


Name:			
Enrolment No:			
UPES End Semester Examination, Dec 2023			
Course: Applied Statistics Program: B.Tech CSE – AI&ML (H & NH) Course Code: CSBA2010		Semester: III Time: 03 hrs. Max. Marks: 100	
Instructions: All questions are Compulsory			
SECTION A (5Qx4M=20Marks)			
S. No.		Marks	CO
Q 1	Distinguish between ‘skewness’ and ‘kurtosis’. Bring out their importance in describing frequency distribution.	4	CO1
Q2	Suppose a company's production line produces light bulbs with a mean lifespan of 800 hours and a standard deviation of 40 hours, following a normal distribution. What is the probability that a randomly selected light bulb from this production line will have a lifespan between 760 and 840 hours?	4	CO2
Q3	In a university with 5,000 students, how would you conduct a simple random sample of 200 students for a survey on campus food preferences? Outline the steps and considerations involved in performing this sampling method. Or Distinguish between discrete and continuous random variables. Provide examples of each type of variable and explain why they fall into their respective category.	4	CO1 CO2
Q4.	Two variables, X and Y, were measured in a study involving 15 participants. The correlation coefficient between X and Y was calculated to be +0.85. Describe what this correlation coefficient value indicates about the relationship between variables X and Y.	4	CO4
Q5.	What is hypothesis testing? What is the Type I and Type II errors in hypothesis testing?	4	CO3
SECTION B (4Qx10M= 40 Marks)			
Q1.	Discuss the utility of a boxplot in identifying outliers within a dataset. Provide an example of a dataset and its corresponding boxplot where outliers are clearly visible and explain how the boxplot aids in outlier detection.	10	CO1

Q2.	<p>A pharmaceutical company claims that a certain drug extends the average lifespan of lab mice to 36 months. A group of researchers believes the drug's effect might be overstated. They administer the drug to a sample of 25 lab mice and find their average lifespan to be 34 months with a standard deviation of 5 months.</p> <p>a) State the null and alternate hypotheses. b) At a 95% confidence level, is there enough evidence to reject the null hypothesis? OR Enumerate and elucidate various approaches utilized to ascertain central tendency within a dataset. Support your explanation with a unique instance for each method.</p>	10	CO2								
Q3.	<p>The average weight of apples in a grocery store is claimed to be 150 grams. A sample of 50 apples yielded an average weight of 145 grams with a standard deviation of 10 grams. At a significance level of 0.05, conduct a Z-test to determine if there is evidence to refute the store's claim about apple weights</p>	10	CO3								
Q4.	<p>Differentiate between linear and nonlinear correlations. Offer examples of scenarios where a linear correlation might be observed and situations where a nonlinear correlation is more likely.</p>	10	CO4								
SECTION-C (2Qx20M=40 Marks)											
Q1.	<p>If you have a dataset consisting of a series of pairs of values (X, Y), how does linear regression determine the line of best fit? Explain the process step by step.</p> <p style="text-align: center;">OR</p> <p>Using the data below, test the claim that there is no difference in the color preferences of men and women. Use $\alpha=.05$</p> <table border="1" data-bbox="240 1398 1162 1566"> <thead> <tr> <th>Color</th> <th>Red</th> <th>Green</th> <th>Blue</th> </tr> </thead> <tbody> <tr> <td>Number of occurrences</td> <td>32</td> <td>45</td> <td>23</td> </tr> </tbody> </table>	Color	Red	Green	Blue	Number of occurrences	32	45	23	20	CO5 CO4
Color	Red	Green	Blue								
Number of occurrences	32	45	23								
Q2.	<p>Elucidate the following and where to apply them with example?</p> <p>a) Cluster Analysis b) Factor analysis c) Multidimensional scaling d) Discriminant functional analysis</p>	4 x 5=20	CO5								