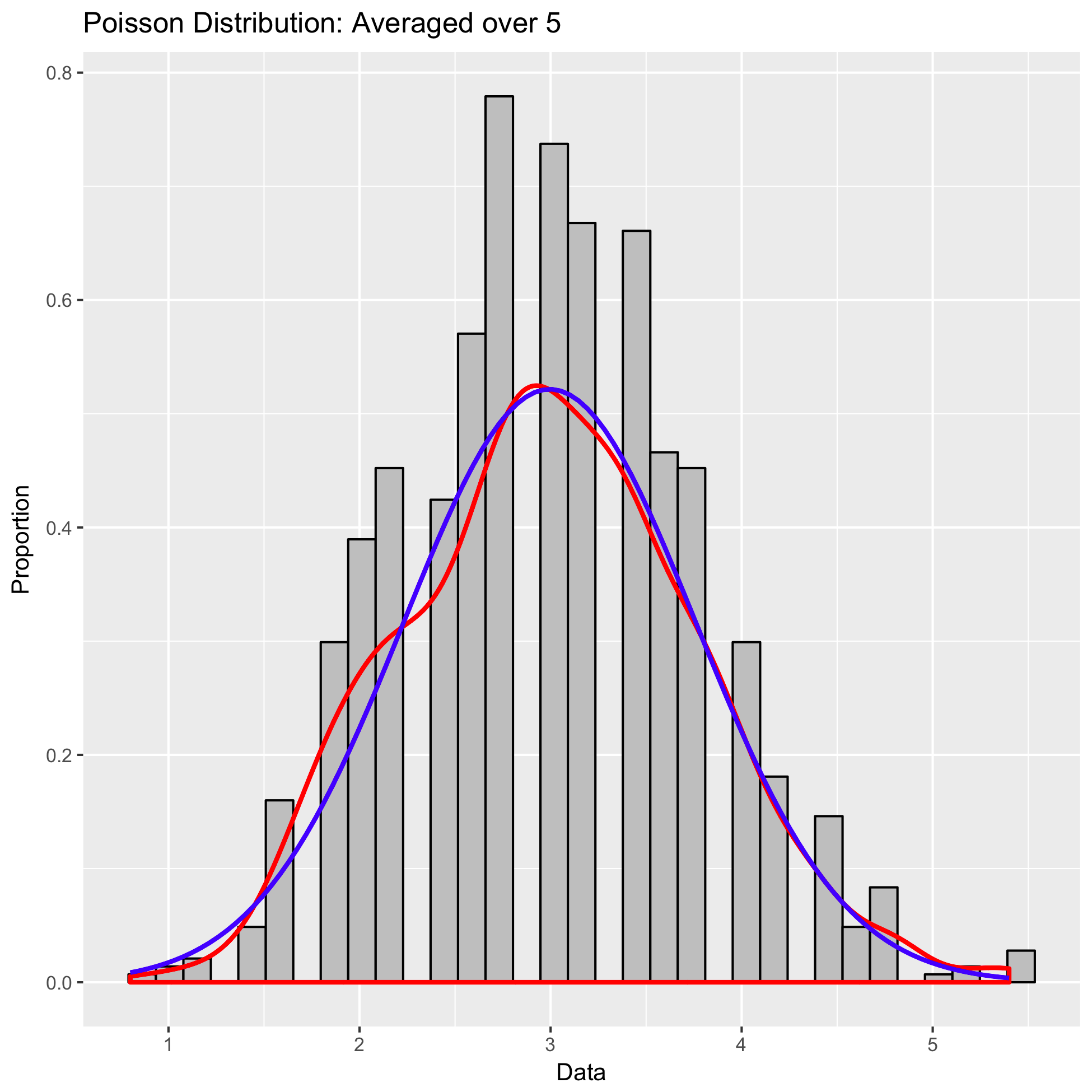
n=1

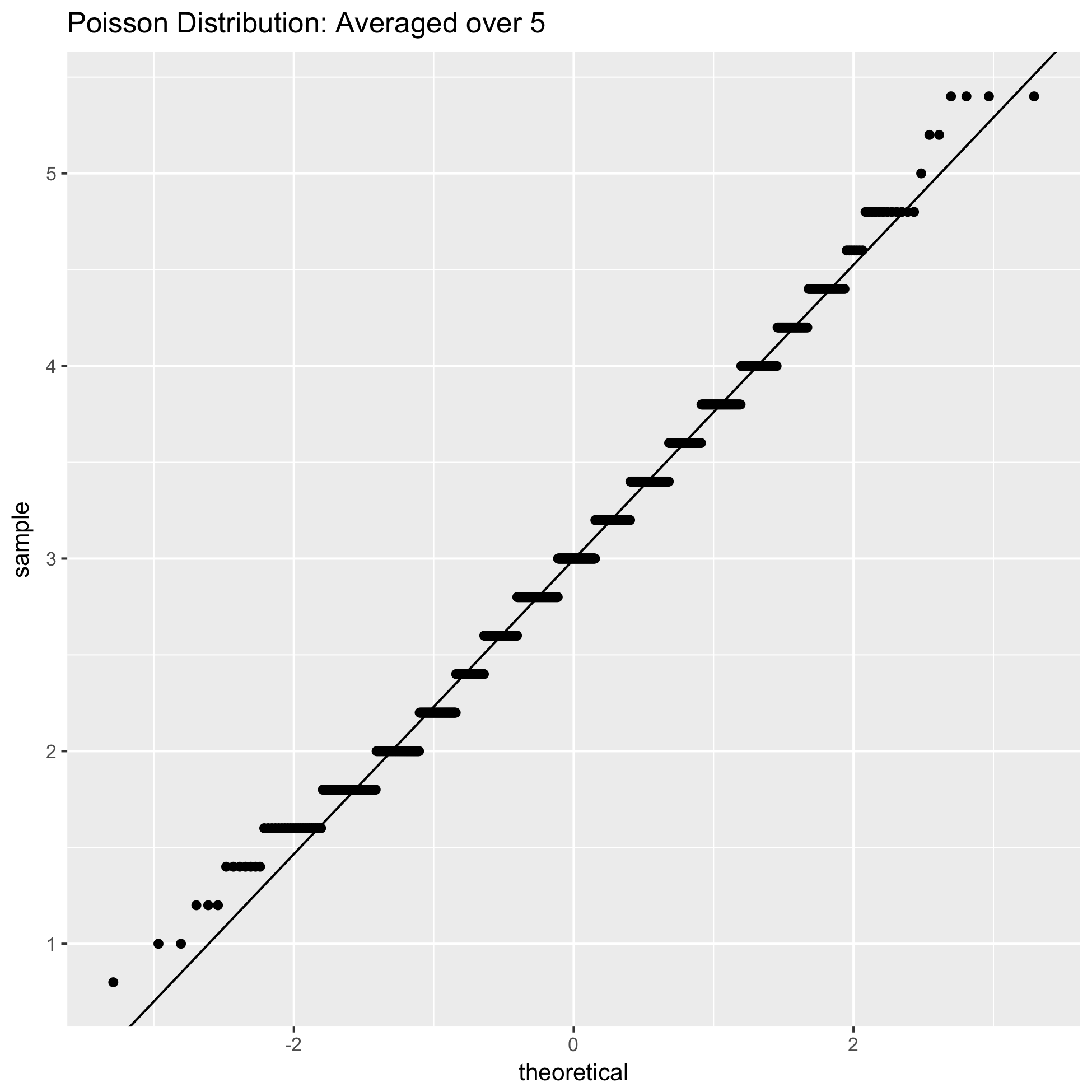




Not sufficiently normal, the QQ plot is very abnormal

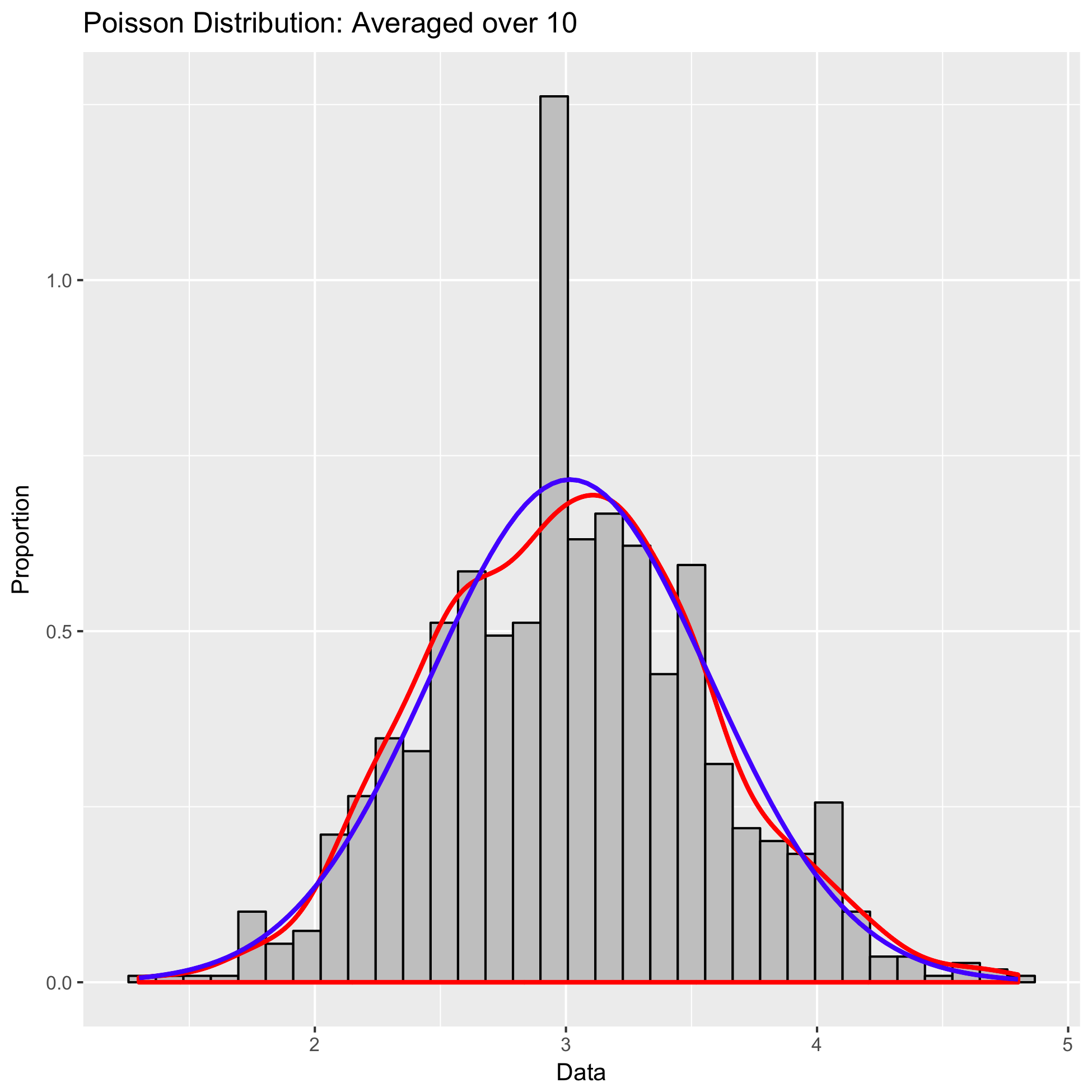
n=5

