Savitribai Phule Pune University

(Formerly University of Pune)

Two Year Degree Program in Statistics

(Faculty of Science & Technology)

Revised Syllabi for

M. A. / M. Sc. Program in Statistics (Under NEP)

(For Colleges Affiliated to Savitribai Phule Pune University)

To be implemented from Academic Year 2023-2024

1. Title of the course: M. A. / M. Sc. in Statistics

2. Preamble of the syllabus: M. A. / M. Sc. Statistics program is of **88** credits spread over four semesters. It has an exit option at the end of the first year (after two semesters) with a Post-Graduate Diploma in Statistics. Also, eligible students can join (entry) directly in the second year of the program for M.A./M.Sc. after the completion of PG diploma or four-year B.A./B.Sc. honours degree in Statistics. This program is offered at the colleges affiliated to the Savitribai Phule Pune University. The program emphasizes theory, practical and modern applications of statistics using practical data analysis and is structured to provide knowledge and skills in depth necessary for the employability of students in industry, other organizations, as well as in academics.

The syllabus has a good balance of theory, methods, practical applications of statistics, research skill development, industrial exposure and two project components.

It is possible for the students to study basic courses from other disciplines such as economics, life sciences, computer science, modelling and simulation and mathematics in place of electives.

3. Introduction: M.A./M.Sc. Statistics program has a semester pattern and a Choice-based Credit System. The program consists of 88 credits.

4. Eligibility Criteria: For M. A. in Statistics the eligibility criteria are as follows:

(i) B. A. (50% marks or equivalent grade) with Statistics as major and Mathematics or Economics as minor level
(ii) B. A. (50% marks or equivalent grade) with Mathematics as major and Statistics at minor level
(iii) B. A. (50% marks or equivalent grade) with Actuarial Science as major and Statistics at minor level
(iv) B. A. (50% marks or equivalent grade) with both Statistics and Mathematics at minor level subjects.

For M. Sc. in Statistics the eligibility criteria are as follows:

(i) B.Sc. (50% marks or equivalent grade) with Statistics as major and Mathematics/ Physics/ Computer Science/Economics at Minor level
(ii) B.Sc. (50% marks or equivalent grade) with Mathematics as major and Statistics at the minor level
(iii) B.Sc. (50% marks or equivalent grade) with Actuarial Science as major and Statistics at the minor level
(iv) B.Sc. (50% marks or equivalent grade) with both Mathematics and Statistics at the minor level

For second year of M.A./M.Sc. admission, the eligibility criteria will be as follows:

(i) Four years B.A./B.Sc. honours in Statistics (50% marks or equivalent grade)
(ii) Post Graduate Diploma in Statistics (50% marks or equivalent grade)

5. Examination

A) Pattern:
(i) **Pattern of examination:** There would be continuous internal assessment (CIA) and an end-of-term examination (ETE) for each course. Both CIA and ETE have 50% weightage. The CIA includes class tests or quizzes, assignments, small projects/ practicals, viva-voce and presentations. There would be assignments or minor projects for someof the elective courses in ETE also.

(ii) **Pattern of the question paper:** For theory/practical courses the duration for the ETE will be three hours for a four-credit course and two hours for a two-credit course.

B) Standard of passing: A student has to obtain 40% marks in the combined grading of the ETE and the CIA for passing the course, with a minimum passing of 30%in both CIA and ETE separately.

C) Award of class: As per the University rules.

D) External students: Not applicable

E) Setting of question paper: As per the University rules.

F) Verification or revaluation: As per the University rules

6. Structure of the Program

The project in Semester III must be carried out as an individual project or a group project with a group size equal to 2 or 3 and each group will be assigned a supervisor.It is expected that students should discuss their project work with their supervisor, for a minimum of **TWO** hours per week. The project involves solving a real-life problem using statistical methods for primary/secondary data.

The project in Semester IV must be carried out as an individual project or a group project with a group size equal to 2 or 3 and each group will be assigned a supervisor. Students are allowed to extend their 3^{rd} semester projects to 4^{th} semester project. It is expected that students should discuss their project work with their supervisor, fora minimum of **THREE** hours per week. **Fourth semester project is expected to have some original contributions, in the form of an algorithm or methodology. The project report/outcome is expected to be converted to one/two publishable research paper(s) after necessary plagiarism checks.**

All the courses offered under M.A./M.Sc. Statistics programs will be available to students from other Departments, whenever they are offered. However, the eligibility for a particular course (Major as well as elective), will be decided by the Teaching and Academic Committee of the Department.

Candidates can get admission directly to Sem III (Year II of the M.A./M.Sc. Program) as per availability of seats and existing intake capacity.

Equivalent courses (Electives) will have to be approved by the DEC and the Head of Department.

After successfully completing Year I (clearing all the necessary papers) of the program, students who wish to take a break will be awarded a **PG Diploma in Statistics**. All those who have been awarded PG Diploma from the University are allowed to rejoin the Master's program only after a gap of minimum one year.

M.A./M.Sc. Statistics Program Structure (as per NEP) from 2023-2024

Savitribai Phule Pune University

Year	Level	Sem	Course Type	Paper Title	Credits
I	6.0	I	Major – Mandatory	STM11: Fundamentals of Analysis & Calculus	4
I	6.0	I	Major - Mandatory	STM12: Linear Algebra	4
I	6.0	I	Major - Mandatory	STM13: Probability Distributions	4
I	6.0	I	Major - Mandatory	STM14: Data Analytics using R (Practical)	2
I	6.0	I	Elective I	STE101: Optimization Techniques STE103: Reliability and Statistical Quality Control STE104: Actuarial Statistics Equivalent course (to be approved bythe DEC/Dept.)*	4
I	6.0	I	Research Methodology	STRM1: Research Methodology (This course should involve topics from Research Methodology, Critical Thinking, Computational Statistics,Data Representation & Visualizationetc.)	4
I	6.0	II	Major – Mandatory	STM21: Modern Statistical Inference	4
I	6.0	II	Major - Mandatory	STM22: Regression Analysis & Applications	4
I	6.0	II	Major - Mandatory	STM23: Multivariate Analysis & Applications	4
I	6.0	II	Major - Mandatory	STM24: Data Analytics using Rand/or Python (Practical)	2
I	6.0	II	Elective II	STE204: Advances in Generalized Linear Models STE205: Statistical Methods in Epidemiology STE208: Discrete Data Analysis Equivalent Course (to be approved by the DEC/Dept.)*	4
I	6.0	II	OJT/FP	ST OJT-FP: Six weeks internship in the industry with a minimum of 25 days (Seven hours per day) working (log sheet required) along with a report / Conduct a field survey with the analysis and report with an equal amount of work / Any other similar activity that requires an equivalent amount of work which can be done in other (nearby) Research Institutes with researchers or Scientists	4
Total					**44**

Year	Level	Sem	Course Type	Paper Title	Credits
II	6.5	III	Major - Mandatory	STM31: Probability Theory	4
II	6.5	III	Major - Mandatory	STM32: Stochastic Processes	4
II	6.5	III	Major - Mandatory	STM33: Design and Analysis of Experiments	4
II	6.5	III	Major - Mandatory	STM34: Advanced Data Analytics using R and/or Python (Practical)	2
II	6.5	III	Elective III	STE301: Statistical Learning & Data Mining STE302: Survival Analysis STE307: Asymptotic Inference Equivalent Course (to be approved by the DEC/Dept.)*	4
II	6.5	III	Research Project	STRP3: Research Project – I	4
II	6.5	IV	Major - Mandatory	STM41: Time Series Analysis	4
II	6.5	IV	Major - Mandatory	STM42: Bayesian Inference	4
II	6.5	IV	Major - Mandatory	STM43: Sampling Theory and Applications	4
II	6.5	IV	Elective IV	STE404: Computer Intensive Statistical Methods STE406: Advanced Statistical Learning Techniques & Applications STE407: Design & Analysis of Clinical Trials Equivalent Course (to be approved by the DEC/Dept.)*	4
II	6.5	IV	Research Project	STRP4: Research Project – II	6
Total					**44**

*'Equivalent Course' is an online course from NPTEL/SWAYAM or an Offline course from another department/Institute. Students have to take permission from the DEC/Dept./Institute before finalizing an elective course which is not listed here.

Detailed Syllabus

STM11: Fundamentals of Analysis & Calculus – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course, the students will be able to	
1. understand the concepts of mathematical analysis	Understand
2. understand the concepts of limits and convergence of sequences and series and solve related problems	Understand Evaluate
3. understand the concepts of limits and continuity of functions solve problems related to these concepts	Understand and Evaluate
4. solve the problems related to univariate differential calculus	Evaluate
5. solve the problems related to multivariate differential calculus	Evaluate
6. apply the techniques for finding the optimum of functions	Evaluate

Unit I

Countability, supremum and infimum of sets of real numbers, denseness property of rational numbers, limit points and interior points of a set, open sets, closed sets, their properties, Compactness

Unit II

Sequences of real numbers, Cauchy sequence, limit superior, limit inferior, limit and convergence of a sequence of real numbers, Cauchy criterion for convergence.

Series of real numbers, convergence of series, tests for convergence of series, absolute convergence, Cauchy product of two series and its convergence. Power series and radius of convergence

Unit III

Functions, continuity, uniform continuity, absolute continuity, functions of bounded variation, calculus of one variable: differentiability, mean value theorem and Taylor series expansion.

Functions of several variables, continuity, partial derivative, the gradient vector, directional derivatives, differentials of functions of several variables, properties, convex and concave functions

Unit IV

Differentials of composite functions, chain rule, the mean value theorem, a sufficient condition for the existence of the differential, partial derivatives of higher order and Taylor's formula. Applications of partial differentiation, Jacobians, extremum problems involving functions of several variables and their applications, constraint optimization of functions with several variables and their applications

Books Recommended:

1. Abbott, S. (2001), *Understanding Analysis*, Springer, New York
2. Apostol T.M. (1975). *Mathematical Analysis: A modern approach to advanced calculus*. Addison-Wesley
3. Bartle R. G. and Sherbert D. R., (2007), *Introduction to Real Analysis*, Wiley
4. Bartle, R. G. (1976). *Elements of Real Analysis*, John Wiley
5. Ghorpade, S. R. and Limaye, B. V. (2006). *A Course in Calculus and Real Analysis*, Springer
6. Ghorpade, S. R. and Limaye, B. V. (2010). A *Course on Multivariable Calculus and Analysis*, Springer
7. Goldberg R. R. (1976). *Methods of Real Analysis*, John Wiley
8. Kreyszig, E. (1975). *Advanced Engineering Mathematics*, Wiley Eastern
9. Kumar, A. and Kumaresan, S. (2014). *A Basic Course in Real Analysis*, CRC Press
10. Radulescu, T. T., Radulescu V. D., Andreescu T., (2009), *Problems in Real Analysis*, Springer, New York
11. Rudin, W. (1985). *Principles of Mathematical Analysis*, McGraw - Hill
12. Trench W. F. (2012). *Introduction to Real Analysis*, E-book.
13. Yau, D. (2013). *A First Course in Analysis*, World Scientific

STM12: Linear Algebra – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course, the students will be able to	
1. solve the problems related to vector spaces	Evaluate
2. solve the problems related to matrix algebra and linear transformations	Evaluate
3. solve problems related to s y s t e m of linear equations	Evaluate
4. understand the concepts of eigenvalue theory solve problems related to eigenvalues of a matrix	Understand and Evaluate
5. understand the concepts of quadratic forms and solve problems related to these topics	Understand Evaluate
6. understand the concepts of matrix derivatives	Understand
7. apply the concept of decomposition of a matrix	Apply

Unit I

Vector spaces, inner product of vector spaces, linear dependence and linear independence of vectors, bases, an orthogonal basis, basis and dimension, properties and uses of a basis.

