Savitribai Phule Pune University

Two Year M. A. / M. Sc. Degree Program in Statistics (Revised Syllabus)

To be implemented in the Department of Statistics,

Savitribai Phule Pune University

(With effect from Academic Year 2021-22)

Submitted by

Prof. T. V. Ramanathan
Head of the Department of Statistics,

Date: 31 May, 2021

Approved by Department Committee
May 31, 2021

Approved by Board of Studies on
June 07, 2021

Dean of Science Faculty

Pro-Vice Chancellor

Vice Chancellor

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

2. Preamble of the syllabus: M. A. / M. Sc. Statistics program is of **minimum** 80 credits spread over four semesters. This program is offered at the Department of Statistics, Savitribai Phule Pune University. The program emphasizes both theory and modern applications of statistics 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 program has some unique features such as independent projects, a large number of elective courses, pre-requisite system and extensive computer training of statistical computations including standard software packages such as MATHEMATICA, MATLAB, MINITAB, R, PYTHON, etc.. The department has the academic autonomy and it has been utilized to add the new and need based elective courses. In the past courses such as *Time Series Analysis, Survival Analysis, Reliability and Quality Control, Actuarial Statistics, Stochastic Models in Finance, Statistical Learning and Data Mining, Statistical Analysis of Clinical Trials, Statistical Analysis of Microarray Data, Expert Systems with Applications, Statistical Methods for Bio-Computing, Computer Intensive Statistical Methods, Statistical Methods in Epidemiology, Advanced Time Series Analysis and Natural Language Processing* have been introduced. The independent project work is one of the important components of this program. The syllabus of the first year (two semesters) covers most of the core courses. In the second year of the syllabus, there are four core courses, five optional courses and one project. The syllabus has been framed to have a good balance of theory, methods and applications of statistics.
It is possible for the students to study basic courses from other disciplines such as economics, life sciences, computer science and mathematics in place of optional/electives.

3. Introduction: M.A./M.Sc. Statistics program has semester pattern and Choice-based Credit System. The program consists of 80 credits.
The program has pre-requisites system. Under this system, a student has to clear pre-requisite courses to take some of the courses in following semesters. Unless a student passes the pre-requisites of a course, (s)he will not be allowed to enroll in the said theorycourses. The Pre-requisites of a course are indicated in square brackets against the course. The Departmental committee may change the pre-requisites of a course.

4. Eligibility: For M. A. in Statistics following candidates are eligible.

(i) B. A. (Second class) with Statistics as principal and Mathematics at subsidiary level,
(ii) B. A. (Second class) with Mathematics as principal and Statistics at subsidiary level,
(iii) M. A. (Second class) in Mathematics
(iv) B. A. (Second class) in Actuarial Science with Mathematics and Statistics at subsidiary level,
(v) B. A. (Second class), with Statistics as one of the three subjects.

For M. Sc. in Statistics following candidates are eligible.

(i) B.Sc. (Second class) with Statistics as principal and Mathematics at subsidiary level,

(ii) B.Sc. (Second class) with Mathematics as principal and Statistics at subsidiary level,

(iii) M.Sc. (Second class) in Mathematics,

(iv) B.Sc. (Second class) in Actuarial Science with Mathematics and Statistics at subsidiary level,

(v) B.A./B.Sc. (Second class), with Statistics and Mathematics as two of the three subjects.

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, assignments, small projects, viva-voce and presentations. There would be assignments or minor projects for some of the elective courses in ETE also.

(ii) **Pattern of the question paper:** For a theory/practical courses the duration for the ETE will be three hours (in offline mode) and one hour in the online mode.

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) **ATKT rules:** A student can register for the third semester, if (s)he completes50% credits of the total credits expected to be completed within first two semesters, subject to the pre-requisite system mentioned above. Once registered, a student must complete M. A./ M. Sc. within a period as specified by the University.

D) **Award of class:** As per the University rules.

E) **External students:** Not applicable

F) **Setting of question paper:** The Department has been given autonomy for setting the question papers and grading the answer papers. A teacher teaching a course sets the question paper of the course and assesses the answer books. Question papers may be moderated by a committee of subject experts.

