Code		Lecture	Tutorial	Practical/ Practice	criteria	the course (if any)
Mathematical Data Science	**4**	**3**	**0**	**1**	**Class XII pass with Mathematics**	**Basic knowledge of R/Python DSC-3: Probability & Statistics**

Learning Objectives: The main objective of this course is to:

- Introduce various types of data and their sources, along with steps involved in data science case-study, including problems with data and their rectification and creation methods.
- Cover dimensionality reduction techniques, clustering algorithms and classification methods.

Learning Outcomes: The course will enable the students to:

- Gain a comprehensive understanding of data science, its mathematical foundations including practical applications of regression, principal component analysis, singular value decomposition, clustering, support vector machines, and *k*-NN classifiers.
- Demonstrate data analysis and exploration, linear regression techniques such as simple, multiple explanatory variables, cross-validation and regularization using R/Python.
- Use real-world datasets to practice dimensionality reduction techniques such as PCA, SVD, and multidimensional scaling using R/Python.

SYLLABUS OF DSE-3(i)

UNIT-I: Principles of Data Science (12 hours)

Types of Data: nominal, ordinal, interval, and ratio; Steps involved in data science case-study: question, procurement, exploration, modeling, and presentation; Structured and unstructured data: streams, frames, series, survey results, scale and source of data – fixed, variable, high velocity, exact and implied/inferred; Overview of problems with data – dirty and missing data in tabular formats – CSV, data frames in R/Pandas, anomaly detection, assessing data quality, rectification and creation methods, data hygiene, meta-data for inline data-description-markups such as XML and JSON; Overview of other data-source formats – SQL, pdf, Yaml, HDF5, and Vaex.

Unit-II: Mathematical Foundations (15 hours)

Model driven data in R^n, Log-likelihoods and MLE, Chebyshev, and Chernoff-Hoeffding inequalities with examples, Importance sampling; Norms in Vector Spaces– Euclidean, and metric choices; Types of distances: Manhattan, Hamming, Mahalanobis, Cosine and angular distances, KL divergence; Distances applied to sets– Jaccard, and edit distances; Modeling text with distances; Linear Regression: Simple, multiple explanatory variables, polynomial, cross-validation, regularized, Lasso, and matching pursuit; Gradient descent.

Unit-III: Dimensionality Reduction, Clustering and Classification (18 hours)

Problem of dimensionality, Principal component analysis, Singular value decomposition (SVD), Best *k*-rank approximation of a matrix, Eigenvector and eigenvalues relation to SVD, Multidimensional scaling, Linear discriminant analysis; Clustering: Voronoi diagrams, Delaunay triangulation, Gonzalez's algorithm for *k*-center clustering, Lloyd's algorithm for *k*-means clustering, Mixture of Gaussians, Hierarchical clustering, Density-based clustering and outliers, Mean shift clustering; Classification: Linear classifiers, Perceptron algorithm, Kernels, Support vector machines, and *k*-nearest neighbors (*k*-NN) classifiers.

Essential Readings

1. Mertz, David. (2021). Cleaning Data for Effective Data Science, Packt Publishing.
2. Ozdemir, Sinan. (2016). Principles of Data Science, Packt Publishing.
3. Phillips, Jeff M. (2021). Mathematical Foundations for Data Analysis, Springer. (https://mathfordata.github.io/).

Suggestive Readings

- Frank Emmert-Streib, et al. (2022). Mathematical Foundations of Data Science Using R. (2nd ed.). De Gruyter Oldenbourg.
- Wes McKinney. (2022). Python for Data Analysis (3rd ed.). O'Reilly.
- Wickham, Hadley, et al. (2023). R for Data Science (2nd ed.). O'Reilly.

Practical (30 hours)- Practical work to be performed in Computer Lab using R/Python:
1. To explore different types data (nominal, ordinal, interval, ratio) and identify their properties.
2. To deal with dirty and missing data, such as imputation, deletion, and data normalization.
3. Use the real-world datasets (https://data.gov.in/) to demonstrate the following:
   a) Data analysis and exploration, linear regression techniques such as simple, multiple explanatory variables, cross-validation, and regularization.
   b) Dimensionality reduction techniques such as principal component analysis, singular value decomposition (SVD), and multidimensional scaling.
   c) Clustering algorithms such as *k*-means, hierarchical, and density-based clustering and evaluate the quality of the clustering results.
   d) Classification methods such as linear classifiers, support vector machines (SVM), and *k*-nearest neighbors (*k*-NN).

DISCIPLINE SPECIFIC ELECTIVE COURSE – 3(ii): LINEAR PROGRAMMING AND APPLICATIONS

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Linear Programming and Applications	4	3	1	0	Class XII pass with Mathematics	DSC-4: Linear Algebra

Learning Objectives: Primary objective of this course is to introduce:
- Simplex Method for linear programming problems.
- Dual linear programming problems.
- The applications of linear Programming to transportation, assignment, and game theory.

Learning Outcomes: The course will enable the students to:

- Learn about the basic feasible solutions of linear programming problems.
- Understand the theory of the simplex method to solve linear programming problems.
- Learn about the relationships between the primal and dual problems.
- Solve transportation and assignment problems.
- Understand two-person zero sum game, games with mixed strategies and formulation of game to primal and dual linear programing problems to solve using duality.

