Homework 4 – Intro. to Computational Statistics

For all problems, please show all your work. As described in the Homework Guidelines, use RMarkdown to write up your work as a .Rmd file, "knit" the result to a PDF file, and submit only that PDF file. Be sure to use R code for all your calculations, and the latex equation format to write up any math. See the Homework Guidelines for more formatting details.

For any of the following calculations, feel free to mix calculations by hand with R calculations. If you do something "by hand," be sure to write out step by step in neatly formatted math (using latex) how you calculated your result. Nothing is required to be solved using R unless specified.

To generate reproducible simulations (and anything else with random numbers generated by R), you should put `set.seed(1)` somewhere at the start of your code. That "seeds" the random number generator with the same initial value (1, though it could be any number), which means that every time you run your code you will get the exact same results.

1.

a. You get back your exam from problem 3.d of Homework 3, and you got a 45. What is your z score?

b. What percentile are you?

c. What is the total chance of getting something at least that far from the mean, in either direction? (Ie, the chance of getting 45 or below or equally far or farther above the mean.)

2.

a. Write a script that creates a vector of 10,000 integers generated from a poisson distribution with lambda = 10, and then draw a sample of 9 integers from that vector of 10,000 integers.

b. Calculate *by hand* your sample's mean. Please show your work using proper mathematical notation using latex. After doing it by hand, verify your result with R.

c. Calculate by hand the sample standard deviation. Verify your result with R.

d. Calculate by hand the standard error. Verify your result with R.

e. Calculate by hand the 95% CI using the normal (z) distribution. (You can use R or tables to get the score.)

f. Calculate by hand the 95% CI using the t distribution. (You can use R or tables to get the score.)

3.

a. Explain why 2.e is incorrect.

b. In a sentence or two each, explain what's wrong with each of the wrong answers in Module 4.4, "Calculating percentiles and scores," and suggest what error in thinking might have led someone to choose that answer. (http://www.nickbeauchamp.com/comp_stats_NB/compstats_04-04.html)

4.

a. Based on 2, calculate how many more individuals you would have to sample from your population to shink your 95% CI by 1/2 (ie, reduce the interval to half the size). Please show your work.

b. Say you want to know the average income in the US. Previous studies have suggested that the standard deviation of your sample will be $20,000. How many people do you need to survey to get a 95% cofidence interval of ± $1,000? How many people do you need to survey to get a 95% CI of ± $100?

5.

Write a script to test the accuracy of the confidence interval calculation as in Module 4.3. But with a few differences: (1) Test the 99% CI, not the 95% CI. (2) Each sample should be only 20 individuals, which means you need to use the t distribution to calculate your 99% CI. (3) Run 1000 complete samples rather than 100. (4) Your population distribution must be something other than a bimodal normal distribution (as used in the lesson), although anything else is fine, including any of the other continuous distributions we've discussed so far. Explain your result and how it validates the theoretical calculation of the 99% CI.