Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

Make-up Examinations – December 2025

Date: 29- 12- 2025 **Time:** 01:00pm – 04:00pm

School: SOCSE	**Program:** B.Tech	
Course Code: CSE3190	**Course Name:** Fundamentals of Data Analytics	
Semester: MK	**Max Marks**:100	**Weightage**: 50%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**26**	**16**	**39**	**19**	**00**

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** any two human generated unstructured data.	**2 Marks**	**L1**	**CO1**
2	**Describe** Dichotomous Variable with examples.	**2 Marks**	**L1**	**CO1**
3	**State** some functions to deal with missing data.	**2 Marks**	**L1**	**CO1**
4	**Recognize** the importance of recording variables in data cleaning?	**2 Marks**	**L1**	**CO2**
5	**Define** the term hypothesis with its types.	**2 Marks**	**L1**	**CO3**
6	**Identify** the R command to generate a vector of integers ranging from 2 to 10.?	**2 Marks**	**L1**	**CO2**
7	**Define** a Line of Best Fit in regression analysis.	**2 Marks**	**L1**	**CO4**
8	**Explain** how the correlation T-test contributes to statistical analysis and its significance in determining relationships between variables.	**2 Marks**	**L1**	**CO3**

9	**Identify** and **describe** the differences between the cbind() and rbind() functions in R, providing an example for each.	**2 Marks**	**L1**	**CO2**
10	**List** the different types of logistic regression methods.	**2 Marks**	**L1**	**CO4**

Part B

Answer the Questions **Total 80 Marks**

11.	**a.**	Define data analysis. **Briefly explain** the types of data analysis.	**10 Marks**	**L2**	**CO1**

or

12.	**a.**	**Describe** the term 'Normalization.' Briefly **explain** min-max normalization. Compute and normalize the following data set: 45, 50, 60, 72, 90, using min-max normalization with a minimum value of 0 and a maximum value of 1.	**10 Marks**	**L2**	**CO1**

13.	**a.**	Describe the concept of a list in R and provide an example. **Demonstrate** how to access and update elements within lists.	**10 Marks**	**L3**	**CO2**

or

14.	**a.**	**Briefly explain** a data frame with its proper syntax. **Examine** the creation of a data frame that consists of the roll numbers of five students along with their names, mid-term marks, and end-term marks.	**10 Marks**	**L3**	**CO2**

15.	**a.**	**Illustrate** the use of R as a calculator application by performing mathematical functions on the console, and write an R script to create R objects for the calculator application.	**10 Marks**	**L3**	**CO1**

Or

16.	**a.**	A large hospital, **CityCare Medical Center**, is struggling with high rates of patient readmissions within 30 days of discharge, particularly among patients with heart failure. The hospital administration aims to reduce readmission rates by leveraging data analytics to understand the situation and improve patient outcomes. **i)** **Describe** the role of each of the four types of data analytics (descriptive, diagnostic, predictive, and prescriptive) in helping CityCare Medical Center reduce patient readmissions. **ii)** **Summarize** specific insights gained from the use of diagnostic and predictive analytics and explain how these insights guided the hospital's decision-making.	**10 Marks**	**L2**	**CO1**

17.	**a.**	Ramesh is working on a statistics paper for his postgraduate course. He meets his friend Amal, a textile engineer. During their conversation, Ramesh, who is currently interning at ABC Researchers, brings up a question that interests him. He asks Amal for his thoughts on the matter. The question is as follows: "The data regarding the sales of soft drinks and cotton clothes in the region over the last 12 months are provided. **Determine** if there is any association between the sales of soft drinks and cotton clothes. If a relationship exists, explain the possible reasons for it."	**15 Marks**	**L2**	**CO3**

Or

18.	**a.**	**Explain** the purpose of a Chi-Squared test with its steps. A researcher wants to determine if there is an association between gender (Male/Female) and preference for a product (Like/Dislike). They conduct a survey with the following results: The researcher wants to test, at a 5% significance level, whether gender is associated with product preference. (Note: At a 5% significance level ($\alpha = 0.05$), the critical value for 1 degree of freedom is 3.841.)	**15 Marks**	**L2**	**CO3**

	Like	Unlike	Total
Male	30	20	50
Female	20	30	50
Total	50	50	100

19.	**a.**	i) **Explain** the concept of the Least Squares Method and its significance in regression analysis. ii) Find the line of best fit for the following data points using the Least Squares Method: (x,y)=(1,3),(2,4),(4,8),(6,10),(8,15). iii) **Describe** the concept of Linear Least Squares by explaining the following key components: a. Data points b. Line of best fit c. Residuals	**15 Marks**	**L2**	**CO4**
		Or			
20.	**a.**	i) **Define** linear regression and classify its types. ii) **Summarize** the evaluation metrics used for regression models. iii) **Explain** logistic regression and its types in detail.	**15 Marks**	**L2**	**CO4**

21.	**a.**	Write an R program to analyze a dataset containing information about 10 students, where each student has the following attributes: • Student: A numeric identifier for each student (1 to 10). • Hours_Studied: The number of hours each student studied (ranging from 2 to 10). • Exam_Score: The exam score obtained by each student (ranging from 42 to 95). **Perform the following tasks:** 1. **Calculate** the mean, median, and standard deviation for both Hours_Studied and Exam_Score. 2. **Determine** the following visualizations using ggplot2: i) A scatter plot of Hours_Studied vs. Exam_Score with a regression line. ii) A histogram for Hours_Studied. iii) A histogram for Exam_Score.	**20 Marks**	**L3**	**CO3**

OR

22.	**a.**	Write an R program to analyze temperature data for six cities (New York, Los Angeles, Chicago, Houston, Phoenix, and Philadelphia) across four seasons (Winter, Spring, Summer, and Fall). The dataset contains 600 temperature observations. Perform the following tasks: 1. Generate the dataset with: i) Cities repeated 100 times each. ii) Seasons distributed equally across the dataset. iii) Temperature values randomly generated with a mean of 60 and a standard deviation of 10. 2. **Examine** the following visualizations using ggplot2: i) A line plot showing the temperature trend for each city across seasons. ii) A bar chart showing the average temperature for each city. iii) A box plot comparing the temperatures across cities.	**20 Marks**	**L3**	**CO3**