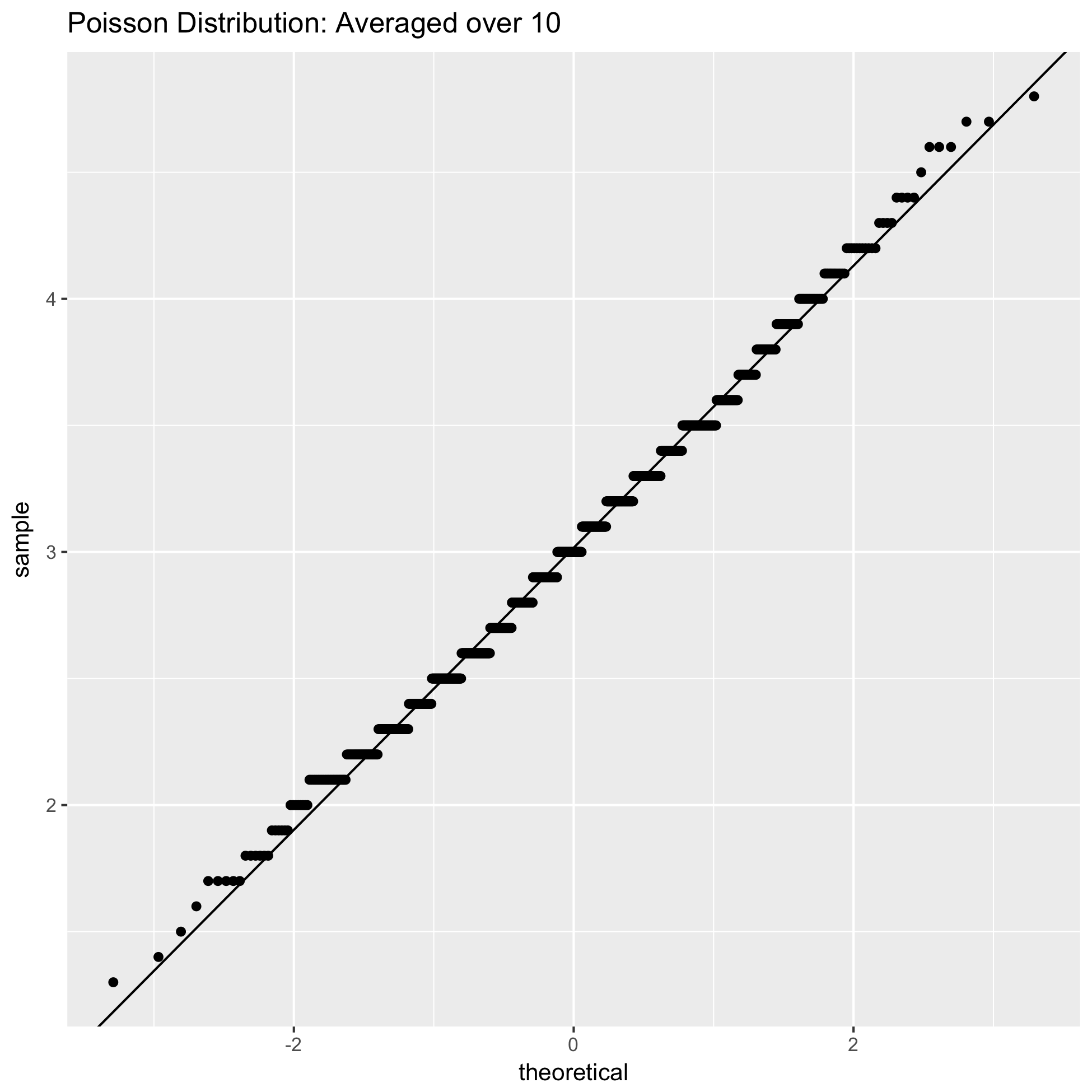




Not sufficiently normal, the red and blue lines are close but the histogram has spaces, the QQ plot is abnormal it has many horizontal steps.

