IIT Mandi
Proposal for a New Course

Course number : MA-611
Course Name : Statistical tools and Computing
Credit Distribution : 3-1-0-4
Intended for : M.Sc. /M.S./PhD/ B.Tech. 3rd and 4th year.
Prerequisite : MA-524 (Probability and Statistics) or any course on probability and
  statistics (like Data Science II) in consultation with the instructor.
Mutual Exclusion : HS550

1. Preamble:

As the scale and scope of data collection continue to increase across virtually all fields, statistical learning has become a critical toolkit for anyone who wishes to understand data. This course on Statistical tools and Computing provides a comprehensive elucidation of key topics in statistical learning. Each module includes a lab component and will involve theoretical topics and programming assignments. The course introduces data analysis as a means of modelling measurements made over an interval of time and space. The course is useful for anyone who wishes to use contemporary tools for data analysis.

2. Course Modules with quantitative lecture hours:

Module 1- Concepts from probability and statistics:
Data (sample vs. Population, histograms, sample mean, median, variance, standard deviation); Probability (axioms, basic rules; and conditional probability); Random variables (discrete vs. Continuous); Review of probability distributions. Some advanced probability distributions: their properties and simulations; Confidence intervals and their significance.
(8 Hours)

Module 2 -Times series methods:
Collection and classification of data; Different types of diagrams to represent statistical data; Frequency distribution and related graphs and charts; Linear and non-linear models.
(6 hour)

Module 3 -Regression, classification and multivariate analysis
Simple regression; Multiple regression; Logistic regression; Generalized linear models; Cross validation; Multicollinearity; Model selection; Prediction and variable selection; Bayesian logistic regression; Principal component analysis; Factors analysis; Discriminant and Classification analysis.
(13 Hours)

Proposal for a New Course

Module 4: Parametric and Nonparametric tests:

Parametric: *Parametric tests* are used only where a normal distribution is assumed. The most widely used tests are the t-test (paired or unpaired); ANOVA (one-way non-repeated, repeated), and Pearson rank correlation.

Nonparametric: *Non-parametric tests* are used when continuous data are not normally distributed or when dealing with discrete variables. Most widely tests used are Chi-squared; Fisher's exact tests; Wilcoxon's matched pairs; Mann–Whitney U-tests; Kruskal–Wallis tests and Spearman rank correlation; Bayesian inference; Kernel Density Estimation.

(15 Hours)

Laboratory/practical/tutorial Modules:

The labs, using programming languages like R/Python/any other, will take place over a two-hour period in alternate weeks. It will run concurrently with the theory course, thus the subjects for the lab will have previously been established in the theory session.

3. Text books:

1. An Introduction to Statistical Learning, with Applications in R (second edition) by James, Witten, Hastie and Tibshirani (Springer, 2021)
2. Introduction to Applied Statistics: A non-Calculus Based Approach by David D. Hanagal (Narosa, 2009)

3. References:

1. Introduction to Statistics by David Lane, Rice University (David Lane, 2003)
2. Introduction to Statistics and Data Analysis by Jay Devore, Roxy Peck, Chris Olsen Third edition (Wadsworth Publishing, 2008)
3. Generalized linear models with examples in R by Peter K. Dunn, Gordon K. Smyth (Springer, 2018)

4. Similarity with the existing courses:

S. No.	Course code and Title	Similarity content	Approx.% of content
1.	MA-524 Probability and Statistics	Introduction to probability, regression, descriptive statistics	~10-15%
2.	MA 605 Statistical data	Regression, classification descriptive statistics	~10-15%

Proposal for a New Course

		analysis		
3.	DS403 Introduction to Statistical Learning	Regression	~5%
4.	HS550	Concepts like probability, regression.	~15-20%

6. Justification of new course proposal if cumulative similarity content is >30%: NA

Proposal for a New Course