



PRESIDENCY UNIVERSITY

BENGALURU

Roll No.													
----------	--	--	--	--	--	--	--	--	--	--	--	--	--

End - Term Examinations – May / June 2026

Date: 23- 05- 2026

Time:09:30am – 12:30pm

School: SOCSE	Program: CSD	
Course Code: CSD1712	Course Name: Statistical Foundations of Data Science	
Semester: IV	Max Marks: 100	Weightage: 50%

CO - Levels	C01	C02	C03	C04	C05
Marks	24	24	24	14	14

Instructions:

- (i) Read all questions carefully and answer accordingly.
- (ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks.

10Q x 2M=20M

1.	What is dimensionality in Big Data?	2 Marks	L1	C01
2.	How is Big Data used in Biological Sciences?	2 Marks	L1	C01
3.	What is Weighted Least Squares (WLS)?	2 Marks	L1	C02
4.	Define the Box-Cox Transformation.	2 Marks	L1	C02
5.	What is inference for the noise level in linear regression?	2 Marks	L1	C03
6.	What is inference in generalized linear models (GLMs)?	2 Marks	L1	C03
7.	What is the main objective of PCA.	2 Marks	L1	C04
8.	Define the Power Method used in PCA.	2 Marks	L1	C04
9.	Define sparse K-means clustering.	2 Marks	L1	C05
10.	What is correlation screening?	2 Marks	L1	C05

Part B**Answer the Questions.****Total Marks 80M**

11.	a.	Explain the role of Big Data in Earth Sciences and Astronomy.	10 Marks	L2	CO1
	b.	Outline the impact of Big Data on dimensionality and data analysis.	10 Marks	L2	CO1
Or					
12.	a.	Explain noise accumulation and its effects in high-dimensional data.	10 Marks	L2	CO1
	b.	Explain spurious correlation, and why is it a problem in Big Data?	10 Marks	L2	CO1

13.	a.	Explain Multiple Linear Regression and the Gauss–Markov Theorem.	10 Marks	L2	CO2
	b.	Explain Spline Regression and Multiple Covariates.	10 Marks	L2	CO2
Or					
14.	a.	Explain Ridge Regression, Penalized Least Squares, and Bias–Variance Tradeoff.	10 Marks	L2	CO2
	b.	Explain Bayesian Interpretation, Ridge Solution Path, and Kernel Ridge Regression.	10 Marks	L2	CO2

15.	a.	Explain asymptotic efficiency and Fisher information.	10 Marks	L2	CO3
	b.	Explain linear regression with random design.	10 Marks	L2	CO3
Or					
16.	a.	Explain partial linear regression.	10 Marks	L2	CO3
	b.	Explain Gaussian graphical models and penalized least squares inference.	10 Marks	L2	CO3

17.	a.	Apply the K-means clustering algorithm to partition a dataset into clusters and interpret the results.	10 Marks	L3	CO4
	b.	Experiment with the concept of factor models and their role in structured covariance learning.	10 Marks	L3	CO4
Or					
18.	a.	Explain the need for variable selection in clustering. Describe the Sparse K-Means algorithm in detail, including its objective function and how feature weights are learned.	10 Marks	L3	CO5
	b.	Define clustering and explain different types of clustering techniques. Provide a detailed explanation of model-based clustering.	10 Marks	L3	CO5