553.283 Introduction to R
Homework 4

Owing to the length of this problem set it's recommended that you do every problem, but I will only grade one of your choosing out of each pair. That is, you must submit one of Problems 1 and 2, one of Problems 3 and 4, etc. Note 1: If a question asks you for a numerical answer, your submission for that question must consist of the R command that produces that answer followed immediately by the output.

Note 2: Please label all axes on any plots you create.

Note 3: When writing functions, please indent each nested block of code further away from the margin than the block in which it is nested. It makes your code much easier for me to read.

Note 4: Properly comment your code where ambiguity may arise. Comments in R are preceded by the # sign.

1. Alter the code for the *hello()* function (slide 13 of Lecture 4) such that it takes one argument, say x, and outputs the greeting "hello x". Test it out via the command *hello("kitty")*.

2. Use preexisting functions in R to write a function called *ourhist* that produces a proportion histogram of a numerical vector x such that the number of breaks is determined by the *"Scott"* method and the histogram is overlaid with the density line. Test it on 100 randomly generated observations from the Exp(3) distribution.

3. Without using the *ifelse()* function, write a function that adds 1 to all the positive entries of a vector and subtracts one from all the non-positive entries. Test it on 100 randomly generated observations from the N(0,1) distribution, and use the function from the previous problem to create a histogram of the modified sample.

4. Write the same function as in the previous problem, only this time use the *ifelse()* function.

5. The classic Fibonacci sequence is defined such that the first two numbers are 0 and 1, and each subsequent number in the sequence is the sum of the two numbers preceding it. I.e.,

$$\begin{aligned} a_1 &= 0 \\ a_2 &= 1 \\ a_k &= a_{k-1} + a_{k-2}. \end{aligned}$$

   It's easy to see that the first nine numbers in the sequence are $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$. Write a recursive function that returns the nth Fibonnacci number.

6. A generalized Fibonacci sequence is one in which the first two numbers are specified (not necessarily 0 and 1), and each subsequent number in the sequence is the sum of the two numbers preceding it. For example, the sequence beginning $3, 7, 10, 17, 37, 54, 91, 145, 236, \ldots$ is a generalized Fibonacci sequence beginning with 3 and 7.

   Write a recursive function that returns the nth number of a generalized Fibonacci sequence such that starting values are chosen by the user or default to 0 and 1 if the user does not specify otherwise.

7. Let $\{a_k\}_{k=1}^{\infty}$ be the classic Fibonacci sequence as in Problem 5, and let $\{b_k\}_{k=1}^{\infty}$ be a generalized Fibonacci sequence starting with two positive numbers of your choosing. The sequence of classic Fibonacci ratios $\{r_k\}_{k=3}^{\infty}$ is defined as the sequence of ratios of sequential Fibonacci numbers. I.e.,

$$r_k = \frac{a_k}{a_{k-1}}, \text{ for } k = 3, 4, 5, \ldots$$

   Write a script that obtains r_3 through r_{20} as well as the corresponding ratios for your own generalized Fibonacci sequence, and plot[1] both sequences on the same set of axes with differently shaped points for each sequence. What do you notice about the sequences of ratios?

[1]Specify only the first argument of *plot()* to plot a sequence of numbers against their indices, and use *points()* to add another sequence in the same manner.

8. A Bernoulli random variable X with success probability p has the probability mass function

$$P(X = k) = \begin{cases} 1 - p, & k = 0 \\ p, & k = 1 \\ 0, & \text{otherwise} \end{cases}.$$

   Without using any of the R functions for the binomial distribution, write *d*-, *p*-, and *r*- functions for the Bernoulli distribution. Verify that they work as intended.

9. Some days Zach finds that he has nothing better to do than sit on a porch and watch cars pass him by. On average, he sees about 15 cars pass in a single day. What is the probability he observes (strictly) more than 20 cars in a day?

   Hint: Recall which random variable mentioned during class is concerned with the distribution of occurrences of an event over a period of time.

10. Simulate a 30-day month of Zach's car-watching pastime, create a histogram, and compute the mean and standard deviation.

11. Simulate an entire year of Zach's car-watching pastime (assume every month is 30 days long), compute the mean for each month, and generate a histogram of the means. Does the distribution of the means look normal? Why or why not?

12. Simulate two decades' worth of Zach watching cars (no leap-years), compute the mean for each year, and generate a histogram of these means. Does the distribution look normal? Why or why not?