COURSE CODE (CREDITS): 19BIWCI637 (2) **MAX. MARKS:** 25

COURSE NAME: Statistics and Exploratory Data Analytics

COURSE INSTRUCTORS: Amol Vasudeva **MAX. TIME:** 1 Hour 30 Minutes

Note: All questions are compulsory. Marks are indicated against each question in square brackets.

1. A survey was given to a random sample of 20 sophomore college students. They were asked, "how many textbooks do you own?" Their responses were:

0	0	2	5	8	8	8	9	9	10	10	10	11	12	12	12	14	15	20	25

   Based on the above data, write down a step-by-step procedure to detect and remove the outliers from the above data using inter quartile range (IQR) method. (CO-2)   **[5 marks]**

2. Assume that you are given a dataset as "iris.csv". Write a Python program to read the data set "iris.csv". Now, use IQR method to detect and remove the outliers from the column "sepal.width". Finally, draw a box plot to verify the removal of the outliers from this column.(CO-2)

   [6 marks]

3. You are given a matrix M. Based on this matrix, write down a Python code to perform the following operations: (CO-3)
   - Use Singular Value Decomposition (SVD) technique to partition the matrix M into three components U, S, and V.
   - Re-combine U, S, and V to get the original matrix M.   **[4 marks]**

4. Assume that you are given an image file as "image.png". Write a Python program and use OpenCV to perform the following: (CO-3)
   - Load and display the original image
   - Split the image into Red, Green, and Blue arrays
   - Initialize PCA with top 50 principal components
   - Apply PCA to red channel to perform the transformation. Next, Apply inverse transformation to the transformed array.
   - Apply PCA to Green channel and then apply the inverse transform to transformed array.
   - Finally, apply PCA to *blue* channel and then apply the inverse transform to transformed array.
   - Display the compressed image with first 50 principal components.   **[6 marks]**

5. Differentiate between the following terms (in 3-4 sentences):
   - Data Standardization and Data Normalization (CO-2)
   - Box-Cox and Yeo-Johnson Power Transformation (CO-2)   **[4 marks]**