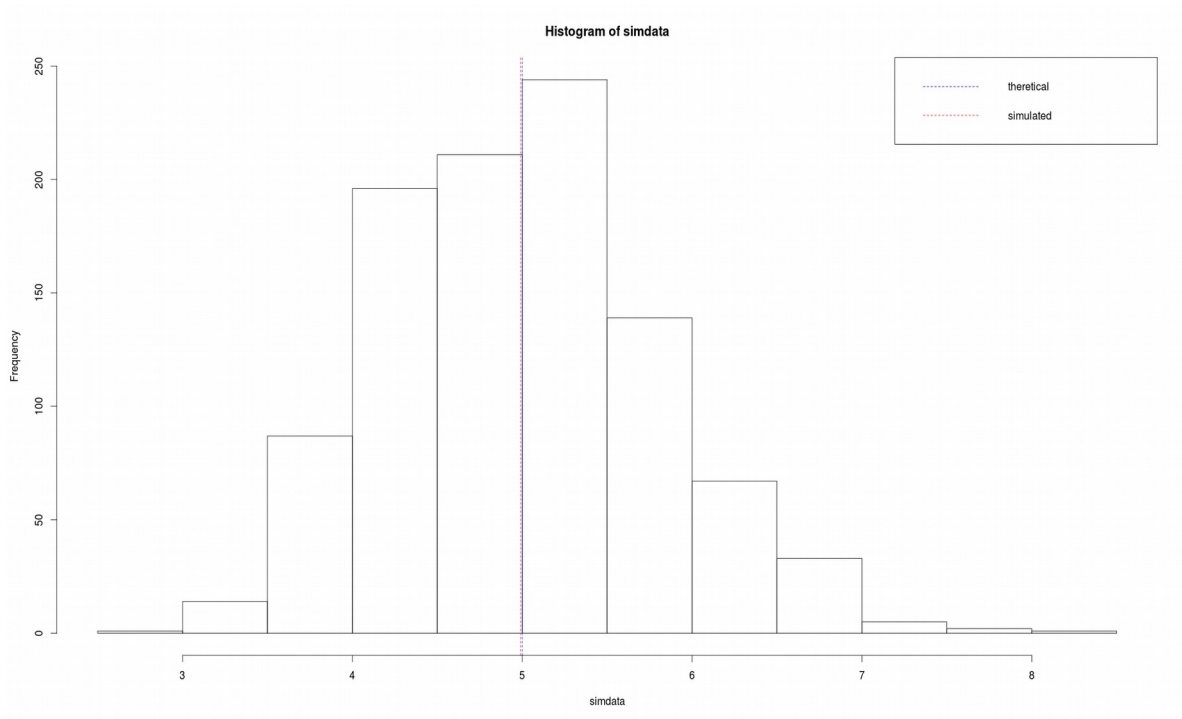
Coursera: Statistical Inference
Project
Q1

The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.
Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should
1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.
Note that for point 3, focus on the difference between the distribution of a large collection of random exponentials and the distribution of a large collection of averages of 40 exponentials.

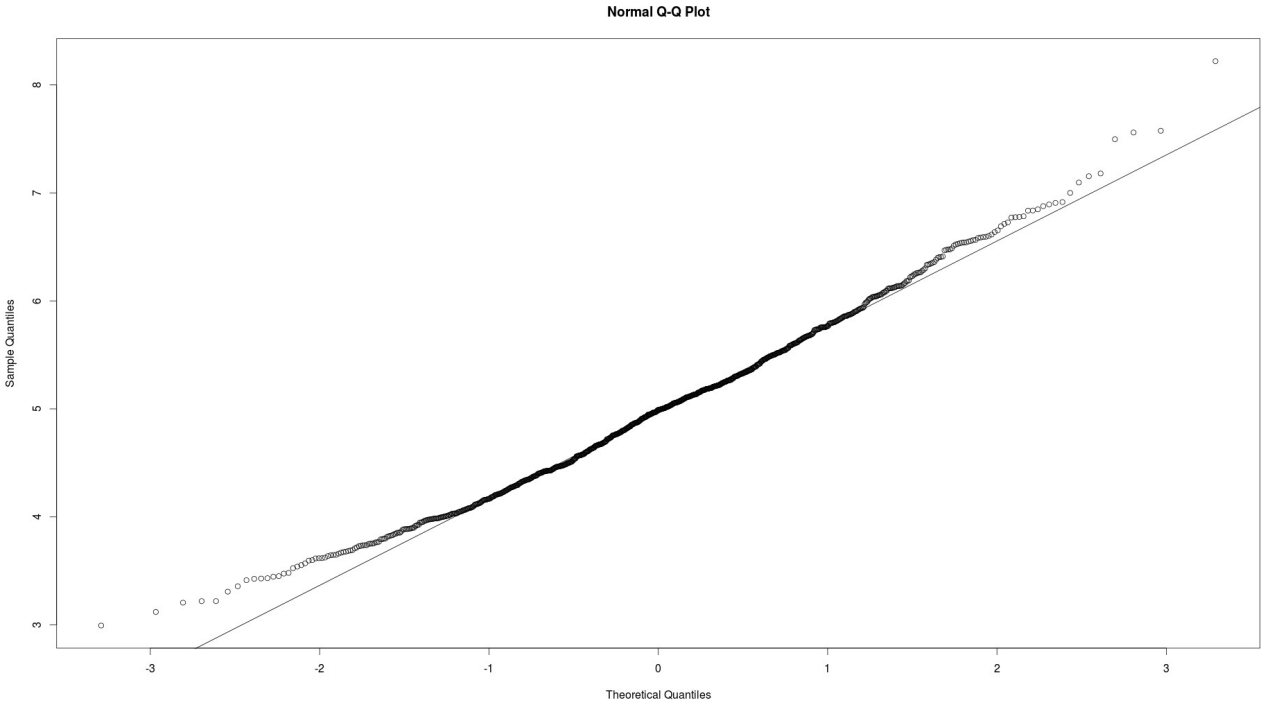
1.
```{r}
set.seed(0)
simdata<-NULL
for ( i in 1:1000){
  simdata<-c(simdata,mean(rexp(40,0.2)))
}
hist(simdata)
abline(v=5,lty=2,col='blue')
abline(v=mean(simdata),lty=2,col='red')
legend('topright',c('theretical','simulated'),lty=2,col=c('blue','red'))
```



The theoretical center is 5, the simulated one is 4.989678. They are close.

2.
The theoretical standard deviation is 5/sqrt(40)=0.7905694, the simulated one is 0.7862304. They are close.

3.
```{r}
qqnorm(simdata)
qqline(simdata)
```



QQ plot does show it's approximately normal.

The distribution of large collection of random exponentials is still exponential while the distribution of a large collection of averages of 40 exponentials is approximately normal.