Part 1: Simulation Exercise for Statistical Inference

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.

1. Show the sample mean and compare it to the theoretical mean of the distribution.

```
lambda = 0.2
n = 40
nsms = 1:1000
set.seed(820)
means <- data.frame(x = sapply(nsms, function(x) {mean(rexp(n,
lambda))}))
head(means)

##          x
## 1 5.750000
## 2 3.808205
## 3 4.058154
## 4 3.999241
## 5 4.312532
## 6 4.418246

mean(means$x)

## [1] 4.998812
```

Mean of the simulation is 4.998812. The expected mean is 1/0.2 = 5.0

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
sd(means$x)

## [1] 0.7909422

(1/lambda)/sqrt(40)

## [1] 0.7905694

var(means$x)

## [1] 0.6255895

((1/lambda)/sqrt(40))^2

## [1] 0.625
```

The standard deviation of the simulation is .7909422. The expected standard deviation is (1/lambda)/sqrt(40) = .7905694.

The variance of the simulation is .6255895. The expected variance is ((1/lambda)/sqrt(40))^2 = .625

3. Show that the distribution is approximately normal.

```
library(ggplot2)
ggplot(data = means, aes(x = x)) + geom_histogram(aes(y = ..density..),
fill = I("blue"),
    binwidth = 0.1, color = I("black")) + stat_function(fun = dnorm,
args = list(mean = 5,      sd = sd(means$x)))
```