Linear transformation and their matrix representations, injective, surjective and inverse linear transformations, rank of a matrix, linear equations, solution space and null space, generalized inverse, echelon forms, canonical forms, Gram-Schmidt orthogonalization, projection theorem.

Unit II

Determinants and their simple properties, partitioned matrices, inverses, vector operator, special types of matrices, orthogonal and idempotent matrices, symmetric and positive definite matrices.

Characteristic roots of real matrices, right and left characteristic vectors, linear independence of characteristic vectors corresponding to distinct characteristic roots, algebraic and geometric multiplicities, Cayley-Hamilton theorem.

Matrix inequalities, rank, determinant, and trace inequalities, eigenvalue inequalities

Generalized inverses: Moore-Penrose inverse, G-inverse.

Unit III

Quadratic forms with symmetric matrices, definiteness of a real quadratic form, reduction of quadratic forms, simultaneous reduction of two quadratic forms, maxima and minima of ratios of two quadratic forms, quadratic form inequalities

Unit IV

Derivatives with respect to vectors and matrices. LU factorization, Cholesky factorization, spectral decomposition, singular value decomposition, applications

Books Recommended:

1. Bapat, R.B. (2011). *Linear Algebra and Linear Models*. Springer and Hindustan Book Agency.
2. Beezer, R. A. (2004). *A First Course in Linear Algebra*, Congruent Press, Washington
3. Gilbert, S. (2014). *Linear Algebra and Its Applications*, 4th Ed., Cengage Learning India Pvt. Ltd.
4. Hohn, F. E. (1973). *Elements of Matrix Algebra*, Macmillan
5. Kollo, T. and Rosen, D. von (2005). *Advanced Multivariate Statistics with Matrices*, Springer, New York.
6. Kumaresan, S. (2000). *Linear Algebra: A Geometric Approach*, Prentice Hall
7. Lay, D. C. Lay, S. R. and Mc Donald, J. J. (2016). *Linear Algebra and Its Applications*, Fifth Edition, Pearson, Boston.
8. Ramachandra Rao, A. and Bhimasankaram, P. (2000). *Linear Algebra*. Hindustan Book Agency
9. Rao, C. R. (1995). *Linear Statistical Inference and Its Applications*, Wiley
10. Searle, S. R. and Khuri, A. I. (2017). *Matrix Algebra Useful for Statistics*, 2nd Ed., John Wiley, New York.

STM13: Probability Distributions – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to class of sets such as fields, sigma fields, Borel fields and solve related problems	Understand Evaluate
2. understand the measure theoretic definition of a random variable random vector and solve problems related to their distributions	Understand and Evaluate
3. solve the problems related to distribution function	Evaluate
4. solve problems related to quantile function	Evaluate
5. understand the concepts such as truncation, symmetry, convolution mixture, compound etc. and solve related problems	Understand Evaluate
6. solve problems related to multiple and partial correlations	Evaluate
7. understand the concepts related to sampling distributions solve problems related to them	Understand and Evaluate
8. understand the theory related to linear and quadratic functions involving normal random vectors and solve related problems	Understand Evaluate
9. understand the concepts related to order statistics and solve problems related to the distributions of order statistics	Understand Evaluate

Unit I

Random experiments, sample spaces, classes of sets, fields and sigma-fields, limit of sequences of subsets, sigma- field generated by a class of subsets, Borel fields, Borel sigma fields on R and (0,1), probability measure on a sigma-field, probability space, continuity of a probability measure. Real valued functions on Ω, properties of inverse images, real and vector-valued random variables

Unit II

Probability spaces, properties of probability measures including monotonicity and continuity, Probability measures on finite and countable infinite sample spaces. Cumulative distribution function (c.d.f.) of a random variable, necessary and sufficient conditions for a function to be a cumulative distribution function, symmetry of a distribution, Quantile functions and their properties, quantile functions as random variables and their c.d.f. Independence of events and random variables, identically distributed random variables,

Continuous, discrete and mixed distribution functions, decomposition theorem. Density function and distribution functions defined in terms of density functions, continuity and differentiability of such distribution functions. Singular and absolutely continuous distribution functions, concept of survival function, hazard rate and cumulative hazard rates. Truncated distributions (binomial, truncated Poisson, normal etc.)

Expectation of random variables, existence and finiteness of expectations, Probability generating function (p.g.f.) and moment generating function (m.g.f.) and their properties, Stieltjes moment problem.

Unit III

Cumulative distribution functions of a random vector, lower dimensional marginal distributions, necessary and sufficient conditions for a function to be a bivariate distribution function, independence of random variables in terms of distribution functions,

Construction of multivariate distribution functions that satisfy the sufficient conditions, bivariate density functions and related distribution functions, marginals do not determine the joint distributions uniquely, conditional densities and conditional distributions.

Expectations and moments of random vectors. mixed moments, variance-covariance matrix, conditional expectation and variances, multiple and partial correlationcoefficients joint m.g.f and relation to marginal m.g.f. and moments, convolutions, mixtures, compound distribution.

Unit IV

Multinomial distribution and joint distributions of order statistics, functions of random vectors and their joint distributions distribution of spacings, normalized spacings with illustration to exponential case, distribution of sample median and sample range.

Sampling distributions of statistics from univariate normal random samples, non-central chi-square, non-central t and F distributions.

Bivariate and multivariate normal distribution, m.g.f. linear and quadratic transformations of multivariate normal vectors, their distributions and properties, Fisher-Cochran theorem.

Multivariate beta, exponential, binomial, Poisson distributions and their properties.

Books Recommended:

1. Berger, R. and Casella G. (2002). *Statistical Inference*, Duxbury Resource Center, Second Edition.
2. Bhat, B. R. (2007). *Modern Probability Theory: An Introductory Text Book*, New Age International
3. Billingsley, P. (1995). *Probability and Measure*, 3rd Ed., John Wiley, New York
4. Dasgupta, A. (2010) *Fundamentals of Probability: A First Course*, Springer, New York.
5. Hogg, R. V., McKean, J. W. and Craig, T. T. (2005*). Introduction to Mathematical Statistics*, Sixth Edition, Pearson Prentice Hall, New Jersey.
6. Rao, C. R. (2002). *Linear Statistical Inference and Its Applications*, Wiley
7. Rohatgi, V. K. & A. K. M. E Saleh (2001). *Introduction to Probability and Statistics*, Wiley, New York.

STM14: Data Analytics Using R (Practical) – 2 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. use R for various statistical computations	Apply
2. understand the theory of random number generation using various methods and apply them to generate random numbers	Understand and Apply
3. apply different search algorithms	Apply
4. use real data sets and perform analysis using R	Apply
5. write programs using R for analyzing data	Apply/Evaluate

Unit I

Basics of R Programming:

The following practical will be conducted using R:

- Calculation of rank and determinant of higher order matrix by partitioning method.
- Calculation of equivalent canonical form by using elementary row and column operations.
- Calculation of inverses of symmetric matrices of higher order by partitioning method.
- Calculation of Inverse, Moore-Penrose inverse, and g-inverse of small and large order matrices.
- Calculation of eigen values and eigen vectors of small and large order matrices Solution of simultaneous system of equations (small and large).
- Spectral decomposition, LU decomposition and SV decomposition of matrix and computation of powers of a matrix.
- Plotting of various probability distributions using R Creating, downloading, manipulating of data files in R.
- Performing descriptive analysis as well as hypothesis testing for real-life data sets

Unit II

Numerical algorithms such as direct search, grid search, interpolation search, gradient search, Bisection and Newton-Raphson methods, Mullers method, Aitkens extrapolation, Mersenne Twister' algorithm. Simple applications of the above methods by writing R codes and demonstrating.

Books Recommended:

1. Bruce, P. and Bruce, A. (2017*). Practical Statistics for Data Scientists*, O'Reilly Media.
2. Kennedy W. J. and Gentle J. E. (1980). *Statistical Computing*, Marcel Dekker
3. Law, A.M. and Kelton, W.D. (2000). *Simulation, Modeling and Analysis*, Tata McGraw Hill, Third Edition
4. Norman Matloff (2011) *The Art of R Programming-A Tour of Statistical Software Design*, No Starch Press, San Francisco
5. Rizzo, M. L. (2007*). Statistical Computing with* R, CRC Press.

6. Tilman M Davis (2016). *The book of R: A First Course in Programming and Statistics*
7. Hadley, W. and Garret, G. (2017) *R for Data Science: Import, Tidy, Transform, Visualize, And Model Data*

STRM1: Research Methodology – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the meaning and scope of doing scientific research.	Understand
2. able to think logically	Apply
3. would be able to use some of the computational algorithms and tools used in modern statistical inference problems.	Apply
4. Would be able to apply several visualization graphical methods	Apply

Unit I: Research Methodology

Objectives and purpose of research, Philosophical foundation for knowledge creation and dissemination, Epistemological, Ontological and other issues in science research, qualitative and quantitative research.

Role of statistics in scientific research, research design, statistical research project.

Types of statistical research: empirical, field experiments, laboratory experiments, and secondary sources of data, exploratory and confirmatory research, planned and ad-hoc methods of data collection, non-response and methods of recovering the missing response.

Unit II: Critical Thinking

Set theory and logic, Theory of Numbers, Constants and Variables, Concept of a sentence in logic, designatory function and a sentential function, Sentential Calculus, logical conjunctions like 'not', 'or', 'and' & 'if…, then…', concepts of argument, premise and conclusion, laws of sentential calculus, Theory of Relations, binary relations, domain and co-domain, algebra of relations: operations on relations, universal relation and the null relations, reflexive relations, transitive relations, symmetric relations etc.

Arguments and conclusions, inductive and deductive logic, Counter examples for the invalidity of arguments

Creativity, Critical Thinking & Problem-Solving

Unit III: Computational Statistics

(a) Introduction to R: Language, variables, data frames, functions, loops, ploting
(b) Theory of random number generation - linear, multiplicative and mixed random number generators. Testing a random number generator- run test, Kolmogrov-Smirnov test, sign test, rank test, gap test, digit frequency test and serial correlation. Selection of a random number generator
(c) Theory of inverse transformation method (ITM) for random variable generation-definition of quantile function, its properties. Quantile function as a random variable and its distribution function. ITM based algorithms to generate random variables from

standard discrete and continuous distributions.