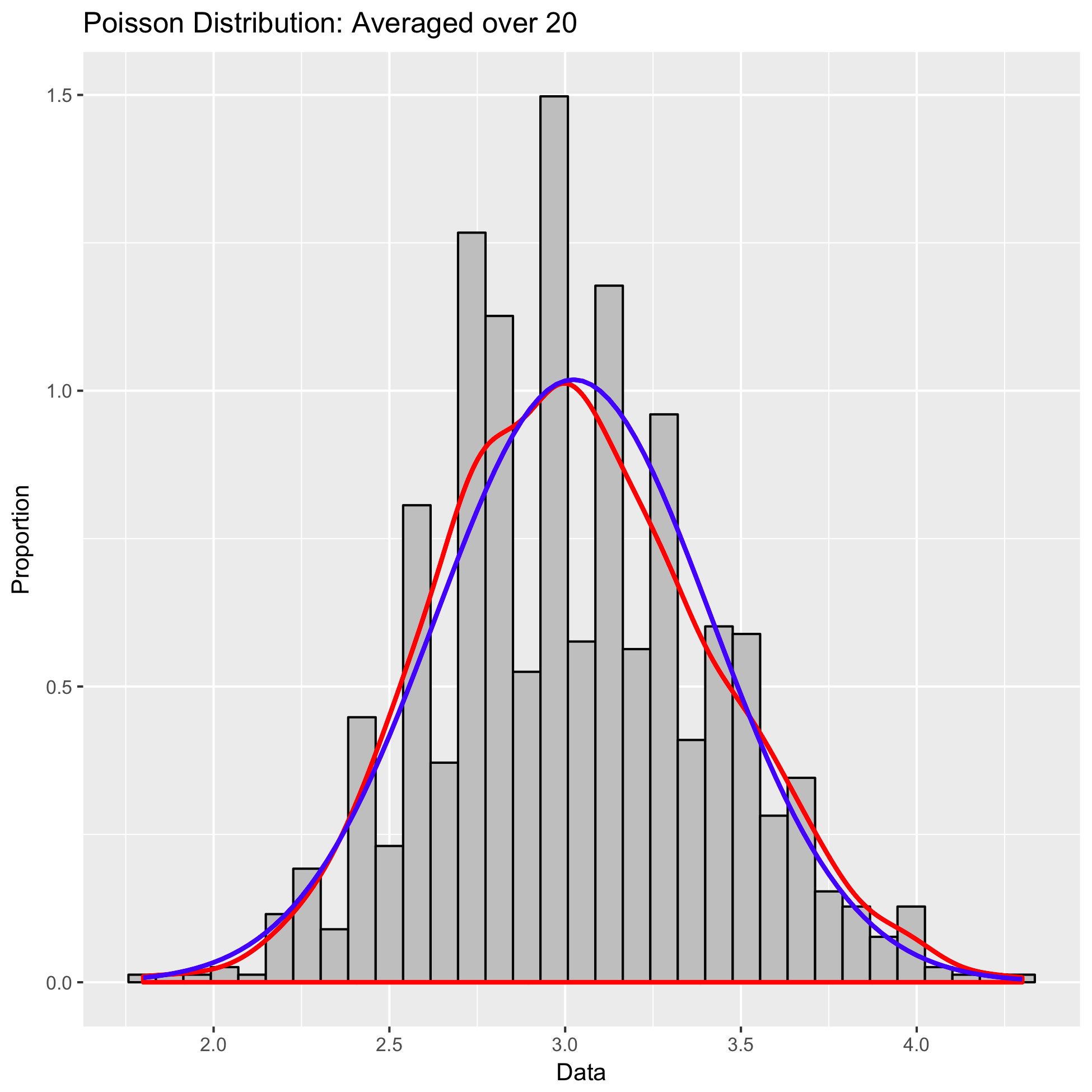
n=10

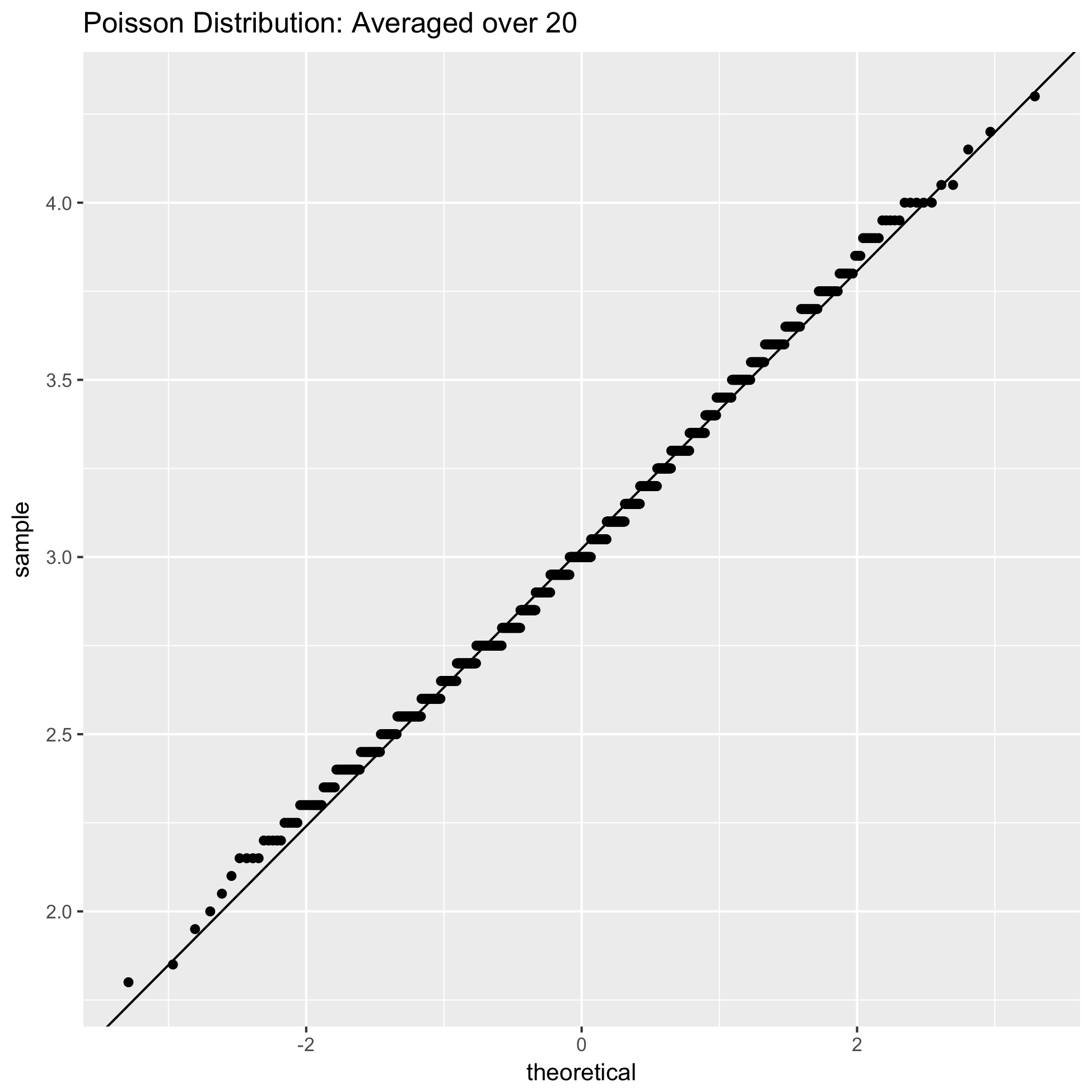




Not sufficiently normal, the red and blue lines are close but the histogram has one abnormally large bar in the middle, the QQ plot is very abnormal it has many horizontal steps which are shorter than before.

