Indian Maritime University

(A Central University, Govt of India)

End Semester Examinations – December 2023

Programme Name: B Tech (ME)

Semester: III

Subject Code: UG11T4305

Subject Name: Statistics and Data Analysis Using Python and R

Date: 13.12.2023 Max Marks: 70

Duration: 03 Hrs Pass Marks: 35

General Instructions

(i) All Sections (A, B & C) are to be attempted.
(ii) Options, if any, are specified in the respective section.
(iii) Scientific calculator is permitted.

Section A

MCQs –All Questions are Compulsory. (10×01 mark = 10 Marks)

1. Variable 'a' is defined as
   a = 'gOOd moRning'
   Command to convert 'a' from 'gOOd moRning' to 'Good Morning' is:-
   a. a.upper()
   b. a.lower()
   c. a.string()
   d. a.title()

2. Consider the tuple,
   t= (5,15,20,10,20,50)
   The output of the code: t.index(20) is
   a. 3
   b. 2
   c. 5
   d. 4

3. Which data structure is ordered and immutable in Python?
   a. List
   b. Tuple
   c. Set
   d. Dictionary

4. What will be the output after the following statements are executed?

```
x=27
y=4
x%=y
print(x)
```

a. 6
b. 6.75
c. 3
d. None of the above

5. What is the primary data structure for storing and manipulating tabular data in Pandas?
   a. Tuple
   b. List
   c. Series
   d. DataFrame

6. In Python, how do you define a function?
   a. Using square brackets
   b. Using curly braces
   c. Using the 'def' keyword
   d. Using the 'function' keyword

7. Which command would you use to find the data type of a variable?
   a. data()
   b. type()
   c. typeof()
   d. str()

8. In NumPy, what is an ndarray?
   a. A numerical data type
   b. A library for data visualization
   c. A multi-dimensional array object
   d. A statistical function

9. What is the median of the following dataset: 8, 5, 12, 35, 15, 22, 30?

   a. 15
   b. 20
   c. 22
   d. 12

10. What is the primary purpose of a box and whisker plot?

    a. To show the distribution of categorical data
    b. To display the correlation between two variables
    c. To visualize the spread and central tendency of numerical data
    d. To represent the probability distribution of a dataset

Section B

Answer all the Questions. **(05×02 marks=10 Marks)**

11. Explain tuple unpacking in Python with an example. 2 Marks

12. What is a NumPy array, and how does it differ from a Python list? 2 Marks

13. Explain dir() function in python. 2 Marks

14. Define "kurtosis" in statistics. 2 Marks

15. Write a R program to print the output as a subtraction of two vectors

    x = (1, 2, 3, 4, 5) and y = (10, 20, 30, 40, 50). 2 Marks

Section C

Answer any 5 of the following 7 questions. **(05×10 marks=50 Marks)**

16.

a) Compare and contrast the list and set data structures in Python with examples. 5 Marks

b) Write Python code to open a file, read its contents, write any sentence, and print them to the console. 5 Marks

17.

a) What are some common types of plots that Matplotlib can create, Explain any two plots in detail. 5 Marks

b) Explain Dictionary data type in Python. How to create, access, and modify dictionary elements? 5 Marks

18.

a) Develop a Python program that implements a function that calculates the factorial of a given number. 5 Marks

b) The joint probability distribution of two random variables X and Y is given by

y \ x →	-2	4
1	0.1	0.1
-3	0.2	0.4
5	0.1	0.1

(i) Evaluate the marginal distributions of y.

(ii) Examine whether X & Y are independent.

(iii) Find P(Y=5/X=4) 5 marks

19. a) Write a python program for the creation of 1-D & 2-D NumPy array.
1-D array containing the values 1,2,3,4,5
2-D array containing two arrays with the values 1,2,3 & 4,5,6.
5 Marks

b) Write an output of the following R program. 5 Marks

```
print("New vector using seq() function-")
v = seq(3, 15, by= 3)
print("Original vector:")
print(v)
print("Check which values are even:\n")
is_even <- vector %% 2 == 0
print(is_even)
```

20.

a) Suppose you have a dataset representing the test scores (out of 100) of a group of students in a math class. The scores are as follows: 85, 92, 78, 88, 95, 90, 82, and 89. Calculate the mean and standard deviation of these test scores.

5 Marks

b) Explain the key data structures in Pandas. 5 Marks

21.

a) Write Python statements that create an empty list, an empty tuple, an empty set, an empty dictionary, and an empty string. 5 marks

b) There are 50 students in a class, the regression equation of marks in Python programming (X) on marks in Mathematics (Y) is 3Y – 5X + 180 = 0. The mean marks of Mathematics is 44 and variance of marks in Python is $(^{9}/_{16})^{th}$ of the variance of marks in Mathematics. Find the mean marks in Python programming and the coefficient of correlation between marks in two subjects. 5 Marks

22.

a) Compare and contrast the use of central tendency measures and dispersion measures in Exploratory Data Analysis. 5 marks

b) Explain for and while loop with syntax, flowchart & examples. 5 marks