INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal-500043, Hyderabad

B.Tech III SEMESTER END EXAMINATIONS (REGULAR/ SUPPLEMENTARY) - FEBRUARY 2024
Regulation: UG20
PRINCIPLES OF DATA SCIENCE

Time: 3 Hours   CSE(DATA SCIENCE)   Max Marks: 70

Answer ALL questions in Module I and II
Answer ONE out of two questions in Modules III, IV and V
All Questions Carry Equal Marks
All parts of the question must be answered in one place only

MODULE – I

1. (a) List five essential steps usually performed in data science. Explain the preprocessing steps needed to transform the data in to a structures form. [BL: Understand| CO: 1|Marks: 7]

   (b) As a city planner, you are tasked with analyzing different types of data related to urban development, including nominal data on land usage, ordinal data on infrastructure priorities, interval data on temperature patterns, and ratio data on population density. How would you use each type of data to inform sustainable urban planning? [BL: Apply| CO: 1|Marks: 7]

MODULE – II

2. (a) Explain the concept of arithmetic and boolean operations in R programming. Discuss the fundamental arithmetic operators and their applications, highlighting the differences between integer and double data types. [BL: Understand| CO: 2|Marks: 7]

   (b) Obtain a function in R that generates a sequence of Fibonacci numbers up to a specified limit. [BL: Apply| CO: 2|Marks: 7]

MODULE – III

3. (a) Explore the philosophical foundations of set theory. Discuss the nature of sets, their existence, and the relationship between sets and mathematical reality. [BL: Understand| CO: 3|Marks: 7]

   (b) In an oral exam you have to solve exactly one problem, which might be one of three types, A, B, or C, which will come up with probabilities 30%, 20%, and 50%, respectively. During your preparation you have solved 9 of 10 problems of type A, 2 of 10 problems of type B and 6 of 10 problems of type C.
   i) What is the probability that you will solve the problem of the exam?
   ii) Given you have solved the problem, what is the probability that it was of type A?
   [BL: Apply| CO: 3|Marks: 7]

4. (a) Discuss parametric equations and their representation on Cartesian graphs. Explore how parametric equations describe curves and paths in the plane and their advantages in certain contexts.
   [BL: Understand| CO: 4|Marks: 7]

(b) Consider a radar station monitoring air traffic. For simplicity we chunk time into periods of five minutes and assume that they are independent of each other. Within each five-minute period, there may be an airplane flying over the radar station with probability 5%, or there is no airplane (we exclude the possibility that there are several airplanes). If there is an airplane, it will be detected by the radar with a probability of 99%. If there is no airplane, the radar will give a false alarm and detect a non-existent airplane with a probability of 10%.

i) How many false alarms (there is an alarm even though there is no airplane) and how many false no-alarms (there is no alarm even though there is an airplane) are there on average per day?

ii) If there is an alarm, what is the probability that there is indeed an airplane?

[BL: Apply| CO: 4|Marks: 7]

MODULE – IV

5. (a) Describe the key differences between observational and experimental ways of obtaining data. How does each approach influence the type of analysis that can be performed?

[BL: Understand| CO: 5|Marks: 7]

(b) Consider any dataset and apply measures of variation (range, variance, standard deviation) to assess the variability of the data. [BL: Apply| CO: 5|Marks: 7].

6. (a) Discuss the step-by-step process of constructing a confidence interval. How is it calculated, and how should it be interpreted? [BL: Understand| CO: 5|Marks: 7]

(b) Demonstrate the fundamental process of hypothesis tests and utilize the five essential steps to conduct hypothesis tests. Also, apply various types of hypothesis tests to make inferences about a population based on a sample of data. [BL: Apply| CO: 5|Marks: 7]

MODULE – V

7. (a) How do scatter plots serve as a fundamental tool for visualizing the relationship between two continuous variables? Discuss the principles of constructing and interpreting scatter plots.

[BL: Understand| CO: 6|Marks: 7]

(b) Consider a dataset representing the monthly sales of two products, A and B, over a year. Create a box plot for each product's sales and interpret the differences in their distributions. Discuss any outliers, if present, and their potential impact on sales analysis.

[BL: Apply| CO: 6|Marks: 7]

8. (a) Design a bar chart to visualize changes in stock prices for multiple companies over a specific period. [BL: Understand| CO: 6|Marks: 7]

(b) Illustrate the applications of line graphs when dealing with categorical variables. How do line graphs effectively capture and communicate temporal trends over time?

[BL: Understand| CO: 6|Marks: 7]

− ∘ ∘ ◯ ∘ ∘ −