StatInfProject

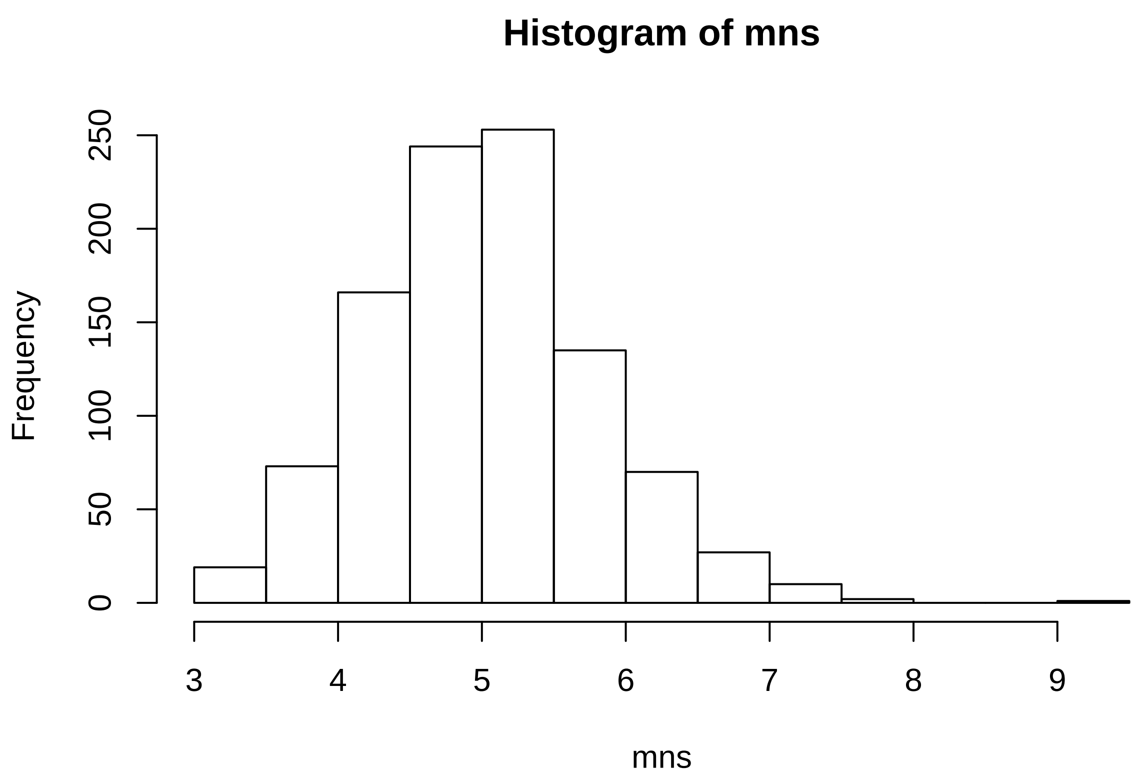
Load libraries:

```r
library(knitr)
```

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.

We will investigate the behavior of the exponential distribution with rate lambda = 0.2. The first thing we will consider is the mean of 40 exponential random variables with rate 0.2, i.e. 40 draws from a exp(0.2). Theoretically, a random variable with this distribution has expected value, or mean, 1/0.2 = 5 and variance = 5^2 = 25. We will now consider the distribution of the mean of 40 independent exp(0.2) random variables. To get an idea of the distribution, we will take 1000 samples of size 40, calculating the mean of each sample, and plotting a histogram of the 1,000 mean values:

```r
mns <- NULL
for (i in 1:1000) mns <- c(mns,mean(rexp(40,0.2)))
hist(mns)
```



The center of the distribution appears to be about 5.

```r
mean(mns)
```

```
## [1] 5.03138
```

As mentioned earlier, the theoretical center of the mean is 1/0.2 = 5. Clearly, the simulation shows the distribution of the mean of 40 exp(0.2) is indeed centered at 5.

2. Show how variable it is and compare it to the theoretical variance of the distribution

The theoretical variance of the mean is (note the variance of one exp(.2) is 25 and these are i.i.d): Var((X_1 + … + X_40)/40) = (40x25)/40x40 = 0.625.

Using the means of 10,000 samples from above, we can calculate the variance:

```
var(mns)
```

