Psychology 613: Multivariate Statistics (Data Analysis III)
Problem Set 2

The following problem set is due at **5pm** on **Thursday, April 14th** via direct deposit to Canvas.

Please complete the following problem in R, save your script (.rmd file) in the format: LASTNAME_FIRSTINITIAL.r (e.g., BERKMAN_E.rmd) then upload it to Canvas.

The central limit theorem states that sample means will tend toward a normal distribution as sample sizes increase, regardless of the distribution from which the individual data points within each sample were drawn. Your homework assignment is to demonstrate this. Please create a script in R demonstrating the CLT by following the algorithm below.

1. Create a variable, "iter" indicating the number of times to iterate through the simulation (that is described in Step 3). Set that variable to be a large number (e.g., 10,000). Create another variable, "sample_size", that contains a single, modest sample size (e.g., 50).

2. Create an empty variable that will be used to store the mean of the sample that will be drawn in each iteration. The variable should be a vector with one row per sample mean that you will collect (i.e., one for each iteration through the loop).

3. Create a "for" loop that starts at 1 and goes through each sample iteration. Call the current iteration number *i*. Inside the loop:
    a. Draw a sample of size "sample_size" from a random uniform distribution, and store this sample in a new variable called "sample".

    b. Calculate the mean of the sample, and then store the mean in a variable called "sample_mean".

    c. Store the sample mean in the *i*th place in the vector that you created in Step #2 (where *i* is the index number of the current sample iteration).

    d. End the loop.

4. After you've looped through all of the samples, create a histogram displaying the distribution of the sample means. Make sure the histogram has 100 bins.

5. Calculate the standard deviation of the sample means using the "sd" function, and then print it to the workspace to two decimal places using the "print" function. In the printout, be sure to include the sample size of each individual sample.

BONUS:

6. Embed the algorithm above in a second loop that repeats the entire procedure for 20 different sample sizes (of your choosing). That is, iterate steps 3-5. Plot the histograms in a 4 x 5 matrix of plots. Plot the standard deviations in a separate figure.