

Name:

Enrolment No:



UPES

End Semester Examination, December 2023

Course: Statistics for Data Science

Program: Bachelor of Computer Application

Course Code: CSBD 2009 P

Instructions: Attempt all the questions. All questions are compulsory.

Semester: III

Time : 03 hrs.

Max. Marks: 100

SECTION A  
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Define the sample and population with examples.	4	CO1
Q 2	Define the discrete random variable with example.	4	CO2
Q 3	Define the Type I and Type II Error.	4	CO3
Q 4	Define the covariance. What are the types of correlation?	4	CO4
Q 5	Define the discriminant function analysis with examples.	4	CO5

SECTION B  
(4Qx10M= 40 Marks)

Q 6	Determine the average marks from the following data:						10	CO1	
	Marks:	0 - 10	10 - 20	20 - 30	30 - 40	40 - 50			50 - 60
	No. of students	42	44	58	35	26			15
Q 7	If three unbiased coins are tossed. Determine the expectation and the variance of the number of heads.						10	CO1	
Q 8	What is factor analysis? What are the types of factor analysis? Write the advantages and disadvantages of the factor analysis.						10	CO5	
Q 9	A sample of 400 male students of a college is found to have a mean height of 171.38 cm. Can it be regarded as a sample from a large population with mean height 171.17 cm and standard deviation 3.30 cm. (Table value of Z at 5% level is 1.96)						10	CO3	
	<b>OR</b>								
A random sample of 900 items is taken from a normal population whose mean and the variance are 4. Can the sample with mean 4.5 be regarded as truly random one at 1% level of significance? (Table value of Z at 1% level is 2.58)									

**SECTION-C**  
**(2Qx20M=40 Marks)**

<b>Q 10 A</b>	The probability that a product manufactured by a company will be defective is $\frac{1}{8}$ . If 20 such products are manufactured, find the probability that (i) none will be defective, (ii) at least two will be defective, and (iii) exactly three will be defective.	<b>10</b>	<b>CO2</b>																																
<b>Q 10 B</b>	Draw a pie diagram to represent the following data of proposed expenditure by a state Government for the year 2001-2002.	<b>10</b>	<b>CO2</b>																																
	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th style="width: 25%;">Items</th> <th style="width: 25%;">Agri. &amp; Rural Development</th> <th style="width: 25%;">Indus.&amp; Urban Development</th> <th style="width: 25%;">Health &amp; Education</th> <th style="width: 20%;">Miscell.</th> </tr> </thead> <tbody> <tr> <td>Proposed Expend. (in million Rs.)</td> <td style="text-align: center;">4,200</td> <td style="text-align: center;">1,500</td> <td style="text-align: center;">1,000</td> <td style="text-align: center;">500</td> </tr> </tbody> </table>	Items	Agri. & Rural Development	Indus.& Urban Development	Health & Education	Miscell.	Proposed Expend. (in million Rs.)	4,200	1,500	1,000	500																								
Items	Agri. & Rural Development	Indus.& Urban Development	Health & Education	Miscell.																															
Proposed Expend. (in million Rs.)	4,200	1,500	1,000	500																															
<b>Q 11 A</b>	Determine the Karl Pearson's coefficient of correlation from the following data:	<b>10</b>	<b>CO4</b>																																
	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tbody> <tr> <td style="width: 15%;">Independent variable (x)</td> <td style="width: 10%;">3</td> <td style="width: 10%;">7</td> <td style="width: 10%;">5</td> <td style="width: 10%;">4</td> <td style="width: 10%;">6</td> <td style="width: 10%;">8</td> <td style="width: 10%;">2</td> <td style="width: 10%;">7</td> </tr> <tr> <td>Dependent variable (y)</td> <td>7</td> <td>12</td> <td>8</td> <td>8</td> <td>10</td> <td>13</td> <td>5</td> <td>10</td> </tr> </tbody> </table> <p style="text-align: center;"><b>OR</b></p> <p>Calculate the covariance of the following observations of the variables X and Y</p> <table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tbody> <tr> <td style="width: 10%;">X:</td> <td style="width: 10%;">15</td> <td style="width: 10%;">20</td> <td style="width: 10%;">25</td> <td style="width: 10%;">30</td> <td style="width: 10%;">40</td> <td style="width: 10%;">50</td> </tr> <tr> <td>Y:</td> <td>44</td> <td>43</td> <td>45</td> <td>37</td> <td>34</td> <td>37</td> </tr> </tbody> </table>	Independent variable (x)	3	7	5	4	6	8	2	7	Dependent variable (y)	7	12	8	8	10	13	5	10	X:	15	20	25	30	40	50	Y:	44	43	45	37	34	37		
Independent variable (x)	3	7	5	4	6	8	2	7																											
Dependent variable (y)	7	12	8	8	10	13	5	10																											
X:	15	20	25	30	40	50																													
Y:	44	43	45	37	34	37																													
<b>Q 11 B</b>	Two variables gave the following data: $\bar{X} = 20$ , $\bar{Y} = 15$ , $\sigma_x = 4$ , $\sigma_y = 3$ , $r = 0.7$ . Determine the regression equations and the most likely value of Y, when $X = 24$ .	<b>10</b>	<b>CO4</b>																																
	<p><b>OR</b></p> <p>Given the regression lines as <math>3x + 2y = 26</math> and <math>6x + y = 31</math>. Determine their point of intersection and interpret it. Also, find the correlation coefficient between x and y.</p>																																		