

CIT 660 – Statistical Analysis and Visualization of Biological Data
Spring 2022
Assignment 01
Deadline: Wednesday - May 18th, 2022, 11:59 pm.

Q1. Define the sample space for:

1. Tossing a fair coin once.
2. Tossing two fair coins simultaneously.
3. Tossing three fair coins simultaneously.
4. A medical researcher is assaying a human tissue in many samples for possible sequence mutations. It is expected that the maximum number of mutations per tissue is 5.

Q2. Compute the probability of:

1. Getting a head in **Q1.1**.
2. Getting a head in **Q1.2**.
3. Getting two tails in **Q1.3**.
4. Not getting three heads in **Q1.3**.
5. Getting a tissue with three or more mutations in **Q1.4** assuming that all the numbers of mutations are equally probable.

Q3. Given a list of numbers, [1, 3, 2, -1, 1, 2, 4, 3, -1, 5], manually compute/obtain:

1. A set of the unique elements in this list and compute the probability of occurrence of each unique element. Is it a PMF or a PDF? Why?
2. The mean value by two different methods: with and without the PMF/PDF.
3. The median value.
4. The mode value.
5. The standard deviation value by two different methods: with and without the PMF/PDF.
6. The range.
7. The median absolute deviation.

Q4. Write an R script called “Q4.R” that takes from the user a list of numbers and outputs the following of the list:

1. Mean and median.
2. Standard deviation and range.

Note: you need to use functions to receive input data from the keyboard in R and to display data on the screen.

Q5. Assume you have the following DNA sequence ACAGTCGACTAGCTTGCACGTAC, write an R **function script** called “DNA_to_RNA.R” to convert this sequence to an RNA sequence, simply by replacing any T nucleotide to a U nucleotide, assuming perfect transcription process without any loss in the DNA sequence bases.

- Do not use any string function that can replace characters.

- You have to loop over the whole sequence and check each character in the DNA sequence separately before converting to an RNA sequence.
- Let the input to “DNA_to_RNA.R” be the given sequence, and let the function return the RNA sequence.
- Write another R script called “main.R” where you can call the “DNA_to_RNA.R” function.

You can hard code (write by hand) the above DNA sequence in the “main.R” script or allow the user to input it from the keyboard into the main script, and then send it as an input to “DNA_to_RNA.R”. The above function should return the RNA sequence to the main script, where the main script displays it on the screen.

Good luck!