



**DEPARTMENT OF DATASCIENCE**

**UNIT TEST-II**

**Class: TE**

**Semester: V**

**Subject: Statistics for AI&DS**

**Date: 21-10-2023**

**Time: 10:00am - 11:30am**

**Max marks: 40**

**Note the following instructions**

1. Attempt all questions.
2. Draw neat diagrams wherever necessary.
3. Write everything in Black ink (no pencil) only.
4. Assume data, if missing, with justification.

Q.N	Questions	MARKS	CO	Blooms Taxono my Level	PO2																								
Q.1.	Attempt any two.																												
	<p>a) The following stem-and-leaf plot shows the ages of a group of people in a room.</p> <div><div>1</div><div>7 8 9</div></div> <div><div>2</div><div>0 2 2 4 5 6</div></div> <div><div>3</div><div></div></div> <div><div>4</div><div>1 2 4</div></div> <p>2   4 means 24 years</p> <p>i) How many people were there in the room? ii) Two people have the same age. What is that age? iii) What is the mode, median and mean of the ages?</p>	[5]	CO4	L3	PO1, PO12																								
	<p>b) Give the use of scatter plots. Below is a table of 11 student's scores out of 100 on their Maths and English tests. Plot a scatter graph from this data and specify its relation.</p> <table><tr><td>Maths mark</td><td>38</td><td>62</td><td>18</td><td>75</td><td>38</td><td>59</td><td>66</td><td>92</td><td>52</td><td>75</td><td>48</td></tr><tr><td>English mark</td><td>74</td><td>44</td><td>85</td><td>19</td><td>88</td><td>69</td><td>50</td><td>33</td><td>29</td><td>32</td><td>56</td></tr></table>	Maths mark	38	62	18	75	38	59	66	92	52	75	48	English mark	74	44	85	19	88	69	50	33	29	32	56	[5]	CO4	L3	PO1, PO12
Maths mark	38	62	18	75	38	59	66	92	52	75	48																		
English mark	74	44	85	19	88	69	50	33	29	32	56																		

	c) Define trimmed mean. Charlie recorded the number of pushups he completed each day for 10 days as follows: 5, 4, 7, 6, 8, 10, 11, 0, 7, 18. Calculate the 20% trimmed mean.	[5]	CO4	L3	PO1, PO12																				
	d) A garden contains 39 plants. The following plants were chosen at random, and their heights were recorded in cm: 38, 51, 46, 79, and 57. Calculate their heights' standard deviation.	[5]	CO4	L3	PO1, PO12																				
Q.2.	a) A trucking company wishes to test the average life of each of the four brands of tyres. The company uses all the brands on randomly selected trucks. The records showing the lives (thousands of miles) of tyres are as given in the table: Test the hypothesis using one-way ANOVA that the average life for each brand of tyres is the same. (Critical value = 5.56) <table border="1"><thead><tr><th>Brand 1</th><th>Brand 2</th><th>Brand 3</th><th>Brand 4</th></tr></thead><tbody><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></tbody></table>	Brand 1	Brand 2	Brand 3	Brand 4	20	19	21	15	23	15	19	17	18	17	20	16	17	20	17	18	[10]	CO5	L3	PO1, PO12
Brand 1	Brand 2	Brand 3	Brand 4																						
20	19	21	15																						
23	15	19	17																						
18	17	20	16																						
17	20	17	18																						
	OR																								
	b) To study the performance of 3 detergents and 3 different water temperature, the following readings were obtained with specially designed equipment: <table border="1"><thead><tr><td></td><td>A</td><td>B</td><td>C</td></tr></thead><tbody><tr><td>Cold Water</td><td>47</td><td>45</td><td>50</td></tr><tr><td>Warm Water</td><td>39</td><td>42</td><td>52</td></tr><tr><td>Hot Water</td><td>44</td><td>36</td><td>48</td></tr></tbody></table> Perform a two way ANOVA using 5% level of significance. (Critical Value = 6.94)		A	B	C	Cold Water	47	45	50	Warm Water	39	42	52	Hot Water	44	36	48	[10]	CO5	L3	PO1, PO12				
	A	B	C																						
Cold Water	47	45	50																						
Warm Water	39	42	52																						
Hot Water	44	36	48																						
	AND																								

