Maulana Abul Kalam Azad University of Technology, West Bengal

M.Sc. in Applied Statistics and Analytics

Effective from 2020-21

Program outcomes for MSc in Applied Statistics and Analytics

Upon completion of this 2-year post-graduate degree program, the learners will be able to

- **PO1**. **Statistical and Analytical Knowledge:** Understand the basic concepts, fundamental principles and the scientific theories related to Statistics and Data Analytics which would converge in preparing the learners ready for scientific decision-making with the extensive use of analytics in various fields.

- **PO2. Problem Analysis**: Develop the ability to absorb and understand various advanced theories in Statistics and Analytics leading to model and analyse the real-world problems.

- **PO3: Modern Software Tool Usages:** Deploy a wide range portfolio of advanced analytical and statistical techniques using contemporary software tools for the purpose of solving real-life problems.

- **PO4: Collaborative Ability:** Interact with fellow professionals of their own and other fields of science, engineering, and humanities through collaborative engagement.

- **PO5. Conduct Investigation of Complex Problem:** Use research-based knowledge and develop research using Optimization, Multivariate analysis and other techniques for interpretation of data, and synthesis of the information to provide valid conclusions.

- **PO6. Individual and Team Work:** Function effectively as an individual as well as a member or leader in diverse teams in multidisciplinary settings as and when necessary.

- **PO7. Contribution to Society and Environment:** The learners will be proficient in deploying a rich *port folio* of tools & techniques that involve collecting, organizing, analyzing, and interpreting data for the purpose of improving the quality of their own lives as well as solving real-life problems of contemporary society and environmental issues. They will be able to ascertain facts from figures, distil insights from indices, and winnow the real from the fake.

- **PO8. Continuous Learning:** Recognize the need for independent and continuous learning in the broadest context of technological requirements and develop corresponding ability to pursue the same.

- **PO9. Development of Expertise:** Write scientific & technical documents, such as essays, research, review articles, project proposals, reports, theses, etc. in the domain of Statistics and Analytics and to cope strongly with the problems of contemporary professional world that has aroused with the advent and expansion of "Data-Science and Technology" in almost all domains of life.

- **PO10. Career Opportunities:** Follow Career path in almost all professional domains like Government, Pharma, Health, Insurance, Banking, Finance, Automobile, Information Technology and others which needs market research, forecasting and predictive analysis in addition to education sector.

Duration**: 2 Years;** Level**: Post graduation;** Type**: Degree**

Semester 1:

				Hours Per Week			**Credit**
Paper Code	**Course Name**	**Course Type**	**Marks**	**L**	**T**	**P**	
Theory							
MS-ASA401	Applied Linear Algebra	CC	100	3	1	0	4
MS-ASA402	Elements of Real Analysis and Probability	CC	100	3	1	0	4
MS-ASA403	Statistical Inference and Introductory Analytics	CC	100	3	1	0	4
MS-ASA404	Analytics Using Software and Programming Language	CC	100	3	1	0	4
MS-ASA405	Research Methodology and IPR	VAC	100	2	1	0	2
Practical							
MS-ASA491 (Lab)	Programming Language Laboratory	CC	100	0	0	4	2
MS-ASA492 (Lab)	Laboratory for Statistics and Linear Algebra	CC	100	0	0	4	2
Sessional							
MS-ASA493	Term Project and Presentation I	SEC	100	2	1	0	2
	Total		800				24

Semester 2:

Code	Course Title	Hours Per Week				Credit
		Course Type	Marks	L	T	P	
Theory							
MS-ASA406	Regression For Predictive Model Building	CC	100	3	1	0	4
MS-ASA407	Optimization Techniques and Soft Computing	CC	100	3	1	0	4
MS-ASA408	Stochastic Processes and its Application	CC	100	3	1	0	4
MS-ASA409	Time Series Analysis and Forecasting Methods	CC	100	3	1	0	4
MS-ASA410 (Audit Course)	Evolution of Statistical Thinking	VAC	-----	2	--	--	0
MS-ASA411	Elective-I	DSE	100	3	1	0	4
Practical							
MS-ASA494 (Lab)	Regression and Time Series Laboratory	CC	100	0	0	4	2
MS-ASA495 (Lab)	Optimization Techniques and Stochastic Process Laboratory	CC	100	0	0	4	2
Sessional							
MS-ASA496	Term Project and Presentation II	SEC	100	2	1	0	2
	Total		800				26

Semester 3:

		Hours Per Week					Credit
Code	**Course Title**	**Course Type**	**Marks**	**L**	**T**	**P**	
Theory							
MS-ASA501	Applied Multivariate Analysis and Data Mining	CC	100	3	1	0	4
MS-ASA502	Machine Learning Algorithms	CC	100	3	1	0	4
MS-ASA503	Advanced Business Analytics and Big Data	CC	100	3	1	0	4
MS-ASA504	Advanced Analytics using Software and Programming Language (SPSS, Hadoop, SAS)	SEC	100	3	1	0	4
MS-ASA505	Elective-II	IDE	100	3	1	0	4
MS-ASA506 (Audit Course)	Emerging Topics in Statistics and Analytics	VAC	-----	2	--	--	0
Practical							
MS-ASA591 (Lab)	Analytics Laboratory	CC	100	0	0	4	2
MS-ASA592 (Lab)	Machine Learning Laboratory	SEC	100	0	0	4	2
Sessional							
MS-ASA593	Term Project and Presentation III	SEC	100	2	1	0	2
	Total		800				26

Semester 4:

		Hours Per Week					Credit
Code	**Course Title**	**Course Type**	**Marks**	**L**	**T**	**P**	
Theory							
MS-ASA507	Biostatistics	CC	100	3	1	0	4
MS-ASA508	Elective -III	DSE	100	3	1	0	4
Practical							
MS-ASA594 (Lab)	Advanced Analytics Laboratory	CC	100	0	0	4	2
Sessional							
MS-ASA595	Capstone Project	SEC	200	0	0	8	8
	Total		500				18

CC: Core Course, **VAC**: Value Added Course, **SEC:** Skill Enhancement Course, **IDE:** Interdisciplinary Course, **DSE:** Discipline Specific Elective Course.

Students are recommended to go for internship/ industrial training during semester break (between II & III).

List of Electives

1. Elective-I (MSASA411)

i. Multiple Testing Problems (MSASA411A)

ii. Robust Statistical Inference and application to predictive model building (MSASA411 B)

iii. Econometrics (MSASA411 C)

Iv. Supply Chain Management (MSASA411D)

2. Elective II (MSASA505)

i. Health Informatics (MSASA505A)

ii. Marketing Analytics (MSASA505B)

Iii. Statistical Fluid Mechanics (MSASA505C)

Iv. Reliability Theory (MSASA505D)

3. Elective III (MSASA508)

i. Aadvanced Bayesian Inference (MSASA508A)

ii. Sports and education analytics (MSASA508B)

iii. Astrostatistics (MSASA508C)

List of non-credit courses

1. Evolution of Statistical Thinking
2. Emerging Topics in Statistics and Analytics

Semester-1

Applied Linear Algebra	**Paper Code: MSASA401**
	4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Make use of computational techniques and algebraic skills essential for the study of systems of linear equations, matrix algebra, vector spaces, eigenvalues and eigenvectors, orthogonality and diagonalization.
2. Develop knowledge and ability for visualization, spatial reasoning, as well as geometric properties and strategies to model, solve problems and view solutions, especially in $\mathbb{R}^2$, $\mathbb{R}^3$ and extending these results to higher dimensions.
3. Apply algorithms of linear algebra to data science and particularly in the domain of machine learning.
4. Use appropriate technology, to enhance and facilitate mathematical understanding of data.
5. Determine acceptable solutions regarding understanding of data.

Module 1: General introduction and review of basic concepts: (5L)

Review of vector spaces, linear independence, bases, dimension, subspaces. Vector space in R^n: System of linear equations, row space, column space and null space. Four fundamental spaces and their significance, relation between rank and nullity, Consistency theorem, Basis from a spanning set and independent set.

Module 2: Linear transformations and Matrices: (6L)

Linear transformations, matrix representations, range and null space, invertibility, solution of linear equations, eigenvalues and eigenvectors, similarity transformations, spectral invariance, diagonalizability, application in webpage ranking.

Module 3: Inner products and Associated Norms: (7L)

Inner products and its properties, Norms, equivalence of norms, orthogonality, orthonormal bases, orthogonalization, Gram-Schmidt orthogonalization process. Linear transformations on inner-product spaces. Orthogonal compliments, orthogonal projection, orthogonal subspace and corresponding representation theorem.

Module 4: Symmetric and Positive-Definite Transformations: (6L)

Symmetric transformations, Symmetric and Hermitian matrices, Linear equations revisited, Orthonormality of eigenvectors, Spectral decomposition theorem, Positive and positive-definite transformations and matrices, Quadratic forms. Unitary transformations, orthogonal and unitary matrices.

