Statistical inference, Course Project, Part 1

Assignment text in *italic*.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should

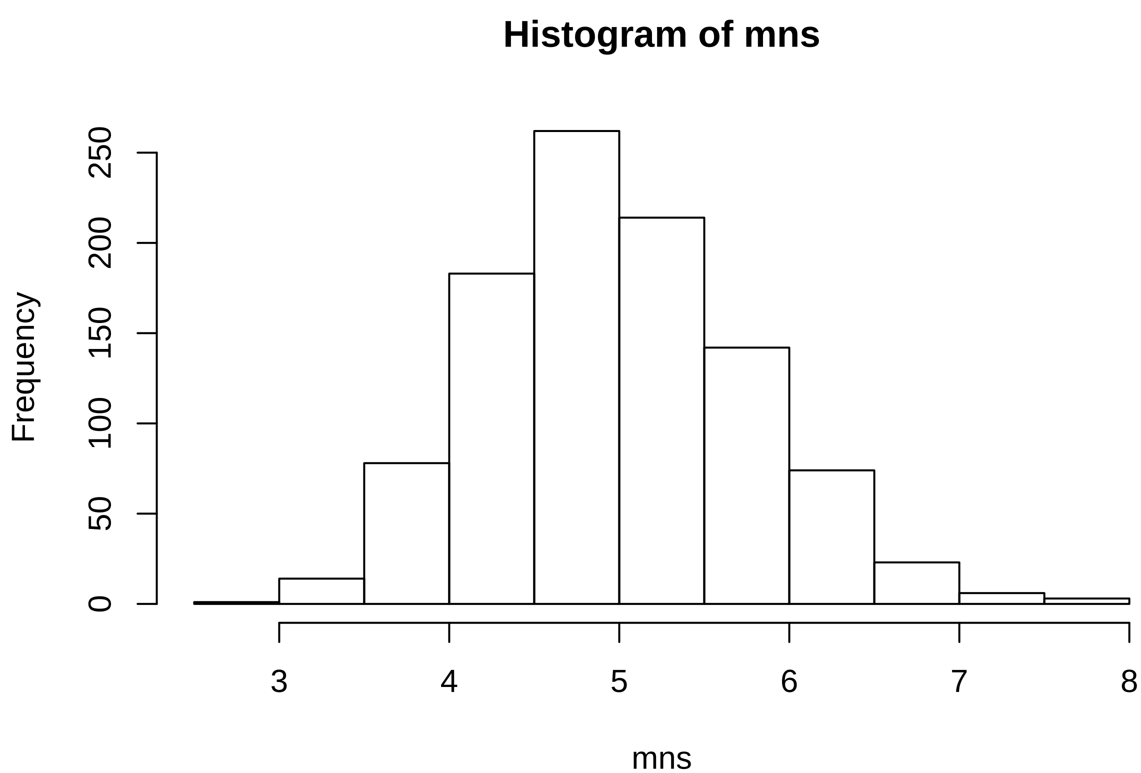
Question 1

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

Generate a distribution containing 1000 sets of 40 exponential(0.2)s and calculate where this distribution is centered.

```
library(rmarkdown)

# Reset the means variable
mns <- NULL
# Set the number of simulations
sim <- 1000
# Simulate 1000 means of 40 exponential(0.2)s and accumulate the mean into mns
for (i in 1 : sim) mns = c(mns, mean(rexp(40, 0.2)))
# Draw a histogram of the means
hist(mns)
```




```
# Print the mean of the means, which is where the distribution is centered
mean(mns)
```

```
## [1] 4.988072
```

The histogram shows the distribution of means. We can see that the distribution is centered around 5, which is also the theoretical center of the distribution 1/lambda = 1/0.2 = 5.

Question 2

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

The variance of a distibution of the type at hand is shown in the histogram above.
We now generate another distribution containing 1000 sets of 40 exponential(0.2)s but calculate the standard deviations as opposed to means. We calculate the standard error of the mean, which is the standard deviation of the sampling distribution of the mean. It should be noted that N here is the size of each sample from which a mean is calculated (40), not the number of simulations (1000), which just defines the granularity of the distribution but not its shape.

```r
# Reset the sds variable
sds <- NULL
# Set the number of simulations
sim <- 1000
# Simulate 1000 standard deviations of 40 exponential(0.2)s and accumulate the mean into sds
for (i in 1 : sim) sds = c(sds, sd(rexp(40, 0.2)))

# Calculate the standard error of the mean
mean(sds)/sqrt(40)
```

```
## [1] 0.7694828
```

```r
# Calculate the theoretical standard error of the mean
(1/0.2)/sqrt(40)
```

```
## [1] 0.7905694
```

The standard error of the mean is roughly 0.78, which is in line with the theoretical (1/lambda)/sqrt(n) = 0.79

Question 3

3. Show that the distribution is approximately normal.

We continue with the originally simulated distribution of means. The Central Limit Theorem states "given a population with a finite mean (mu) and a finite non-zero variance (sigma)^2, the sampling distribution of the mean approaches a normal distribution with a mean of (mu) and a variance of (sigma)^2/N as N increases." We test to what extent the distribution of means is normal with the Shapiro-Wilk test of normality.

```r
shapiro.test(mns)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mns
## W = 0.9898, p-value = 2.011e-06
```

The test is commonly said to be adequate for p.value < 0.1. This p.value is $<$ 1e-4, which means the distribution of means is very close to normal.