	<p>c) In order to the following data represents the number of units of tablet production(in thousands) per day by five different technicians by using 3 different type of machines.</p> <table><tr><th>Technician s</th><th>Machine X</th><th>Machine Y</th><th>Machine Z</th></tr><tr><td>A</td><td>54</td><td>48</td><td>57</td></tr><tr><td>B</td><td>56</td><td>50</td><td>62</td></tr><tr><td>C</td><td>44</td><td>46</td><td>54</td></tr><tr><td>D</td><td>53</td><td>48</td><td>56</td></tr><tr><td>E</td><td>48</td><td>52</td><td>59</td></tr></table>	Technician s	Machine X	Machine Y	Machine Z	A	54	48	57	B	56	50	62	C	44	46	54	D	53	48	56	E	48	52	59	[5]	CO5	L3	PO1, PO12
Technician s	Machine X	Machine Y	Machine Z																										
A	54	48	57																										
B	56	50	62																										
C	44	46	54																										
D	53	48	56																										
E	48	52	59																										
	OR																												
	<p>d) Consider there are three groups and their reaction time is measured. Check whether there is difference between the groups using Kruskal Wallis Test at 5% level of significance.</p> <table><tr><th>A</th><th>B</th><th>C</th></tr><tr><td>34</td><td>44</td><td>35</td></tr><tr><td>36</td><td>37</td><td>39</td></tr><tr><td>41</td><td>45</td><td>42</td></tr><tr><td>43</td><td>33</td><td>46</td></tr></table> <p>Conduct a Friedman Test with the given data and judge whether there is any difference among the machines. (Tabulated value = 5.99)</p>	A	B	C	34	44	35	36	37	39	41	45	42	43	33	46	[5]	CO5	L3	PO1, PO12									
A	B	C																											
34	44	35																											
36	37	39																											
41	45	42																											
43	33	46																											
Q.3.	<p>Find the linear regression of the data of weekend product sales(in Thousands) given in table. Use Linear regression in matrix form. Predict the 7th week sale.</p> <table><tr><th>X(Week)</th><th>Y(Sales in thousands)</th></tr><tr><td>1</td><td>1</td></tr><tr><td>2</td><td>3</td></tr><tr><td>3</td><td>4</td></tr></table>	X(Week)	Y(Sales in thousands)	1	1	2	3	3	4	[10]	CO6	L3	PO1, PO12																
X(Week)	Y(Sales in thousands)																												
1	1																												
2	3																												
3	4																												

	4	8																												
	<b>OR</b>																													
	b) Find the multiple regression equation using the below data:			[10]	CO6	L3  PO1, PO12																								
	<table><tr><td>Subject</td><td>Y</td><td>X1</td><td>X2</td></tr><tr><td>1</td><td>-3.7</td><td>3</td><td>8</td></tr><tr><td>2</td><td>3.5</td><td>4</td><td>5</td></tr><tr><td>3</td><td>2.5</td><td>5</td><td>7</td></tr><tr><td>4</td><td>11.5</td><td>6</td><td>3</td></tr><tr><td>5</td><td>5.7</td><td>2</td><td>1</td></tr></table>			Subject	Y	X1	X2	1	-3.7	3	8	2	3.5	4	5	3	2.5	5	7	4	11.5	6	3	5	5.7	2	1			
Subject	Y	X1	X2																											
1	-3.7	3	8																											
2	3.5	4	5																											
3	2.5	5	7																											
4	11.5	6	3																											
5	5.7	2	1																											
	<b>AND</b>																													
	c) Find the simple linear regression equation using the below data:			[5]	CO6	L3  PO1, PO12																								
	<table><tr><td>Hour</td><td>Temp</td></tr><tr><td>2</td><td>21</td></tr><tr><td>4</td><td>27</td></tr><tr><td>6</td><td>29</td></tr><tr><td>8</td><td>86</td></tr><tr><td>10</td><td>86</td></tr><tr><td>12</td><td>92</td></tr></table>			Hour	Temp	2	21	4	27	6	29	8	86	10	86	12	92													
Hour	Temp																													
2	21																													
4	27																													
6	29																													
8	86																													
10	86																													
12	92																													
	<b>OR</b>																													
	d) Find the value of the correlation coefficient from the data given in the following table:			[5]	CO6	L3																								
	<table><tr><td>Subject</td><td>Age(X)</td><td>Glucose Level(B)</td></tr><tr><td>1</td><td>43</td><td>99</td></tr></table>			Subject	Age(X)	Glucose Level(B)	1	43	99																					
Subject	Age(X)	Glucose Level(B)																												
1	43	99																												

	2	21	65				
	3	25	79				
	4	42	75				
	5	57	87				
	6	59	81				