Module 5: Matrix Decompositions: (6L)

LU, QR, Cholesky, Schur and Singular value decompositions with applications, non-negative matrix factorization and its and Applications in Clustering and Recommender Systems.

Module 6: Least Square Problems: **(6L)**

Solution of linear least squares problems using matrix method, Constrained least squares applications, Moore-Penrose inverse; Solution of Rank deficient least squares problems.

Module 7: Linear Models: **(4L)**

Gauss Markov Model, Estimable function, error function, BLUE, Gauss Markov theorem. Correlated set-up, least squares estimate with restriction on parameters.

References:

1. Linear Algebra and its Application by Gilbert Strang, 4th edition, Cengage Learning along with MIT Lecture videos by the Author.

2. Introduction to Applied Linear Algebra: Vector, Matrices and Least Squares by Stephen Boyd and Lieven Vandenberghe, 2018 edition, Cambridge University Press.

3. Applied Linear Algebra by Peter J. Olver and Chehrzad Shakiban, second Edition, Springer.

4. Matrix Computations, G. H. Golub and C. F. Van Loan, 3rd Edition, John Hopkins University Press, 1996.

5. Linear Regression Analysis, George A.F. Seber, 2nd Edition, Wiley, 2003.

6. Linear Models, Shayle R. Searle, 2nd Edition, Wiley.

7. Linear Statistical Inference and its applications, CR Rao, 2nd Edition, Wiley.

Elements of Real Analysis and Probability	**Paper Code: MSASA402**
	4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Illustrate the knowledge of real numbers, least upper bounds, and the triangle inequality, define functions between sets; equivalent sets; finite, countable and uncountable sets,
2. Apply concepts of real analysis towards statistical arena, e.g., limit theorems, large sample inferences and other areas.
3. Use appropriate knowledge to deal with random variables and their distributions.
4. Understand abstract ideas and rigorous methods for convergence of sequence and series to apply in a wide variety of domains like financial analysis, predicting behavior of market and other related areas.
5. Apply probability as a tool of sophisticated data analysis.
6. Explain the importance of paradigm shift from classical to Bayesian inference.

Elements of Real Analysis:

Module-1 **(4L)**

Theory of real Number: - Concepts: Archimedean Principle, Concepts and related theorems, Concept of real number, Rational number, Properties of Sets. Limit points, Compactness, Bolzano-Weierstrass Theorem, Heine-Borel Theorem (Statement only) and its illustration with the help of applications.

Module-2 **(4L)**

Sequence and Series, Power series. Review of the concepts of Limit, Continuity, Derivatives & Intermediate Value Theorems, Taylor's Expansion. Sequence of Functions and associated results –with examples and applications.

Module-3 **(4L)**

Function of more than one variable, limits, continuity and partial derivatives, Evaluation of multiple integrals.

Module-4 **(3L)**

Solving numerical problems using Riemann integral and Stieltjes integral. Improper integrals and related theorems.

Elements of Probability

Module-1 **(10L)**

Genesis of probability Theory, Emergence of probability as an analytics tool, Sample space, Events, Classical definition of probability and its limitations, Kolmogorov's Axiomatic definition of probability, Classical

definition as a special case of axiomatic definition, Bayes' theorem and Bayesian inference paradigm, Prior and Posterior distributions, Bayesian Statistics vs Classical vs frequentist approach.

Module-2 **(10L)**

Random Variables and its distributions, Concept of location, spread, shape of a probability distribution, Stochastic inequalities and their applications, Generating functions, Emergence of random variable in real life, Probability models, Discrete and Continuous distributions, Functions of random variables , Derived random variable and their distributions.

Module-3 **(5L)**

Order statistics, Concept of ranking, Drawing of random samples, Law of large numbers, Central limit theorem.

References (Real Analysis):

1. Principles of Mathematical Analysis, Walter Rudin, 3rd Edition, McGaw Hill.
2. Mathematical Analysis, Tom M. Apostol, 2nd Edition, Wiley.
3. Mathematical Analysis, SC Malik and Arora, Multi Color Edition, New Age International Publishers.
4. Introduction to Real Analysis, Bartle, Sherbert, 4th Edition, Wiley.

References (Probability):

1. Introduction to Probability Models, Sheldon M Ross, 9th Edition, Academic Publishers.
2. Mathematical Statistics, Jun Shao, 2nd Edition, Springer.
3. Statistical Decision Theory and Bayesian Analysis, James O Berger, 2nd Edition, Springer.
4. Introductory Statistics with R, Peter Dalgaard, 2nd Edition, Springer.
5. Exercises in Probability, T Cacoullos, 1st Edition, Springer.

Statistical Inference and Introductory Analytics **Paper Code: MSASA403**

4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Apply computational skills to implement various statistical inferential approaches.
2. Demonstrate the plausibility/validity of pre-specified ideas regarding the parameters of the model through hypothesis testing.
3. Apply non parametric methods as a distribution free statistical tool.
4. Illustrate the areas of applicability of analytics and statistics specially in the domain of machine learning.

Module-1 **(10L)**

Emergence of analytics, Comparison with classical statistical analysis, Statistical inference as a business analytics tool, Basics of machine learning, Supervised learning, Unsupervised learning, Semi supervised learning.

Module-2 **(20L)**

Basics of point estimation and its application in real scenario (e.g. catch-re catch problem), Setting up null and alternative hypothesis based on real problems, Power and size of a statistical test and their interpretation, P-value and its implications, Testing under large sample, Application of hypothesis testing across different real situations, Student's t, Fisher's t, Pearsonian Chi Square Statistics, ANOVA.

Module-3 **(10L)**

Non-Parametric methods, Sign Test, Paired sign test, Mann Whitney test, Run Test, Kruskal Walli's Test, Matched Pair Test, Friedman's test, Ansari Bradley Test, Mann Kendall trend test, Confidence Interval, Sample size determination in real life, Stein's two stage procedure.

References:

1. Testing Statistical Hypotheses, E.L Lehman, Joseph P Romano, 3rd Edition, Springer.
2. Probability and Statistical Inference, Nitis Mukhopadhyay, 1st Edition, Chapman and Hall.
3. Introduction to Mathematical Statistics, Robert V Hogg, Allen Craig, 7th Edition, Pearson.
4. Statistical Inference, George Casella, Roger L Berger, 2nd Edition, Cengage Learning.
5. Linear Statistical Inference and its Applications, CR Rao, 2nd Edition, Wiley.

Analytics using Software and Programming Language **Paper Code: MSASA404**

4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Develop a meaningful pattern in data through its graphical interpretation using different programming languages.
2. Learn the implementation of analytic algorithms through coding.
3. Compare the utility and usefulness of different analytic tools including coding.
4. Experiment with large scale projects from different domains based on analytics.

Module-1 **(6L)**

Fundamental concepts of data visualization, Tableau interface, various Tableau tools for data visualization, Construction of Charts, graphs, table and maps using Tableau.

Module-2 **(6L)**

Reading Data into excel and data manipulation, Basics of data cleaning in excel, Basic functions in excel, Data organizations, Diagrams in excel, Pivot Tables, Statistical analysis with excel.

Module-3 **(14L)**

Basics Syntax in Python, Data structures in Python, Reading and manipulating csv files, Use of Numpy library, Data Processing, Data visualization in Python, Basic Statistical analysis by Python.

Module-4 **(14L)**

R as an open-source software, R packages, Basic syntaxes in R, Data import and export in R, Data visualizations in R, Linear Algebra in R, Monte Carlo Techniques in R, Advanced Statistical Analysis in R.

References:

1. Communicating Data with Tableau, Ben Jones, O'Reilly, 1st Edition.
2. Statistics for managers using Microsoft Excel, David Levine, David Stephan, Pearson, 5th Edition.
3. Python for data analysis, Wes McKinney, O'Reilly, 2nd Edition.
4. R for everyone: Advanced Analytics and Graphics, Jared P. Lander, Addison-Wesley, 2nd Edition.

Research Methodology & IPR	**Paper Code: MSASA405**
	2 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Adapt research methodology and explore the technique of defining a research problem.
2. Develop ability for literature review and carrying out a literature search.
3. Explain various research designs and their characteristics
4. Make use of the details of sampling designs, measurement and scaling techniques along with different methods of data collections.
5. Determine applicability of various forms of the intellectual property with reference to its relevance and business impact in the changing scenario of global business environment in appropriate cases.

Module-1 **(4L)**

Meaning of research problem, Sources of research problem, Criteria and characteristics of a good research problem, Errors in selecting a research problem, Scope and objectives of research problem. Approaches of investigation of solutions for research problem, Data collection, Analysis, Interpretation, Necessary instrumentations.

Module-2 **(3L)**

Effective literature studies, Approaches, Analysis. Plagiarism and research ethics.

Module-3 **(4L)**

Effective technical writing, how to write report, Paper. Developing a Research Proposal, Format of research proposal, a presentation and assessment by a review committee.