(d) Theory of Acceptance-Rejection method (ARM) for random variate generation - the conditional distribution of Y given that [U<= f(Y)/ M g(Y)] when Y ~ g and U ~ U(0; 1); where f and g are density functions. Interpretation and optimal choice of M using exponential tilting, ARM based algorithms for random variable generation.

(e) Generation of random variables using the relationships between distributions, composition and convolution methods. Algorithms for random variable generation from mixture distributions, chi-square, t and F-distributions.

(f) Random variable generation from bivariate, multivariate and conditional distributions.

Methods to compute integrals- quadrature formula, double integration, Gaussian integration, Monte Carlo methods: Monte Carlo integration and its application to compute expected values and probabilities, Theory of Importance Sampling with applications to reduce Monte Carlo error and rare-event simulation, verification of WLLN, CLT and other approximations through simulation. Empirical computation of level of significance and power of tests

Unit IV: Data Representation & Visualization

Methods of Data Visualization,why we visualize data, Visualization as a cognitive aid, Six Meta-Rules for Data Visualization.

Basics of ggplot, Power BI and their applications for visualization.

Books Recommended

1. Bruce, P. and Bruce, A. (2017*). Practical Statistics for Data Scientists*, O'Reilly Media.
2. *Few, S. (2009). Now You See It: Sample Visualization Techniques for Quantitative Analysis,* Oakland Press: CA: Analytics Press
3. Hadley, W. and Garret, G. (2017) *R for Data Science: Import, Tidy, Transform, Visualize, And Model Data*
4. Kennedy W. J. and Gentle J. E. (1980). *Statistical Computing*, Marcel Dekker
5. Law, A.M. and Kelton, W.D. (2000). *Simulation, Modeling and Analysis*, Tata McGraw Hill, Third Edition
6. Norman Matloff (2011) *The Art of R Programming-A Tour of Statistical Software Design*, No Starch Press, San Francisco
7. Rizzo, M. L. (2007*). Statistical Computing with* R, CRC Press.
8. Steel, J. and Iliinsky, (2010). *Beautiful Visualization*, O'Reilly Media
9. Tilman M Davis (2016). *The book of R: A First Course in Programming and Statistics*
10. Tufte, E. (2001). *The visual display of quantitative information*, 2nd Edition, Graphics Press
11. Richard Hammack, https://www.people.vcu.edu/~rhammack/BookOfProof/
12. A Gentle Introduction to the Art of Mathematics, Joseph E. Fields, https://osj1961.github.io/giam/
13. Ted Sundstrom, Mathematical Reasoning: Writing and Proof, https://www.tedsundstrom.com/mathematical-reasoning-writing-and-proof

STM21: Modern Statistical Inference – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the principles of data reduction	Understand
2. understand different family of distributions	Understand
3. demonstrate the conceptual understanding of minimum variance unbiased estimation	Apply
4. evaluate estimates with optimal properties from a given sample with appropriate distributional assumptions	Evaluate
5. obtain tests and confidence intervals with some with optimal property	Evaluate
6. understand the properties of MLE	Understand

Unit I

Data reduction, sufficiency, sufficient partition, Neyman factorization theorem, minimal sufficiency, completeness, ancillary statistics and Basu's theorem

One-parameter exponential family, multi-parameter exponential family and Pitman family of distributions, canonical form, convexity property, minimal sufficiency and completeness

Unit II

Unbiased estimator, estimability of parametric functions, Cramer-Rao inequality, uniformly minimum variance unbiased estimators, Rao-Blackwell and Lehmann-Scheffe theorems

Estimation Methods: Method of moments, Maximum likelihood estimation, restricted parameter space, inconsistent MLEs, MLEs in irregular cases, Introduction to nonparametric estimation and Bayesian estimation

Unit III

Testing of Hypothesis, Test function, MP tests, Neyman- Pearson lemma, UMP tests, nonexistence of UMP tests MLR property, Introduction to multiple testing problems

Unit IV

Confidence sets and intervals, shortest expected length confidence intervals, relation with testing of hypotheses, UMA confidence intervals

Books Recommended:

1. Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series, Second Edition.
2. Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*. Cambridge University Press
3. Kale, B.K. & Muralidharan, K. (2015) *Parametric Inference: An Introduction*, Alpha Science International Ltd.
4. Lehmann, E. L. and Romano, J. (2005). *Testing Statistical Hypotheses*, Springer
5. Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, New York
6. Rao, C. R. (1995). *Linear Statistical Inference and its Applications*, Wiley

7. Rohatgi, V. K. and Saleh, A.K. Md. E. (2001). *Introduction to Probability andStatistics*, John Wiley & Sons, New York.
8. Shao, J. (2003). *Mathematical Statistics*, Springer-Verlag, New York,

STM22: Regression Analysis & Applications – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. solve problems involving simple and multiple linear regression	Evaluate
2. carry out regression analysis given the data	Analyze
3. carry out binary and multiple logistic regression	Analyze
4. analyze nonnormal data using GLM	Analyze
5. understand the concepts of semiparametric and nonparametric regression models including GAM	Understand

Unit I

Simple linear regression, assumptions, inference, diagnostic checks and testing, polynomial regression, transformations, method of weighted least squares, inverse regression.

Multiple linear regression: Gauss Markov setup, inference, restricted parameter estimation, Residual Analysis, Assumptions verification.

Unit II

Variable selection problems, different methods of variable selection such as forward, backward, best subset etc., multicollinearity and ridge regression, penalized methods, least absolute selection and shrinkage operator (LASSO).

Unit III

Logistic regression: Logit, ML estimation, tests of hypotheses, Wald test, LR test, score test, test for overall regression, multiple logistic regression, forward, backward method, interpretation of parameters, relation with categorical data analysis, odds ratio, inference on odds ratio, model over fitting coupled with bias and variance trade-off, logistic regression as a classifier, evaluation of models metric - sensitivity, specificity, Precision, recall, MAPE etc.

Unit IV

Generalized linear model (GLM): Link functions, Poisson, binomial, inverse binomial, inverse Gaussian, gamma etc., theory and applications to various data sets.

Books Recommended:

1. Cameron, A. C. and P. K. Trivedi (1998). *Regression Analysis of Count Data*, Cambridge
2. Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis*, John Wiley, Third Edition.
3. Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*, Wiley.
4. Kleinbaum, D. G. & Klein, M. (2002). *Logistic Regression: A Self-Learning Text*,

Springer
5. McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman& Hall.
6. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). *Introduction toLinear Regression Analysis*, Wiley.
7. Neter, J., W., and Kutner, M. H. (1985). *Applied Linear Statistical Models*, Wiley.
8. Weisberg, S. (2005). *Applied Liner Regression*, Wiley.
9. Yan, X. and Su, X. G. (2009). *Linear Regression Analysis: Theory & Computing*, World Scientific.

STM23: Multivariate Analysis & Applications – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. carry out an extensive exploratory multivariate analysis for a given multivariate data	Analyze
2. carry out cluster analysis of given multivariate data	Analyze
3. solve problems involving multivariate normal distribution	Evaluate
4. carry out statistical inference procedures using the data from a multivariate normal distribution.	Analyze
5. carry out classification of given multivariate data	Analyze

Unit I

Exploratory multivariate data analysis, sample mean vector, sample dispersion matrix, correlation matrix, graphical representation, means, variances, covariances, correlations of linear transforms, introduction to principal component analysis, correspondence analysis, factor analysis, canonical correlation coefficients and canonical variables.

Unit II

Cluster analysis and multidimensional scaling.

Multivariate normal distribution, pdf and mgf, singular and nonsingular normal distributions, distribution of a linear form and a quadratic form of normal variables, marginal and conditional distributions.

Unit III

MLE's of the parameters of multivariate normal distribution and their sampling distributions, properties of the Wishart Distribution, tests of hypothesis about the mean vector of a multivariate normal distribution, Hotelling's T^2-statistic and its distribution, applications of Hotelling's T^2-statistic. Goodness of fit of multivariate normal distribution. simultaneous confidence interval for the linear functions of the mean, Tests of significance for multiple and partial correlation coefficients.

Unit IV

Classification problem, discriminant analysis, Mahalanobis D^2-statistic, methods and applications of MANOVA (without derivation of the distribution of Wilks' lambda).

Likelihood ratio tests, introduction to non-Gaussian multivariate distributions such as multivariate beta, t, F distributions, Introduction to copula and its applications.

Directional and circular data and introduction to their analysis (exploratory analysis).

Books Recommended:

1. Anderson, T. W. (1984). *Introduction to Multivariate Analysis*, John Wiley
2. Fang ,K., Kotz, S., Ng K. W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman and Hall
3. Härdle, W. K. & Simar, L. (2012). *Applied Multivariate Statistical Analysis*, Springer, New York
4. Härdle, W. K., Hlávka, Z. (2007). *Multivariate Statistics: Exercises and Solutions*, Springer, New York
5. Johnson R.A. & Wichern, D.W. (2007). *Applied Multivariate Statistical Analysis*, 6th Ed., Pearson Education
6. Kotz, S., Balakrishnan N. and Johnson N. L. (2000). *Continuous Multivariate Distributions, Volume 1, Models and Applications*, John Wiley & Sons,
7. Kshirsagar, A. M. (1983). *Multivariate Analysis*, Marcel Dekker
8. Manly, B. F. J., (2004), *Multivariate Statistical Methods - A primer*, Chapman and Hall / CRC Florida.
9. Mardia,K. V. and Jupp, P. E. (2000), *Directional Statistics*, John Wiley & Sons
10. Morrison, D.F. (1990). *Multivariate Statistical Methods*, McGraw Hill Co.
11. Rao, C. R. (1995). *Linear Statistical Inference and its Applications*, Wiley Eastern
12. Timm, N. H. (2002), *Applied Multivariate Analysis*, Springer, New York

STM24: Data Analytics using R and/or Python (Practical) – 2 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to do data analysis using R and Python:	
1. carry out regression analysis given the data using R and Python	Analyze
2. carry out binary and multiple logistic regression using R & Python	Analyze
3. analyze non-normal data using GLM (Poisson, NB etc.)	Analyze
4. analyze multivariate data which uses PCA, FA, MDS etc.	Analyze
5. carry out clustering/classification given multivariate data	Analyze
6. carry out statistical inference related to multivariate normal data (estimation, testing, confidence interval)	Analyze

Unit I:

1. Simple & Multiple Linear Regression.
2. Variable Selection Problem.
3. Multicollinearity and Ridge Regression
4. Regularized Methods (LASSO)
5. Logistic regression (binary and multiple)
6. Poisson/Negative binomial regression
7. GLM

Unit II

8. Graphical representation of multivariate data
9. Principal Component Analysis, Correspondence analysis
10. Factor Analysis
11. Cluster Analysis
12. Canonical Correlations
13. BIAS and MSE of estimators, Power and Size of tests and Coverage of confidenceinterval comparison in univariate inference problems
14. Model Sampling from multivariate normal distribution
15. Applications of Hotelling's T^2
16. MANOVA
17. Discriminant Analysis

Books Recommended:

1. All Recommended books in STM21, STM22 and STM23 courses

STOJT-FP: On Job Training / Field Practice – 4 Credits

6 weeks of internship in the industry with a minimum of 25 days (7 hours per day) working (log sheet required) along with an activity report / Conduct a field survey with the analysis and report with an equal amount of work/ Any other similar activity that requires an equivalent amount of work which has to be done in other Research Institutes under the guidance of researchers or Scientists.

