Roll No.											

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – December 2025

Date: 15-12-2025 **Time:** 1.00pm to 04.00pm

School: SOCSE	**Program:** B.Tech	
Course Code: CSE2500	**Course Name:** Data Analytics	
Semester: V	**Max Marks**: 100	**Weightage**: 50%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**26**	**24**	**26**	**24**	**-**

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.	Define data analysis.	**2 Marks**	**L1**	**CO1**
2.	Define structured and unstructured data with examples.	**2 Marks**	**L1**	**CO1**
3.	Define Data Transformation.	**2 Marks**	**L1**	**CO1**
4.	What is a sparse matrix?	**2 Marks**	**L1**	**CO2**
5.	Write any two advantages of data visualization.	**2 Marks**	**L1**	**CO2**
6.	What is meant by the correlation coefficient (r)?	**2 Marks**	**L1**	**CO3**
7.	Write the assumptions of Fisher’s exact test.	**2 Marks**	**L1**	**CO3**
8.	What is the main difference between one-tailed and two-tailed tests?	**2 Marks**	**L1**	**CO3**

9.	What is Least Square Method?	**2 Marks**	**L1**	**CO4**
10.	List two real-world applications of regression analysis.	**2 Marks**	**L1**	**CO4**

Part B

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

11.	**a.**	Compare structured, semi-structured, and unstructured data with suitable use cases.	**10 Marks**	**L2**	**CO1**
	b.	Explain the scales of data. Provide examples of each and discuss their applications in statistical analysis.	**10 Marks**	**L3**	**CO1**
		Or			
12.	**a.**	Compute and normalize following group of data: 1000,2000,3000,9000 using min-max normalization by setting min:0 and max:1	**10 Marks**	**L2**	**CO1**
	b.	Illustrate the data analysis process and its objectives in business decision-making.	**10 Marks**	**L3**	**CO1**

13.	**a.**	Illustrate the use of the ggplot2 package in R for data visualization. Explain its layered structure and create two suitable plots (a bar chart and a scatter plot) with example data.	**10 Marks**	**L2**	**CO2**
	b.	Identify the missing data in a datasets using functions.	**10 Marks**	**L2**	**CO2**
		Or			
14.	**a.**	Explain in detail about the data manipulation techniques.	**10 Marks**	**L2**	**CO2**
	b.	Briefly explain about the recoding variables and the R functions used for.	**10 Marks**	**L2**	**CO2**

15.	**a.**	A bottling machine manufacturer claims that only 2% of bottles are underfilled (defective). A quality engineer inspects 500 bottles and finds 15 underfilled. At $\alpha = 0.05$ (two-tailed), test the claim. [Hint: CV = 1.96]	**20 Marks**	**L3**	**CO3**

		Or			
16.	**a.**	A researcher wants to determine whether a new teaching method improves students' performance compared to the traditional method. The following data were collected from 50 students, divided equally into two groups of 25 each: Teaching Method: Traditional — Pass 18, Fail 7; New Method — Pass 24, Fail 1 **i.** State the null (H_0) and alternative (H_1) hypotheses for the study. **ii.** Using the Fisher's Exact Test, determine whether there is a significant association between the teaching method and the exam result. **iii.** Use a significance level (α) = 0.05, p value = 0.02 **iv.** Interpret your result in the context of the problem.	**20 Marks**	**L3**	**CO3**

Teaching Method	Pass	Fail
Traditional	18	7
New Method	24	1

17.	**a.**	By the method of least squares, find the straight line that best fits the following data:	**20 Marks**	**L2**	**CO4**
		Or			
18.	**a.**	Demonstrate on Linear regression analysis, types, performance, with its evaluation metrics.	**20 Marks**	**L2**	**CO4**

x	**1**	**3**	**4**	**6**	**8**	**9**	**11**	**14**
y	**1**	**2**	**4**	**4**	**5**	**7**	**8**	**9**