Module-4 **(3L)**

Nature of Intellectual Property: Patents, Designs, Trade and Copyright. Process of Patenting and Development: technological research, innovation, patenting, development. International Scenario: International cooperation on Intellectual Property. Procedure for grants of patents, Patenting under PCT.

Module-5 **(3L)**

Patent Rights: Scope of Patent Rights. Licensing and transfer of technology. Patent information and databases. Geographical Indications.

Module-6 **(3L)**

New Developments in IPR: Administration of Patent System. New developments in IPR; IPR of Biological Systems, Computer Software etc. Traditional knowledge Case Studies, IPR and IITs.

References:

1 “Research methodology: an introduction for science & engineering students”, . Stuart Melville and Wayne Goddard.

2. “Research Methodology: An Introduction”, Wayne Goddard and Stuart Melville.

3.“Research Methodology: A Step-by-Step Guide for beginners”, Ranjit Kumar, 2nd Edition.

4. Halbert, “Resisting Intellectual Property”, Taylor & Francis Ltd, 2007.

5.“Industrial Design”, Mayall, McGraw Hill, 1992.

6. “Product Design”, Niebel, McGraw Hill, 1974.

7.“Introduction to Design”, Asimov, Prentice Hall, 1962.

8.“Intellectual Property in New Technological Age”, Robert P. Merges, Peter S. Menell, Mark A. Lemley 2016.

9. “Intellectual Property Rights Under WTO”, T. Ramappa, S. Chand, 2008.

Programming Language Laboratory	**Paper Code: MSASA491** **2 Credits, 100 marks**

Course outcomes:

After completion of the course the students will be able to,

1. Develop a meaningful pattern in data through its graphical interpretation using different programming languages.
2. Learn the implementation of analytic algorithms through coding.
3. Compare the utility and usefulness of different analytic tools including coding.

Modules	**Name of the Topic**	**Number of Lab hours**
1	Tableau laboratory	8
2	Excel laboratory	8
3	R laboratory	12
4	Python laboratory	12

Laboratory for Statistics and Linear Algebra	**Paper Code: MSASA492** **2 Credits, 100 marks**

Course Outcome:

After completion of the course, the student will be able to

1. Understand the principles of basic statistics and acquire skills in statistical programming.
2. Interpret the fundamental of exploratory data analysis and interpret the relevant diagrams
3. Implement Python programs for different techniques of linear algebra for basic matrix operations, determination of eigenvalues and eigenvectors, matrix factorization with relevant applications in Machine learning/Data Science
4. Implement the python/R programming in practical applications using compound data.

Chapter	Name of the topic	Hours
01	Direct data collection, Data Collection from web sources, Data cleaning, Data visualizations.	4
02	Basic Matrix operations, Various decompositions, Application of matrix algebra in real life problems.	4
03	Gram-Schmidt orthogonalization, Eigenvalues and Eigenvectors and Eigen decomposition of a square matrix.	4
04	Singular value decomposition and non-negative matrix factorization of any matrix.	4
05	Solution of a set of linear equations AX = B where B does not belong to C(A), by the method of least squares.	4
06	Graphical representation of data, Problems based on measures of central tendency and dispersion.	4
07	Determination of Karl Pearson correlation coefficient and Correlation coefficient for a bivariate frequency distribution.	2
08	Application of Statistical inference tools in real-life data.	6
09	Handling data sets for basic analytics, training and testing data sets.	4
10	Probability computation using software, Selection of probability model for real-life data, Visual representation of law of large numbers and central limit theorems.	4

Term Project and Presentation I	**Paper Code: MSASA493** **2 Credits, 100 marks**

Term projects are to be done by different group of students throughout the semester followed by submission of project report and presentation.

Semester-2

Regression for Predictive Model Building	**Paper Code: MSASA406** **4 Credits, 100 marks**

Course Outcome:

After completion of the course, the student will be able to

1. Understand notion of statistical model building.
2. Demonstrate different aspects of regression diagnostics and their remedies.
3. Make use of different kinds of statistical models for suitable data.
4. Apply modern era regression based on decision tree and random forest.

Module 1 **(20L)**

Building a regression model: Transformations of data, Stepwise regression, Model selection (adjusted R2, cross validation and Cp criteria, AIC, BIC). Multicollinearity – detection and remedial measures. Dummy variables, piecewise regression, splines and scatter plot smoothing.

Detection of outliers and influential observations: residuals and leverages. Checking for normality: Q-Q plots, Normal Probability plot, Shapiro-Wilks's test. Heteroscedasticity and Autocorrelation – detection and remedies. Bayesian Regression, Robustification of regression model.

Module 2 **(8L)**

Generalized Linear Models: Introduction, Components of a GLM, Maximum Likelihood estimation, Deviance, Introduction to categorical data, Contingency tables and discrete data analysis, Odds ratio in 2x2 table, Partial and conditional odds ratio in multiway tables.

Module 3 **(8L)**

Different regression models (Logistic regression, Count regression, Polytomous regression), context, goodness of fit tests to real data.

Module 4 **(8L)**

Decision tree, algorithm for tree-based models, random forest, variable interest, interaction among predictors.

References

1. Applied Regression: An Introduction, Colin Lewis-Beck and Michael Lewis-Beck, Second Edition, 2015, SAGE.
2. Regression Modelling Strategies, Frank E. Harrell, Jr., Second Education, Springer, 2015.
3. Regression Diagnostics – identifving Influential data & sources of collinearity, David A. Belsley, Edwin Kuh, Roy E. Welsch 2013.
4. Econometric Methods, Johnston J, Edition 4, McGraw-Hill, 1996.

Optimization Techniques and Soft Computing	**Paper Code: MSASA407**
	4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Relate real life minima/maxima problems into optimization framework.
2. Apply efficient computational procedures to solve linear and non-linear optimization problems.
3. Analyze the difference between ideas of various direct and indirect search methods.
4. Choose applicability of different queuing models in real life cases.
5. Apply different soft computing techniques including Fuzzy logic and evolutionary algorithms.

Module-1 **(6L)**

Introduction: Concept of optimization – Classification of optimization problems.
Linear Programming: Examples of linear programming problems – Formulation, Simplex methods including Charne's Penalty Method, Two Phase Method and Degeneracy, Dual simplex method, Sensitivity analysis. Solution of the Transportation and Assignment problems, Shortest route problem.

Module-2 **(9L)**

Non-Linear Optimization Techniques:
Unconstrained Optimization: Maximization and minimization of convex functions, Necessary and sufficient conditions for local minima, Speed and order of convergence.

Univariate search methods: Direct Search Methods-Fibonacci and Golden Section Search, Descent Methods-Steepest Descent, Fletcher Reeves, Newton's and Conjugate Gradient methods.

Unconstrained Optimization Techniques: Necessary and sufficient condition for Kuhn – Tucker conditions for optimality, Lagrange's Multiplier Method, Gradient projection method, Penalty function methods.

Module-3 **(5L)**

Queueing Theory: Introduction of the queuing system, Various components of a queueing system. Pure Birth Process; Pure Death Process, Birth and Death Process, M/M/1 and M/M/c models, M/M/c model with finite waiting space, and models with a finite source of customers (machine interference problem). Semi-Markovian queueing systems: M/G/1, M/G/1 with service vacations and G/M/1.

Module-4 **(6L)**

Introduction to Soft Computing: Concept of computing systems, "Soft" computing versus "Hard" computing, Characteristics of Soft computing, Some applications of Soft computing techniques.

Fuzzy logic**:** Introduction to Fuzzy logic, Fuzzy sets and membership functions, Operations on Fuzzy sets, Fuzzy relations, rules, propositions, implications and inferences, Some applications of Fuzzy logic, Fuzzy Inference Systems.

Module-5 **(8L)**

Genetic Algorithm (GA): Differences and similarities between conventional and evolutionary algorithms, working principle, Genetic Operators- reproduction, crossover, mutation, Solving single-objective optimization problems using GAs.

Artificial Neural Networks: Biological neurons and its working, Simulation of biological neurons to problem solving, Different ANNs architectures, Applications of ANNs to solve some real-life problems.

Module-6 **(6L)**

Optimization in Finance: Short term financing, asset pricing and arbitrage, portfolio selection and asset allocation, the fundamental theorem of asset pricing, arbitrage detection using LP.

Supply chain optimization problems:
Introduction to Supply-Chain Optimization, Sequencing and scheduling problems in production planning, Classical machine scheduling problems.

Formulation of n -period supply-chain problem under certainty. Minimization of n period costs subject to stock conservation constraints.

Myopic Supply Policy with Stochastic Demands.

