Course Project Inference

The purpose of this project is to perform a simple simulation using R and comparing the results to the theoretical basis.

To begin we'll define a few variables for the simulation. This includes setting the seed(42), number of simulations(10000) and the lambda value to 1/5.

```
set.seed(42)

nSim <- 10000
samples <- 40
lambda <- 0.2

sim <-matrix(rexp(nSim*samples, lambda), nSim, samples)
```

Also, we'll compute the theoretical values for the mean, std. dev., variance and the variance of the means(the squared SE):

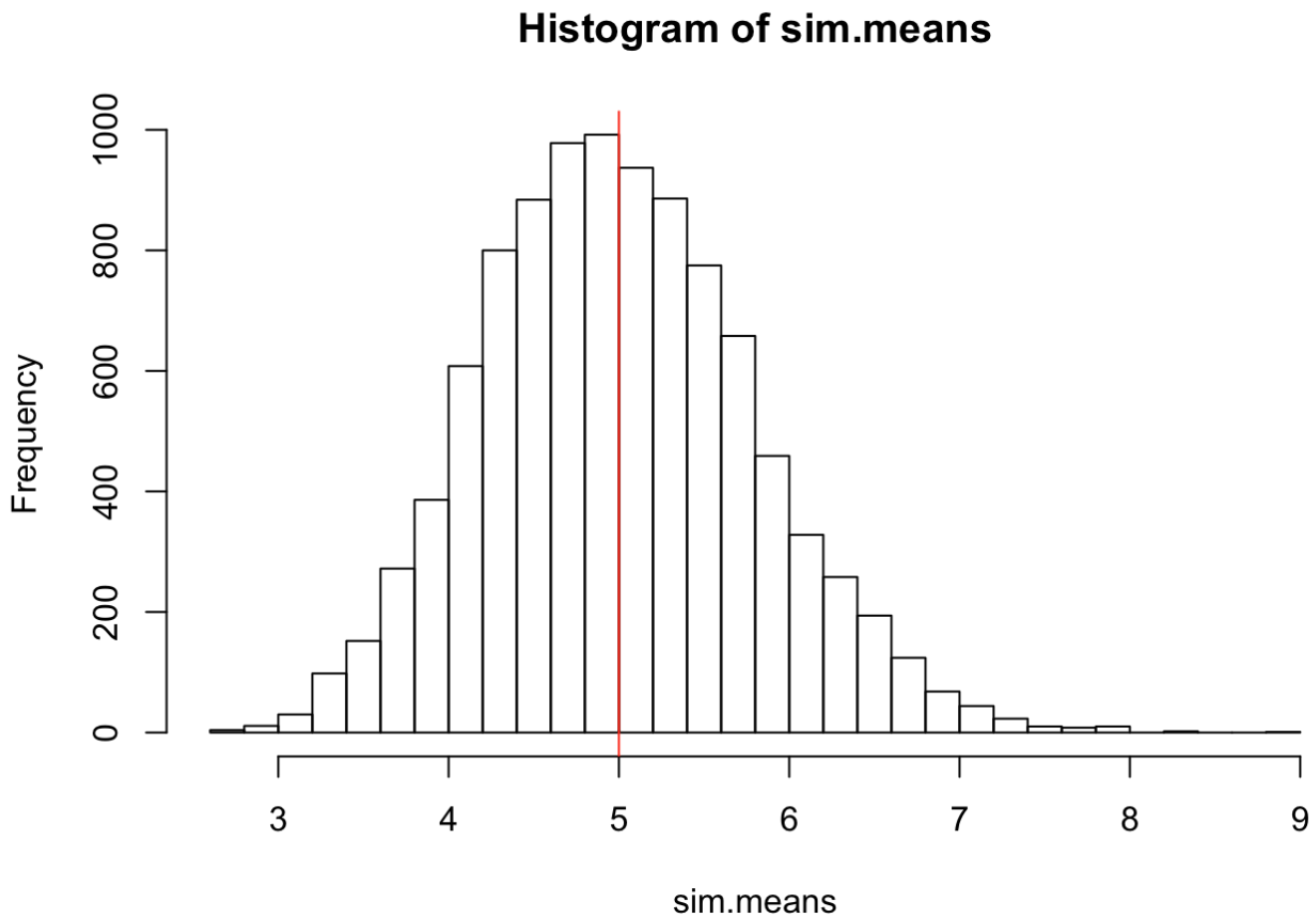
```
theoretical.mean <- 1/lambda
theoretical.sd <- 1/lambda
theoretical.var <- theoretical.sd**2
theoretical.var.of.means <- theoretical.var/samples
```

Question 1 - Show where the distribution is centered and compare to the theoretical center

The sample means are calculated as:

```
sim.means <- rowMeans(sim)
```

And the distribution of the means is seen below with the theoretical mean indicated by the reference line:



The theoretical mean is 1/lambda = 5 and the difference between the avg of sample means and the theoretical mean is:

```
mean(sim.means) - theoretical.mean
```

```
## [1] -0.002893226
```

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

The variance of the sample means is computed below along with the difference between the variance of the sample means and the theoretical variance of the means:

```
sim.var.of.means <- var(sim.means)
sim.var.of.means
```

```
## [1] 0.6310692
```

```
theoretical.var.of.means
```