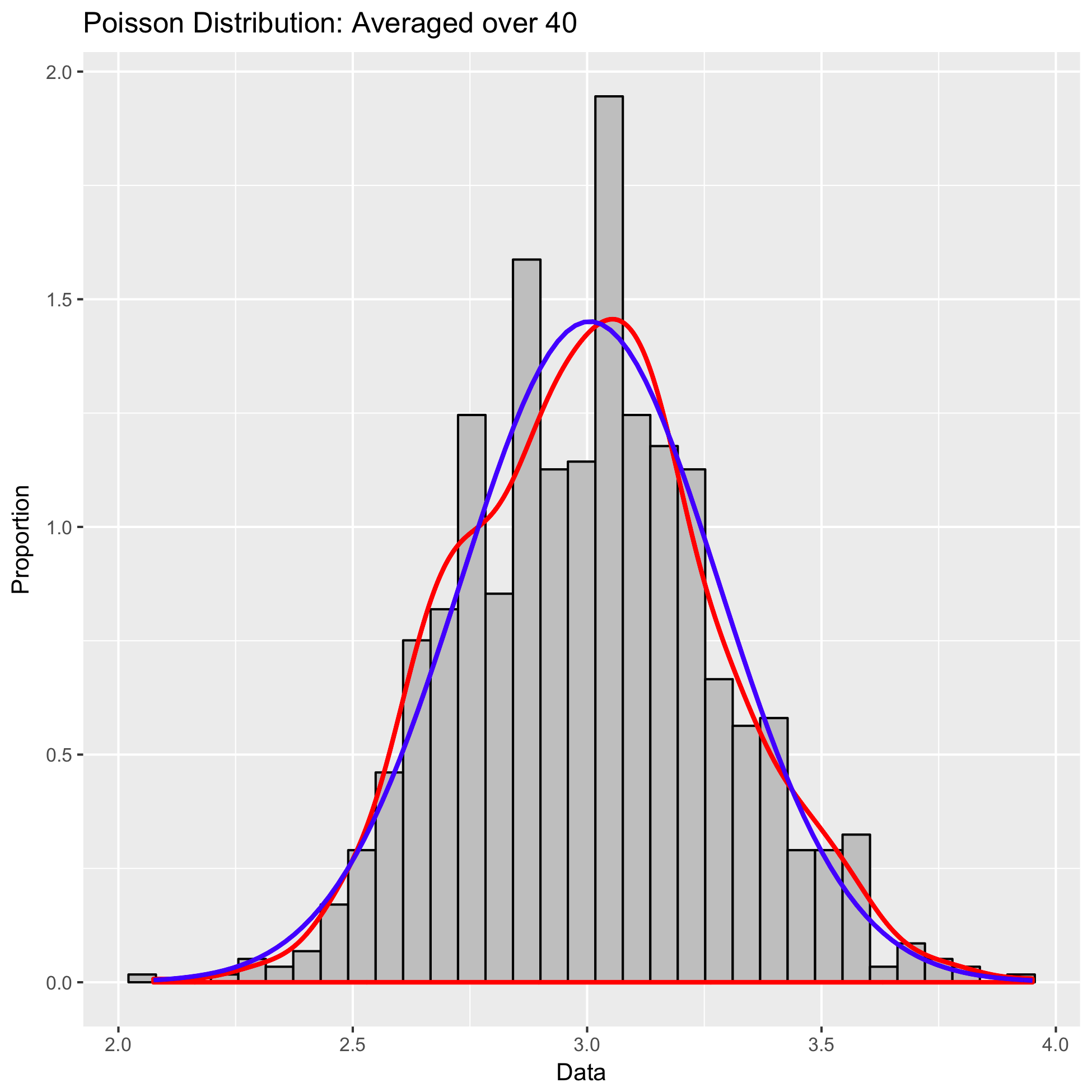
n=20

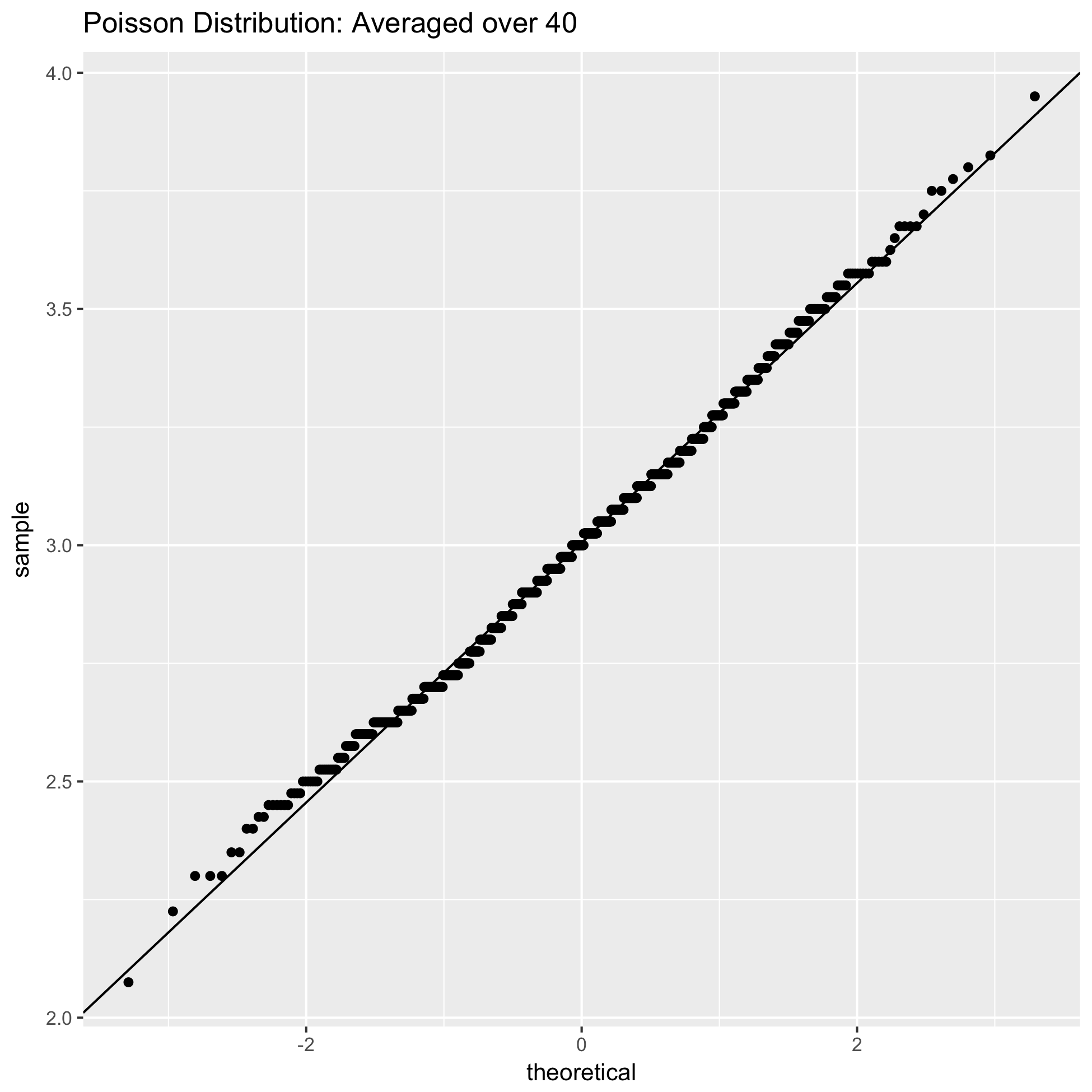




Not sufficiently normal, the red and blue lines are close but the histogram has alternating high and low bars, the QQ plot is very abnormal it has many horizontal steps with the data points more separated on either side of the line.