References

1. "Operations Research, An Introduction", H. A. Taha, PHI, India, 2002.
2. "Engineering Optimization: Theory and Practice", S. S. Rao, 4th Edition, John Wiley & Sons (2009). Kwang Y. Lee.
3. Fundamentals of Queueing Theory, D. Gross and C. M. Harris, Wiley, 2018.
4. Fuzzy Logic with Engineering Applications (3rd Edn.), Timothy J. Ross, Willey, 2010.
5. Foundations of Neural Networks, Fuzzy Systems, and Knowldge Engineering, Nikola K. Kasabov, MIT Press, 1998.
6. An Introduction to Genetic Algorithms, Melanie Mitchell, MIT Press, 2000.
7. Genetic Algorithms in Search, Optimization and Machine Learning, David E. Goldberg, Pearson Education, 2002.
8. Soft Computing, D. K. Pratihar, Narosa, 2008.
9. Neural Networks and Learning Machines, (3rd Edn.), Simon Haykin, PHI Learning, 2011.
10. Optimization Methods in Finance, Gerard Cornuejols and Reha Tutuncu,Carnegie Mellon University, Pittsburgh, PA 15213 USA, January 2006.
11. Lectures in Supply-Chain Optimization Arthur F. Veinott, Jr., Department of Management Science and Engineering, Stanford University, Stanford, California 94305.

Stochastic Processes and its Application	**Paper Code: MSASA408**
	4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Recall the knowledge of probabilistic methods in engineering and scientific application.
2. Demonstrate applicability of essential mathematical tools for handling random processes.
3. Make use of the stochastic simulation techniques.
4. Apply probabilistic and stochastic methods in modern engineering problems.
5. Classify different types of stochastic processes based on their properties.

Module-1 **(10L)**

Branching Processes: Brief overview of modelling -- deterministic/stochastic; discrete time / continuous time. review of discrete time branching process, extinction probabilities and asymptotic behavior, Brief excursion to continuous time branching process, general two-type birth-death branching process, Branching process with general lifetime variable (Bellman-Harris process).

Module-2 **(10L)**

Modelling in Genetics: Brief review of genetics, including the Hardy-Wienberg laws, their ramifications including mutation and fitness coefficient, Inbreeding and changes of coefficient of inbreeding over generations, Markovian models: Sib mating, Wright-Fisher, Moran, Kimura models, Wright-Fisher model with varying generation sizes, Hidden Markov models.

Module-3 **(12L)**

Random Processes: Introduction and classification of Random Processes, Stationarity and Independence of random processes, First and higher order Stationary Processes.

Markov Processes: Discrete Markov chains and Markov processes, the Markovian property, Chapman-Kolmogorov's theorem and the classification of Markov processes. Transition probability, Transition intensity, Forward and backward equations, Stationary and asymptotic distributions, Convergence of Markov chains, birth-and-death processes, absorption probabilities, time to absorption, renewal theory, martingales, Brownian motion and diffusion processes.

Module-4 **(8L)**

Epidemic Modelling: Simple and general epidemics - both deterministic as well as stochastic. Threshold theorems (without proof). Greenwood, Reed-Frost models, Neyman-Scott models of spatial spread of epidemics.

References:

1. A First Course in Stochastic Processes, S. Karlin, H. M. Taylor, Academic Press, 2nd edition 1975.
2. The Elements of Stochastic Processes, N. T. J. Bailey, Wiley-Interscience, 1991.
3. The Theory of Stochastic Processes, D. R. Cox and H. D. Miller, Chapman and Hall/CRC, 1st edition 1977.
4. The mathematical theory of infectious diseases and its applications, N. T. J. Bailey, London: Griffin, 1975.
5. Stochastic Processes, J. Medhi, Wiley Eastern Limited, 2nd edition 1994.
6. Stochastic Processes, S. M. Ross, Wiley, 2nd Edition 1996.

Time Series Analysis and Forecasting Methods **Paper Code: MSASA409**

4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Interpret time series data in the context of analytics.
2. Apply non stochastic and stochastic aspects of time series through different models.
3. Apply time series modelling in share market or other arenas where volatility is frequent.
4. Demonstrate combined analysis both involving time series as well as regression.

Module-1 **(10L)**

Introduction: Classical Models, Smoothing Techniques – exponential and Holt-Winters methods.

Module-2 **(10L)**

Evolutionary and Stationary time series. Autocorrelation and partial autocorrelation functions. AR, MA, ARMA, ARIMA, SARIMA.

Module-3 **(14L)**

Box-Jenkins Model. Volatility: ARCH, GARCH models and their variants. Multivariate Time Series Models: VAR and VARMA models. Forecasting Methods, Time series regression models.

Module-4 **(6L)**

Analysis in the Frequency Domain: The Spectrum and Periodogram Analysis.

References:

1. The Analysis of Time Series – An Introduction, C. Chatfield, CRC 1999.
2. Time Series Analysis – Forecasting & Control, G.E.P. Box, G.M. Jenkins & G.C.Reinsel Wiley, 5th edition, 2015.
3. Introduction to Time Series Analysis and Forecasting, P.J. Brockwell & R.A. Davis, Wiley, 2nd edition, 2002.
4. Analysis of Financial Time Series, Ruey S. Tsay, Wiley India (P) ltd, 2010.

Evolution of Statistical Thinking **Paper Code: MSASA410**

Audit Course

Course Objectives and outlines:

After completion of the course, the student will be able to

1. Develop awareness regarding inheritance of chance factors in different areas of life,
2. Learn the life and work of "Stalwarts of Statistics",
3. Build ideas on basic principles of probability and how they may be abused,
4. Relate risk and uncertainty with real life problems.

We all begin life with naïve realism: the doctrine that things are how they appear. But things are often not what they seem. We like to believe that effects are completely determined by comprehensible causes. But a lot of what happens in life is as much a result of random factors as is due to known ones. Thus, our past is not so easy to understand, nor is our future easy to predict; and in both pursuits we benefit from looking beyond the superficial explanation. This course will survey the history of the struggle of the human mind to comprehend and formulate the randomness and uncertainty inherent in life, and in doing so, underscore the mathematics invented in this process. Like archaeology, it will present the tools needed to identify the footprints of chance.

The contribution of Jakob Bernoulli will be a central point of reference in this journey. He had shown that through mathematical analysis one could learn how the inner hidden probabilities that underlie natural systems are reflected in the data those systems produce. However, Bernoulli left for his posterity to solve the question of how to infer, from the data produced, the underlying probability of events.

The fact that human intuition is ill suited to situations involving uncertainty was known as early as the 1930s. However, that knowledge for the most part has not trickled down from academic circles to the popular psyche. This course is an attempt to remedy that. It aspires to enable and inspire the students to integrate probabilistic and statistical thinking in the conduct of their daily lives, and to not be fooled by randomness.

This course is about the principles that govern 'chance' inherent in life, the development of those principles, and how they play out in politics, business, medicine, economics, sports, leisure, and other indispensable areas of human affairs. It is also about the way we make choices and the processes that lead us to make mistaken judgments and poor decisions when confronted with the unavoidable uncertainty and randomness of life. The course proceeds by taking the mind to the following tourist spots of inquiry:

- The hidden role of chance … when human beings can be outperformed by a rat.
- The basic principles of probability and how they are abused … why a good story is less likely to be true than a flimsy explanation.
- A framework for thinking about random situations … from a gambler in plague-ridden Italy to Let's Make a Deal.

- How to count the number of ways in which events can happen, and why it matters … the mathematical meaning of expectation.
- The extent to which probabilities are reflected in the results we observe … Zeno's paradox, the concept of limits, and beating the casino at roulette.
- How to adjust expectations in light of past events or new knowledge … mistakes in conditional probability from medical screening to the OJ Simpson trial and the prosecutor's fallacy.
- The meaning and lack of meaning in measurements … the bell curve and wine ratings, political polls, grades, and the position of planets.
- How large numbers can wash out the disorder of randomness … or why 200,000,000 drivers form a creature of habit.
- Fooled by the regularities in chance events … can the success of Wall Street gurus be random?
- Risk and uncertainty … and the deep danger of confusing the two.

References

1. The Drunkard's Walk – How Randomness Rules Our Lives" by Leonard Mlodinow, Penguin, 2009.

2. The Art of Statistics – Learning from Data" by David Spiegel halter, Pelican, 2020.

Regression and Time Series Laboratory	**Paper Code: MSASA494** **2 Credits, 100 marks**

After the completion of the course the students will be able to

1. Understand the situation for applying regression and time series to a data set,

2. Apply predictive analytics for a response variable based on a set of covariates,

3. Make use of different software for conducting predictive analytics.

Sl No	Name of the topic	Hours
01	Regression diagnostics	8
02	Remedial measures and model validation	12
03	Analysis of Non-Stochastic time series data	4
04	Analysis of Stochastic time series data	8
05	Forecasting and regression time series	8

Optimization Techniques and Stochastic Process Laboratory	**Paper Code: MSASA495** **2 Credits, 100 marks**

Course Outcome:

After completion of the course, the student will be able to

1. Demonstrate different techniques of optimization.
2. Apply available software packages for obtaining the solution of different optimization problems.
3. Demonstrate applications of stochastic processes including Markov processes and Epidemic modelling.

