Name:

Enrolment No:

UPES

End Semester Examination, May 2024

Course: Applied Statistical Analysis | **Semester: VI**

Program: B.Tech. (CSE + All) | **Time : 03 hrs.**

Course Code: CSBA3016P | **Max. Marks: 100**

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

SECTION A
(5Qx4M=20Marks)

S. No.		Marks	CO
Q 1	Explain the experimental data and survey data with examples.	4	CO1
Q 2	Comment on the statement "a binomial variate has mean 4 and *s.d.* 3".	4	CO2
Q 3	Explain the Type I and Type II Errors with examples.	4	CO3
Q 4	Define the dependent and independent variables in terms of regression analysis with examples.	4	CO4
Q 5	Define the discriminant function analysis with examples.	4	CO5

SECTION B
(4Qx10M= 40 Marks)

Q 6	Draw the histogram and frequency curve for the following distribution: Age (Yrs): 17 -19 | 19 -21 | 21-23 | 23-25 | 25 -27 | 27 -29 | 29 - 31 No. of Students: 7 | 13 | 24 | 30 | 22 | 15 | 6	10	CO1
Q 7	The probability of a man hitting a target is $\frac{1}{3}$. (i) If he fires 5 times, what is the probability of his hitting the target at least twice? (ii) How many times must he fire so that the probability of his hitting the target at least once is more than 90%.	10	CO2
Q 8	What is factor analysis? What are the types of factor analysis? Write the advantages and disadvantages of the factor analysis.	10	CO5

<table>
<tr><td>Q 9</td><td>A die was thrown 9000 times and of these 3220 yielded a 3 or 4. Can the die be regarded as unbiased?
OR
A random sample of 10 women is selected and their weights (in lbs.), before and after they are put on a new diet, are recorded as given below:
<table><tr><td>Weight before diet (X)</td><td>180</td><td>178</td><td>165</td><td>200</td><td>160</td><td>145</td><td>170</td><td>210</td><td>185</td><td>155</td></tr><tr><td>Weight after diet (Y)</td><td>174</td><td>181</td><td>157</td><td>198</td><td>152</td><td>150</td><td>160</td><td>205</td><td>178</td><td>160</td></tr></table>Use the Wilcoxon Signed-rank test at a 0.05 level of significance to test the claim that the new diet is effective in reducing weight by P-value method.</td><td>10</td><td>CO3</td></tr>
<tr><td colspan="3">SECTION-C
(2Qx20M=40 Marks)</td></tr>
<tr><td>Q 10 A</td><td>Two judges in a beauty competition rank the 12 entries as follows:
<table><tr><td>X:</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td></tr><tr><td>Y:</td><td>12</td><td>9</td><td>6</td><td>10</td><td>3</td><td>5</td><td>4</td><td>7</td><td>8</td><td>2</td><td>11</td><td>1</td></tr></table>Evaluate the Spearman’s rank correlation coefficient between X and Y.</td><td>10</td><td>CO4</td></tr>
<tr><td>Q 10 B</td><td>Given the regression lines as $3x + 2y = 26$ and $6x + y = 31$. Determine their point of interaction and interpret it. Also, find the correlation coefficient between x and y.</td><td>10</td><td>CO4</td></tr>
<tr><td>Q 11 A</td><td>A trucking company wishes to test the average life of each of the four brands of tyres. The company uses all brands on randomly selected trucks. The records showing the lives (thousands of miles) of tyres are as given the table:
<table><tr><th>Brand 1</th><th>Brand 2</th><th>Brand 3</th><th>Brand 4</th></tr><tr><td>20</td><td>19</td><td>21</td><td>15</td></tr><tr><td>23</td><td>15</td><td>19</td><td>17</td></tr><tr><td>18</td><td>17</td><td>20</td><td>16</td></tr><tr><td>17</td><td>20</td><td>17</td><td>18</td></tr><tr><td></td><td>16</td><td>16</td><td></td></tr></table>Test the hypothesis that the average life for each brand of types is the same at 1% level of significance.
OR</td><td>20</td><td>CO3</td></tr>
</table>

To test the hypothesis that the average number of days a patient is kept in the three local hospitals say, A, B and C is the same, a random check on the number of days those seven patients stayed in each hospital reveals the following:

Hospital A:	8	5	9	2	7	8	2
Hospital B:	4	3	8	7	7	1	5
Hospital C:	1	4	9	8	7	2	3

Test the hypothesis that the average stay at each of the hospitals are same at 5% level of significance.

Table values of *t, Z* and *F*:

1	Table value of F at 5% level of for $v_1 = 2$ and $v_2 = 18$ is 3.55.
2	Table value of F at 1% level of for $v_1 = 3$ and $v_2 = 14$ is 5.56.
3	Table value of F at 5% level of for $v_1 = 2$ and $v_2 = 12$ is 3.88.
4	Table value of F at 5% level of for $v_1 = 3$ and $v_2 = 6$ is 4.76.
5	Table value of F at 5% level of for $v_1 = 2$ and $v_2 = 6$ is 5.14.
6	Table value of t at 1% level of for 48 *d.f.* for two-tailed test is 2.58.
7	Table value of Z at 5% level of significance for two-tailed is 1.96.
8	Table value of Z at 5% level of significance for right-tailed is 1.645.
9	The critical value of T (Signed-Rank Test) for n=10 at 5% level of significance for single tail test is 11.
10	The critical value of T (Signed-Rank Test) for n=9 at 5% level of significance for single tail test is 8.