SYLLABUS OF DSE-3(ii)

UNIT– I: Introduction to Linear Programming (12 hours)

Linear programming problem: Standard, Canonical and matrix forms, Geometric solution; Convex and polyhedral sets, Hyperplanes, Extreme points; Basic solutions, Basic feasible solutions, Correspondence between basic feasible solutions and extreme points.

UNIT– II: Optimality and Duality Theory of Linear Programming Problem (18 hours)

Simplex method: Optimal solution, Termination criteria for optimal solution of the linear programming problem, Unique and alternate optimal solutions, Unboundedness; Simplex algorithm and its tableau format; Artificial variables, Two-phase method, Big-M method. Duality Theory: Motivation and formulation of dual problem, Primal-Dual relationships, Fundamental theorem of duality; Complementary slackness.

UNIT – III: Applications (15 hours)

Transportation Problem: Definition and formulation, Northwest-corner, Least- cost, and Vogel's approximation methods of finding initial basic feasible solutions; Algorithm for solving transportation problem.

Assignment Problem: Mathematical formulation and Hungarian method of solving.

Game Theory: Two-person zero sum game, Games with mixed strategies, Formulation of game to primal and dual linear programming problems, Solution of games using duality.

Essential Readings

1. Bazaraa, Mokhtar S., Jarvis, John J., & Sherali, Hanif D. (2010). Linear Programming and Network Flows (4th ed.). John Wiley and Sons. Indian Reprint.
2. Hillier, Frederick S. & Lieberman, Gerald J. (2021). Introduction to Operations Research (11th ed.). McGraw-Hill Education (India) Pvt. Ltd.
3. Taha, Hamdy A. (2017). Operations Research: An Introduction (10th ed.). Pearson.

Suggestive Readings

- Hadley, G. (1997). Linear Programming. Narosa Publishing House. New Delhi.
- Thie, Paul R., & Keough, G. E. (2008). An Introduction to Linear Programming and Game Theory. (3rd ed.). Wiley India Pvt. Ltd. Indian Reprint 2014.

DISCIPLINE SPECIFIC ELECTIVE COURSE – 3(iii): MATHEMATICAL STATISTICS

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/		

				Practice		
Mathematical Statistics	**4**	**3**	**1**	**0**	**Class XII pass with Mathematics**	**DSC-3: Probability & Statistics DSC-11: Multivariate Calculus**

Learning Objectives: The main objective of this course is to introduce:

- The joint behavior of several random variables theoretically and through illustrative practical examples.
- The theory underlying modern statistics to give the student a solid grounding in (mathematical) statistics and the principles of statistical inference.
- The application of the theory to the statistical modeling of data from real applications, including model identification, estimation, and interpretation.
- The idea of Fisher information to find the minimum possible variance for an unbiased estimator, and to show that the MLE is asymptotically unbiased and normal.

Learning Outcomes: The course will enable the students to:

- Understand joint distributions of random variables including the bivariate normal distribution.
- Estimate model parameters from the statistical inference based on point estimation and hypothesis testing.
- Apply Rao-Blackwell theorem for improving an estimator, and Cramér-Rao inequality to find lower bound on the variance of unbiased estimators of a parameter.
- Understand the theory of linear regression models and contingency tables.

SYLLABUS OF DSE - 3(iii)

UNIT–I: Joint Probability Distributions (15 hours)

Joint probability mass function for two discrete random variables, Marginal probability mass function, Joint probability density function for two continuous random variables, Marginal probability density function, Independent random variables; Expected values, covariance, and correlation; Linear combination of random variables and their moment generating functions; Conditional distributions and conditional expectation, Laws of total expectation and variance; Bivariate normal distribution.

UNIT-II: Sampling Distributions and Point Estimation (15 hours)

Distribution of important statistics such as the sample totals, sample means, and sample proportions, Central limit theorem, Law of large numbers; Chi-squared, *t*, and *F* distributions; Distributions based on normal random samples; Concepts and criteria for point estimation, The methods of moments and maximum likelihood estimation (MLE); Assessing estimators: Accuracy and precision, Unbiased estimation, Consistency and sufficiency, The Neyman factorization theorem, Rao-Blackwell theorem, Fisher Information, The Cramér-Rao inequality, Efficiency,

UNIT-III: Confidence Intervals, Tests of Hypotheses and Linear Regression Analysis (15 hours)

Interval estimation and basic properties of confidence intervals, One-sample t confidence interval, Confidence intervals for a population proportion and population variance.
Statistical hypotheses and test procedures, One-sample tests about a population mean and a population proportion, P-values for tests; The simple linear regression model and its estimating parameters; Chi-squared goodness-of-fit tests, Two-way contingency tables.

Essential Reading

1. Devore, Jay L., Berk, Kenneth N. & Carlton Matthew A. (2021). Modern Mathematical Statistics with Applications. (3rd ed.). Springer.

Suggestive Readings

- Devore, Jay L. (2016). Probability and Statistics for Engineering and the Sciences. Ninth edition, Cengage Learning India Private Limited, Delhi. Fourth impression 2022.
- Hogg, Robert V., McKean, Joseph W., & Craig, Allen T. (2019). Introduction to Mathematical Statistics. Eighth edition, Pearson. Indian Reprint 2020.
- Mood, A.M., Graybill, F.A., & Boes, D.C. (1974). Introduction the Theory of Statistics (3rd ed.). Tata McGraw Hill Pub. Co. Ltd. Reprinted 2017.
- Wackerly, Dennis D., Mendenhall III, William & Scheaffer, Richard L. (2008). Mathematical Statistics with Applications. 7th edition, Cengage Learning.