COURSE CODE (CREDITS): 19B1WCI637 (2)
COURSE NAME: Statistics and Exploratory Data Analytics
COURSE INSTRUCTORS: Dr. Amol Vasudeva

MAX. MARKS: 35

MAX. TIME: 2 Hours

Note: (a) All questions are compulsory.
(b)Marks are indicated against each question in square brackets.
(c) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

1. Calculate the median of the following frequency distribution. (CO-2) **[5 marks]**

Class Interval	Frequency
0-8	8
8-16	10
16-24	16
24-32	24
32-40	15
40-48	7

2. Find 27% trimmed mean on the data 2, 4, 6, 7, 11, 21, 81, 90, 105, 121. (CO-2) **[5 marks]**

3. Write the steps to compute the Eigen vectors and Eigen values from the matrix A. (CO-3) **[6 marks]**

$$A = \begin{bmatrix} 9 & 6 & 9 \\ 9 & 9 & 3 \\ 6 & 6 & 6 \\ 6 & 6 & 9 \\ 3 & 3 & 3 \end{bmatrix}$$

4. Write the steps to find the optimal number of clusters in a K-Mean clustering using Silhouette Coefficient Method? (CO-4) **[5 marks]**

5. What does a PCA do? How is the first principal component axis selected? What does a principal component in a PCA represent? (CO-3) **[2+1+1 = 4 marks]**

6. There are groups of students belonging either to gifted class or regular class. The students who had high scores in the last year's exam were assigned to gifted class. The rest of the students were allocated to the regular class. The details of the students' ranks in the last year are given in a file student_marks.csv with the following two fields. (CO-4)

Score out of 100	Number of Students
40	10
45	14
......	......
100	12

Write a Python program to draw the histogram (number vs. score).
Write a Python program to split the histogram into two equal portions. **[6 marks]**

7. Differentiate between the following terms: (CO-4) **[4 marks]**
   - Hard Clustering and Soft Clustering
   - K-Mean Clustering and Gaussian Mixture Model