Sl No	Name of the topic	Hours
01	Optimization technique through software	10
02	Applications on queuing, soft computing and fuzzy logic	10
03	Applications on Markov Chain and hidden Markov models	10
04	Epidemic Modelling and Black Scholes model	10

Semester 3

Applied Multivariate and Data Mining	**Paper Code: MS ASA 501**
	4 credits, 100 marks

Course outcome:

After completion of the course, the student will be able to

1. Develop idea about multivariate data and different multivariate probability distributions.
2. Relate the inferential aspects of multivariate models.
3. Illustrate principles of handling and mining of raw industrial data.
4. Apply different statistical technique to deal with multivariate regression and ANOVA.

Module 1 **(10L)**

Concept of multivariate data with example, multiple linear regression, multiple and partial correlation coefficient, multivariate normal distribution, Wishart distribution, Hoteling's T square, Mahalanabish D square.

Module 2 **(8L)**

Clustering techniques, hierarchical clustering for categorical and continuous data. Agglomerative and divisive clustering, k means clustering, determination of number of clusters.

Module 3 **(8L)**

Discriminant analysis for classification, Bayes, Minimax and Likelihood Ratio procedures, misclassification probability and its estimation, Fisher's method for discriminating among several classes. Classification using regression techniques.

Module 4 **(5L)**

Multivariate multiple regression, test of regression parameters, multivariate ANOVA (MANOVA), Principal Component and Factor Analysis, Structural Equation Modelling.

Module 5 **(12L)**

Requirement of data mining in industry, principal steps of data mining, data mining techniques- pattern recognition, classification, association, outlier detection, clustering, regression, prediction.

References

1. Elements of Statistical Learning, Hastie T, Tibshirani R, Friedman J, 2009, Springer Text.
2. Applied Multivariate analysis, Jonson RA, Wichern DW, 2012, Pearson.

Machine Learning Algorithms	**Paper Code: MSASA502**
	4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Identify patterns and concepts from data.
2. Apply methods of Machine Learning in various functional areas & industries.
3. Analyze various machine learning algorithms and techniques with a modern outlook focusing on recent advances.
4. Differentiate supervised and unsupervised learning paradigms of machine learning.
5. Use reinforcement learning and evolutionary algorithms.
6. Apply Scalable Machine Learning techniques and various feature extraction strategies.

Module 1 **(9L)**

Supervised Learning (Regression/Classification)

Basic methods: Distance-based methods, Nearest-Neighbors, Decision Trees, Naive Bayes.

Linear models: Linear Regression, Logistic and Multinomial Logistic Regression: Logistic function, estimation of probability using logistic regression, Deviance, Wald's test, Hosmer Lemeshow test. Feature selection in logistic regression, Generalized Linear Models Support Vector Machines, Nonlinearity and Kernel Methods.

Beyond Binary Classification: Multi-class/StructuredOutputs, Ranking.
Basic building blocks of Naive Bayes classifier and learn how to build an SMS Spam Ham Classifier using Naive Bayes technique.

Module 2 **(7L)**

Unsupervised Learning

Clustering: K-means/Kernel K-means, Dimensionality Reduction: PCA and kernel PCA.
Matrix Factorization and Matrix Completion, Generative Models (mixture models and latent factormodels).

Module 3 **(6L)**

Evaluating Machine Learning algorithms and Model Selection, Introduction to Statistical Learning Theory, Ensemble Methods: Boosting, Bagging, Random Forests.

Module 4

(8L)

Feature Representation Learning, Deep Learning: Introduction to Neural Network, Convolutional Neural

Network, Recurrent Neural Network and their applications.

Module 5 **(12L)**

Scalable Machine Learning (Online and Distributed Learning) and its applications: Social recommender systems, Real time analytics, Spam filtering, Topic modeling, Document analysis.
Semi- supervised Learning, Active Learning, Inference in Graphical Models, Introduction to Bayesian Learning andInference.
Reinforcement Learning Algorithms: Markov Chains, Markov Decision Process, Policy Iteration and Value Iteration Algorithms with applications in marketing.

- Each module is to be completed with associated algorithms.

References:

1. Machine Learning: Tom M Michell, McGraw-Hill.
2. Machine Learning: A Probabilistic Perspective, Kevin P. Murphy, The MIT Press, Cambridge, Massachusetts London, England.
3. Machine Learning: Step-by-Step Guide to Implement Machine Learning Algorithms with Python, Rudolph Russell-Online Resource, Copyright 2018.
4. Machine Learning: Algorithms and Applications, M. Mohammed, M.B. Khan and E.B.M. Bashier, 2017, CRC press.

Advanced Business Analytics and Big Data	Paper Code: MS ASA 503
	4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Demonstrate the notion of big data and its application to analytics.
2. Apply the concept of advanced regression analysis.
3. Explain the ways to perform sentiment analysis.
4. Apply the methods for industrial raw data preprocessing.

Module 1: **(5L)**

Introduction to big data, concept of variety, velocity and veracity, source of big data.

Module 2: **(10L)**

Generalization of Linear Regression- Ridge Regression, LASSO, Kernel Regression, Partial least squares.

Module 3

Data preprocessing, data cleaning, data transformation, data reduction, missing data generation.

Module 4: **(10L)**

Opinion mining and sentiment analytics, aspects-based opinion summary, document sentiment classification, sentence subjectivity and sentiment classification, aspect-based opinion mining, mining comparative opinion, opinion spam detection.

Module 5 **(10L)**

Resampling methods, cross validation, Bootstrap, Jackknifing, EM algorithm.

Reference

1. Web data mining, Bing Liu, 2nd edition, Springer text, 2011.
2. Elements of Statistical Learning, Hastie T, Tibshirani R, Friedman J, Springer Text, 2009.
3. The Jackknife and Bootstrap, Shao, Jun, Tu, Dongsheng, Springer Text, 1st edition, 1995.

Advanced Analytics using Software and Programming Language (SPSS, Hadoop, SAS)	**Paper Code: MS ASA 504** **4 Credits 100 MARKS**

Course Outcome:

After completion of the course, the student will be able to

1. Experiment with big data using Hadoop.
2. Develop solution of different statistical problems using SPSS.
3. Analyze and interpret the output of SPSS and AMOS.
4. Apply statistical modelling using SAS.

Module 1: **(15L)**

Introduction to big data and Hadoop, Hadoop Architecture Distributed Storage (HDFS) and YARN, Data Ingestion into Big Data Systems and ETL, Hadoop configuration and setting up Hadoop cluster.

Module 2: **(15L)**

Getting to Know SPSS: Starting SPSS, Working with data file, SPSS windows, Menus, Dialogue boxes. Preparing the Data file: Creating data file and entering data, Defining the variables, Entering data, modifying data file, import file. Screening and cleaning data, Manipulation of data. Applications in correlation, regression model building and data mining techniques.

Structural Equation Modelling (SEM) and its applications.

Module 3: **(15L)**

Introduction of SAS System & Architecture, History and Various Modules of SAS, Features of SAS, Variables & SAS Syntax Rules, SAS Data Sets, Data Set Options, Operators, Reading Raw Data, Infile Statement With Options, Working with External Files & Options, Multiple Observations, Input Styles, Select Statement, Leave and Continue Statements, Creating & Redefining Variables, Where Statement, If – Then Else Statement, Goto, Stop And Error Statements, Output Statement, Put Statement, Do Loops, Modifying And Combining Data Sets, Updating Master Data Set, Display Manager Commands, SAS Functions, An Introduction to Arrays and Array Processing, Overview Of Methods for Combining SAS Data Set, Proc commands.

References

1. Step by Step Programming with base SAS software, SAS Institute, 2007.
2. Discovering Statistics Using IBM SPSS Statistics, Andy Field, 2018.
3. Big Data and Hadoop, V K Jain, 2018.

Analytics Laboratory	**Paper Code: MS-ASA 591**
	2 Credits, 100 Marks

Course Outcome:

After completion of the course, the student will be able to

1. Understand the idea of modern business intelligence.
2. Determine analytical solution to business problems.
3. Interpret people's behavioural pattern based on analytics.
4. Use analytics in every day's activity.

List of Problems (Indicative)

1. A company wants to open a new restaurant outlet. Give them a brief idea about tentative location, cuisine type and pricing.

2. An FMCG wants to launch a new cola drink. Give them an idea about flavour of the drink, pricing and target audience.

3. Think beyond superstition. How to decide statistically, whether a particular day is auspicious or not.

Students are encouraged to collect live data for this course. All groups are expected to provide different solutions. Students may use any software tool or any statistical technique.

Machine Learning Lab	**Paper Code:** MS-ASA592 **2 Credits, 100 marks**

Course Outcome:

After completion of the course, the student will be able to

1. Use basic techniques of Machine Learning.
2. Develop skills of using recent machine learning software for solving practical problems.
3. Develop the use of state-of-the-art methods and modern programming tools for data analysis using machine learning programs and algorithms.
4. Build the capacity for carrying independent study and research in the domain of Machine Learning.