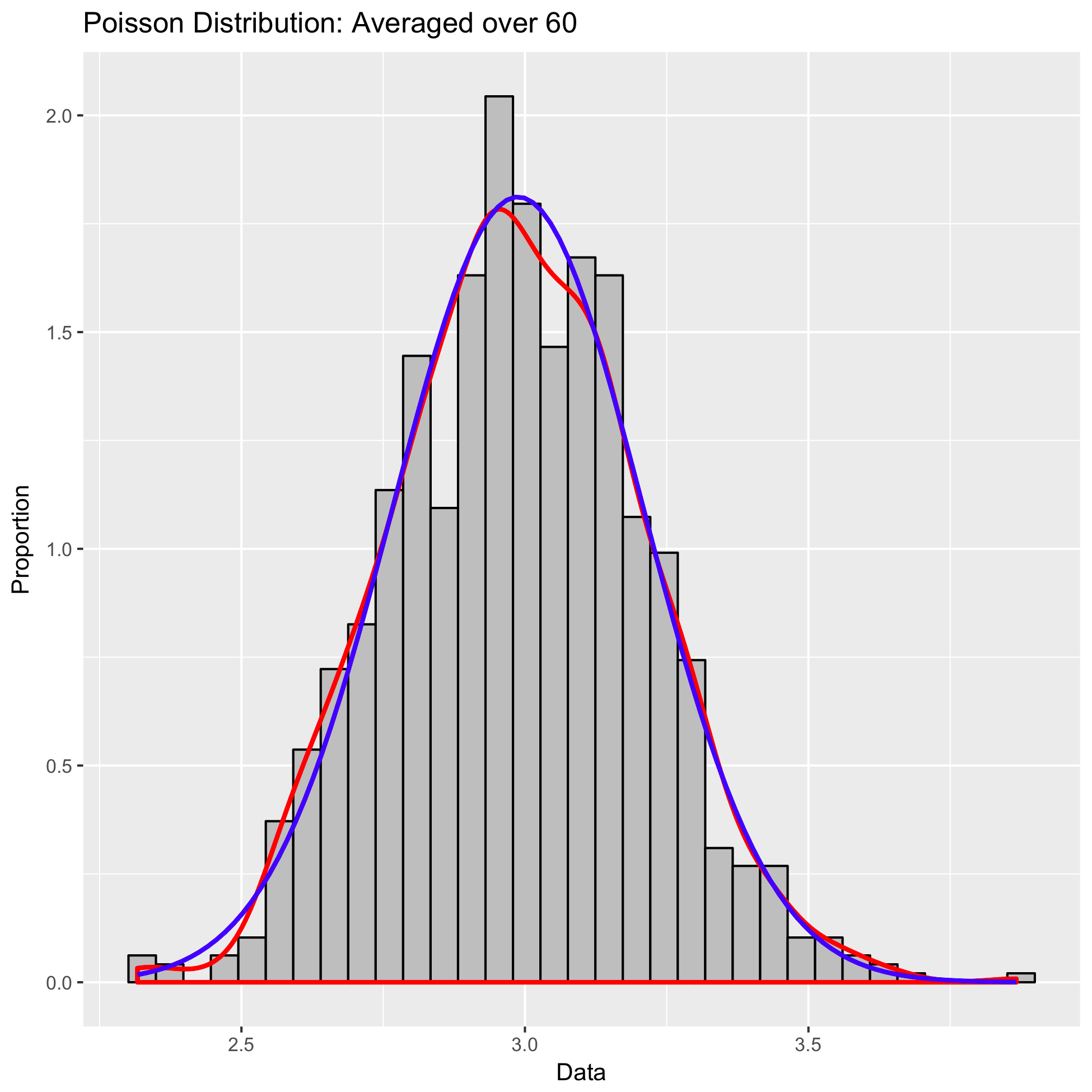
n=40

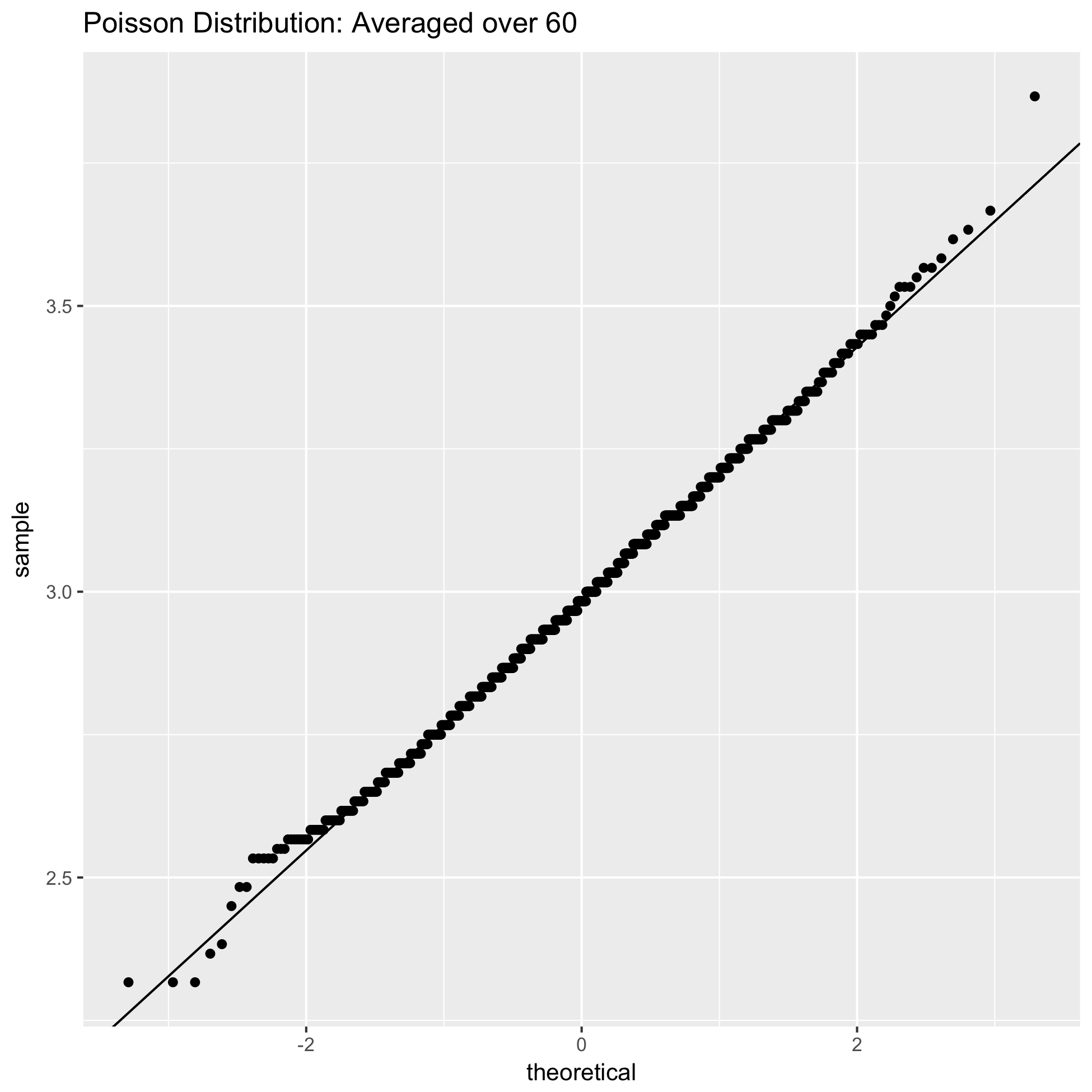




Both the QQ plot and histogram look normal, with a slight variation from normal in the histogram and some spaces between horizontal steps still in the QQ plot.

n=60





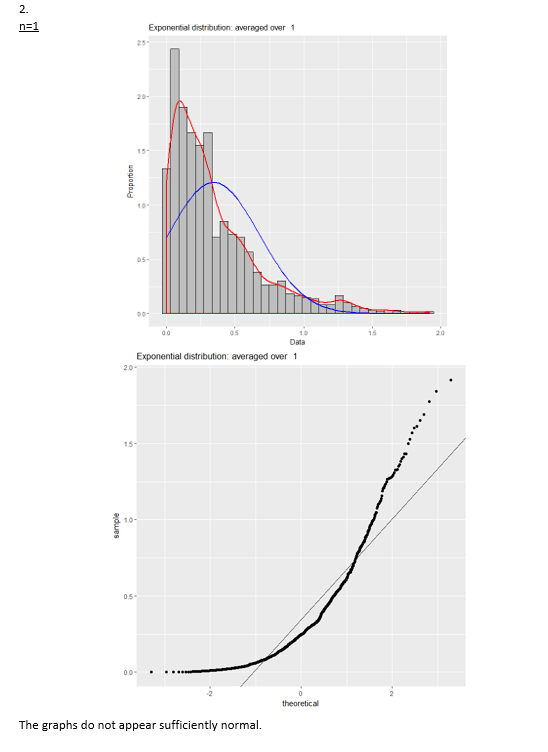
Both the QQ plot and histogram look normal