Students can approach the industry/Research Institute directly to get the OJT-FP.

There should be a supervisor from the organization/Institute from which the OJT-FP is being done apart from the internal supervisor. The completion certificate as well as the activity report should be signed by both supervisors.

The external supervisor will grade for 80% (40 CIA and 40 ETE) and the internal supervisor for 20% (10 CIA and 10 ETE) of the OJT-FP.

STM31: Probability Theory – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course, the students will be able to	
1. understand the basics of measure-theoretic approach to probability	Understand
2. solve problems related to probability measure and distribution function	Evaluate
3. solve problems involving expectations of random variables	Evaluate
4. examine the convergence of a sequence of random variables	Evaluate
5. understand the law of large numbers and the central limit theorem related to the sequence of random variables	Understand

Unit I

Review of real and vector-valued random variables, distribution functions (d.f.), discrete and continuous random variables, vector random variables and their distribution functions, Jordan decomposition of a d.f.

Expectation, linear properties of expectations, Inequalities involving expectations and probability.

Unit II

Independence of two events and n (> 2) events, sequence of independent events, independent classes of events, π-system and λ-system of events.

Dynkin's theorem (without proof) independence of random variables

Unit III

Convergence of a sequence of random variables, Various types of convergence and their interrelationships, Cramer's theorem (Slutsky's theorem), Fatou's lemma, monotone convergence theorem and dominated convergence theorem

Borel zero-one law, Borel-Cantelli Lemma, Kolmogorov zero-one law.

Laws of large numbers, weak (with proof) and strong (without proof) laws of large numbers,

Unit IV

Characteristic functions, Continuity theorem for characteristic functions, Inversion theorem (without proof).

Central Limit Theorem, Liapounov's and Lindeberg's central limit theorems (without proof), Implications and applications

Books Recommended:

1. Athreya, K. B. and Lahiri S. (2006). *Probability Theory*, Hindustan Book Agency.
2. Bhat, B. R. (2007). *Modern Probability Theory: An Introductory Text Book*, New Age International.
3. Billingsley, P. (1995). *Probability and Measure*, 3rd Ed., John Wiley, New York
4. Chung, K. L. (2001). *A Course in Probability Theory*, Third Ed., Academic Press, London.
5. Gut, Allan (2005), *Probability: A Graduate Course*. Springer, New York

STM32: Stochastic Processes – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course, the students will be able to	
1. understand the concepts related to the Markov chain and solve problems related to the Markov chain model	Understand Evaluate
2. understand the concepts related to Branching processes and solve problems related to branching process models	Understand Evaluate
3. understand the concepts related to birth-death processes solve problems related to these models	Understand Evaluate
4. understand the concepts related to Poisson processes, Renewal processes etc. and solve problems related to these models	Understand Evaluate
5. understand the concepts related to Gaussian and related processes and solve problems related to these models	Understand Evaluate
6. Generate all the processes/models mentioned in the syllabus and visualize the process and sample path	Visualize

Unit I

The notion of stochastic processes, Markov chain, one-step transition probabilities, Chapman-Kolmogorov equations, evaluation of higher-step transition probabilities, classification of states, periodicity of a Markov chain, concept of closed class, minimal closed class, stationary distribution. Some examples such as gamblers's ruin problems and one-dimensional random walks. Concept of absorption probabilities, use of these to compute the probability of winning the game by a gambler having initial capital 'c'

Unit II

Branching process, classification of states, identification of criticality parameter, extinction probability, the relationship between criticality parameter and extinction probability of the process, Expression for mean and variance of the process. Extinction probability, some epidemiological applications,

Introduction to Markov chain in continuous time, concept of intensity rate, relationship between intensity matrix and matrix of transition probability function. Kolmogorov's forward and backward equations

Unit III

Introduction to the birth process, birth and death process, linear birth and death process, Growth model with immigration and related results, Expression for mean and variance of a birth process and, birth and death process, Applications of these processes.

Unit IV

Poisson process, two definitions and their equivalence, distribution of inter-arrival times, conditional joint distribution of inter-arrival times, compound Poisson process, some applications. Introduction to t h e renewal process, relationship with Poisson process, key and elementary renewal theorems (without proof) associated with renewal processes, some applications.

Brownian motion, hitting times, maximum variable and the Gambler's ruin problem

Gaussian Processes, Ornstein-Uhlenbeck process, Brownian bridge, geometric Brownian motion,

Books Recommended:

1. Bhat B.R. (2000). *Stochastic Models: Analysis and Applications*, New Age International.
2. Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. 1, Wiley Eastern.
3. Hoel, P.G. Port, S.C. & Stone, C.J. (1972). *Introduction to Stochastic Processes*, Houghton Mifflin
4. Karlin, S & Taylor, H.M. (1975). *A First Course in Stochastic Processes* (Second. Edition), Academic Press.
5. Medhi, J. (2010) *Stochastic Processes*, New Age Science Ltd.
6. Pinsky M. A. and Karlin, S. (2010). *An Introduction to Stochastic Modeling*, 4thEdn. Academic Press.
7. Ross, S. (2014). *Introduction to Probability Models*, 11th Edn. Academic Press.
8. Serfozo, R. (2009). *Basics of Applied Stochastic Processes*, Springer.

STM33: Design and Analysis of Experiments – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to different designs including BIBD and solve problems related to them	Understand Evaluate
2. understand the concepts related to different factorial designs solve problems related to them	Understand Evaluate
3. understand the concepts related various advanced designs solve problems related them	Understand and Evaluate
4. understand the concepts related to response surface methodology solve problems related to them	Understand and Evaluate
5. understand the concepts related to Taguchi methods solve problems related to them	Understand and Evaluate
6. analyze the data using all the designs discussed in the course	Apply &Analyze

Unit I

Review of randomization, replication and local control, Analysis of one way classification model. Analysis of two-way classification model with equal number of observations per cell with and without interactions. Analysis of two-way classification model with unequal number of observations per cell without interactions, notion of connectedness, balance and orthogonality, analysis of BIBD and PBIBD

Analysis of covariance in one way and two-way classification models, testing of hypotheses for estimable parametric functions.

Unit II

Analysis of 2^k full factorial experiments: diagrammatic presentation of main effects and first and second order interactions, model, analysis of single as well as more than one replicates, using ANOVA. Total confounding of 2^k design in 2^p blocks, $p \geq 2$. Partial confounding in 2^p blocks, $p = 2, 3.$, factional factorial experiments, Resolution of a design, (III, IV & V), aberration of a design.

Unit III

Response surface methodology (RSM): linear and quadratic model, stationary point, central composite designs (CCD), ridge systems, multiple responses, concept of rotatable designs, optimality of designs, simplex lattice designs, simplex centroid designs.

Unit IV

Taguchi methods: concept of noise factors, concept of loss function, S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays. Random effect models and mixed models.

Books Recommended:

1. Bapat, R.B. (2011). *Linear Algebra and Linear Models*. Springer and Hindustan Book Agency.
2. Dean, A. and Voss, D. (1999). *Design and Analysis of Experiments*, Springer.
3. George E. P. Box, Draper N.R. (1987). *Empirical Model-Building and Response Surfaces*, Wiley.
4. Hicks, C.R., Kenneth V. and Turner, Jr. (1999). *Fundamental Concepts in the Design of Experiments*, Oxford University Press.
5. Kshirsagar A.M. (1983). *Linear Models*, Marcel Dekker
6. Montgomery, D.C. (2001). *Design and Analysis of Experiments*, Wiley.
7. Ogawa, J. (1974). *Statistical Theory of the Analysis of Experimental Design*, Marcel Dekker.
8. Phadke, M.S. (1989). *Quality Engineering using Robust Design*, Prentice Hall, Englewood Cliffs, New Jersey
9. Wu, C.F. Jeff and Hamada M. (2000). *Experiments: Planning, Analysis and Parameter Design Optimization*, John Wiley and Sons

STM34: Advanced Data Analytics using R and/or Python (Practical)

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. simulate various stochastic models discussed in ST 301	Visualize
2. carry out data analysis related to all the designs in ST 302	Analyze

Unit I

1. Simulation of Markov chain and computing the stationary distribution of an ergodic Markov chain.
2. Simulation of branching process and estimating its mean and variance.
3. Simulation of Poisson and related processes.
4. Generating birth-death process and its limiting distribution.

Unit II

5. One way classification. Multiple comparison tests.
6. Two way classification with equal number of observations per cell (Model with interaction). Two way classification with unequal number of observations per cell (Model without interaction).
7. Analysis of LSD and BIBD.
8. Analysis of covariance in one way and two way model.
9. 2^k factorial experiments, Analysis of single replicate of 2^k.
10. Total and partial confounding in 2^k factorial experiments.
11. Random effect and mixed models.
12. Analysis of first and second order response surface model.
13. Central composite design. Contour and surface plots, Box-Behnken design, Small.
14. Taguchi methods: S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays.

STRP3: Research Project I – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. read research papers	Understand
2. Formulate a statistical data analysis project involving, collection, coding, analysis (using elementary as well as advance statistical methods), and interpretation of results	Apply Analyze
3. Prepare presentation and report of a project using LaTeX	Apply

STRP3: Research Project I Guidelines

1. STRP3 Research Project I is an individual or group activity with a maximum of **THREE** students in a group.
2. As a part of this course, students should learn LaTeX document preparation and Beamer Presentation. (This can be done as a part of skill-based course as well).
3. Use real data sets for project problems, as far as possible.
4. There will one presentation and one viva-voce (each graded out of 25 points by the Supervisor) for the continuous internal assessment (CIA). In the presentation, students

   Are expected describe their project problem, the data they are going to analyze and the objectives of their project. In addition to this, they should also mention their methodology. Students are expected to read at least THREE research papers which addresses similar kind of problems and they should include main contents of the papers in their first presentation as well as in final report. In the second presentation, students should discuss the results of their analysis, findings and new methodology they have introduced (if any). Students should make sure that they have something innovative in their project work.
5. The completed project report should be submitted to the Project coordinator on or before the last day of the semester.
6. All project groups are expected to make the final presentation as per schedule. Project draft report as well as the final presentation will be evaluated by an external examiner (out of 50 = Report – 20, Presentation – 15, Viva - 15). When external examiner is not available, the Head may appoint an external examiner from the Department.

