END TERM EXAMINATION

Data Analysis Using R and Python

Time: 3 Hours **Maximum Marks: 70**

SECTION A (Compulsory)

7 × 2 = 14 Marks

1. Define categorical data with an example.
2. Write any two R data types.
3. What is a scatter diagram?
4. Define tuple in Python.
5. What does the function `apply()` do in R?
6. Mention one use of NumPy in data analysis.
7. What is meant by univariate data?

SECTION B

Attempt any four questions. Each question carries 14 marks.

1. Explain the process of importing, understanding, and cleaning data. Illustrate with appropriate examples.
2. Discuss in detail the different data structures in R such as vectors, lists, data frames, and matrices.
3. Explain the concepts of univariate, bivariate, and multivariate visualizations using `ggplot`.
4. Describe descriptive statistics for data frames in R, including the use of `summary()` and other relevant functions.
5. Explain Python data types such as strings, lists, tuples, and dictionaries with examples.

6. Discuss data loading, manipulation, and saving using Pandas in Python with proper syntax.

7. Explain the use of Apply family functions (`apply`, `sapply`, `tapply`) with examples.

8. Describe array-oriented programming in NumPy and discuss its relevance in data cleaning and preparation.