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

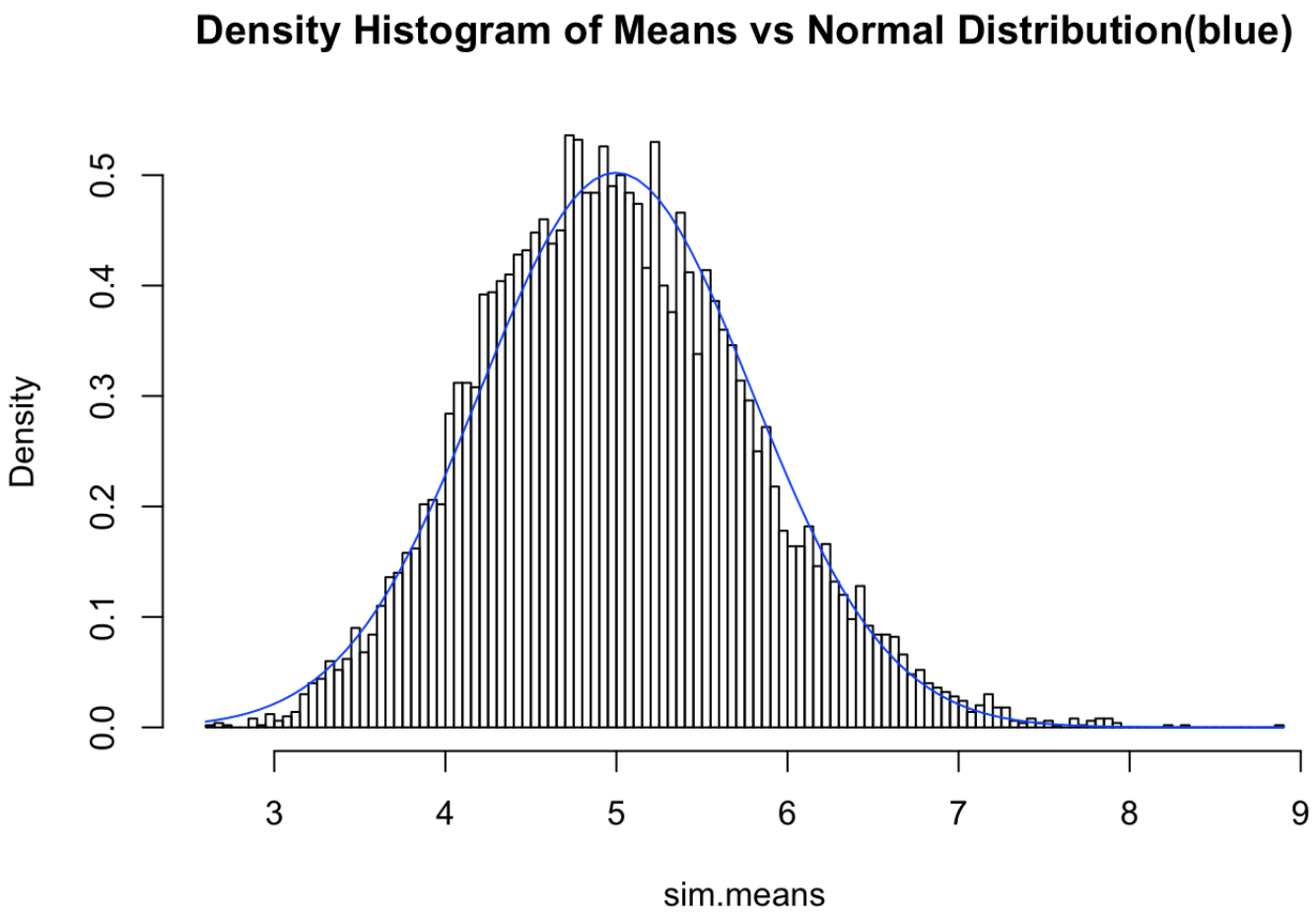
```
sim.var.of.means - theoretical.var.of.means
```

```
## [1] 0.006069216
```

Question 3 - Show that the distribution is approximately normal.

In order to show that the distribution is approximately normal we'll plot the distribution of sample means based on the probability(as opposed to frequency) and super impose a normal curve:

```
hist(sim.means, breaks = 100, prob = T, main = "Density Histogram of Means vs Normal Distribu
tion(blue)")
curve(dnorm(x, mean = mean(sim.means), sd = sd(sim.means)), add = T, col = "blue")
```



Question 4 - Evaluate the coverage of the confidence interval for 1/lambda using 1.96

The coverage is computed by taking the Sample Mean and Standard Error for each simulation and determining the proportion of cases where Confidence Interval (sample mean + 1.96 * SE) contains the true population mean of 1/lambda:

```
sim.sd <-apply(sim, 1, sd)
sim.se <- sim.sd/sqrt(samples)
CI.ub <- sim.means + (1.96*sim.se)
CI.lb <- sim.means - (1.96*sim.se)
coverage <- sum((CI.ub >= 1/lambda)&(CI.lb <= 1/lambda))
coverage/length(sim.means)
```

```
## [1] 0.9229
```

The results show that the coverage is approximately 92%, which differs from the expected coverage of 95%. this seems at least in part due to the small sample size and trying large popualtions seemed to correct for it.