Math 4311 Lab: Confidence Intervals

1. **Critical Values of z in R.** Suppose Z is a standard normal random variable. Use R to find the following.

   (a) $P(0.83 < Z < 1.74)$. Here you will use the `pnorm` command. You could type `help(pnorm)` to see the documentation of this command. Typing `pnorm(1.96)` should give you an idea of what this function does.

   (b) Find a number z_0 such that $P(-z_0 < Z < z_0) = 0.9$. Use the command `qnorm` on this problem. Try typing `qnorm(0.025)` to see what it does.

2. **Computing Confidence Intervals.** Import the file `sat-data.txt` that was used for the regression lab. As a reminder, this data set contains math and verbal SAT scores for 1000 hypothetical students. Find 95% confidence intervals for the following:

   (a) μ_X, the average math SAT score

   (b) μ_Y, the average verbal SAT score

   (c) $\mu_X - \mu_Y$, the difference in average math and verbal SAT scores. (*Hint: Are you dealing with two independent samples or two paired samples?*)

3. **Writing Functions.** Write a function called `myprod` that accepts two numbers x and y as inputs and returns their product xy as output. Here is a similar example:

```
###mysum is a function that accepts two numbers x and y as inputs
###and returns their sum.

mysum=function(x,y){
         sum=x+y
         return(sum)}
```

4. **Creating a Confidence Interval Function.**

   (a) Write a function called `myconfint` that accepts a vector x of sample observations and returns the corresponding 95% confidence interval for μ. The output should be a vector of length two whose first component is the lower bound of the confidence interval and whose second component is the upper bound. Test your function using the imported SAT data. (To create a vector in R, use the `c` command, e.g., type `c(1,-4,9)` for the vector $(1, 4, 9)$.)

   (b) **Optional Improvement.** Rewrite myconfint so that the confidence level is a variable accepted by the function.

Simulating Confidence Intervals

5. Consider the distribution $U(0, 70)$.

   (a) What is the population mean μ for this distribution?

   (b) Generate a sample of size 100 from this distribution, and compute the 95% confidence interval. Does this confidence interval contain μ?

(c) Generate 1000 samples of size 100 from this distribution, and compute the 95% confidence interval for each one. What percentage of these confidence intervals contain μ? (*This problem sounds daunting but is surprisingly easy to do in R. The code below will help you.*)

```
#First, we need to create a 100x1000 matrix to store these samples.
#Each column will then be a sample of size 100.

sample.matrix=matrix(runif(100*1000,0,70),nrow=100)


#Now, we apply the myconfint function to every single column and store
#the results in a matrix.  This matrix will be 2x1000, where each column
#is a vector of length two representing a confidence interval.

results=apply(sample.matrix,2,myconfint)


#The  apply  command is very powerful.  The general syntax is

#apply(Some Matrix, 1 or 2, Some Function)

#This will apply the chosen function to the given matrix.
#The middle argument is 1 if you are applying the function to rows
#and it is 2 if you're applying to columns.

#We have our confidence intervals in the results matrix.  The first row
#contains all of the lower bounds, and the second row contains the upper bounds.
#Let's create separate vectors for these.

L=results[1,]
U=results[2,]

#Now, let's create a vector of length 1000 called success.  It will be true
#if the confidence interval was successful (if L<35<U) and false otherwise.

success = (L<35)&(35<U)

#Look at the entries in success, and you will see true/false values.
#What percentage of them are true?  Hint:  you can do arithmetic on
#true/false values, and R will treat them as ones and zeroes.
```

(d) Does the percentage of successful confidence intervals match your expectations? Explain.