G) **Verification or revaluation:** As per the University rules

6. Structure of the Program

(a) Compulsory papers

Compulsory courses in semester I-IV are listed below. In the square brackets **pre-requisite** subjects have been mentioned. **Each course is of 4 credits**.

Semester I

ST-101: Mathematical Analysis
ST-102: Linear Algebra
ST-103: Probability Distributions
ST-104: Calculus
ST-105: Statistical Computing (Theory 3 credits + Practical 1 credit)

Total credits 20

Semester II

ST-201: Probability Theory [ST-101]
ST 202: Multivariate Analysis [ST-102]
ST-203: Modern Statistical Inference [ST-103, ST-104]
ST-204: Regression Analysis [ST-102]
ST-205: Practical- I (based on topics from ST-202, ST-203 and ST-204)

Total credits 20

Semester III

ST-301: Stochastic Processes
ST-302: Design and Analysis of Experiments [ST-204]
ST-303: Practical - II (based on ST-301, ST-302)
Optional - I
Optional - II

Total credits 20

Semester IV

ST-401: Sampling Methods and Applications
ST-402: Project
Optional - III
Optional - IV
Optional - V

Total credits 20

The core courses have 60 (47 (theory) + 9 (practical) + 4 (project)) credits in all (75% of total credits) and the five optional courses have 20 credits (25% of total credits).

A student may be allotted optional courses of Semester III by the Department at the beginning of the second year of the program. The allotment is based on the performance of the student as indicated by her/his rank among all the students admitted in the same year and the courses that the student preferred.

A student can opt for a theoretical project in lieu of an optional course or two optional courses with the permission of the Head of the Department.

A list of optional courses is given below. **Each optional course is of 4 credits.**

.

List of 25 Optional Courses

ST O1 Advanced Probability [ST 201]
ST O2 Advanced Stochastic Processes [ST 301]
ST O3 Inference in Stochastic Processes [ST 203, ST301
ST O4 Spatial Processes and their Applications [ST 301]
ST O5 Stochastic Models in Finance [ST 301]
ST O6 Asymptotic Theory of Statistical Inference [ST201, ST 203]
ST O7 Time Series Analysis [ST 204]
ST O8 Advances in Generalized Linear Models [ST 204]
ST O9 Astrostatistics [ST 202, ST 203, ST 204]
ST O10 Optimization Techniques [ST 102, ST 104]
ST O11 Reliability and Statistical Quality Control [ST 203]
ST O12 Computer-Intensive Statistical Methods [ST 105 ST 203]
ST O13 Statistical Learning and Data Mining [ST 202, ST 204]
ST O14 Survival Analysis [ST 203, ST 204]
ST O15 Actuarial Statistics [ST 103]
ST O16 Medical and Health Statistics [ST 103, ST 204]
ST O17 Design and Analysis of Clinical Trials [ST 204, ST 302]
ST O18 Statistical Methods in Micro-array Data Analysis [ST 202]
ST O19 Statistical Methods for Bio-computing [ST 105, ST 301]
ST O20 Bayesian Inference [ST203]
ST O21 Expert Systems with Applications [ST 105]
ST O22 Financial Econometrics [ST 204, ST 301]
ST O23 Advanced Time Series Analysis [ST 204]
ST O24 Statistical Methods in Epidemiology
ST O25 Statistical Foundations of Data Science [ST204]

The Head of the Department may introduce additional optional courses on recommendations of the Departmental Committee. The syllabus of the optional courses will be prepared by the concerned teacher and will be flexible to accommodate new developments in that area. Whenever such an optional course is floated, the concerned syllabus will be discussed and approved in the Departmental committee.

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

Service Courses:
In addition to above, Department of Statistics will also offer two services courses for M.A./M.Sc./ M.E./M.Tech./M.Com./M.B.A./Integrated M.Sc. (4^{th} & 5^{th} year) students from the departments and centers in SPPU campus.

