JHU Statistical Inference Course Course Project Part 1

Part 1: Simulation Exercise Instructions

The Instructions for part 1 of the project are as follows:

> In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. 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 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Following the above instructions, we are going to simulate 1000 exponential distributions each of 40 samples. We are then going to compare the simulated mean and variance to the expected theoretical mean and variance for the distribution.

Question 1

Define constants for the simulation:

```
set.seed(47) ##Set a seed value for reproducibility

t <- 1000 ##Number of trials to perform
n <- 40 ##Number of samples per trial
lambda<- 0.2 ##Value of lambda for simulations
```

The following code performs the simulations of the exponential distribution. The replicate function is used to perform the simulation using the rexp function $\boldsymbol{t}$ times with a sample size of $\boldsymbol{n}$ and lambda of $\boldsymbol{lambda}$. The mean of each of the $\boldsymbol{t}$ simulations is than calculated.

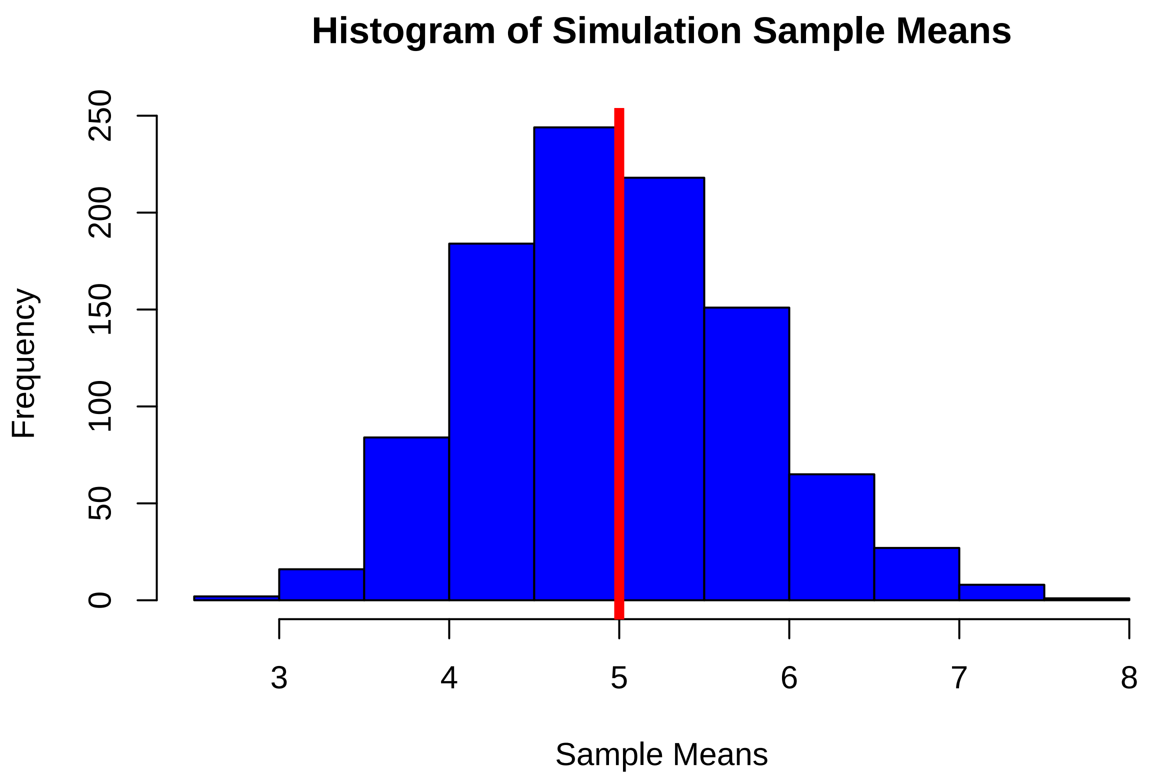
```
##simulations
simulations<-replicate(t, rexp(n,lambda))

##calculate mean
means<-colMeans(simulations)
```

