Lab on the Central Limit Theorem
by Prof. T. Fiore

Preparation: At home, before lab, prepare for this lab by working through the website

`http://mse.redwoods.edu/darnold/math15/UsingRInStatistics/CentralLimit2.php`

by reading it, pasting the commands into RStudio, and looking at the output. Be aware of the following points concerning proper pasting and several typos on the website.

- When copying and pasting a command, do *not* include the command line prompt symbol `>` in your copying and pasting! So for instance, where the website says

  ```
  > curve(dexp(x,rate=1),0,4,lwd=2,col="red",ylab="p")
  ```

  you would only copy and paste

  ```
  curve(dexp(x,rate=1),0,4,lwd=2,col="red",ylab="p")
  ```

  without the `>` .

- Similarly, when there are two commands displayed on the website with prompts at the beginning, you will need to copy them individually and evaluate them individually, without the prompt symbols. But, if no prompt symbols are there, you can copy several lines at once and evaluate them at once.

- Several of the commands use previous command definitions or outputs that are stored in the short term memory, so you must do the commands in order, and you cannot start in the middle, otherwise there will be an error.

- If your plot window is too small, and you do some kind of plot command, you will get an error that says "figure margins too large," even when there is no error in the command! When that happens, slide the vertical border line between the console and the plot window to the left to make the plot window bigger, and redo the command.

- About 1/3 down the website page, in the section Producing a Vector of Sample Means, everywhere it says

  ```
  exp(n,rate=lambda)
  ```

  it should actually have an `r` preceding it to say the following.

  ```
  rexp(n,rate=lambda)
  ```

In particular, the "For loop" will also need this correction when you paste it in, and a few other spots.

Be sure to actually *read the website* and reflect on what it is saying as you go through and evaluate the commands. The website is explaining the theorem and illustrating it graphically with the commands. **Do not simply cut and paste the commands without reading the website because you will not learn anything that way!**

Assignment: In Lab, open RStudio and open a Word File. In the Word file, put your names in large font at the beginning, and then label 1, 2, 3, 4 with vertical space in between. Paste and type your answers to problems 1, 2, 3, 4 below in the word file, and print the word file when you are done.

1. **Idea: We can approximate a density empirically by taking a large number of draws from a population with that density and plotting a probability histogram.**

   Let's illustrate this for the Beta density with $\alpha = 5$ and $\beta = 3$.

   ```
   xbar=rep(0,500)

   for (i in 1:500) { xbar[i]=rbeta(1,5,3) }

   hist(xbar,prob=TRUE,breaks = seq(-.5, 1.5, by = .05),
        main="Prob. Histogram for Beta")
   ```

   Now superimpose the Beta density with $\alpha = 5$ and $\beta = 3$ onto the probability histogram.

   ```
   curve(dbeta(x,5,3),-.5,1.5,add=TRUE,lwd=2,col="red")
   ```

   Export the final figure as a picture to the desktop and paste it into your Word file. Shrink it in the word file so it doesn't take up too much space, but make sure it is large enough to read when you print out the assignment. Discuss with your lab partner what the code is doing and *how this is an illustration of the first sentence of this exercise*. In the word file, write one sentence that *says what you just did in terms of the first sentence idea above and use the word Beta*. In other words, restate the first idea sentence above but tailored to what you just did. Your answer must be a complete sentence, with proper grammar, ending in a period.

2. Plot the exponential density with mean 2 and standard deviation 2. Do this by replacing the `...` in the following command with the appropriate number.

   ```
   curve(dexp(x,rate=...),-1,7,lwd=2,col="red",
        main="Exponential Density with Mean 2")
   ```

Did you get an error? If so, read the sentence preceding this command again. (Remember the advice above not to simply cut and paste the commands without reading because...)

Another hint: what you should write in `...` is not the number 2. Check your graph: how high should the value of the density function be above input zero, and how high should it be slightly to the left of input zero?

Export the final figure as a picture to the desktop and paste it into your Word file. Shrink it in the word file so it doesn't take up too much space, but make sure that it is readable.

3. Now we approximate the density of an average of n random samples from a population with an exponential density with mean 2 (this is the density you just plotted). This average of n random samples corresponds to the random variable we called $\overline{Y_n}$ in class. We do this for $n = 25, 36, 100, 400$. Here are the commands for $n = 25$. Cut and paste all at once.

```
xbar=rep(0,500)
for (i in 1:500) { xbar[i]=mean(rexp(25,rate=1/2) )}
hist(xbar,prob=TRUE,breaks = seq(0, 4, by = .05),
     main="Prob. Hist. of Mean of 25 Samples")
curve(dnorm(x,mean=2,sd=2/sqrt(25)),0,4,add=TRUE,lwd=2,col="red")
```

The probability histogram has a normal density superimposed. Discuss with your lab partner what the code is doing.

Repeat this command for $n = 36, 100, 400$. For this you must change all the 25's to the new number. You can use the "up" button to avoid cutting and pasting. (**Hint:** Did you remember to change *all* the 25's each time? If your plots aren't verifying the Central Limit Theorem, then you forgot some 25's, go back and fix it each time and re-plot.)
Export all **four** figures as pictures to the desktop and paste them into your Word file, in order, **on one page**. You can use the toggle in the plot window to get back to the previous plots and export them one by one. Shrink them in the word file so they don't take up too much space, but make sure they are readable.

4. (a) In your Word file, state the the Central Limit Theorem and its Corollary from class. **Type** directly from the class notes (do not paste in a picture of my notes, type it out!). Remember to type the words "Theorem" and "Corollary" before the respective statements.

   (b) In your Word file, explain **how problems 2 and 3 illustrate the Corollary to the Central Limit Theorem (also use the idea of problem 1)**. For this you should say what is the population density (is it normal?), mention the mean of the population density, mention the means and standard deviations of the superimposed normal curves **for each** n, and say the general feature of what is happening to your histograms and normal curves as n gets larger (25, 36, 100, 400). *Look at your graphs to see the means and standard deviations, toggle through the RStudio Plot window from 25 to 400 and see the effect of the Central Limit Theorem!*