Sl. No.	Topics	Hours
1.	Linear regression, Decision trees, overfitting	6
2.	Instance based learning, Feature reduction	6
3.	Bayes Decision Theory	6
4.	Logistic Regression, Support Vector Machine, Kernel function and Kernel SVM.	8
5.	Neural network: Perceptron, Multilayer network, Back propagation, Introduction to deep neural network.	8
6.	Clustering: k-means, Adaptive hierarchical clustering, Gaussian mixture model.	6

Use of any language (R or Python) or packages such as tensor flow, Matlab is encouraged

References:

1. Machine Learning: Step-by-Step Guide to Implement Machine Learning Algorithms with Python, Rudolph Russell-Online Resource, Copyright 2018.
2. Machine Learning: Algorithms and Applications, M. Mohammed, M.B. Khan and E.B.M. Bashier, 2017, CRC press.
3. Introduction to Machine learning with python by Andreas C. Müller and Sarah Guido, O'REILLY, 2016.
4. Introduction to Machine Learning using Python, Jeeva Jose, Khanna Publishing House.

Semester 4

Biostatistics	**Paper Code: MS ASA 507** **4 Credits , 100 Marks**

Course Outcome:

After completion of the course, the student will be able to

1. Understand different biostatistical problems.
2. Apply statistical tools in medical science.
3. Design different stages of clinical trial.
4. Infer about the root cause of disease and its spread.

Module 1 **(15L)**

Introduction. Basic functions and Models. Censoring and Truncation. Parametric univariate estimation: Standard models – exponential, Weibull, log-logistic, lognormal and Gamma. Nonparametric univariate estimation: Actuarial, Kaplan-Meier and Nelson-Aalen estimators. Tests of equality of survival functions: Gehan's and Mantel-Haenszel tests. Semiparametric regression models: Cox proportional hazard model – estimation, tests, diagnostics.

Module 2 **(15L)**

Introduction, Ethical issues in clinical trials, Types of clinical trials, Sample size determination, Group sequential monitoring. Randomized clinical trials: Randomization for balancing treatment assignments (random allocation rule, truncated binomial design, biased coin designs Incorporating covariate information. Dose response studies. Determination of optimum dose in phase 2 clinical trials.

Module 3 **(10 L)**

Definition of epidemiology. Case study on John Snow and the Lambeth cholera epidemic. Study designs: Ecological, Cross-sectional, Cohort, Case-Control and its variants, SIR model, Prospective and Retrospective studies. Standard measures of Disease frequency and association based on rates and proportions. Special emphasis on covid 19 and other epidemic case studies.

References

1. W. F. Rosenberger & J.M. Lachin: Randomization in Clinical Trials- Theory and Practice, Wiley Publications, 2002.
2. .J. Kleinbaum & M. Klein : Survival Analysis – A Self-Learning Text, Springer Text, 2005.

Advanced Analytics Laboratory	Paper Code: MS ASA 594
	2 Credits, Marks 100

Course Outcome:

After completion of the course, the student will be able to

1. Identify statistics as a tool in biomedical research.
2. Understand the application of statistics in modern technological research.
3. Use analytics in psephology.
4. Use analytics in sentiment analysis.

List of problems (Indicative)

Sl. No	Topic	Hours
1	Analysis of a clinical trial.	10
2	Analysis of astronomical data.	10
3	Analysis of demographical data.	10
4	Analysis of poll data and poll prediction.	10

Elective Topics

Multiple Testing Problems	**Paper Code: MSASA 411 A**
	4 Credits, 100 marks

Course outcome:

After the completion of the course, the student will be able to

1. Examine the problems of multiple testing and statistical inference from a modern point of view.
2. Analyse High-dimensional data in the field of biological, physical, and social sciences.
3. Apply a range of modern methods that provide statistical inference tools in the context of modern large-scale data analysis.

Module 1:

Replicability and reproducibility News coverage on replicability in medicine, psychology, political science, & other fields Intro to multiple testing: testing the global null hypothesis.

Module 2:

Global null hypothesis, Family-wise error rate, false discovery rate, Methods for multiple testing to control FWER & FDR, FWER and FDR without independence.

Module 3:

The empirical Bayes perspective on FDR Local FDR Adaptive & structured testing: hierarchies / groups of hypotheses, ordered / sequential / online testing, Adaptive & structured testing, Sign error, magnitude error Confidence intervals, false coverage rate, Introduction to selective inference, conformal prediction.

Reference

1. Testing Statistical Hypothesis, Lehmann E L, Romano J P, Springer Text.
2. Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment, S. Stanley Young, Peter H. Westfall, Wiley.

Robust Statistical Inference and application to predictive model building.	**Paper Code: MS-ASA 411 B** **4 Credits, 100 marks**

Course Outcome:
After completion of the course, the student will be able to

1. Understand the impact of outlier and influential observation on regression data.
2. Explain the notions of robust statistical inference.
3. Apply the method involving robust statistical procedure.

Module 1 **(10 L)**

Introduction to Robust Statistical Inference, The trade-off between efficiency and robustness, Historical note on robust Inference, Influence function, Breakdown point.

Module 2 **(15L)**

Introduction to minimum distance inference, M estimator, Hellinger distance estimation, concept of weighted likelihood estimator, Bregman divergence, density power divergence estimator, generalized density power divergence estimator, MAD estimator.

Module 3 **(15L)**

Alternative techniques of ordinary least square, iteratively reweighted least square, MAD estimator of regression parameters, robust testing of regression parameters, minimum distance estimation for regression parameters in linear model and generalized linear model.

References:

1. Robust Regression and Outlier Detection, Rousseew, P, Leroy A, John Wiley, 1987.
2. Statistical Inference: The Minimum Distance Approach , , Basu A, Shiyoa H, Park C, Taylor and Francis,2010.

Econometrics	**Paper Code: MSASA 411 C** **4 Credits**, **Marks 100**

Course Outcome:

After completion of the course, the student will be able to

1. Make use of Statistical Methods in Economics.
2. Demonstrate basic concepts in probability theory and statistical inference.
3. Experiment with empirical economic research as well as independent research projects in the relevant domain.
4. Implement the techniques and criticize empirical studies in economics.

Module 1: Single-equation models (**5L)**

Censored data, measurement errors, lagged variables, Meaning of Econometrics, Objectives of Econometrics, The Sources of Hypothesis used in Econometrics, The Raw Materials of Econometrics, time series and Cross section data: the problem of their pooling together.

Module 2: Simultaneous Equations **(6L)**

Identification & estimation, the use of structural models, Simultaneous equations bias, jointly dependent and predetermined variables, structural form reduced form, final form.

Module 3: The identification problem **(5L)**

Identification problem- Rank and order conditioned, methods of Estimation- Method of Indirect least squares 2 LS, method of instrumental variable MLIML, 3 SLS and FIMLM, the structure and the reduced form, Indirect least squares, instrumental variable techniques, two-stage least square.

Module 4: Unrelated regression models **(6L)**

Details of model, Estimation of model parameters, equivalence to OLS.

Module 5: Analysis of Panel Data **(5L)**

Granger causality, Exogeneity testing. Error Correction Model.

References

1. Econometric Methods, Johnston J, Dinardo J, Mcgraw Hill Education.
2. Introduction to econometrics, G.S Maddala, Macmillan, 1992.

Supply Chain Management	Paper Code: MSASA 411 D
	4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Apply the basic framework of Supply Chain Management.
2. Illustrate different aspects of SCM Analytics and its implementation.
3. Analyse the role of transportation, SC metrics, and designing automated dashboards.
4. Demonstrate modern dimensions of Supply Chain Management in terms of production planning & control.

Module-I (10L)

1. **Overview of SCM:** Value Chain, Value System and Supply Chain, Supply Chain Drivers and Obstacles, Concept of Supply Chain Management (SCM), Push-based Supply Chain vs. Pull-based Supply Chain, Trade-off between Push and Pull Strategies, Agile Supply Chain.

   (6L)
2. **Operational Aspects in SCM:** JIT and Kanban in SCM, Green Logistics, Reverse Logistics, Ethical SCM, Overview on application of SQC, SPC, TQM and TPM.

   (4L)

Module-II (7L)

3. **Role of Transportation in SCM –** Key Role Players in Transportation, Transportation Modes, Performance Characteristics and Selection, Vehicle Scheduling and Routing, Distribution Management.

   (4L)

4. **Supply Chain Analytics:** Overview of supply chain analytics, supply chain decisions, purchasing and e-commerce in supply chain, Types of supply chain, Facilities layout: Heuristic algorithm.

   (3L)

Module-III (6L)

5. **Supply chain Metrics**: Definition, key metrics – cash to cash cycle time, perfect order measurement, customer order cycle time, fill rate, supply chain cycle time, freight bill accuracy, freight cost per unit, inventory turnover, inventory days of supply, days sales outstanding, on time shipping rate, average payment period, gross margin return on investment.

