Indian Maritime University

(A Central University, Govt of India)

Mar/Apr'26 SE

Programme Name: B Tech (ME)

Semester: II

Subject Code: UG11T5202

Subject Name: Introduction to Computers & Python Programming

Date: 12.03.2026	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 respective section.

Section A

Ten MCQs/Fill in the Blanks of 01 Mark each – Choose the correct answer as applicable.

1. Which Pandas function is used to read a CSV file?
   a) read_file()
   b) open_csv()
   c) read_csv()
   d) load_csv()

2. In R, what is a vectorized operation?
   a) Operation on single elements
   b) For loop
   c) If statement
   d) Operation over an entire vector

3. What does the str() function do in R?
   a) Sorts a vector
   b) Changes a string to lower case
   c) Displays the structure of an R object
   d) Summarizes a data frame

4. Which Python library is primarily used for numerical computations and working with arrays?

a) pandas
b) matplotlib
c) NumPy
d) seaborn

5. What function is used to generate binomially distributed random numbers in Python?

a) np.random.binomial(n, p, size)
b) np.random.poisson(n, size)
c) np.random.normal(n, size)
d) np.random.uniform(n, size)

6. Which of the following is used to calculate the cumulative distribution function (CDF) of a normal distribution in R?

a) dnorm(x, mean, sd)
b) pnorm(x, mean, sd)
c) qnorm(x, mean, sd)
d) rnorm(x, mean, sd)

7. Which of the following is NOT a component of a neural network?
a) Input layer
b) Hidden layers
c) Output layer
d) Residual layer

8. Which activation function is commonly used for binary classification?
a) ReLU
b) Softmax
c) Sigmoid
d) Tanh

9. Which interpolation method is best suited for estimating values at unsampled locations based on the assumption that closer points have more influence?

a) Kriging
b) Inverse Distance Weighting (IDW)

c) Linear Regression

d) Ordinary Least Squares (OLS)

10. Which of the following is a non-parametric method used to estimate survival curves in survival analysis?

a) Kaplan-Meier Estimator

b) Exponential Model

c) Weibull Model

d) Log-Normal Model

Section B

Five Questions of 02 Marks each

11. Write two differences between Python lists and NumPy arrays.
12. Explain the function of .describe() in Pandas.
13. List two key differences between data frames and matrices in R.
14. What are control structures in R? Give an example of any one.
15. Write the output of the following R-code

```
x <- c(12, 15, 14, 10, 18, 21, 24)
quantile(x)
```

Section C

Seven Questions of 10 Marks each of which any 05 questions to be answered.

16. a) Write Python code using Pandas to load a CSV, handle missing data, and compute descriptive statistics. [5M]

    b) What is data wrangling? Explain data merging and reshaping with suitable Pandas functions and examples. [5M]

17. a) Describe different types of plots available in base R. Write R code to generate a scatter plot and customize it with title, colour, and axis labels.

    [5M]

b) Discuss control structures in R. Write a function using if-else and for loop to calculate the factorial of a number. [5M]

18. a) Write a R/Python-commands to solve the following problem: In a particular restaurant, an average of 3 out of every 5 customers ask for water with their meal. A random sample of 10 customers is selected. Find the probability that

(i) Exactly 6 ask for water with their meal,
(ii) Less than 9 ask for water with their meal.
(iii) No one asks for water with their meal?
(iv) At most 2 ask for water with their meal?
(v) At least 3 ask for water with their meal? [5M]

b) Define and differentiate between discrete and continuous probability distributions. Provide examples of each type and explain the key functions such as Probability Density Function (PDF), Cumulative Distribution Function (CDF), and Quantile Function. Provide R/Python commands for these functions. [5M]

19. a) Define linear regression and explain its model formulae. How are model matrices used in regression analysis? Discuss the importance of regression diagnostics, including residual analysis, leverage, and Cook's Distance. Provide R/Python command for linear regression. [5M]

b) Define Analysis of Variance (ANOVA) and factorial designs. Why are they important in experimental studies? Explain the concept of Main effect and interaction effects. [5M]

20. a) Describe the structure of a neural network. Discuss the role of input layers, hidden layers, and output layers. [5M]

b) Explain the concept of partitioning in decision trees and its role in classification and regression. Describe how decision trees partition the input space . Explain how splits are chosen and their effect on model performance. [5M]

21. a) Discuss the role and challenges of visualization in multivariate data analysis. Explain how scatterplot matrices, parallel coordinate plots, and principal component analysis (PCA) are used for visualization. [5M]

b) Define cluster analysis and explain its purpose in multivariate data analysis. Differentiate between k-means clustering, and Gaussian Mixture Model-based clustering. Discuss the challenges and methods of choosing the number of clusters. [5M]

22. a) Describe time series and it's components. Explain the importance of stationarity. Additionally, define and differentiate between autocorrelation and partial autocorrelation, including their significance in time series analysis. [5M]

b) Define univariate functions and their role in optimization. Discuss Golden Section Search method used for optimizing univariate functions. Also name the Python library for implementing this optimization method.

[5M]