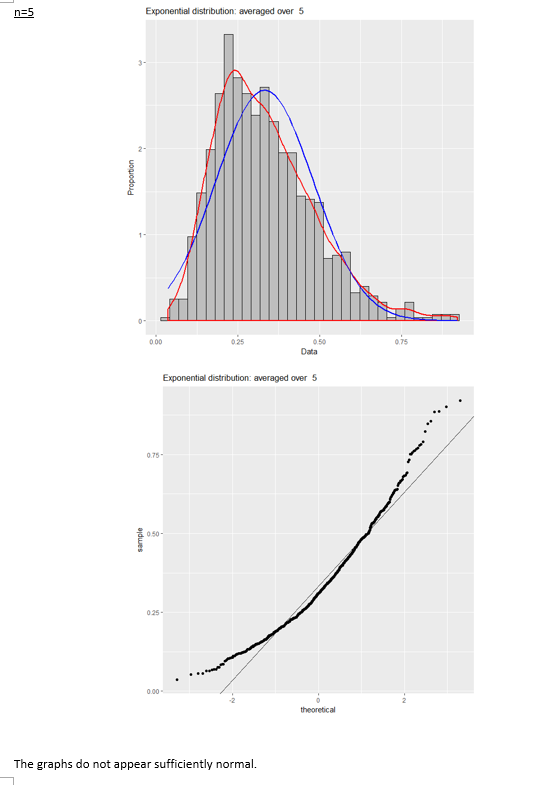
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n | experimental mean of your 1000 (from output) | theoretical mean (Equations 1) | experimental standard deviation of your 1000 (from output) | theoretical standard deviation (Equations 1) |
| 1 | 3.093 | 3 | 1.783124 | 1.732 |
| 5 | 2.9954 | ==3 | 0.7649078 | ==0.7746 |
| 10 | 3.0166 | 3 | 0.5572028 | 0.5477 |
| 20 | 3.0236 | 3 | 0.3916266 | 0.3873 |
| 40 | 3.0053 | 3 | 0.2748954 | 0.2739 |
| 60 | 2.9877 | 3 | 0.2201791 | 0.2236 |

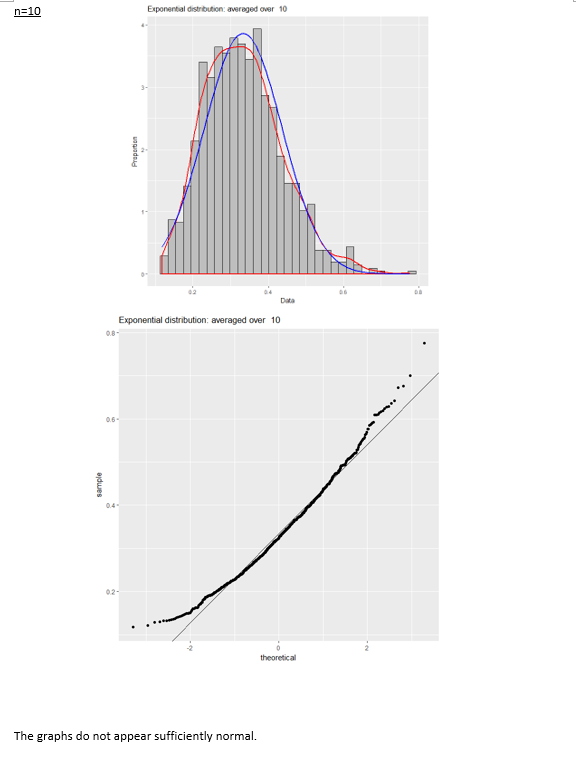
For Poisson Distribution:

**F. (BONUS: 20 points) Exponential distribution**

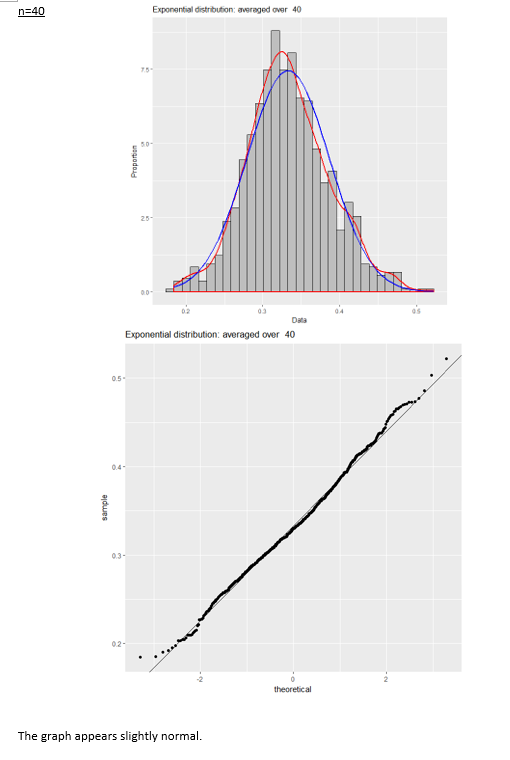
****

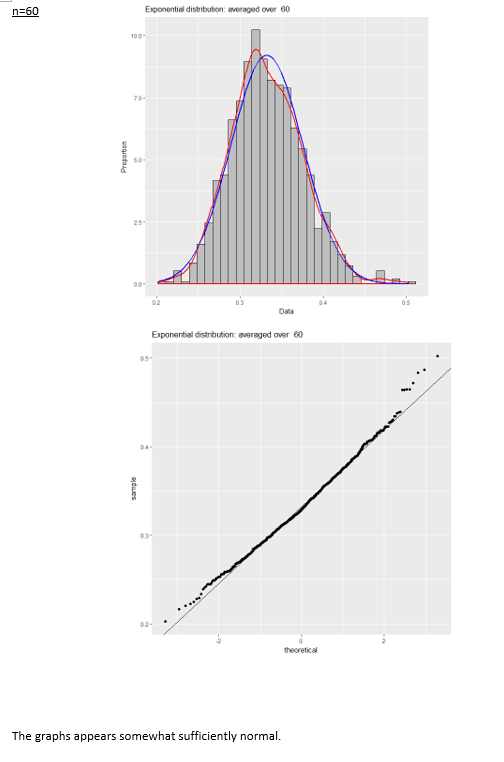
****

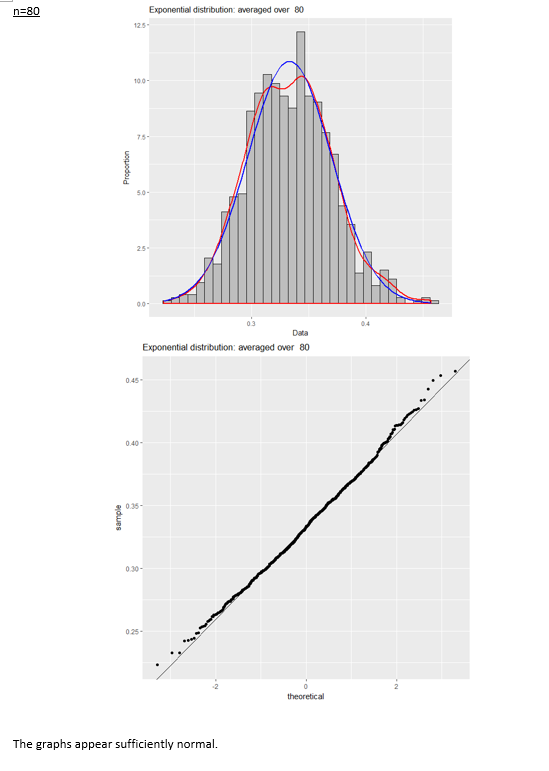
****

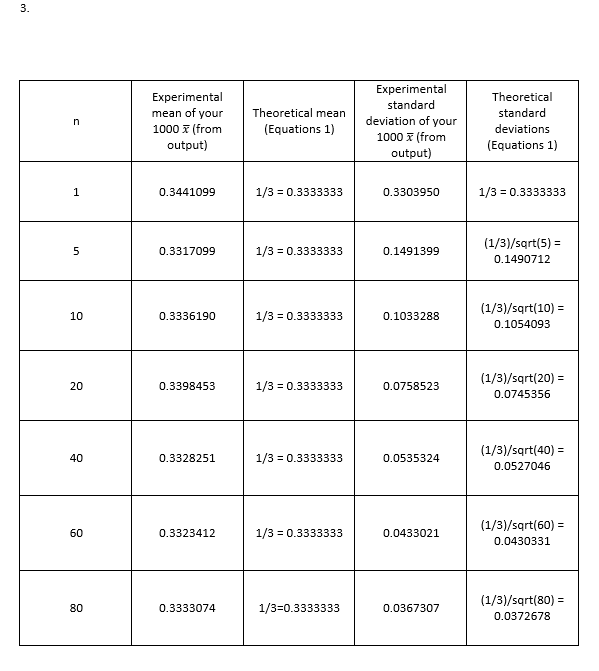
****

****

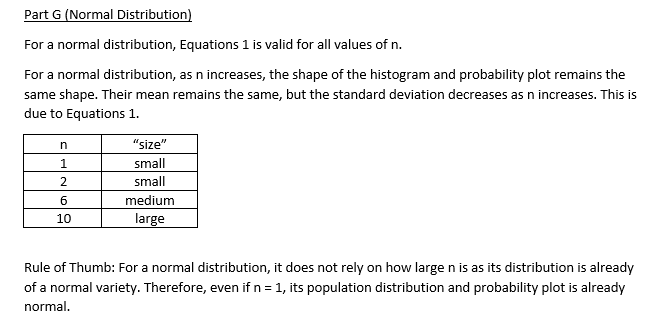
****

****

****

****

**G. (10 points) Concluding remarks**

****

Uniform distribution:

1.Equations are valid.

2. As n increases, the histogram and probability plot gain more normality. The mean remains the same, standard deviation decreases as n increases.

3.N = 15 is considered large.

|  |  |
| --- | --- |
| n | “size” |
| 1 | Small |
| 2 | Small |
| 10 | Medium |
| 15 | large |

Rule of thumb: An n value of 15 or above should be considered “large enough” for X bar to become approximately normal.

G (Gamma distribution)

1. Equations 1 appear to be valid for all values of *n*, with experimental and theoretical values being approximately equal.
2. As *n* increases, the distribution of the sample mean becomes closer to normal. Numerically, the standard deviation decreases, while the mean remains relatively unchanged.
3. N = 60 can be considered large for gamma distribution.

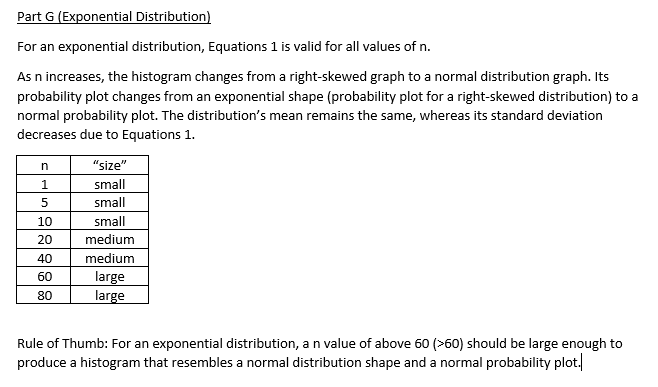
|  |  |
| --- | --- |
| n | “Size” |
| 1 | small |
| 5 | small |
| 10 | medium |
| 20 | medium |
| 40 | almost large |
| 60 | large |

4. As a rule of thumb, N should be larger than approximately 60 before we can say the sample mean will have an approximately normal distribution for a population with a gamma distribution.

E.Poisson Distribution:

1. Equation 1 is valid for all values of n for poisson distribution.
2. For Poisson's distribution as the value of n increases, the histogram and probability plot become more normal. The mean remains around a value of 3 for all values of n, while the standard deviation decreases from 1.78 to 0.22 as the value of n increases.
3. n=60 is considered large, the QQ plots are abnormal for all values of n below this.

|  |  |
| --- | --- |
| n | “Size” |
| 1 | small |
| 5 | medium |
| 10 | medium |
| 20 | medium |
| 40 | almost large enough |
| 60 | large |

1. Rule of Thumb: An n value of 60 is large enough for the distribution to becomes approximately normal producing a histogram and probability plot resembling those for the normal distribution. 

Overall

|  |  |
| --- | --- |
| **Distribution** | **Value of *n* for approximately normal sample mean distribution** |
| Standard normal | 1 |
| Uniform | 15 |
| Gamma | 60 |
| Poisson | 60 |
| Exponential | 60 |

As a general rule of thumb: the closer a population distribution is to a normal distribution, the higher *n* should be in order to ensure that the sample mean distribution becomes approximately normal. As *n* increases, every population distribution will have a sample mean distribution approaching normality; some population distributions will simply achieve approximately normal sample mean distributions faster than others.