   (3L)

6. **Integrated models for facility location selection and capacity allocation**: p-Median location models, fixed charge location-allocation models, gravity models, locating plants and warehouses simultaneously, location and allocation in multiple stages.

   (3L)

Module-IV (9L)

7. **Resource planning and control**: Deterministic inventory models – EOQ, EBQ, quantity discounts, and multiple items inventory models, buffer/safety stock model, intentional shortages. Probabilistic inventory models (lot sizing for time varying demand), periodic review model.

(3L)

8. **Production planning decisions**: Aggregate planning in a supply chain – role of aggregate planning in a supply chain, aggregate planning strategies, role of IT in aggregate planning in a supply chain, aggregate planning in practice. Tools and methods for aggregate planning – graphical approach, tabular method, linear programming, transportation problem, dynamic programming.

(6L)

Module-V (8L)

9. **Production control decisions**: Scheduling – shop loading and sequencing, branch and bound algorithm, heuristics, line balancing, JIT, TOC (Theory of Constraints).

(4L)

10. **Designing automated dashboards** with relevant KPIs for production planning and control, resource planning and control, material planning and control, capacity planning and control, quality planning and control and distribution system optimization.

(4L)

Suggested Readings

1. Chopra,S., Meindl, P., VirKalra, D., Supply Chain Management: Strategy, Planning and Operation, Pearson.

2. Palekar, A. & Shiralkar, S.W., Supply Chain Analytics with SAP NetWeaver Business Warehouse, Tata McGraw Hill.

3. Krajewski, L.J., Malhotra, M.K., Ritzman, L. P., & Srivastava, S.K., Operations Management: Processes and Supply Chains, Pearson.

4. Srinivasan, G., Quantitative Models in Operations and Supply Chain Management, PHI.

5. Mathirajan, M., Rajendran, C., Sadagopan, S., Ravindran, A., & Balasubramanian, P.(Eds.), Analytics in Operations/Supply Chain Management, I.K International Publishing House Pvt. Ltd.

Health Informatics	Paper Code: MSASA 505A
	4 Credits , 100 Marks

Course Description- This course provides an introduction to health informatics, the field devoted to the optimal use of data, information, and knowledge to advance individual health, health care, public health, and health-related research. It introduces the concepts of population health and precision medicine and the information systems that support them. It covers basic principles of knowledge management systems in biomedicine, various aspects of Health Information Technology standards, and IT aspects of clinical process modeling.

Course Outcome:

After completion of the course, the student will be able to

1. Learn the basic definitions, key concepts, terminology, and historical context of Health Informatics.
2. Understand fundamental characteristics of data, information, and knowledge in the Health Informatics domain.
3. Apply common algorithms for health applications and IT components in representative clinical processes.
4. Develop understanding of population health and precision medicine.
5. Understand basic principles of knowledge management systems in biomedicine.
6. Develop understanding of various aspects of Health Information Technology standards.
7. Relate IT aspects of clinical process modeling with health information systems.

Module-1

Introduction to Health Informatics & Data, Information, and Knowledge: (8L)

Introduction to health informatics and its significance, Definitions and key concepts in health informatics, Background disciplines, Historical overview and future challenges.

Introduction to knowledge hierarchy: Data, information, and knowledge, The definitions of healthcare data and information, Types of healthcare information (internal versus external data and information), The major purposes of maintain patient records, The content and uses of patient records and claim content, The common issues related to healthcare data quality, The challenges associated with measuring and ensuring healthcare data quality, Quality assessment including total quality management and data quality, Introduction to biomedical research and publicly available resources.

Module-2

The National Landscape of Healthcare It & History of Healthcare Information System: **(6L)**

The major influences shaping the health IT landscape in India, The roles played by the major government initiatives and private sectors in advancing health IT in India, The major events that have influences the adoption of health IT and systems.

History and evolution of healthcare information systems (HCIS), The major advances in information technology and significant federal initiatives that influenced the adoption of healthcare information systems, The major types of administrative and clinical information systems used in healthcare Current issues pertaining to the use HCIS, HIM Application and System.

Module-3

Medical Algorithms & Medical Decision Making (8L)

Various ways to describe algorithms, such as flowchart, pseudo code, and conceptual graph, Introduction to medical algorithms, Algorithms in computer science, such as decision tree and regression, Calculation of measurements of classification performance—sensitivity and specificity.

Decision-making process, Medical decision-making process (diagnosis, treatment, monitoring, prognosis), Informatics in clinical decision-making, Introduction to evidence-based medicine.

Compare and contrast electronic medical record (EMR) with electronic health record (EHR), Identify the more common controlled vocabularies in use today: ICD, CPT, NDC, RxNorm, LOINC and SNOMED CT.

Module-4

Modeling and Simulations & Population Health and Precision Medicine (6L)

Develop understanding of modeling and simulation, Become familiar with applications of modeling and simulation in biomedicine, Medical Coding, International Classification of Disease (ICD-10).

Health Information System and Health Information Analysis:
Data and information need of health systems in managing population health, Key health
IT tools and strategies for population health management, Concepts of precision medicine.

Module 5:

Standards in Health Informatics (6L)

Introduction to standards, The Need for Health Informatics Standards, The role of federal initiative and legislation that that have significant impact on the adoption of healthcare information standards in India.

Healthcare Law and Ethics, Major types of healthcare information standards and the organizations that develop or approve them.

The importance of healthcare IT standards to the future of the Indian health care delivery system, Health Information Governance.

Module 6:

Knowledge management system & Organizing Health IT services (6L)

Introduction to knowledge management, Knowledge discovery, Data mining and text mining, Knowledge management and decision-making support in biomedicine.

The roles, responsibilities, and major functions of the IT department in healthcare organizations.

Database Management Systems and Data Quality, Database Programming, Statistics for Health Sciences, Electronic Medical Record.

References:

1. Healthcare Information Systems and Informatics: Research and Practices by Joseph Tan, Information Science Publishing, 2008.
2. Implementing an Electronic Health Record System (Health Informatics), edited by James M. Walker, Eric J. Bieber, Frank Richards and Sandra Buckley, 2006.
3. Health Informatics: Practical Guide for Healthcare and Information Technology Professionals by Robert E. Hoyt and Ann K. Yoshihashi, Informatics Education, (Sixth Edition), 2014.
4. Innovation in Health Informatics: A Smart Healthcare Primer, Edited by Miltiadis D. Lytras and Akila Sarirete, Academic Press, 2020.

Marketing Analytics	**Paper Code: MS ASA 505B**
	4 Credits,100 marks

Course Outcomes:

After completion of the course, the student will be able to

1. Demonstrate the basic framework of Marketing Analytics.
2. Explain different aspects of Marketing Analytics and its implementation.
3. Analyze the impact of segmentation, targeting and positioning using marketing metrics.
4. Use modern dimensions of Digital Marketing including SEO & SEM.

Module-1:

1. **Overview of Retail & Marketing Analytics:** Need for Data Driven Marketing Approach, Model Building in Marketing Engineering, Basic Principles of Marketing Analytics to Business Problems, Slicing and Dicing Marketing data with Pivot Tables, Excel Charts to summarize data.

(5L)

2. **Positioning Strategies:** Concept of Product positioning, Conduct a Positioning Study, Perceptual Mapping using Principal Component Analysis (PCA), Multidimensional Scaling (MDS).

(5L)

Module 2:

3. **Customer Lifetime Value (CLV):** Concept of CLV, Comparison of CLV with related metrics, Analyzing CLV, Extensions of CLV Analysis, Drivers of CLV, Uses of CLV metrics. **(5L)**

4. **Product Designing:** Concept of Product Designing, Conjoint Analysis as a decompositional preference model, Steps in Conjoint Analysis, Uses of Conjoint Analysis. **(5L)**

Module 3:

Strategic Marketing Analytics: The STP framework, Value generation through STP framework, Managing the segmentation process, Segmentation in Real world: Cluster Analysis, Hierarchical and Non-Hierarchical - K Means Clustering, Prediction of Customer's segment membership: Discriminant Analysis (DA), two Group DA. **(5L)**

5. **Marketing Forecasting:** Simple Regression and Correlation, Multiple Regression to forecast sales, Modeling trend and Seasonality, Ratio to Moving Average Method, Winter's Method.

(5L)

Module 4:

6. **Fundamentals of Digital Marketing**: concept, history, types, implementation and benefits of digital marketing, application of Markov Chain using Google Analytics. **(5L)**

7. **SEO & SEM**: concept of Search Engines optimization, how SEO operates, website domain, file name, design layouts, optimized keywords, keyword frequency weightage, prominence, placement of keywords, finding keyword, word stemming, metatag optimization, title optimization, anchor optimization, mobile SEO techniques. Concept, as a marketing tool, importance of social media marketing, social marketing strategy (SMO) for business, SMO key concepts, business profile creation, brand awareness, social engagement; Viral marketing, tools of measurement of popularity, traffic, analytics and statistics. **(5L)**

References:

1. Marketing Analytics - Data-driven Techniques with Microsoft Excel by Wayne L. Winston.
2. Marketing and Sales Analytics: Proven Techniques and Powerful Applications from Industry Leaders by Cesar A. Brea.
3. Marketing Analytics by Winston.
4. Digital Marketing Analytics: Making sense of Consumer Data in digital world, Chuck Hemann, Ken Burbary; Que Publishing.
5. Advanced Customer Analytics: Targeting, Valuing, Segmenting and Loyalty Techniques' Mike Grigsby.