Calculate the expected mean of the distribution:

```
expectedmean<-1/lambda
```

Plot out the histogram of sample means with the sample mean indicated with a red vertical line:

```
hist(means, main="Histogram of Simulation Sample Means", xlab="Sample Means", col="blue")
abline(v=expectedmean, col="red",lwd="5")
```



The theoretical expected mean is 5 and the mean from the simulation results is 4.9769381

Question 2

Calculate the variance of the sample means:

```
samplevariance<-(sd(means))^2
```

Calculate the expected variance of the distribution:

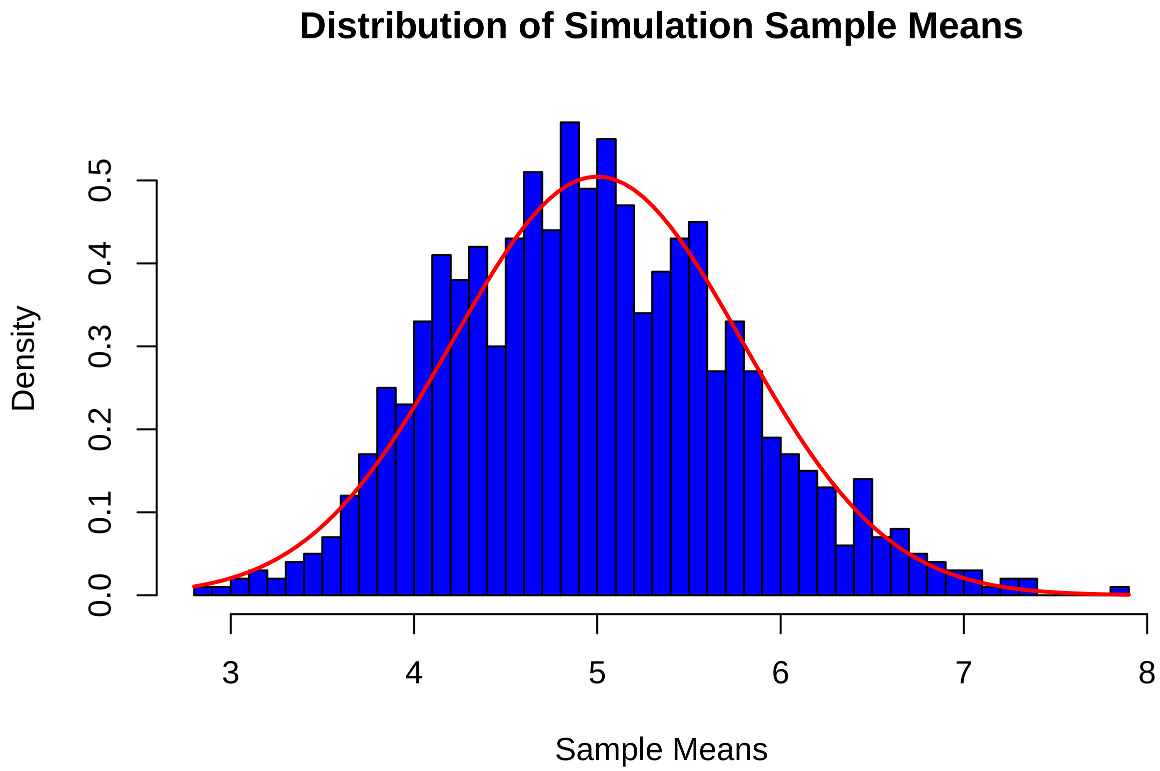
```
expectedvariance<-((1/lambda)*(1/sqrt(n)))^2
```

The theoretical expected variance is 0.625 and the variance from the simulation results is 0.6262234

Question 3

Below is a plot of the probability distribution of the simulated data overlayed with the theoretical expected normal curve for the results. If the simulated data is normal we would expect it to match the shape of the expected normal distribution.

```
#plot distribution
hist(means, freq = FALSE, breaks=1.5*n, main="Distribution of Simulation Sample Means",
     xlab="Sample Means", col="blue")

#Overlay Normal Curve
curve(dnorm(x, mean=expectedmean, sd=sqrt(expectedvariance)),
          col="red", lwd=2, add=TRUE, yaxt="n")
```



As shown above the simulated appears to be normal and to match the overlayed expected normal curve.