Name:

Enrolment No:

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES (UPES)
End Semester Examination, May 2023

Course: Data Analysis and Modelling Technique **Semester: VIII**
Program: B.Tech (CSE-AIML) **Time : 03 hrs.**
Course Code: CSBA 4014 **Max. Marks: 100**

Instructions: All the questions are compulsory.

SECTION A
(5Qx4M=20Marks)
Explain max by 50-60 words wherever required. Attempt all questions from Sec A.

S. No.		Marks	CO
Q1	How do you test a small sample hypothesis? And state the basic difference between null hypothesis and alternative hypothesis.	2+2=4	CO3
Q2	Differentiate between discrete and continuous random variable.	4	CO1
Q3	A bag contains 5 white and 3 black balls. Two balls are drawn at random one after the other without replacement. Find the probability that both the balls drawn are black.	4	CO1
Q4	You got a dataset depicting the popularity of two graphic novels given by a critic which contains three variables. 1) Time of survey (in dd-mm-yy format) 2) Rating of 'Marvel' (in range between 0 to 10) 3) Rating of 'DC' (in range between 0 to 10) The data is collected every day since 1970. You need to graphically represent the data in a chart. What will you use? And why?	4	CO2
Q5	What is the probability of throwing a number greater than 3 with an ordinary dice?	4	CO1

SECTION B
(4Qx10M= 40 Marks)
Each question will carry 10 marks. Write short / brief notes (Explain max by 100-150 words wherever required).

Q6	**a)** What do you understand by the term descriptive statistics. **(3 marks)** **b)** Provide an example of descriptive statistics with diagrammatic representation? **(7 marks)**	10	CO2
Q7	**a)** Analyzing the Mid-Sem marks for students. The following data was observed.	10	CO2

Sl no	Total Students
0-10	5
10-20	3
20-30	5

	<table><tr><td>30-40</td><td>8</td></tr><tr><td>40-50</td><td>16</td></tr><tr><td>50-60</td><td>18</td></tr><tr><td>60-70</td><td>5</td></tr><tr><td>70-80</td><td>3</td></tr><tr><td>80-90</td><td>2</td></tr><tr><td>90-100</td><td>0</td></tr></table> (i) Compute the Skewness present in the data? What can you conclude? **(4 Marks)** (ii) Compute the kurtosis. What is the observation indicating? **(4 Marks)** **b)** A distribution has Q1= 31.3 and median = 35, and Q3 = 36.4. Calculate the co efficient of skewness. **(2 Marks)**		
Q8	Write short note on: (**Attempt any two**) **(2*5)** **(i)** Z Test. **(ii)** T Test. **(iii)** Bayesian Network. **(iv)** Maximum likelihood estimation.	**10**	**CO3**
Q9	Explain the Central Limit theorem and state the merits ,demerits and uses of standard deviation with a basic example. **(6+4)**	**10**	**CO1**
SECTION-C **(2Qx20M=40 Marks) Each Question carries 20 Marks. Instruction: Write long answer. Explain max by 200 words wherever required. Make diagrams wherever needed.**			
Q 10	**<u>Attempt 10(a) or 10(b)</u>** **a)** Explain the concept and working principle of the Monte Carlo simulation along with their advantages and disadvantages. **(20 marks)** **OR** **b)** Explain the basic concepts of Hidden Markov Model(HMM) including **i)** Markov chain, **ii)** definition of HMM, **iii)** HMM assumptions, **iv)** Computing Likelihood: The Forward Algorithm, **v)** Learning in HMM, **vi)** Advantages and Disadvantages of HMM). **(3+2+4+5+2+4)**	**20**	**CO4**
Q 11	**a)** Given the following statistics, what is the probability that a woman has cancer if she has a positive mammogram result? 1. 1% of women have cancer. 2. 90% of women who have cancer test positive on mammograms. 3. 8% of women will have false positives. **(8 marks)** **b)** How to find f test and t test p values? **(6 marks)** **c)** Difference between Bayesian Network and Markov model? **(6 marks)**	**20**	**CO3, CO4**