ST S1: Introduction to Statistical Analysis
ST S2: Advanced Statistical Analysis

Question papers: In view of academic autonomy given to the Department, question papers are set by the teacher who teaches the course and these may be moderated by a committee of experts, as indicated earlier.

Medium of instructions: English

7) Equivalence: The Head will take the decisions as and when the case arises.

8) University Terms: Time table for the beginning and end of the terms as announced by the University will be followed.

11) Qualification of Teacher: As per the University rules.

9) and 10) Subject wise detailed syllabus and recommended books: Detailed syllabi, along with the list of recommended books of the compulsory courses in the first year of M.Sc. are given in the attached document.

Detailed Syllabus

CORE COURSES

The syllabus of all the **CORE COURSES** is given below. All these core courses are of **4 credits**.

ST 101: Mathematical Analysis

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. solve the problems related to set theory and point set topology	Evaluate
3. understand the concepts of limits and convergence of sequences and series and solve problems related these topics	Understand and Evaluate
4. understand the concepts of limits and continuity of functions and solve problems related to these concepts	Understand and Evaluate
5. understand the concepts of compactness, completeness and connectedness of metric spaces	Understand

Unit I
Countability, supremum and infimum of sets of real numbers, Archimedean property, denseness property of rational numbers, metric spaces, limit points and interior points of a set, open sets, closed sets, their properties etc. (These concepts will be introduced through metric spaces and R^n will be considered as a special case),

Unit II
Compactness, Bolzano-Weirstrass theorem, Heine-Borel theorem, sequences of real numbers, Cauchy sequence, limit superior, limit inferior, limit and convergence of a sequence of real numbers, Cauchy criterion for convergence, Completeness of R.

Unit III
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, examples and problems on these concepts.

Unit IV
Continuous functions, definition using open sets, uniform continuity, Lipschitz continuity, uniform convergence of sequences and series of functions, applications to power series.

Introduction to topological spaces, completeness, compactness and connectedness (using metric spaces)

Books Recommended

1. Abbott, S. (2001), *Understanding Analysis*, Springer, New York
2. Apostol, T. M. (1975). *Mathematical Analysis: A Modern Approach toAdvanced 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. Goldberg R. R. (1976). *Methods of Real Analysis*, John Wiley
7. Kumar, A. and Kumaresan, S. (2014). *A Basic Course in Real Analysis*, CRC Press
8. Radulescu, T. T., Radulescu V. D., Andreescu T., (2009), *Problems in Real Analysis*, Springer, New York
9. Rudin, W. (1985). *Principles of Mathematical Analysis*, McGraw - Hill
10. Trench W. F. (2012). *Introduction to Real Analysis*, E-book.
11. Yau, D. (2013). *A First Course in Analysis*, World Scientific

ST 102: Linear Algebra

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 system of linear equations	Evaluate
4. understand the concepts of eigen value theory and solve problems related to eigen values of a matrix	Understand and Evaluate
5. understand the concepts of quadratic forms and solve problems related to these topics	Understand and 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,* 2^{nd} Ed., John Wiley, New York.

ST 103: Probability Distributions

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 and Evaluate
2. understand the measure theoretic definition of a random variable and 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 and Evaluate
6. solve problems related to multiple and partial correlations	Evaluate
7. understand the concepts related to sampling distributions and solve problems related to them	Understand and Evaluate
8. understand the theory related to linear and quadratic functions of normal random vectors and solve related problems	Understand and Evaluate
9. understand the concepts related to order statistics and solve problems related to the distributions of order statistics	Understand and 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 correlation coefficients 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*, DuxburyResource 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*, 3rdEd., 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 toMathematical 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.

ST 104: Calculus

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. solve the problems related to univariate differential and integral calculus	Evaluate
2. solve the problems related to multivariate differential and integral calculus	Evaluate
3. apply the techniques for finding optimum of functions	Evaluate
compute multiple line and surface integrals	Evaluate

Unit I
Review of calculus of one variable: differentiability, mean value theorem and Taylor series expansion. Functions of several variables: Continuity, uniform continuity, absolute continuity, functions of bounded variation.

