Course Title: Statistical Data Analysis (SDA)

Institute Elective (UG3/UG4)
By: Dr. Mainak Thakur

Course Level: L1 **L-T-P-C:** 3 - 1 - 0 – 4

No. of Students: TBD

Course Objective: The objective of this course is to introduce the students with the world of statistics as a tool for data science so that they can analyse data and interpret knowledge out of it. After taking this course students would be able to form a statistical problem from a data driven practical problem and would be able to solve the problem using univariate and/or multivariate data coming from different domains.

Pre-requisite: Good knowledge in Probability

Course Outline:

The following list of topics is tentative. Based on available time slots, some topics may be dropped or added or reordered.

This course is divided into the following modules:

1) Introduction to Statistical Thinking
2) Review of Probability
3) Overview of Exploratory Data Analysis
4) Overview of Sampling Theory
5) Important Univariate Distributions
6) Estimation Theory
7) Hypothesis Testing
8) Multivariate Descriptive Statistics
9) Multivariate Normal Distribution and properties
10) Regression Analysis: Multiple Linear Regression
11) Overview of Different Multivariate Techniques:
    - Cluster Analysis
    - Principal Component Analysis
    - Factor Analysis
11) Time Series Analysis
12) Application using Matlab/R/Python

BOOKS:

i) Applied Multivariate Statistical Analysis by Richard Johnson & Dean Wichern, Pearson
ii) Basic Econometrics by Damodar Gujarati
iii) Applied Statistics and Probability for Engineers (3rd ed.) by Montgomery, D. C., 2002.
iv) Statistical Methods (Combined Volume) by N G Das, Tata McGraw-Hill.
v) An Introduction to Multivariate Statistical Analysis by T W Anderson, John Wiley & Sons Inc.
vi) Time Series Analysis (4th ed.) by George E. P. Box, Gwilym M. Jenkins and Gregory C. Reinsel, John Wiley & Sons, 2013.

Evaluation Policy:

Mid Sem1: 10%
Mid Sem2: 10%
Project: 30%
Quiz: 20%
Assignment: 30%

Project should include a complete report of statistical analysis of the data provided.

Laptop/mobile use policy in classes: Strictly Not allowed

Industry Relevance: This course is highly relevant to the Data Analytics industry. The syllabus is designed keeping in mind the current industry activities and their demand of expertise in the field of data science.

Course Plan Submitted By

Dr. Mainak Thakur
Assistant Professor
Indian Institute of Information Technology, Sri City
Andhra Pradesh - 517646, India.

e-mail: mainak.t@iiits.in
Phone: +91 8697947300