Statistical Fluid Mechanics	**Paper Code: MS ASA 505C**
	4 Credits, 100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Relate with the properties of fluids and the applications of fluid mechanics.
2. Analyze problems related to calculation of forces in fluid structure interaction.
3. Understand the concept of fluid measurement, types of flows and dimensional analysis.
4. Develop understanding about hydrostatic law, principle of buoyancy and stability of a floating body and application of mass, momentum and energy equation in fluid flow.
5. Understand fundamental knowledge of fluid, its properties and behavior under various conditions of internal and external flows.

Module 1 **(12L)**

Mathematical Description of Turbulence, Mean Values and Correlation Functions: Statistical formulation of Fundamental Turbulence Problem, Moments of Fluid Dynamic fields, Random fields with a Normal Probability Distribution (Gaussian Fields), Laminar flows, Statistical analysis of Boundary layer flow, Statistical study of hydromagnetic boundary layer flow, Statistical paradigm, Statistical probable error, Statistical analysis of Motion of circular and elliptic cylinders. Steady streaming with circulation. Navier-Stokes equation of motion of a viscous fluid.

Module 2 **(10L)**

Kinematics of Fluid in Motion: Eulers equation of motion, Lagrange motion. Boundary conditions and boundary surface stream lines and paths of particles. Irrotationaland rotational flows. Impulsive action. Equations of motion, Some symmetrical forms of the equation of continuity, cylindrical and spherical.

Module 3 **(10L)**

One-dimensional inviscid incompressible flow: Theory of irrotational motion flow and circulation. irrotational motion, Application of Bernoulli's equation and theorem, Energy theorem. Uniqueness theorem. Complex potential, sources. sinks, doublets and their images circle theorem. Theorem of Blasius.

Module4 **(12L)**

Vortex motion. Vortex line and filament equation of surface formed by stream lines and vortex lines in case of steady motion. Strength of a filament. Velocity field and kinetic energy of a vortex system. Uniqueness theorem rectilinear vortices. Vortex pair. Vortex doublet.
vortex sheet. Laminar flow of viscous incompressible fluids. Steady and unsteady incompressible flow with constant fluid properties.

References:

1. Statistical Fluid Mechanics: Mechanics of Turbulence-A.S. Monin and A.M.Yaglom, MIT Press
2. Basic Development in Fluid Mechanics-Reid, W.H., Academic Press
3. Fluid Dynamics-M. D. Raisinghania, S. Chand & company PVT. LTD.

Reliability Analysis	**Paper Code: MS-ASA505D**
	4 Credits,100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Explain major concepts of reliability prediction Reliability and Apply in the field of Software Testing.
2. Apply in Software Testing Reliability: -Feature Testing, Regression Testing, Load Testing.
3. Analyze statistical experiments leading to reliability modelling.

Module 1: **(12L)**

Reliability of System of independent component, Example of series, Parallel, k out n system
Properties of System Reliability, Computing system reliability: Bridge System, Bridge Structure
Reliability importance of components: Series, parallel, 2 out of 3 system , Problems.

Module 2: (**8L)**

Modular Decomposition of the coherent system: series, parallel.
Simple problems related to modular decomposition. Modular Decomposition of a coherent system: stereo, bridge, 2 out of 3 systems, 3 out of 4 systems.

Module 3: (**4L)**

Dual of a system., properties of Dual system, Dual of Series system, Dual of k out of n system, Dual of stereo system, Properties of coherent Dual system.

Module 4: **(6L)**

Path vector, Cut vector, minimum path vector, minimum cut vector, Finding MPS, MCS of different systems. Expressing a system in terms of series and parallel combination of MPS and MCS.

Module 5: (**10L)**

Time to failure, Hazard rate, Reliability function, Failure rate, Relationship between f(t), F(t), R(t),

Notion of ageing, IFR and DFR distribution, Bathtub Curve, System Reliability with exponential distribution, series system, parallel system, Reliability of coherent system, K out of n system, system reliability of some special cases.

Reference Books:

1. Reliability in Engineering Design, Kapur and Lamberson, Wiley, 1977.
2. Statistical Theory of Reliability and Life Testing, Richard E Barlow and Frank Proshan, New York: Holt, Rinehart & Winston, Inc., 1975

Advanced Bayesian Inference **Paper Code: MS ASA 508 A**

4 Credits,100 marks

Course Outcome:

After completion of the course, the student will be able to

1. Identify the paradigm shift from classical inference to Bayesian Inference.
2. Demonstrate different aspects of Bayesian Inference.
3. Apply Bayesian methods in predictive modelling.

Module 1 **(5L)**

Bayesian Linear Regression with Conjugate priors, Bayesian Model Selection, Bayesian Information Criterion, Bayesian econometrics.

Module 2 **(10L)**

Monte Carlo Method, Markov chains and MCMC, Gibbs Sampling with examples in R and WinBUGS, The Metropolis-Hastings Algorithm.

Module 3 **(10L)**

Generalized linear models and categorical data, longitudinal models, Bayesian multiple imputation. Bayesian Hypothesis Testing (One-sided and Two-sided Example), The Bayes Factor, A Test for Comparing Two Population Means.

Module 3 **(15L)**

Hierarchical Bayes Examples, Exchangeability, Hierarchical Bayesian Analysis Shrinkage and Bayesian Estimation, Empirical Bayes Estimation (with examples), Comparison of Hierarchical vs. Empirical Bayes. Bayesian Probit and Logistic Regression (Multi-category Ordinal Response). Multinomial-Dirichlet model.

References

1. An Introduction to Bayesian Analysis , Ghosh JK, Delampady M, Samanta T, Springer Text.
2. Statistical Decision Theory and Bayesian Analysis, Berger,J, 1985, Springer Text.

Sports and Education Analytics	**Paper Code: MSASA 508 B**
	4 Credits,100 MARKS

Course Outcome:

After completion of the course, the student will be able to

1. Learn the importance of analytics in gaming strategy.
2. Understand important factors in sports and to create appropriate performance metric.
3. Learn the importance of analytics in education sector.
4. Apply better teaching strategy based on analytics of students' response.

Module 1 **(15L)**

Illustration of ERP as an education analytics repository, analytics for career selection, analytics for betterment of teaching, analytics for betterment of student-teacher relationship.

Module 2: **(25L)**

Different performance measures of cricketers or soccer player, analytics on performance of a player in a game, prediction of a game based on players' statistics, predictive analytics of soccer leagues or cricket tournament.

References

1. Sports Analytics and Data Science: Winning the Game with Methods and Models, Thomas W Miller, Pearson FT Press, 1st edition, 2015.
2. Machine Learning and Data Mining for Sports Analytics, edited by Ulf Brefeld, Jesse Davis, Jan Van Haaren and Albrecht Zimmermann, Springer Text, 2020.

Astrostatistics	**Paper Code: MS ASA 508 C**
	4 Credits, 100 MARKS

Course Outcome:

After completion of the course, the student will be able to

1. Understand the classification of stellar objects,
2. Demonstrate different astronomical problems and emergence of astronomical data,
3. Perform Classification of different Steller objects and galaxies by means of statistical Techniques.

Module 1 **(12L)**

Spectral Classification of Stars: Saha's equation, Harvard System, Absolute and apparent magnitude, Mass luminousity relation, Parallax. Stellar Structure, Polytropic and homologous models.

Module 2 **(15L)**

Evolution of Stars: Observational basis, Sources of stellar energy, Hertzsprung-Russell diagram, evolution of low and high mass stars, Chandrasekhar limit. (5) Stellar populations- Galactic and Globular Clusters. Galaxies-Classification-surface brightness profile-fundamental plane and its significance- Rotation curves-Missing mass and dark matters. Astronomical Measurement Errors: Statistical issues and problems.

Module 3 **(14L)**

Data archives and Virtual Observatories. Applications of PCA and ICA for Steller object and galaxy classification.

References:

1. Astro statistics, Babu, G.J., Feigelson E. D., Chapman and Hall,1996.
2. Data Analysis in Astronomy, Ettore Majorana International Science Series, Di Gesu, V, Searsi L, Crane P, Friedman, J.H. and Levialdi S. (Editors), Springer-Verlag New York Inc.,2012.
3. Statistical Methods for Astronomical Data Analysis, Chattopadhyay, A. K. and Chattopadhyay, T., Springer-Verlag New York Inc.,2014.