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



 **PRESIDENCY UNIVERSITY**

  **Bengaluru**

|  |
| --- |
| **End - Term Examinations – JANUARY 2025** |
| **Date:** 03-01-2025 **Time:** 09:30 am – 12:30 pm |

|  |  |
| --- | --- |
| **School:** SOCSE | **Program:** B.Tech (COM/CEI) |
| **Course Code :** CSE2027 | **Course Name :** FUNDAMENTALS OF DATA ANALYSIS |
| **Semester**: V | **Max Marks**: 100 | **Weightage**: 50% |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **CO - Levels** | **CO1** | **CO2** | **CO3** | **CO4** | **CO5** |
| **Marks** | **14** | **14** | **24** | **24** | **24** |

**Instructions:**

1. *Read all questions carefully and answer accordingly.*
2. *Do not write anything on the question paper other than roll number.*

**Part A**

|  |
| --- |
| **Answer ALL the Questions. 10 x 2 Marks=20 Marks** |
| **1** | Define the term "Veracity" in the context of big data. | 2 Marks | L1 | CO1 |
| **2** | Mention any two challenges faced in real-world data collection. | 2 Marks | L1 | CO1 |
| **3** | Define standard error and explain its importance in inferential statistics. How does the standard error change with sample size? | 2 Marks | L1 | CO2 |
| **4** | Define Skewness with its types. | 2 Marks | L1 | CO2 |
| **5** | In what ways can focus group discussions serve as an effective method of primary data collection? | 2 Marks | L1 | CO3 |
| **6** | Explain how a questionnaire can be structured to minimize bias in responses. | 2 Marks | L1 | CO3 |
| **7** | What type of chart is ideal for showing relationships between two variables? Precisely explain. | 2 Marks | L1 | CO4 |
| **8** | What is a pivot table used for? | 2 Marks | L1 | CO4 |
| **9** | What is the primary purpose of a regression model? | 2 Marks | L1 | CO5 |
| **10** | Why is non-linear regression more complex than linear regression? | 2 Marks | L1 | CO5 |

**Part B**

|  |
| --- |
| **Answer the Questions Total 80 Marks.** |
| **11.** | **a.** | Explain the differences between nominal, ordinal, interval, and ratio variables. Provide examples of each and discuss their applications in statistical analysis. | **10 Marks** | **L1** | **CO1** |
|  | **b.** | A fitness trainer claims that their new workout program helps people lose more weight. Two groups of 8 participants each were tested: one followed the new program (Group 1), and the other followed the standard program (Group 2). The weight loss (in kg) is recorded as:Group 1: 5, 6, 7, 8, 6, 5, 7, 6Group 2: 4, 5, 6, 4, 5, 4, 5, 4At α=0.05 test if there is a significant difference in weight loss between the two programs.[Note: Critical Value = 2.145; p value = 0.0018] | **10 Marks** | **L3** | **CO2** |
| **Or** |
| **12.** | **a.** | Discuss the importance of data preparation in the data analysis pipeline. Explain the steps of removing irrelevant variables and applying data transformations with examples. | **10 Marks** | **L1** | **CO1** |
|  | **b.** | A survey is conducted to check the preference for two brands, Brand X and Brand Y, among customers of two age groups: Below 30 and 30 or above. The partially filled table is given below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Preference** | **Below 30** | **30 or Above** | **Total** |
| **Brand X** | 40 | ? | 100 |
| **Brand Y** | ? | 60 | 120 |
| **Total** | 80 | 100 | ? |

1. Fill in the missing values in the table.
2. Calculate P(Below 30), P(Brand X), and P(Below 30 ∩ Brand X).
3. Are the events "Below 30" and "Prefers Brand X" independent?
4. Compute P(Below 30 | Brand X).

Compute P(Below 30∪Brand X). | **10 Marks** | **L3** | **CO2** |
|  |  |  |  |  |  |
| **13.** | **a.** | Compare and contrast the effectiveness of observation versus survey methods for collecting data on employee engagement during virtual meetings. Which method would you recommend for this research topic, and why? | **10 Marks** | **L1** | **CO3** |
|  | **b.** | A management team wants feedback from leaders on the effectiveness of current remote work policies.·Hint: Conduct structured interviews with department heads, asking them the same set of questions about the impact of remote work on team performance and employee morale. This provides consistent feedback that can be analyzed quantitatively. | **10 Marks** | **L1** | **CO3** |
| **or** |
| **14.** | **a.** | What is the observation method of data collection? Discuss its types, strengths, and weaknesses. Provide examples of situations where this method is particularly effective. | **10 Marks** | **L1** | **CO3** |
|  | **b.** | Define secondary data and explain the methods used for its collection. Discuss the advantages and limitations of using secondary data in research studies. | **10 Marks** | **L1** | **CO3** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **15.** | **a.** | Discuss various types of charts and their significance in data visualization. Provide examples of when each type is most effectively used. | **10 Marks** | **L2** | **CO4** |
|  | **b.** | A neighborhood tracks electricity consumption (in kWh) for 20 households:215, 330, 400, 245, 310, 275, 380, 420, 260, 290, 305, 375, 395, 255, 240, 265, 335, 410, 285, 3001. . Use intervals 200−249, 250−299, 300−349, 350−399, 400−449

(ii). Calculate the relative and cumulative frequencies. | **10 Marks** | **L3** | **CO4** |
| **Or** |
| **16.** | **a.** | Explain the concept of a pivot table and its applications in data analysis. Provide an example to illustrate its use in summarizing large datasets. | **10 Marks** | **L2** | **CO4** |
|  | **b.** | A meteorologist records daily rainfall (in mm) over 20 days:12, 25, 30, 18, 40, 15, 35, 20, 50, 28, 32, 22, 48, 26, 42, 24, 38, 45, 17, 36(i). Organize the data into intervals 10−19, 20−29, 30−39, 40−49, 50−59.(ii). Calculate the relative and cumulative frequencies. | **10 Marks** | **L3** | **CO4** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **17.** | **a.** | If is the pull required to lift a load by means of a pulley block, find a linear law of the form connecting , using the following data.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **P** | **12** | **15** | **21** | **25** |
| **W** | **50** | **70** | **100** | **120** |

where P & W are taken in kg-wt. Compute P when W = 150kg-wt. | **10 Marks** | **L3** | **CO5** |
|  | **b.** | Discuss the concept of classification in machine learning. Explain its process and provide examples of algorithms used for classification tasks. | **10 Marks** | **L1** | **CO5** |
| **Or** |
| **18.** | **a.** | The fuel consumption F (in liters) of a car is given for various distances traveled D (in kilometers) as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **D** | **50** | **100** | **150** | **200** |
| **F** | **4** | **8** | **12** | **16** |

(i). Determine the linear relationship between F and D.(ii). Predict the fuel consumption when the distance traveled is 250 km. | **10 Marks** | **L3** | **CO5** |
|  | **b** | Describe in detail about the Prediction model and its phases. | **10 Marks** | **L1** | **CO5** |

**\*\*\*\*\* BEST WISHES \*\*\*\*\***