

MAHATMA GANDHI UNIVERSITY, KOTTAYAM
MGU-UGP (HONOURS)
THIRD SEMESTER
B.Sc(Hons) DEGREE EXAMINATION MATHEMATICS
(2025 ADMISSION ONWARDS)
MG3DSCMATA11–STATISTICAL COMPUTING FOR DATA SCIENCE

Duration: 2 hrs

Maximum Marks: 70

Students should attempt at least one question from each course outcome to enhance their overall outcome attainability.

Part A

Answer any five questions. Each question carries 2 marks

1. State the major stages of the Data Science workflow. [K][1].
2. How is a new key-value pair added to a dictionary? [U][1].
3. Differentiate between ndim and size. [U][2]
4. What is an index in a Pandas Series? [K][2]
5. Name the three common measures of central tendency. [K][3]
6. Give one example where correlation does not imply causation. [U][3]
7. Write the Matplotlib command used to display a graph. [K][4]
8. How are outliers represented in a box plot? [U][4]

Part B

Answer any five questions. Each question carries 6 marks

9. Explain conditional statements and iteration structures in Python with suitable examples. [U][1]

10. Explain how control is transferred between multiple functions during the execution of a Python program. [U][1]
11. Write code to create a 3×3 NumPy array and display its properties. [A][2]
12. Compare statistical functions available in NumPy. [An][2]
13. Explain Pandas Series with an example. [U][2]
14. Explain the effect of outliers on mean and median. [An][3]
15. Calculate the standard deviation for the following dataset: 12, 15, 18, 20, 15, 22, 17, 19, 16, 21. [A][3]
16. Explain the roles of Q_1 , median and Q_3 in a box plot. [U][4]

Part C

Answer any three questions. Each question carries 10 marks

17. Explain Python lists in detail, including creation, indexing, modification, insertion, deletion and aggregation operations. [An][1]
18. Given the following two lists:

$$A = [10, 20, 30, 40, 50], \quad B = [5, 10, 15, 20, 25]$$

Compute the element-wise sum of the two datasets using

(i) Python lists and **(ii) NumPy arrays**. Write the Python code for both approaches, display the results, and compare their implementation. [An][2].

19. Compare two datasets having the same mean but different dispersion and interpret the results. [An][2].
20. Given the following data, calculate the Pearson correlation coefficient and interpret the nature and strength of the relationship between study hours and examination marks.

Study Hours (X)	2	3	4	5	6	7	8	9
Marks (Y)	40	45	50	55	65	70	75	80

[An][3].

21. Given the following marks:

[45, 50, 55, 58, 60, 62, 65, 68, 70, 72, 75, 80, 82, 85, 90]

Write a complete Python program to:

- (a) Plot a histogram with five bins and explain what the histogram reveals about the distribution of marks.
- (b) Plot a box plot for the given marks and explain what the box plot reveals about the distribution, including its central tendency, spread, and possible outliers.

[An][4].

22. Discuss the complete Exploratory Data Analysis (EDA) process as presented in the notes. Explain how visualization acts as a “bridge between data and human understanding”, with reference to the height–weight example.

[An][4]