Statistical Inference - Course Project - Part 1

Introduction

This is the project for the statistical inference class. In this project we perform a simulation exercise using exponential distribution samples and then perform basic inferential data analysis.

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. In this project, we set lambda = 0.2 for all of the simulations. We also set the number of observations to 40.

Question 1

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

```
#Set the appropriate parameters for the simulation
n = 40
nosims = 1000
expLambda = 0.2

#Set seed for reprodcibility
set.seed(1)

#Simulate the exponential distributions
simulatedExpValues = replicate(nosims, rexp (n, expLambda))

#Calculate the individual means of the simulated exponential simulations
simulatedExponentialMeans = colMeans (simulatedExpValues)

#Calculate the mean of the sinulated means.
#This tells us where the simulation means are centered.
meanOfSimExpMeans = mean(simulatedExponentialMeans)
meanOfSimExpMeans
```

```
## [1] 4.990025
```

```
#Theoretical mean
meanTheoretical = 1/expLambda
meanTheoretical
```

```
## [1] 5
```

Based on the value above, we see that the distribution of the means of the simulated distributions is 4.9900252. The actual mean of the distribution is 5

Question 2

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

```r
#Calculate the variance of the simulated means
varianceSimExponMeans = var (simulatedExponentialMeans)
varianceSimExponMeans
```

```
## [1] 0.6111165
```

```r
#Calculate the theoretical variance
varianceTheoretical = ((1/expLambda)/sqrt (n))^2
varianceTheoretical
```

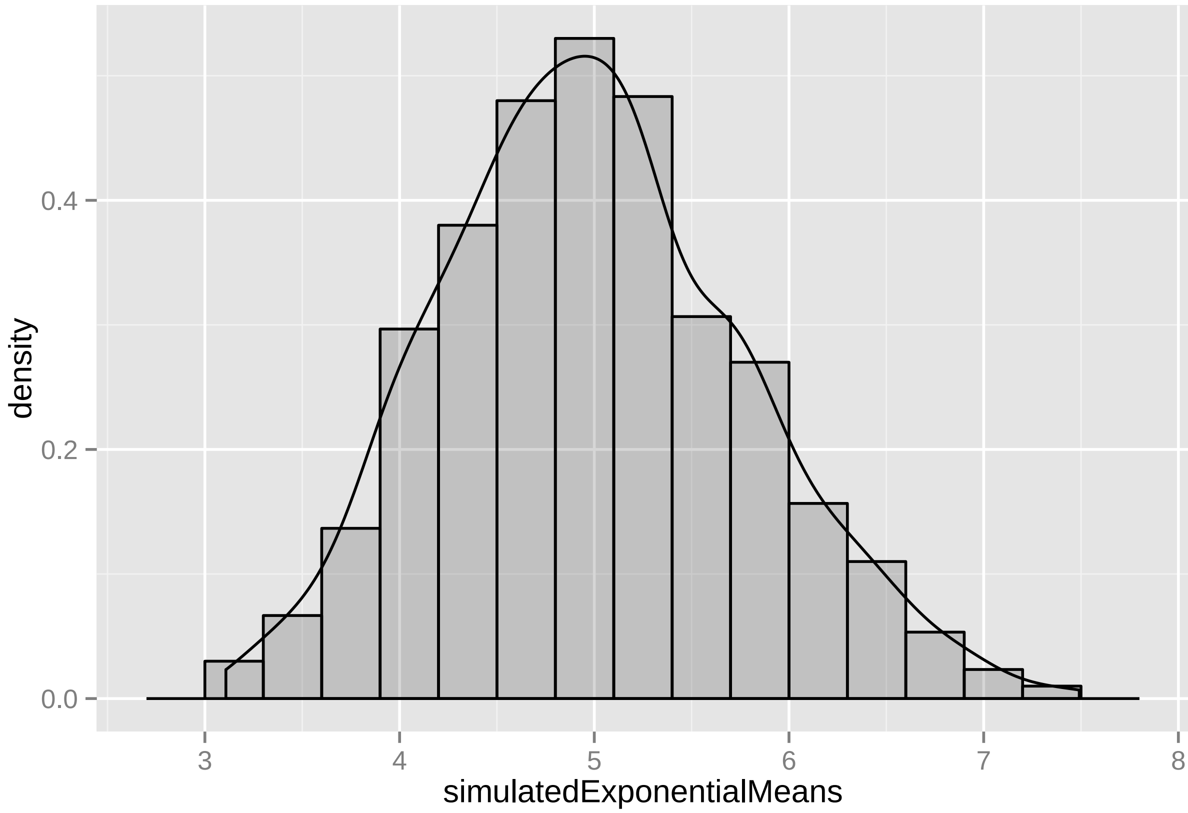
```
## [1] 0.625
```

Based on the value above, we see that the variance of the means of the simulated distributions is 0.6111165. The actual variance of the distribution is 0.625

