R22 Regulation *Subject code:4O5FA*

TKR COLLEGE OF ENGINEERING AND TECHNOLOGY

(Autonomous, Accredited by NAAC with 'A+' Grade)

B.Tech V Semester Supplementary Examinations, April 2026

FUNDAMENTALS OF DATA SCIENCE

(IT)

Maximum Marks: 60 Date:27.04.2026 Duration: 3 hours

Note: **1.This question paper contains two parts A and B.**
2. Part A is compulsory which carries 10 marks. Answer all questions in Part A.
3. Part B consists of 5 Units. Answer any one full question from each unit.
4. Each question carries 10 marks and may have a, b, c, d as sub questions.

Part-A

All the following questions carry equal marks (10X1M=10 Marks)		Marks	CO	BTL
1.a)	What is Datafication?	1M	1	L1
b)	What is Overfitting in statistical modeling?	1M	1	L2
c)	Name the three measures of central tendency.	1M	2	L2
d)	What is a binary attribute?	1M	2	L1
e)	Define a data frame in R.	1M	3	L1
f)	What is the purpose of a factor in R?	1M	3	L2
g)	What is the purpose of relational operators in R?	1M	4	L2
h)	Which loop in R executes statements repeatedly while a condition is TRUE?	1M	4	L4
i)	Name any two types of regression analysis used in R.	1M	5	L2
j)	What is the purpose of a box plot?	1M	5	L3

Part-B

Answer All the following questions. (5X10M=50Marks)		Marks	CO	BTL
2	a) Explain the difference between Big Data and Data Science. b) Summarize the steps involved in setting up the R environment and discuss the basic structure of an R program.	5M 5M	1	L1 L2
	OR			
3	a) Describe the concept of Statistical Inference and explain the terms populations and samples with examples. b) Apply the concept of model fitting in a simple statistical problem and explain how overfitting can be detected and avoided.	5M 5M	1	L2 L3
4	a) List and describe the different types of attributes used in data analysis. b) Explain the measures of dispersion such as range, variance, standard deviation, and interquartile range.	5M 5M	2	L1 L2
	OR			
5	a) Differentiate between nominal and ordinal attributes with suitable examples.	5M	2	L2

	b) Illustrate how graphical displays (such as histograms, box plots, and scatter plots) can be used to describe basic statistical properties of data.	5M		L3
6	a) Describe how to create and name matrices in R and mention how matrix sub setting works. b) Differentiate between a matrix, an array, and a data frame in R.	5M 5M	3	L2 L3
	OR			
7	a) Explain the concept of factors in R, including factor levels and ordered factors, with suitable examples. b) Demonstrate how to convert a list to a vector and explain when such a conversion is useful in data manipulation.	5M 5M	3	L2 L3
8	a) Describe the syntax and usage of conditional statements (if, else if, else) in R with examples. b) Explain the concept of control flow in R and how relational and logical operators influence it.	5M 5M	4	L4 L3
	OR			
9	a) Explain the process of writing a function in R, including function definition and calling. b) Write an R program using a **for** loop to find the sum of all even numbers from 1 to 50.	5M 5M	4	L2 L4
10	a) List and describe different types of charts available in R for data visualization. b) Explain the purpose and interpretation of box plots and histograms in analyzing data distribution.	5M 5M	5	L5 L3
	OR			
11	a) Describe the concept of linear regression and how it models the relationship between two variables. b) Perform multiple linear regression analysis in R and explain how to interpret the regression coefficients.	5M 5M	5	L3 L5