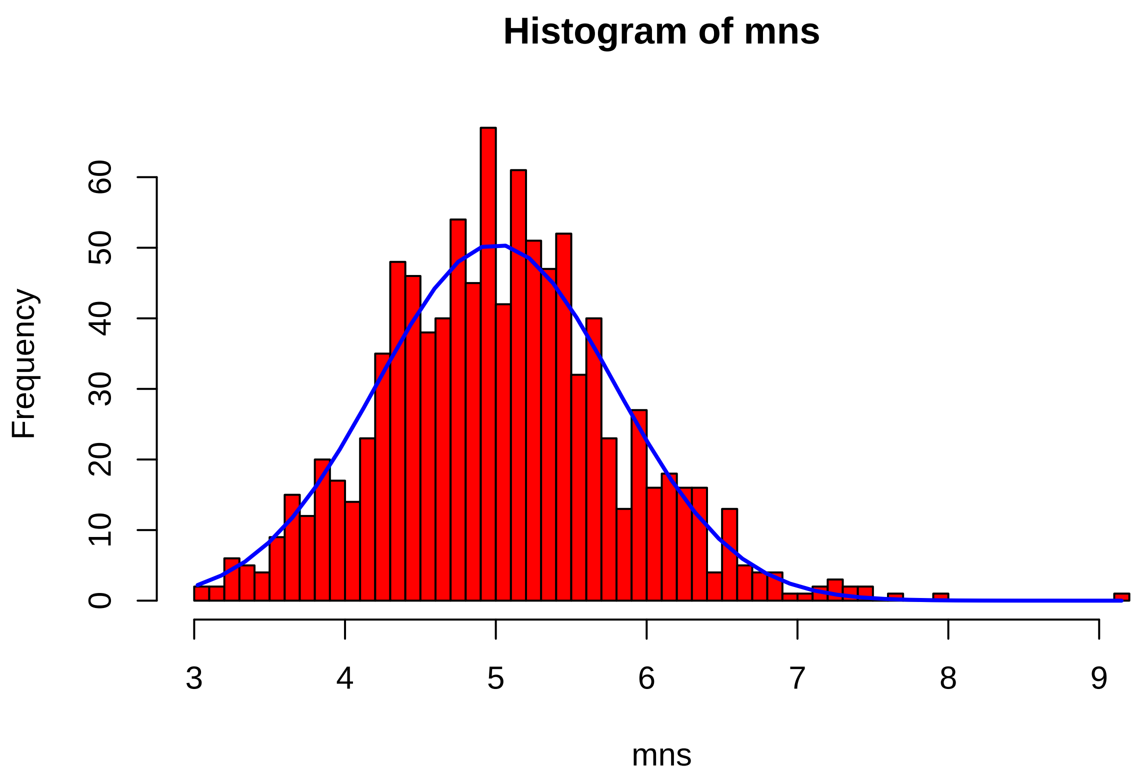
```
## [1] 0.6230529
```

We can see this is very close to the theoretical variance.

3. Show that the distribution is approximately normal

The Central Limit Theorem tells us that the distribution of the mean of 40 i.i.d exp(0.2) random variables converges in distribution to a normal distribution with mean 5 and variance 25/40 = 0.625. Let us again plot the histogram of the 10,000 samples of 40 exp(0.2), with more bins this time, and overlay a plot of a normal distribution with mean 5 and variance 0.625:

```
h <- hist(mns, breaks=50, col="red")
xfit<-seq(min(mns),max(mns),length=40)
yfit<-dnorm(xfit,mean=mean(5),sd=sqrt(0.625))
yfit <- yfit*diff(h$mids[1:2])*length(mns)
lines(xfit, yfit, col="blue", lwd=2)
```



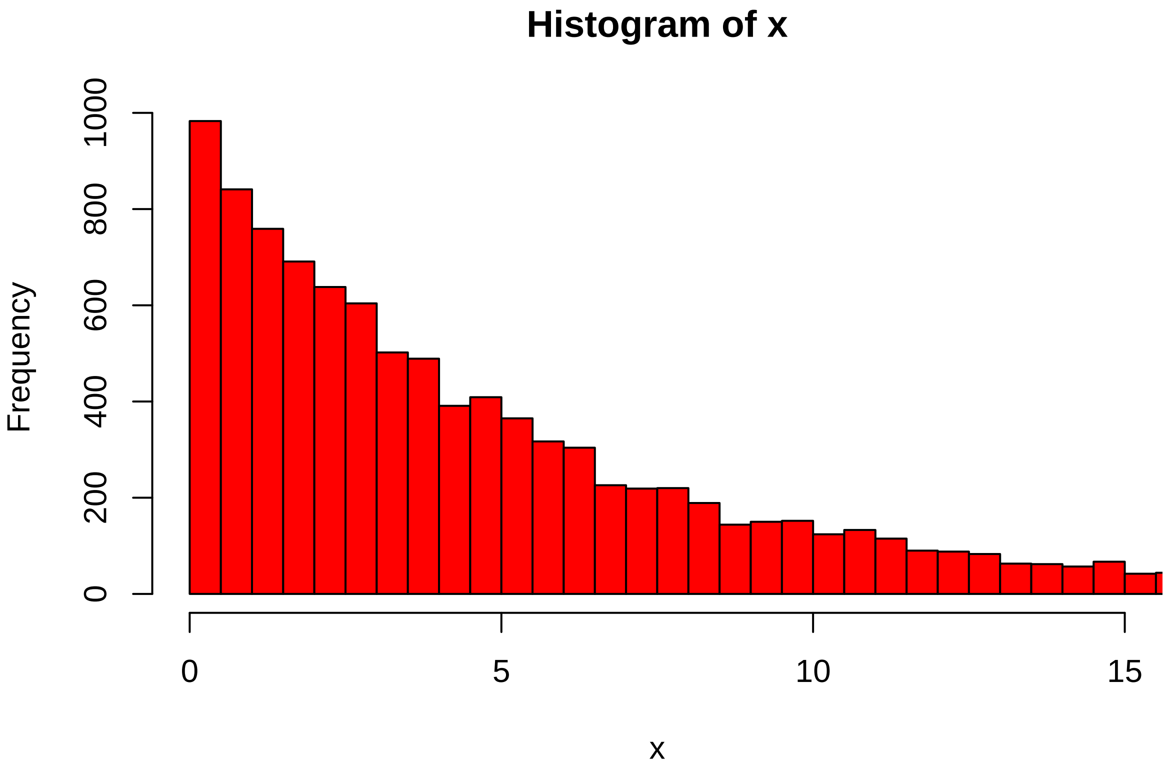
As can be seen, the distribution of the mean is approximately normal with mean 5 and variance 0.625.

Finally, we compare the above distribution to a large collection of random exponential with rate 0.2:

```
x <- rexp(10000,rate=0.2)
mean(x)
```

```
## [1] 5.050163
```

```r
hist(x,xlim=c(0,15),breaks=100, col="red")
```



Clearly, this is not normal, but follows an exponential distribution with rate 0.2.