Homework 3 of STAT 540
Section 001, Fall 2024
Due: Wednesday Sep 18 (before class)
Total Points: 106

Please hand in a hard copy of your homework (compiled pdf file from R markdown) in class and email your R code to Kaniz Fatema (KFATEMA@email.sc.edu). Please use the R markdown Homework template (HWtemplate.Rmd) to write your homework solutions. Work on the homework independently.

Problem 1. The function nchar tells you how many characters long a character vector is. Write a line of code that assigns to the object new_names the state abbreviation when the state name is longer than 8 characters. (5 points)

Problem 2. Write a function compute_s_n that for any given n computes the sum $S_n = 1^2 + 2^2 + \cdots n^2$. Report the value of the sum when n=10. (5 points)

Problem 3. Perform the following steps in R:

(a) Simulate 30 samples from Normal(mean=0, sd=1) (2 points)
(b) Randomly assign 15 samples into control and 15 into treatment group (15 points) [Hint: Use sample]
(c) Perform two sample T-test and report the p value. (2 points)
(d) Randomly generate 1000 samples from uniform distribution, and plot the histogram of the 1000 samples. [Hint: Use hist(x) to plot a histogram of x.] (2 points)
(e) Repeat (a) (b) (c) 1000 times, and stored the corresponding 1000 p values in a vector, plot a histogram using these 1000 p values. What is the distribution of p values? (15 points)

Problem 4. Perform the following steps in R:

(a) Simulate a string of 10,000 characters drawn uniformly and independently from the set {A, C, G, T} [Hint: sample] (7 points)
(b) Create a frequency table of the string [Hint: table] (3 points)
(c) Write a function to create a contingency table of adjacent k-tuples. For

example, with k=3 and with the string “CAGACAAAAC”, you would want to produce the following table: [Only use for loops and paste(, collapse=“”), Do not use embed, substr or do.call] (30 points)

AAA	AAC	ACA	AGA	CAA	CAG	GAC
2	1	1	1	1	1	1

Problem 5. x! = 1 × 2 ×3…×x; 0!=1. x is an integer ≥ 0. Write your own function to perform the calculation. (20 points) [Do not use the function **prod** and **factorial** in R]