STM41: Time Series Analysis – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. carry out an exploratory analysis of time series	Analyze
2. understand the concepts of stationarity of a time series and solve related problems	Understand and Evaluate
3. test the stationarity of a time series	Analyze
4. understand the theory related to linear time series models and fit an appropriate linear time series model for the data	Understand and Analyze
5. understand the theory related to estimation and forecasting using a time series model and apply them for a time series data	Understand and Apply
6. understand the theory related to ARCH/GARCH models and analyze data using ARCH/GARCH models	Understand and Analyze
7. use information criteria for the selection of models	Analyze
8. understand the theory of INAR models and analyze count data using Poisson INAR models	Understand and Analyze

Unit I

Exploratory time series analysis, tests for trend and seasonality. Exponential and Moving average smoothing. Holt -Winters smoothing. Forecasting based on smoothing, adaptive smoothing.

Time - series as a discrete parameter stochastic process. Auto covariance and autocorrelation functions and their properties, Portmanteau tests for noise sequences, transformation to obtain Gaussian series.

Unit II

Stationary processes: General linear processes, moving average (MA), auto regressive (AR), and autoregressive moving average (ARMA), Stationarity andinvertibility conditions. Nonstationary and seasonal time series models: Auto regressive integrated moving average (ARIMA) models, Seasonal ARIMA (SARIMA) models, Transfer function models (Time series regression).

Unit III

Forecasting in time series models, Durbin-Levinson algorithm, innovation algorithm (without proof), Estimation of mean, auto covariance and autocorrelation functions, Yule- Walker estimation, Estimation of ARIMA model parameters, maximum likelihood method, large sample theory (without proofs). Choice of AR and MA periods, FPE, AIC, BIC, residual analysis and diagnostic checking, Unit-root non stationarity, unit-root tests.

Unit IV

Multivariate Time series model, VAR models, Vector ARMA models.

Conditional heteroschedastic models, ARCH and GARCH, properties, examples, estimation & forecasting, extensions of ARCH & GARCH to asymmetric models.

Count time series models, INAR models, Poisson INAR models, Coherent forecasting

Books Recommended:

1. Brockwell, P.J. and Davis, R. A. (2003). *Introduction to Time Series Analysis*, Springer
2. Chatfield, C. and Xing, H. (2019).*The Analysis of Time Series: An Introduction with R*. Chapman & Hall.
3. Cryer, J. D. and Chan, K-S. (2008). *Time Series Analysis with Applications in R*., Springer
4. Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd Ed. Wiley.
5. Hamilton N. Y. (1994). *Time Series Analysis*, Princeton University press.
6. Kendall, M. and Ord, J. K. (1990). *Time Series*, 3rd Ed. Edward Arnold.
7. Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer
8. Shumway, R. H.and Stoffer, D. S. (2010). *Time Series Analysis & Its Applications*, Springer.
9. Tsay, R. S. (2010). *Analysis of Financial Time Series*, Wiley.
10. Tsay, R. S. (2012). *An Introduction to Financial Time Series Data with R*, Wiley.

STM42: Bayesian Inference – 4 Credits

Course Outcome (CO)	**Cognitive level**

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

1. understand the concepts such as HPD, credible intervals, Bayesian prediction and solve related problems	Understand Evaluate
2. understand the concepts related to loss functions, posterior loss solve related problems	Understand and Evaluate
3. understand the concepts of choosing an appropriate prior and solve prior-posterior related problems	Understand and Evaluate
4. understand the concepts related to Bayesian model selection solve related problem	Understand and Evaluate
5. understand the asymptotics related to posterior distribution	Understand
6. compute estimators using Bayesian computing techniques such MH, GIBBS, EM and MCMC algorithms and carry out the convergence diagnostics procedures using R/WinBUGS	Understand and Apply/Evaluate

Unit I

Basics of minimaxity, subjective and frequentist probability, Bayesian inference, prior distributions, posterior distribution, loss function, principle of minimum expected posterior loss, quadratic and other common loss functions, advantages of being a Bayesian. Improper priors, common problems of Bayesian inference, point estimators, Bayesian HPDconfidence intervals, testing, credible intervals, prediction of a future observation

Unit II

Bayesian analysis with subjective prior, robustness and sensitivity, classes of priors, conjugate class, neighbourhood class, density ratio class, different methods of construction ofobjective priors: Jeffrey's prior, probability matching prior, conjugate priors and mixtures, posterior robustness: measures and techniques.

Unit III

Model selection and hypothesis testing based on objective probabilities and Bayes factors large sample methods:
Limit of posterior distribution, consistency of posterior distribution, asymptotic normality of posterior distribution.

Unit IV

Bayesian computations: Analytic approximation, E-M Algorithm, Monte Carlo sampling, Markov Chain Monte Carlo Methods, Metropolis-Hastings Algorithm, Gibbs sampling, examples, convergence and diagnostic issues

Books Recommended

1. Albert, J. (2009). *Bayesian Computation with R*, Springer
2. Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer
3. Bolstad, W. M. (2007). *Introduction to Bayesian Statistics*, 2nd Edn. Wiley,
4. Christensen R, Johnson, W., Branscum, A. and Hanson T. E. (2011). *Bayesian Ideas*

and Data Analysis: An Introduction for Scientists and Statisticians, Chapman & Hall.
5. Congdon, P. (2006). *Bayesian Statistical Modeling*, Wiley
6. Gelman, A., Crlin, J. B., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian Data Analysis*, CRC Press.
7. Ghosh, J. K., Delampady M. and T.Samantha (2006). *An Introduction to Bayesian Analysis: Theory & Methods*, Springer.
8. Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*, Springer
9. Jim, A. (2009). *Bayesian Computation with R*, 2nd Edn, Springer.
10. Lee, P. M. (2012). *Bayesian Statistics: An Introduction*, 4thEdn., Wiley
11. Marin, J-M. and Robert, C. P. (2014). *Bayesian Essentials with R*. Springer
12. Ntzoufras, I. (2008). Bayesian Modeling Using WinBUGS, Wiley.
13. Rao. C.R. and Day. D. (2006). *Bayesian Thinking, Modeling & Computation, Handbook of Statistics*, Vol. 25. Elsevier
14. Turkman, M. A. A., Paulino, C. D. and Muller, P. (2019). *Computational Bayesian Statistics: An Introduction*, CUP

STM43: Sampling Methods and Applications – 4 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related various standard sampling designs and solve problems related to them	Understand Evaluate
2. understand the concepts related to cluster, double and multi-stage sampling and solve problems related to them	Understand and Evaluate
3. understand the concepts related to various methods of imputing the missing data and solve related problems	Understand and Evaluate
4. understand the concept of super population model and solve related problems	Understand and Evaluate
5. understand the concepts of network and adaptive sampling solve related problems	Understand and Evaluate
6. design an appropriate survey and provide the related analysis	Apply and Analyze

Unit I

Review of basic\ methods of simple random sampling and stratified random sampling, Use of supplementary information for estimation, ratio and regression estimators with their properties and generalizations,

Systematic sampling, PPS sampling, Estimation problems, Hansen-Horwitz estimator and its properties, Horwitz-Thompson estimator and its properties, Midzuno-Sen method

Unit II

Cluster sampling, multistage sampling, Double sampling procedures and their ratio and regression estimators, stratification estimator, Multiphase sampling.

Non-sampling errors, response and non-response errors, Hansen and Hurwitz's model and their treatments, methods of imputation, randomized response, Warner's model, Franklin's model, Jackknife technique.

Unit III

Inference under super population model, model-assisted and model-based inference, Robustness of designs and predictors, Bayesian inference, Spatial Smoothing, Sampling on Successive Occasions: Panel Rotation, Non-Response and Not-at-Homes, Weighting Adjustments and Imputation, Time Series Approach in Repeated Sampling, Comparison of strategies under super population models

Unit IV

Network and Adaptive Procedures, Estimation by Network and by Adaptive Sampling, Constraining Network Sampling and Constraining Adaptive Sampling

Design an appropriate survey with a group of not more than FOUR students, collect the relevant data using the sampling design adopted, carry out the analysis of the data collected. (The entire exercise should be carried out under the supervision of the concerned teacher. This exercise can be considered as a CIA component).

Note: Students should undertake a sample survey and the related analysis during the tenure of this course as a part of the practical application.

Books Recommended

1. Arnab, R. (2017). *Survey Sampling: Theory & Applications,* Academic Press
2. Chaudhuri, A. (2014). *Modern Survey Sampling,* CRC Press
3. Cochran, W.G. (1984). *Sampling Techniques,* Wiley.
4. Des Raj and Chandhok, P. (1998). *Sample Survey Theory*, Narosa.
5. Gal, I. and Ograjensˇek, I. (2017). *Official Statistics and Statistics Education: Bridging the Gap,* Journal of Official Statistics, Vol. 33, No. 1, pp. 79–100
6. Latpate, R., Kshirsagar, J.,Gupta V. and Chandra, G. (2021). *Advanced Sampling Methods,* Springer.
7. Okafor, C (2002). *Sample survey Theory with Applications,* Snaap Press Ltd.
8. Singh, D. and Chaudhary F.S (1986).*Theory and Analysis of Sample Survey Designs,* Wiley Eastern Limited.
9. Singh, S. (2003). *Advance Sampling Theory and Applications* (Volume I and II), Kluwer Academic Publishers.
10. Sukhatme, P.V, Suktatme, B.V.,Sukhatme, S. and Asok, C. (1984).*Sampling Theory of Surveys with Applications,* Indian Society for Agricultural Statistics, NewDelhi.
11. Thmpson, S. K. (2012). *Samplig,* 3rdEdn., Wiley

STRP4: Research Project II – 6 Credits

Course Outcome (CO)	**Cognitive level**
After completion of this course, the students will be able to	
1. Formulate a statistical research problem and solve it	Apply/Develop
2. Write One/Two Research papers and publish them in a Scopus Indexed Journal	Research/Publish
3. Prepare presentation and project report	Apply
4. Prepare Research Papers	Research

Project Guidelines:

1. STRP4 Research Project II is an individual or group activity with a maximum of **THREE** students in a group.
2. Formulate a problem and develop new methodology/algorithm for solving the proposed problem.
3. Prepare a report which is equivalent to the old M.Phil degree
4. **Preparation of ONE/TWO research papers is mandatory for this project. Thus, the original contribution in the form of an algorithm or methodology is mandatory.**
5. There will be one presentation and one viva-voce (each graded out of 25 points by the Supervisor) for the continuous internal assessment (CIA). In the presentation, students are expected to describe their project problem, the data they are going to analyze and the objectives of their project. In addition to this, they should also mention their methodology. Students are expected to read at least THREE research papers which address similar kinds of problems and they should include the main contents of the papers in their first presentation as well as in the final report. In the second presentation, students should discuss the results of their analysis, findings and new methodology they have introduced (if any). Students should make sure that they have something innovative in their project work.
6. The completed project report should be submitted to the Project Coordinator on or before the last day of the semester.
7. All project groups are expected to make the final presentation as per schedule. The project draft report as well as the final presentation will be evaluated by an external examiner (out of 50 = Report – 20, Presentation – 15, Viva - 15). When an external examiner is not available, the Head may appoint an external examiner from the Department.

