

BTUESD007



INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, NAGPUR

Department of Computer Science and Engineering

Branch: CSE-DS

Subject: Introduction to Data & Analytics (CSL 109)

Sessional-2 Examination

Semester -I

Date: 19/01/2023 (Thursday)

Time: 4:00 PM to 5:00 PM

Duration: 1 Hours

Max. Marks: 15

Important Instructions:

- All the questions are compulsory.

Q. No.	Questions	Marks & CO																							
1(a)	For the given dataset: 13, 15, 16, 20, 20, 21, 22, 23, 25, 30, 33, 35, 20, 38, 30, 16, 40, 33, 19, 45, 46, 52, 78, 12, 28, 46, 28. (a) Compute the five-number summary of the data. (b) Show a boxplot of the data. (c) Find out the outliers in the dataset.	3 [CO2]																							
(b)	How to measure the correlation between two variables. Take scenario of student data that include Age and IQ level. Calculate coefficient of correlation between them.	2 [CO2]																							
2(a)	Suppose that the values for a given set of data are grouped into intervals. The intervals and corresponding frequencies are as follows. Compute Karl Pearson coefficient of skewness. <table><tr><th>Age</th><th>Frequency</th></tr><tr><td>1-5</td><td>200</td></tr><tr><td>5-15</td><td>450</td></tr><tr><td>15-20</td><td>300</td></tr><tr><td>20-50</td><td>1500</td></tr><tr><td>50-80</td><td>700</td></tr><tr><td>80-110</td><td>44</td></tr></table>	Age	Frequency	1-5	200	5-15	450	15-20	300	20-50	1500	50-80	700	80-110	44	3 [CO3]									
Age	Frequency																								
1-5	200																								
5-15	450																								
15-20	300																								
20-50	1500																								
50-80	700																								
80-110	44																								
(b)	How to measure the kurtosis of the data. Using the data (12 13 54 56 25), determine the type of kurtosis present.	2 [CO3]																							
3	A reputed marketing agency in India has three different training programs for its salesmen. The three programs are Method – A, B, C. To assess the success of the programs, 4 salesmen from each of the programs were sent to the field. Their performances in terms of sales quantity are given in the following table. Verify whether their significant difference among methods and among salesmen are present or not. <table><tr><th rowspan="2">Salesmen</th><th colspan="3">Methods</th></tr><tr><th>A</th><th>B</th><th>C</th></tr><tr><td>1</td><td>4</td><td>6</td><td>2</td></tr><tr><td>2</td><td>6</td><td>10</td><td>6</td></tr><tr><td>3</td><td>5</td><td>7</td><td>4</td></tr><tr><td>4</td><td>7</td><td>5</td><td>4</td></tr></table> <p>Note: The value of F Distribution table at 5% significance level is given as: F (3, 6) = 4.7571, F (2, 6) = 5.1456.</p>	Salesmen	Methods			A	B	C	1	4	6	2	2	6	10	6	3	5	7	4	4	7	5	4	5 [CO3]
Salesmen	Methods																								
	A	B	C																						
1	4	6	2																						
2	6	10	6																						
3	5	7	4																						
4	7	5	4																						

\*\*\*\*\*

BTUESD007  
 8  
 Square 11 PDF  
 (3,6)  
 (2,6)  
 12