

Name:	
Enrolment No:	

UPES
End Semester Examination, May 2023

Course: Applied Statistical Analysis	Semester: 6
Program: B.Tech CSE H & NH	Time: 03 hrs.
Course Code: CSBA3016P	Max. Marks: 100

Instructions: All questions are Compulsory

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO											
Q1.	Explain the research process methodology.	4	CO1											
Q2.	Create a box plot for the following dataset that shows the height of ten plants: <table border="1" style="margin: 10px auto; border-collapse: collapse;"> <thead> <tr> <th style="background-color: #90EE90;">Plant height (inches)</th> </tr> </thead> <tbody> <tr><td style="text-align: center;">14</td></tr> <tr><td style="text-align: center;">16</td></tr> <tr><td style="text-align: center;">12</td></tr> <tr><td style="text-align: center;">11</td></tr> <tr><td style="text-align: center;">24</td></tr> <tr><td style="text-align: center;">19</td></tr> <tr><td style="text-align: center;">13</td></tr> <tr><td style="text-align: center;">12</td></tr> <tr><td style="text-align: center;">20</td></tr> <tr><td style="text-align: center;">10</td></tr> </tbody> </table>	Plant height (inches)	14	16	12	11	24	19	13	12	20	10	4	CO1
Plant height (inches)														
14														
16														
12														
11														
24														
19														
13														
12														
20														
10														
Q3.	A personal computer has the length of time between charges of the battery is normally distributed with a mean of 66 hours and a standard deviation of 20 hours. Find the probability when the length of time will be between 58 and 75 hours.	4	CO2											
Q4.	Find the t-test value for the following two sets of values: 7, 2, 9, 8 and 1, 2, 3, 4?	4	CO2											

Q5.	Consider the time series data given below:										4	CO5	
	x_i	8	3	2	10	11	3	6	5	6			8
	y_i	4	12	1	12	9	4	9	6	1			14
Use the least square method to determine the equation of line of best fit for the data.													

SECTION B
(4Qx10M= 40 Marks)

Q 1.	Elaborate non-probability Sampling. Discuss any 4 types of Non-Probability Sampling Methods. OR Discuss the different types of Analytics with suitable examples.	10	CO1
------	--	----	-----

Q2.	A survey was conducted in your city. Given is the following sample data containing a person's age and their corresponding income. Find out whether the increase in age has an effect on income using the correlation coefficient formula. (Use $1/\sqrt{1811181}$ as 0.074 and $1/\sqrt{2091209}$ as 0.07)	10	CO2										
	<table border="1" style="width: 100%; text-align: center;"> <tr> <td>Age</td> <td>25</td> <td>30</td> <td>36</td> <td>43</td> </tr> <tr> <td>Income</td> <td>30000</td> <td>44000</td> <td>52000</td> <td>7000</td> </tr> </table>			Age	25	30	36	43	Income	30000	44000	52000	7000
Age	25			30	36	43							
Income	30000	44000	52000	7000									

Q3.	Three types of fertilizers are used on three groups of plants for 5 weeks. We want to check if there is a difference in the mean growth of each group. Using the data given below apply a one way ANOVA test at 0.05 significant level.	10	CO4															
	<table border="1" style="width: 100%; text-align: center;"> <tr> <th>Fertilizer 1</th> <th>Fertilizer 2</th> <th>Fertilizer 3</th> </tr> <tr> <td>6</td> <td>8</td> <td>13</td> </tr> <tr> <td>8</td> <td>12</td> <td>9</td> </tr> <tr> <td>4</td> <td>9</td> <td>11</td> </tr> <tr> <td>5</td> <td>11</td> <td>8</td> </tr> </table>			Fertilizer 1	Fertilizer 2	Fertilizer 3	6	8	13	8	12	9	4	9	11	5	11	8
Fertilizer 1	Fertilizer 2			Fertilizer 3														
6	8			13														
8	12			9														
4	9	11																
5	11	8																

	3	6	7					
	4	8	12					
Q4.	A random sample of 30 apples was taken from a large population. On measuring their diameter the mean diameter of the sample was 91 millimeters with a standard deviation of 8 mm. Calculate the 85% confidence limits for the mean diameter of the whole population of apples.				10	CO3		
SECTION-C (2Qx20M=40 Marks)								
Q1.	With the help of a suitable diagram explain the following terms: (4 marks each) 1. Significance level 2. Critical Region 3. Critical Value 4. Confidence Interval 5. P-Value				20	CO3		
Q2.	The following data shows the sales (in million dollars) of a company.				20	CO5		
	x	2015	2016	2017			2018	2019
	y	12	19	29			37	45
	Estimate the sales in the year 2020 using the regression line.							
	OR							
	Calculate the Chi-square value for the following data of incidences of water-borne diseases in three tropical regions.							
		India	Equador	South America	Total			
	Typhoid	31	14	45	90			
	Cholera	2	5	53	60			
	Diarrhoea	53	45	2	100			

		86	64	100	250			
--	--	----	----	-----	-----	--	--	--