ELECTIVE COURSES

All elective courses are of 4 credits

Electives – Semester I

STE101: Optimization Techniques

Course Outcome (CO)	Cognitive level
After completion of this course, the students will be able to	
1. formulate and solve Linear, Integer and dynamic programming problems using advanced methods	Understand Apply
2. understand the concepts related to assignment and transportation and solve related problems	Understand Apply
3. understand the theory related to nonlinear programming problems and solve them	Understand and Apply
4. understand the concepts related to network models and solve problems of network analysis	Understand and Apply

Unit I

Linear programming: Review

Computational complexity of LPP, ellipsoid method, polynomial time algorithm, Karmarkar's polynomial time algorithm, convergence and complexity, duality theory and dual simplex method

Integer linear programming problem: pure and mixed integer programming problem, Gomory's all integer programming method. fractional cut method- all integer and mixed integer linear programming problem, branch and bound method, cutting planes

Dynamic programming, sensitivity, Bellman's optimality principle, stochastic dynamic programming

Unit II

Transportation and assignment problems, recent algorithms for solving these problems

Unit III

Nonlinear programming: Karush-Kuhn-Tucker conditions, convexity, quadratic programming, Wolfes, Beales and Fletchers algorithms for solving quadratic programming problems. Convex problems, duality in nonlinear programming, mixed integer models.

Unit IV

Networking models: Network flows, maximal flow in the network, transportation problems, transshipment problems and assignment problems as networking problems. Network scheduling by PERT/CPM techniques, resource analysis in network scheduling.

Books Recommended:

1. Bertsekas, D. (1999). *Nonlinear Programming*, 2nd Edn. Athena Scientific.
2. Chong, E. K. P. and Zak, S. (2004). *An Introduction to Optimization*, Wiley.
3. Fletcher, R. (2000). *Practical Methods of Optimization*, Wiley
4. Hadley, G. (1987). *Linear Programming*. Addison-Wesley.
5. Hiller, F.S. and Lieberman, G.J., (2009). *Introduction to Operations Research* (9th ed.), McGraw-Hill
6. Kambo, N.S. (1991). *Mathematical Programming Techniques*. Affiliated East-West press.
7. Panneerselvam, R. (2012). *Operations Research*, 2nd Edn. Prentice Hall of India.
8. Sinha, S. M. (2006) *Mathematical Programming: Theory and Methods*, Elsevier's
9. Taha, H. A. (2016) *Operations Research: An Introduction*, 10th edition, Prentice Hall
10. Winston, W.L., (2003*)* *Introduction to Mathematical Programming* (4th ed.), Duxbury Press

STE103: Reliability & Statistical Quality Control

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand different types of systems and evaluate the reliability of such systems	understand and Evaluate
2. understand the concepts of ageing of systems and classify them based on ageing properties	Understand and Evaluate
3. evaluate the bounds of reliability of simple systems	Evaluate
4. understand the concepts related to replacement models and evaluate or compare replacement policies	Understand and Evaluate
5. understand the concepts related to CUSUM and EWMA charts and evaluate measures associated with these charts	Understand and Evaluate
6. make economic design of control charts	Evaluate
7. carry out process capability analysis	Evaluate
8. construct control charts for vector-valued quality characteristics	Evaluate
9. design sampling plans	Evaluate

Unit I

Coherent structures, representation of coherent systems in terms of paths and cuts, modules of coherent systems. Reliability of system of independent components, association of random variables, bounds on system reliability, improved bounds on system reliability using modular decompositions.

Shape of the system reliability function, applications to relay circuits and safety monitoring systems, notion of aging and life distributions of coherent systems, distributions with increasing failure rate average arising from shock models, preservation of life distribution classes under reliability operations. Reliability bounds, mean life series and parallel systems.

Unit II

Classes of life distributions applicable in replacement models, shock models, age replacement and block replacement policies, renewal theory useful in replacement models, replacement policy comparisons, preservation of life distribution classes under reliability operations.

Unit III

CUSUM chart for process mean, CUSUM chart for process variability, tabular CUSUM. EWMA chart for process mean. EWMA chart for process variability. Comparison of Shewhart control charts with CUSUM chart and EWMA chart.

Economic designing of control charts, Duncan's model, concepts of conforming run length (CRL), CRL chart, properties of CRL chart, average run length (ARL), average time to signal (ATS), ARL and ATS models to obtain the design parameters.

Unit IV

Process capability analysis

Synthetic and 'Group Runs' (GR) control charts, multi-attribute control charts, multivariate control charts for mean vector and covariance matrix.

Acceptance sampling plans, chain sampling plans, Bayesian sampling plans

Books Recommended:

1. Barlow, R. E. and Proschan, F. (1975*). Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart and Winston Inc.
2. Barlow, R. E. and Proschan, F. (1996). *Mathematical Theory of Reliability.* John Wiley.
3. Guenther, W. C. (1977). *Sampling Inspection in Statistical Quality Control*, Alan Stuart.
4. Levenson, W. (2011). *Statistical Process Control for Real-World Applications*. CRC Press.
5. Montgomery, D. C. (2005*). Introduction to Statistical Quality Control*, Wiley.
6. Tobias, P. A. and Trindane, D. C. (1995). *Applied Reliability*, Second edition. CRC Press.

STE104: Actuarial Statistics

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to force mortality, future life random variables and solve associated problems	Understand and Evaluate
2. understand the concepts of present value of money, related theory and solve associated problems	Understand and Evaluate
3. understand the concepts of annuity, related theory and	understand and

	solve associated problems	Evaluate
4.	understand the theory associated with the premium calculation of insurance products evaluate premium in such cases	Understand and Evaluate

Unit I

Future life time random variable, its distribution function and density function, concept of force of mortality, curtate future life time random variable its probability mass function, deferred probabilities, all these functions in terms of international actuarial notation. Analytical laws of mortality such as Gompertz' law and Makeham's law, single decrement life table, select and ultimate life table.

Unit II

Concept of compound interest rate, discount factor, present value of the money, nominal rate of interest, force of interest, assurance contracts with level and varying benefits, such as whole life insurance, term insurance endowment insurance. Means and variances of the present value random variables of the payments under these contracts under the assumption of constant force of interest, when the benefit payments are made at the end of year of death (discrete set up) or when it is paid at the epoch of death (continuous set up). Actuarial present value of the benefit, net single premiums.

Unit III

Annuity contracts, annuity certain, discrete annuity, monthly annuity, continuous annuity, deferred annuity, present values and accumulated values of these annuities. Continuous life annuity, discrete life annuity, such as whole life annuity, temporary life annuity, n-year certain and life annuity, life annuities with mthly payments, present value random variables for these annuity payments, their means and variances, actuarial present value of the annuity

Unit IV

Loss at issue random variable, various principles to decide net premiums for insurance products and annuity schemes defined in unit II and III, fully continuous premiums and fully discrete premiums, True monthly payment premiums. Extended equivalence principle to decide gross premiums, concept of reserve, prospective & retrospective approach, fully continuous reserve, fully discrete reserve

Books Recommended:

1. Bowers, JR. N.L., Gerber, H.U., Hickman, J.C., Jones, D.A. and Nesbitt, C.J. (1997). *Actuarial Mathematics*, 2nd Edn., The Society of Actuaries.
2. Deshmukh S.R. (2009). *Actuarial Statistics: An Introduction Using R*, Universities Press.
3. Harriett, E.J. and Dani, L. L.(1999). *Principles of Insurance: Life, Health, and Annuities*, 2nd Edn., Life Office Management Association.
4. Neill, Alistair (1977). *Life Contingencies*, The Institute of Actuaries.
5. Palande, P. S., Shah, R. S. and Lunawat, M. L. (2003). *Insurance in India -Changing Policies and Emerging Opportunities, Response Books.*

Electives – Semester II

STE204: Advances in Generalized Linear Models

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the general theory of GLM	Understand
2. apply GLM for to data sets and arrive at meaningful conclusions	Apply/Analyze
3. understand the concepts related to binary and multinomial logistic models and apply them for various data sets	Understand and Apply/Analyze
4. understand the concepts related to count data GLM and apply them for various count data sets	Understand and Apply/Analyze
5. apply GLM for correlated data sets	Apply/Analyze
6. apply Bayesian methods for data analysis using GLM	Apply/Analyze
7. apply GLM for the analysis related to data sets in various domains	Apply/Analyze

Unit I

Generalized linear models: model fitting and inference, exponential dispersion family distributions, lilkelihood and asymptotic distributions, likelihood-ratio/Wald/Score methods of inference, parameters, deviance, model comparison, and model checking, goodness of fit.

Unit II

Binary logistic models, nominal responses: baseline-category logit models, ordinalresponses: cumulative logit and probit models, probit and complementary log–log models, multinomial response models.

Unit III

Models for count data, Poisson GLMs for counts and rates, Poisson/multinomial models for contingency tables, negative Binomial GLMS, models for zero-inflated data.

Quasi-likelihood methods, variance inflation for over dispersed Poisson and Binomial GLMs, Beta-Binomial models and Quasi-likelihood alternatives, Quasi-likelihood and model misspecification.

Unit IV

Modeling correlated responses, marginal models and models with random effects normal linear mixed models, fitting and prediction for normal linear mixed models, Binomial and Poisson GLMMs, GLMM fitting, inference and prediction.

Marginal modeling and generalized estimating equations (GEE).

Bayesian generalized linear models, Empirical Bayes and hierarchical Bayes modeling.

Applications in survival analysis, insurance, engineering, Correlated survey responses etc.

