MSDC-103

Sri Sathya Sai Institute of Higher Learning

(Deemed to be University)

DMACS, MDH Campus

Course: **Data Science and Computing** Date: **September 13, 2019**

Title: **Multivariate Statistical Analysis** Test: **Mini-Project/ Assignment**

For given data set (`National Track Records for Men.csv`), answer the following questions. Maximum marks will be **25**.

(1) Check the normality of the data. [2 Marks]

(2) Obtain the sample correlation matrix $\mathbf{R}$ for given data, and determine its eigen values and eigenvectors. [3 Marks]

(3) Determine the the number k of principal components for the standardized variables. Obtain the k components. [4 Marks]

(4) Prepare a table showing the correlations of the standardized variables with the k principal components. [2 Marks]

(5) Rank the nations based on their score on the fist principal component. Find India's rank? [2 marks]

(6) Perform graphical analysis on k principal components to verify question (a). [2 Marks]

(7) Convert the National Track Records for Men to speeds measured in metres per second, then perform the cluster analysis by using K-means method with suitable (need to determine) number of clusters. [5M]

(8) Find the multilinear regression model for the first seven varibles $x_1, \ldots, x_7$ (independent) and x_8 as response varibale. [5 Marks]

Assignment submission last date: **20th September 2019, 4 pm**.