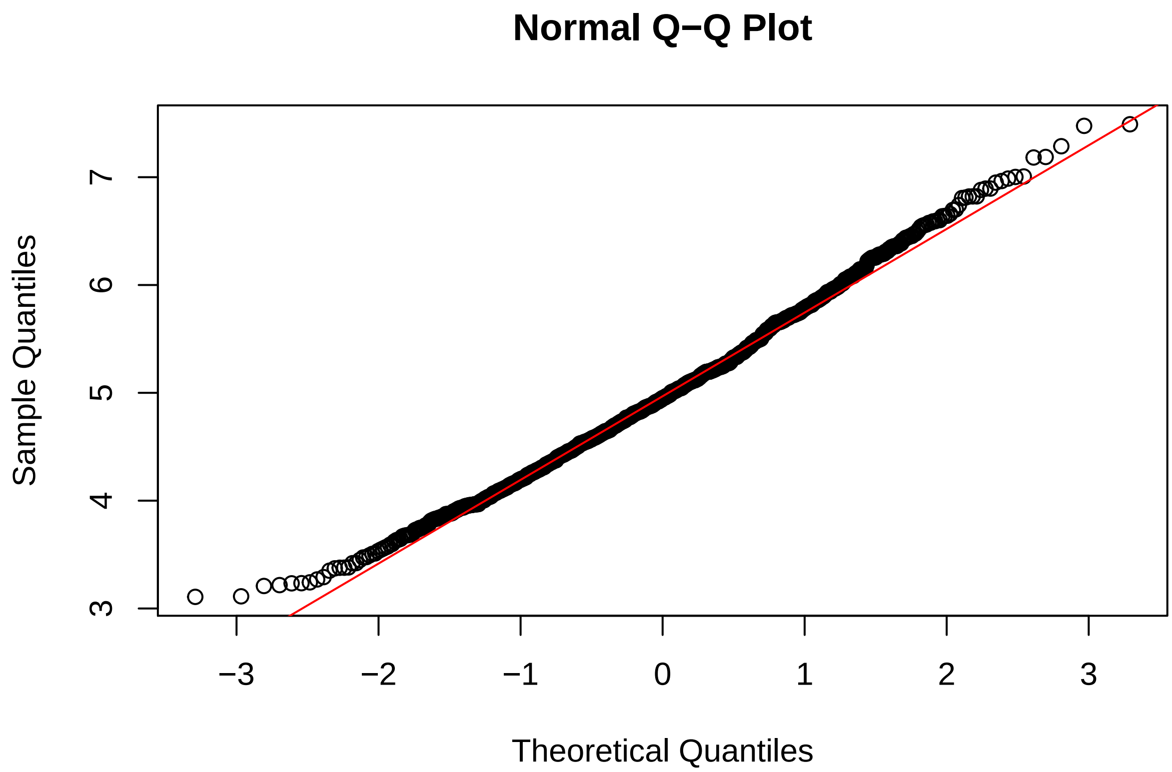
Question 3

Show that the distribution is approximately normal. From the plots below, it is clear that as we increase the number of simulations (in this case: 1000), the distribution of the simulated sample means becomes normal.

```r
library (ggplot2)
par(mfrow = c(1,1))
distributionCurve <- ggplot(data.frame(simulatedExponentialMeans),
                            aes(x = simulatedExponentialMeans)) +
                            geom_histogram (alpha = 0.2, binwidth=0.3, colour = "black",
                                            aes(y = ..density..)) +
                            geom_density()
distributionCurve
```




```
par(mfrow = c(1,1))
qqnorm(simulatedExponentialMeans)
qqline(simulatedExponentialMeans, col = 2)
```




```
#Calculate the 95% Confidence Interval of the simulated sample means
meanOfSimExpMeans + c(-1,1) * qt (0.975, 40) * sd (simulatedExponentialMeans)
```

```
## [1] 3.410071 6.569979
```