Books Recommended:

1. Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*, Wiley
2. Dobson, A. J. (2002). *An Introduction to Generalized Linear Models*, 2nd Ed. Chapman & Hall
3. Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and theirApplications*, Springer
4. Jong, P. and Heller, G. Z. (2008) *Generalized Linear Models for Insurance Data*, Cambridge University Press.
5. Lindsey, J. K. (1997). *Applying Generalized Linear Models*, Springer
6. McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman & Hall
7. McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear and Mixed Models*, Wiley
8. Stroup, W. W. (2013). Generalized *Linear Mixed Models, Modern Concepts, Methods and Applications*, CRC Press

STE205: Statistical Methods in Epidemiology

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. utilize the basic terminology and definitions of epidemiology	Understand
2. learn key features and applications of descriptive and analytic epidemiology,	Understand and Apply
3. use statistical techniques in the analysis, predictions and presentation of epidemiological data,	Understand and Apply
4. calculate and interpret ratios, proportions, incidence rates, mortality rates, prevalence rate etc., to model the spread and related aspects of a given epidemics so as to gain insight into its management,	Understand, Apply and Analyze
5. to use statistical methods for analyzing the shut down strategy, testing strategy, vaccination strategy etc.	Apply and Analyze

Unit I

Epidemiologic terms and parameters: Infection period, incubation period, latent period, number of asymptomatic carriers, disease frequency, disease frequency association, concept of prevalence, measures of risk, reproduction numbers, preventive reproduction numbers, infection rate, fatality rate, transmission intensity, doubling time, flattening of the curve, prevention strategies.

Unit II

Concepts of disease occurrence, chains of infections, disease occurrence patterns, SIR epidemic models, Reed-Frost chain binomial epidemic models, SIR and SEIR models,random networks for epidemics, models for spatiotemporal spread, incorporating the effects of interventions, predicting the course of the spread.

Mathematical models developed for epidemics such as H1N1, COVID 19 spread, applications of SIR and SEIR models, assessment of lock down effect, introduction to spatial epidemiology such as spatial exploration of epidemiological data, quantification of spatial patterns and clusters, spatial exposure assessment, methods for assessing risk with examples/models from H1N1 and COVID 19.

Unit III

Epidemiological study designs, cohort studies, case-control studies, randomized control studies, intervention, statistical inference for the epidemiological parameters, Bayesian inference for latent (unobserved) variables (MCMC, adaptive MCMC); Inference for outbreaks when virus sequence data is available, Methods based on evolution of the virus by comparing virus sequences of diagnosed cases.

Unit IV

Testing, sensitivity, specificity and ROC curve related to diagnostic testing, sample size determination, pooled testing, composite sampling.

Books Recommended:

1. Diekmann, O., Heesterbeek, H. and Britton, T. (2013) *Mathematical Tools for Understanding Infectious Disease Dynamics*, Princeton University Press
2. Held, L., Hens, N., O'Neill, P.D. and Wallinga, J. (Eds). (2019). *Handbook ofInfectious Disease Data Analysis*. CRC Press.
3. Yang, Z. (2014). *Molecular Evolution: A Statistical Approach*, Oxford University Press.
4. Armitage, P., Berry, G. and Matthews, J. N. S. (2002). *Statistical Methods inMedical Research*, Wiley.
5. Becker, N. G. (2015). *Modeling to Inform Infectious Disease Control*, CRC Press.
6. Elston, R. C. and Johnson W. D. (2008). *Basic Biostatistics for Geneticists and Epidemiologists*: *A Practical Approach*, Wiley
7. HardeoSahai and Khushid, A. (2009). *Statistics in Epidemiology: Methods,Techniques and Applications*, CRC Press
8. Krämer, A. Kretzschmar, M. and Krickeberg, K. (Editors) (2010). *Modern Infectious Disease Epidemiology: Concepts, Methods, Mathematical Models, and Public Health*, Springer
9. Lawson, A. B. (2006). *Statistical Methods for Spatial Epidemiology*, Wiley
10. Lawson, A. B. (2018). *Bayesian disease mapping: Hierarchical Modeling in Spatial Epidemiology*, CRC Press
11. Marschner, I. C. (2014). *Inference Principles for Biostatisticians*, CRC Press
12. Merril, R. M. (2015*). Statistical Methods in Epidemiologic Research*,Jones &Bartlett Publishers
13. Merrill, R. M. (2012). *Fundamentals of Epidemiology and Biostatistics*, Jones & Bartlett Publishers
14. Nigel, B., Daniel, P. and Debbi, S. (2018). *Quantitative Methods for HealthResearch : A Practical Interactive Guide to Epidemiology and Statistics*, Wiley
15. Pagano, M. and Gauvreau, K. (2018). *Principles of Biostatistics*, CRC Press
16. Stewart, A. (2016). *Basic Statistics and Epidemiology: A Practical Guide*, Fourth

Edition, CRC Press
17. Sullivan, L. M. (2018*). Essentials of Biostatistics in Public Health*, 3rd Edition, Jones & Bartlett Learning

STE208: Discrete Data Analysis

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. Able to develop a critical approach to the analysis of contingency tables	Analyze
2. Understand the the basic ideas and methods of generalized linear models	Understand
3. Able to link logit and log-linear methods with generalized linear models	Understand
4. To develop basic facility in the analysis of discrete data	Analyze

Unit I
Review of discrete probability distributions: binomial, multinomial, and Poisson. Likelihood, Tests for one-way tables using Pearson's chi-square and likelihood ratio statistics.

Unit II
Contingency tables 2 × 2 and $r \times c$ tables, tests for independence and homogeneity of proportions, Fishers exact test, odds ratio and logit, other measures of association.
Using 3-way tables with full independence and conditional independence, collapsing and Simpson's paradox.

Unit III
Generalized linear models in Poisson regression and logistic regression contexts for dichotomous response, modelling binary clusteted data, interpretation of coefficients, Generalized estimating equations, main effects and interactions, model selection, diagnostics, and assessing goodness of fit.

Unit IV
Polytomous logit models for ordinal and nominal response. Loglinear models (and graphical models) for multi-way tables. Inference in log-linear models with sparse data.

Causality, repeated measures, generalized least squares, mixed models, latent-class models, missing data, and/or algebraic statistics approaches.

Books Recommended:

1. Agresti, A. (2013). *Categorical Data Analysis*, 3rd Edition, Wiley
2. Anderson F. (2020). *Categorical Data Analysis by Examples: Hands on ApproachUsing R.*
3. Azen R and Walker, C. M. (2021). *Categorical Data Analysis for the Behaviouraland*

Social Sciences, 2nd Edn. Routledge
4. Friendly, M. and Meyer, D. (2016). *Discrete Data Analysis with R: Visualizationand Modeling Techniques for Categorical and Count Data*, CRC Press
5. Hirji, K. F. (2006). *Exact Analysis of Discrete Data*, Routledge
6. Rudas, T. (2018). *Lectures on Categorical Data Analysis*, Springer
7. Santner, T. J and Duffy, D. E. (1989) *The Statistical Analysis of Discrete Data*, Springer
8. Tang, W. He, H. and Tu, X. M. (2012). *Applied Categorical and Count Data Analysis*.
9. Upton, G. J. G. (2017) *Categorical Data Analysis by Examples*, Wiley.

Electives – Semester III

STE301: Statistical Learning & Data Mining

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to supervised and unsupervised learning methods and apply them for different data	Understand Apply/Analyze
2. understand the concepts of feature selection and feature extraction	Understand and Apply/Analyze
3. understand and apply the concepts of Regression Trees, Random Forests, Bagging and boosting	Understand and Apply/Analyze
4. understand the concepts related to SVM, Neural Networks, etc. and apply them for analyzing data	Understand and Apply/Analyze
5. understand the concepts related to text mining and apply them in various contexts	Understand and Apply/Analyze
6. apply clustering algorithms and related methods	Apply

Unit I

Concept of statistical learning, inference vs prediction, types of errors in modelling (reducible, irreducible), assessing model accuracy via train-test and cross-validation approaches, bias-variance tradeoff, concepts of supervised and unsupervised learning, classifier performance via confusion matrix and related measures, ROC, K - nearest neighbourhood algorithm for classification and regression, naïve Bayes and Bayesian networks.

Unit II:

Feature selection and feature extraction, feature selection using exploratory and inferential statistical techniques, multiplicity issues, missing data imputation techniques, classification and regression trees, Classification error measures such as misclassification probability, Gini Index, Cross-Entropy, Cost-Complexity pruning, missing data in trees, bagging and boosting, random Forests, algorithms like Adaboost, XGBoostetc.

Unit III

Optimal separating hyperplane, soft-margin classifiers, support vector machines, handling nonlinear class boundaries via kernels, neural networks for classification and regression, projection pursuit regression, activation functions, hidden layers, feed forward and back propagation techniques in fitting.

Unit IV

Unsupervised learning methods such as self-organizing maps, association rule mining, K medoids clustering, methods for determining optimal number of clusters, introduction to text analytics, Zipf's law, concepts such as stop words, document term matrix, N-grams, sentiment analysis, topic modelling.

Books Recommended:

1. Alpaydin, E. (2014), *Introduction to Machine Learning*, 3rd Ed. MIT Press.
2. Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth and Brooks.
3. Hastie T., Tibshirani R. and Friedman J. H., (2008). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.
4. James G., Witten, D., Hastie, T. Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*, Springer
5. Larose, D. T. and Laros, C. (2015). *Data Mining and Predictive Analytics*. Wiley.
6. Mohammad J. Zaki and Wagner Meira. (2014). *Data Mining and Analysis. Fundamental Concepts and Algorithms*. Cambridge University Press, New York.
7. Ripley,vB. D. (1996). *Pattern Recognition and Neural Networks.* Cambridge University Press
8. Shmueli, G., Patel, N. Bruce, P. (2010). *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XL Miner*, Wiley.
9. Silge J. and Robinson D. (2017), *Text Mining with R - A Tidy Approach*, OReilly Publication

STE302: Survival Analysis

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. evaluate the survival probability with respect to various ageing models	Understand and Evaluate
2. estimate the survival function parametrically using various parametric models from the given survival data	Understand and Evaluate/Apply
3. estimate the survival function nonparametrically from a given survival data (Kaplan-Meir estimation)	Understand and Evaluate/Apply
4. understand the concepts related to the estimation of survival function under a parametric regression set up	Understand and Apply/Analyze
5. understand the concepts related to estimation of survival function	Understand and

	under a semi-parametric regression set up (Cox PH model)	Apply/Analyze
6.	understand the theory related to competing risk model and apply them for the estimation of survival function under that set up	Understand and Apply/Analyze
7.	understand the concepts related to the point-process approach of survival function estimation and analysis	Understand and Apply/Analyze
8.	understand the concepts related to frailty modeling and Apply them for survival data	Understand and Apply/Analyze

Unit I

Survival data, Concepts of time, order and random and hybrid censoring, Life distributions - exponential, gamma, lognormal, Pareto, linear failure rate, ageing classes- IFR, IFRA, NBU, NBUE, HNBUE and their duals, bathtub failure rate.

Parametric inference, point estimation, confidence intervals, scores, tests based on LR, MLE.

Unit II

Life tables, failure rate, mean residual life and their elementary properties.

Estimation of survival function - actuarial estimator, Kaplan - Meier estimator, estimation under the assumption of IFR/DFR.

Unit III

Semi-parametric regression for failure rate - Cox's proportional hazards model, partial likelihood, estimation and inference methods for the Cox models, time-dependentcovariates, residuals and model diagnosis, functional forms of the Cox models, goodness-of-fit tests for the Cox models.

