

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

April – May 2023

B.Tech Program: Computer Engineering Scheme II

Examination: LY Semester: VIII

Course Code: CEDLC8021 and Course Name: Applied Data Science

Date of Exam: 16/05/2023

Duration: 2.5 Hours

Max. Marks: 60

Instructions:

- (1) All questions are compulsory.
- (2) Draw neat diagrams wherever applicable.
- (3) Assume suitable data, if necessary.

		Max. Marks	CO	BT level
Q 1	Solve any six questions out of eight:	12		
i)	Calculate the Mean, Median and Mode of the following data: 5, 10, 10, 12, 13. Are these three equal ?	2	1	Ap
ii)	Find out the determinant of the following matrix $\begin{bmatrix} 3 & -1 \\ 4 & 3 \end{bmatrix}$.	2	2	Ap
iii)	What is mean by term 'estimate of location' with respect to exploratory data analysis?	2	3	U
iv)	Define the following terms a) Sample Bias b) Central Limit theorem.	2	4	U
v)	What is 'P-Value' with respect to Hypothesis testing?	2	5	U
vi)	Describe in short about unconstrained optimization.	2	6	U
vii)	Differentiate between binomial and Poisson distribution.	2	4	U
viii)	Illustrate the use of confusion matrix in data science.	2	6	U
Q.2	Solve any four questions out of six.	16		
i)	Suppose a healthcare organization is studying the effectiveness of a new medication for treating a specific condition. For this type of scenario which type of statistics is better a) Descriptive b) Prescriptive or c) both. Justify your answer.	4	1	An
ii)	Calculate the rank of a Matrix using echelon form. $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 3 \\ 3 & 5 & 7 \end{bmatrix}$	4	2	Ap
iii)	Explain the key terms of exploring the data distribution with respect to EDA.	4	3	U

April – May 2023

B.Tech Program: Computer Engineering Scheme II

Examination: LY Semester: VIII

Course Code: CEDLC8021 and Course Name: Applied Data Science

Date of Exam: 16/05/2023

Duration: 2.5 Hours

Max. Marks: 60

iv)	Why to use Student's t test for this type of scenario- Suppose a researcher is interested in comparing the average scores of two groups of students on a standardized test. Justify your answer?	4	4	An																																																																							
v)	Explain the Chi-Square test with example.	4	5	U																																																																							
vi)	Compare and contrast between precision and recall.	4	6	U																																																																							
Q.3	Solve any two questions out of three.	16																																																																									
i)	How Skewness is used for this type of scenario- consider the distribution of returns for a particular stock over a certain period of time. The distribution may not be symmetric, with a longer tail to one side indicating greater variability in the returns. Explain the different types of skewness with its rule of thumb.	8	1	An																																																																							
ii)	Write a short note on a) Exponential Distribution b) Weibull Distribution	8	4	U																																																																							
iii)	Data set is given and its predicted output is given. Determine which type of test is used here and explain each parameter given in output and also analyse whether reject or accept the null hypothesis	8	5	An																																																																							
	<table border="1"> <thead> <tr> <th>Group 1</th> <th>Group2</th> <th>Group3</th> </tr> </thead> <tbody> <tr> <td>51</td> <td>23</td> <td>56</td> </tr> <tr> <td>45</td> <td>43</td> <td>76</td> </tr> <tr> <td>33</td> <td>23</td> <td>74</td> </tr> <tr> <td>45</td> <td>43</td> <td>87</td> </tr> <tr> <td>67</td> <td>45</td> <td>56</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="5">SUMMARY</th> </tr> <tr> <th>Groups</th> <th>Count</th> <th>Sum</th> <th>Average</th> <th>Variance</th> </tr> </thead> <tbody> <tr> <td>Group 1</td> <td>5</td> <td>241</td> <td>48.2</td> <td>153.2</td> </tr> <tr> <td>Group2</td> <td>5</td> <td>177</td> <td>35.4</td> <td>128.8</td> </tr> <tr> <td>Group3</td> <td>5</td> <td>349</td> <td>69.8</td> <td>183.2</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Source of Varia</th> <th>SS</th> <th>df</th> <th>MS</th> <th>F</th> <th>P-value</th> <th>F crit</th> </tr> </thead> <tbody> <tr> <td>Between Grc</td> <td>3022.933</td> <td>2</td> <td>1511.467</td> <td>9.747206</td> <td>0.00306</td> <td>3.885294</td> </tr> <tr> <td>Within Grc</td> <td>1860.8</td> <td>12</td> <td>155.0667</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Total</td> <td>4883.733</td> <td>14</td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Group 1	Group2	Group3	51	23	56	45	43	76	33	23	74	45	43	87	67	45	56	SUMMARY					Groups	Count	Sum	Average	Variance	Group 1	5	241	48.2	153.2	Group2	5	177	35.4	128.8	Group3	5	349	69.8	183.2	Source of Varia	SS	df	MS	F	P-value	F crit	Between Grc	3022.933	2	1511.467	9.747206	0.00306	3.885294	Within Grc	1860.8	12	155.0667				Total	4883.733	14							
Group 1	Group2	Group3																																																																									
51	23	56																																																																									
45	43	76																																																																									
33	23	74																																																																									
45	43	87																																																																									
67	45	56																																																																									
SUMMARY																																																																											
Groups	Count	Sum	Average	Variance																																																																							
Group 1	5	241	48.2	153.2																																																																							
Group2	5	177	35.4	128.8																																																																							
Group3	5	349	69.8	183.2																																																																							
Source of Varia	SS	df	MS	F	P-value	F crit																																																																					
Between Grc	3022.933	2	1511.467	9.747206	0.00306	3.885294																																																																					
Within Grc	1860.8	12	155.0667																																																																								
Total	4883.733	14																																																																									
Q.4	Solve any two questions out of three.	16																																																																									

K. J. Somaiya Institute of Technology, Sion, Mumbai-22
(Autonomous College Affiliated to University of Mumbai)

April – May 2023

B.Tech Program: Computer Engineering Scheme II

Examination: LY Semester: VIII

Course Code: CEDLC8021 and Course Name: Applied Data Science

Date of Exam: 16/05/2023

Duration: 2.5 Hours

Max. Marks: 60

i)	Differentiate between rank and nullity of a matrix. Calculate the nullity of a matrix 1 2 0 2 4 0 3 6 1	8	2	Ap
ii)	Write a short note on a) Long tailed distribution b) Normal Distribution	8	4	U
iii)	Discuss the following terms with example a) ROC b)AUC c)Lift	8	6	U
