R Homework Assignment 2

Visualizing the Central Limit Theorem: The distribution of the population of sample means (from all samples of the same size n), are approximately normal for large enough n (usually n ≥ 30 if the distribution of the population from where the samples is not normal, and any n when this distribution is normal).

It is a mathematical fact, that has nothing to do with the central limit theorem that the mean of the population of all possible sample means of the same size (regardless, the n and regardless of the distribution of the population where the sample means come) is the same as the mean of the population where the samples are obtained: $\mu_{\bar{X}} = \mu_X$

It is also a mathematical fact, that has nothing to do with the central limit theorem that the mean of the standard deviation of the population of all possible sample means of the same size (regardless, the n, regardless of the distribution of the population where the sample means come) is equal to the standard deviation of the population where the samples are obtained divided by the squared root of n: $\sigma_{\bar{X}} = \sigma_X/\sqrt{n}$.

Using R, generate 1000 samples of size n = 5 from an Exponential distribution X with λ= 1/2 (remember that the mean of X is μ_X= $1/\lambda$, and its standard deviation is σ_X= $1/\lambda^2$) and place this 1000 samples in a 1000 x 5 matrix called M:

```
M<- matrix(NA, nrow = 100, ncol = 5)
for (i in 1:1000){
  M[i,] <- rexp(5, rate = 2)
}

# Vector to assign names to the columns of matrix A

a=c()
for (i in 1:5) {
  d=paste("X", i, sep ="")
  a=c(a,d)
}
a
colnames(M) <- a
M
```

a. Obtain a 1000 x 1 vector containing the average of each sample. (An approximate population of sample means from X). Then obtain the mean standard deviation of this population of sample means, and a histogram of this population of sample means of size n = 5. Do these results match what is stated in the Central Limit Theorem?

b. Obtain a 1000 x 1 vector containing the standard deviation of each sample.

c. Obtain two 1000 x 1 vectors containing the upper and lower bounds, respectively, of 95% CI's for μ_X using each sample assuming that you know σ_X (i.e. use the formula $\bar{x} \pm 1.96\sigma_X/\sqrt{n}$). Then obtain the percentage of the 1000 intervals that contain the mean μ_X. Is this number close to 95%? ______. Finally, use **your results** to complete this sentence:
Approximately, ______ percent of intervals constructed using the formula above contain the population mean μ_X.

d. Obtain two 1000 x 1 vectors containing the upper and lower bounds, respectively, of 95% CI's for using each sample assuming that you **don't know** sigma (that is, for each sample, construct a CI using the sample standard deviation of each sample mean $\bar{x}$: $\bar{x} \pm 1.96s/\sqrt{n}$. Then obtain the percentage of the 1000 intervals that contain the population mean μ_X. Is this number close to 95%?______. Finally, use **your results** to complete this sentence:
Approximately, ______ percent of intervals constructed using the formula above contain the population mean μ_X.

e. Obtain a 1000 x 1 vector with the z-scores of 1000 sample means. Then obtain the average of the z-scores, their standard deviation, and their histogram. What distribution do you think the z-scores have?

For next week: Repeat a to d above using sample sizes n = 10, 20, 30, and 40 and write your conclusions.