Competing risk models, repair models, probabilistic models, joint distribution of failure times unconditional tests for the time truncated case.

Tests for exponentiality, two sample non-parametric problem

Unit IV

Nelson-Aalen estimators, counting processes and martingales, modeling counting processes, regression models for modeling multiple events.

Frailty models, shared frailty models, identifiability of frailty models, frailty regression models, Bivariate and correlated frailty models, additive frailty models.

Books Recommended:

1. Collett, D. (2003). *Modelling Survival data in Medical Research*, Second Edition, Chapman & Hall/CRC
2. Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data*, Chapman and Hall.
3. Deshpande, J.V. and Purohit, S.G. (2005). *Life Time Data: Statistical Models and Methods*, Word Scientific.
4. Duchateau, L. and Johnson, P. (2008). *The Frailty Mode*l. Springer: New York.
5. Hanagal, D. D. (2011). *Modeling Survival Data Using Frailty Models*. CRC Press.

6. Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer: New York.
7. Kalbfleish, J. D. and Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
8. Klein, J. P. and Moeschberger, M. L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data* , Springer, New York
9. Liu Xan (2012). *Survival Analysis: Models and Applications*, Wiley.
10. Moore, D. F. (2016). *Applied Survival Analysis Using R*, Springer
11. Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data, Extending the Cox Model*, Springer, New York.
12. Wienke, A. (2011). *Frailty Models in Survival Analysis*, CRC Press: New York.

STE307: Asymptotic Inference

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. obtain CAN estimators under various situations	Evaluate
2. understand the symptotic properties of MLE	Understand
3. obtain asymptotic tests for various testing problems	Evaluate
4. understand the concepts related to asymptotic efficiency in testing and evaluate the efficiency of tests	Understand Evaluate
5. understand the concepts related to local asymptotic normality and examine the LAN for various cases	Understand Evaluate

Unit I

Consistent and asymptotically normal (CAN) estimators for real and vector valued parameters, invariance property under continuous transformation, methods for generating CAN estimators for real and vector valued parameters using method of moments and method of percentiles. Comparison of consistent estimators, minimum sample size required by the estimator to attain certain level of accuracy.

Unit II

Asymptotic properties of Maximum likelihood estimates, inconsistent MLEs, Asymptotic distribution of MLE in special class of distributions: Cramer regularity conditions, Cramer-Huzurbazar theorem, extension to vector-valued parameters.

Unit III

Asymptotic theory of tests of hypotheses: Tests based on MLEs. Likelihood ratio tests, asymptotic distribution of log likelihood ratio, Wald test, score test, Pearson's chi-square test and LR test, consistent test.

Asymptotic efficiency in testing, Pitman efficiencies, Bahadur slopes and Bahadur efficiency.

Asymptotic confidence intervals: construction and examples.

Unit IV

Contiguity of probability measures, local asymptotic normality (LAN), efficiency of estimators, Le Cam's lemmas, limitations and redundancy of Cramer's regularity conditions.

Books Recommended:

1. Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series, Second Edition.
2. Das Gupta, A. (2008), *Asymptotic Theory of Statistics & Probability*, Springer, New York
3. Deshmukh S. R. and Kulkarni M. G. (2021). *Asymptotic Statistical Inference - A BasicCourse Using R*, Springer
4. Ferguson, T. S. (1996), *A Course in Large Sample Theory*, Chapman & Hall,London
5. Kale, B.K. & Muralidharan, K. (2015) *Parametric Inference: An Introduction*, AlphaScience International Ltd.
6. Le Cam, L. M. and Yang, G. (1990), *Asymptotics in Statistics: Some Basic Concepts*, Springer, New York
7. Lehmann, E. L. (1999), *Elements of Large Sample Theory*, Springer, New York
8. Lehmann, E. L. and Romano, J. (2005). *Testing Statistical Hypotheses*, Springer
9. Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, New York
10. Rao, C. R. (1995). *Linear Statistical Inference and its Applications*, Wiley
11. Rohatgi, V. K. and Saleh, A.K. Md. E. (2001). *Introduction to Probability and Statistics*, John Wiley & Sons, New York.
12. Roussas, G. G. (1972), *Contiguity of Probability Measures: Some Applications in Statistics*, Cambridge University Press, London
13. van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press, London
14. Shao, J. (2003). *Mathematical Statistics*, Springer-Verlag, New York,

Electives - Semester IV

STE404: Computer Intensive Statistical Methods

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand and apply different sampling methods such as AR, importance, MH, Gibbs, etc.	Understand Apply
2. handle missing observation cases with methods related to EM algorithm	Understand and Apply
3. carry out bootstrap and jackknife methods for bias, standard error estimation and confidence interval construction	Understand and Apply
4. understand the concepts related to bagging and boosting and apply them	Understand and Apply

5. understand the concepts related to smoothing techniques and apply them	Understand and Apply

Unit I

Review of ARM, importance sampling, Metropolis-Hastings and Gibbs sampling algorithms, rejection algorithms for approximate Bayes computation (ABC-Rejection), Particle filtering, Inference in Hidden Markov Models (HMM).

Unit II

Missing values and imputation techniques: Missing values and types of missing, imputation methods for missing values, single and multiple imputations. MCMC methods for missing values, EM algorithm and applications: EM algorithm for incomplete data, EM algorithm for mixture models, EM algorithm for missing values, stochastic EM algorithm.

Unit III

Bootstrap methods, estimation of sampling distribution, various types of confidence intervals, variance stabilizing transformation, Jackknife and cross-validation, permutation tests. Bagging and boosting methods with applications. Cross validation analysis

Unit IV

Smoothing techniques: kernel estimators, nearest neighbor estimators, orthogonal and local polynomial estimators, wavelet estimators, splines, choice of bandwidth and other smoothing parameters. statistical methods for big data analytics

Books Recommended:

1. Buuren, Stef van (2012). *Flexible Imputation of Missing Data*. Chapman and Hall.
2. Chihara, L. and Hesterberg, T. (2011) *Mathematical Statistics with Resampling and R*. Wiley.
3. Davison, A.C. and Hinkley, D.V. (1997) *Bootstrap methods and their Applications*. Chapman and Hall.
4. Effron, B and Hastie, T (2016). *Computer-Age Statistical Inference-Algorithms, Evidence and Data Science*, Cambridge University Press.
5. Gilks, W. R., Richardson, S., and Spiegelhalter, D. (eds.) (1995) *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
6. Good, P. I. (2005) *Resampling Methods: APractical Guide to Data Analysis*. Birkhauser Bosel.
7. Jim, A. (2009). *Bayesian Computation with R*, 2nd Edn, Springer.
8. McLachlan, G.J. and Krishnan, T. (2008) *The EM Algorithms and Extensions*.Wiley.

STE406: Advanced Statistical Learning Techniques & Applications

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand different deep learning methods	Understand
2. understand semi supervised learning methods	Understand
3. apply them for solving problems in different domains	Apply

Unit I
Deep Learning Architectures: Logistic regression Neural Networks - Perceptron, multilayer network, backpropagation, RBF Neural Network, CNN, RNN, LSTM, AlexNet, VGGNet, GoogleNet, Backpropagation, Deep networks Regularization, Dropout, Batch Normalization.

Unit II
Deep Learning for Computer Vision: Popular CNN architectures Transfer learning, autoencoders and relation to PCA, Object detection, image segmentation RNN and LSTM for image captioning/video.

Unit III
Deep Reinforcement Learning: Introduction to sequential decision making under uncertainty Implementing RL algorithms with deep neural networks. Value functions, Finite and infinite Problems.

Unit IV
Representation Learning: Deep Generative Models, Semi and Self-supervised Learning.

Books Recommended:

1. Alpaydin, E. (2015). Introduction to Machine Learning, 3rd Edition, Prentice Hall (India)
2. Duda, R. O. Hart, P. E. and Stork, D. G. (2007). Pattern Classification, 2nd Edn., Wiley India,
3. Bishop, C. M. (2006). Pattern Recognition and Machine Learning (Information Science and Statistics),Springer
4. Bhuyan, M. K. (2019). Computer Vision and Image Processing: Fundamentals and Applications, Published by CRC
5. Haykin, S. O. (2016). Neural Networks and Learning Machines, 3rd Edition, Pearson Education (India),
6 Goodfellow, I., Bengio, Y. and Courville, A. (2016). Deep Learning, MIT Press
7. Nielsen, M. A. (2015). Neural Networks and Deep Learning, Determination Press,
8. Bengio, Y. (2009). Learning Deep Architectures for AI, Now Publishers Inc.,

STE407: Design & Analysis of Clinical Trials

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. Understand different phases of clinical trials	Understand
2. Understand data management in clinical trials	Understand
3. Understand various aspects associated with designing a clinical trials (cross-over design, Balaam.s design etc.)	Understand and Apply
4. Apply different statistical procedures useful in testing Bioequivalence of more than two drugs	Apply
5. Carry out drug interaction, dose proportionality etc.	Apply

Unit I

Introduction to clinical trials: need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multi-center trials. Data management: data definitions, case report forms, database design, data collectionsystems for good clinical practice. Bioavailibility, pharmacokinetics and pharmacodynamics, two-compartment model.

Unit II

Design of clinical trials: parallel vs. cross-over designs, cross-sectional vs. longitudinal designs, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials. Design and monitoring of Phase III trials with sequential stopping, design of bio-equivalence trials, Inference for 2x2 crossover design: Classical methods of intervalhypothesis testing for bioequivalence, Bayesian methods, nonparametric methods.

Unit III

Estimands in clinical trials, power and sample size determination, multiplicative (or log-transformed) model, ML method of estimation, assessment of inter and intra subject variabilities, detection of outlying subjects. Optimal crossover designs: Balaam's design, two-sequence dual design. Optimal four period designs. Assessment of bioequivalence for more than two drugs, Williams design.

Unit IV

Designs based on clinical endpoints: Weighted least squares method, log-linear models, generalized estimating equations, drug interaction study, dose proportionality study, steady state analysis. Meta analysis, analysis of categorical data.

Books Recommended:

1. Chow S.C. and Liu J.P.(2009). *Design and Analysis of Bioavailability and bioequivalence*. 3rd Ed. CRC Press.
2. Chow S.C. and Liu J.P. (2004). *Design and Analysis of Clinical Trials*. 2nd Ed. Marcel Dekkar.

3. Fleiss J. L.(1989). *The Design and Analysis of Clinical Experiments*, Wiley.
4. Friedman L. M.Furburg C. Demets D. L.(1998). *Fundamentals of Clinical Trials*, Springer.
5. ICHE9 Guideline: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf
6. Jennison .C. and Turnbull B. W. (1999). *Group Sequential Methods with Applications to Clinical Trails*, CRC Press.
7. Marubeni .E. and Valsecchi M. G. (1994). *Analyzing Survival Data from Clinical Trials and Observational Studies*, Wiley.