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

Unit II
Differentials of composite functions (of several variables) and the 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

Unit III
Inverse function theorem (without proof), implicit function theorem (without proof), extremum problems involving functions of several variables and their applications, constraint optimization of functions with several variables and their applications

Riemann and Riemann–Stieltjes integrals, integration by parts, mean value theorem. Jordan content and Lebesque integrals

Unit IV
Improper Riemann – Stieltjes integrals: Improper integrals of first and second kind for one variable. Uniform convergence of improper integrals, differentiation under the sign of integral, Leibnitz rule, Multiple integrals and evaluation of multiple integrals by repeated integration, Mean-value theorem for multiple integrals.

Books Recommended

1. Apostol T.M. (1975). *Mathematical Analysis: A modern approach to advanced calculus*. Addison-Wesley
2. Bartle, R. G. (1976). *Elements of Real Analysis*, John Wiley
3. Ghorpade, S. R. and Limaye, B. V. (2006). *A Course in Calculus and Real Analysis*, Springer

4. Ghorpade, S. R. and Limaye, B. V. (2010). A *Course on Multivariable Calculus and Analysis*, Springer
5. Kreyszig, E. (1975). *Advanced Engineering Mathematics*, Wiley Eastern
6. Rudin, W. (1985). *Principles of Mathematical Analysis*, McGraw-Hill
7. Trench W. F. (2012). *Introduction to Real Analysis*. E-book

ST 105: Statistical Computing

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. handle different statistical software	Understand
2. use R and PYTHON for various statistical computations	Apply
3. understand the theory of random number generation using various methods and apply them to generate random numbers	Understand and Apply
4. apply different search algorithms	Apply
5. evaluate different integrals using numerical as well as Monte Carlo method.	Evaluate
6. verify different basic concepts in probability and statistics empirically	Evaluate
7. evaluate rank and determinants of matrices of small and large order	Evaluate
8. compute inverse, g-inverse, M-P inverse of small and large order matrices with the help of R or PYTHON	Evaluate
9. solve system of linear equations using R and PYTHON	Apply

This course is partly theory and mostly lab oriented. There will be 2 hours of lectures per week and a minimum of 4 hours of lab. Good programming practices and efficient use of software packages such as R, PYTHON, MINITAB, MATLAB, MATHEMATICA should be taught during lab sessions.

Unit I

(a) 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

(b) 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.

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

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

Unit II

Numerical algorithms such as direct search, grid search, interpolation search, gradient search, Bisection and Newton-Raphson methods, Mullers method, Aitkens extrapolation, simple applications of the above methods.

Unit III

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

Practical on topics from linear algebra:

1. Calculation of rank and determinant of higher order matrix by partitioning method.
2. Calculation of equivalent canonical form by using elementary row and column operations
3. Calculation of inverses of symmetric matrices of higher order by partitioning method
4. Calculation of Inverse, Moore-Penrose inverse, and g-inverse of small and large order matrices.
5. Calculation of eigen values and eigen vectors of small and large order matrices
6. Solution of simultaneous system of equations (small and large)
7. Spectral decomposition, LU decomposition and SV decomposition of matrix and computation of powers of a matrix

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. Rajaraman, V. (1993). *Computer Oriented Numerical Methods*, Prentice-Hall, Fourth edition
6. Rizzo, M. L. (2007*). Statistical Computing with* R, CRC Press.
7. Schilling, R.J. and Harris, S.L. (2002). *Applied Numerical Methods for Engineers Using MATLAB and C*, Thomson and Brooks/Cole, Singapore

8. Thomas Haslwanter (2016). *An Introduction to Statistics with Python: With Applications in the Life Sciences*, Springer

ST 201: Probability Theory

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 convergence of sequence of random variables	Evaluate
5. understand the law of large numbers and the central limit theorem related to 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) law 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

ST 202: Multivariate Analysis

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*, 6^{th} 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

ST 203: Statistical Inference

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 property from a given sample with appropriate distributional assumptions	Evaluate
5. obtain tests and confidence intervals with some	

	with optimal property	Evaluate
6.	obtain CAN estimators under various situations	Evaluate
7.	understand the properties of MLE	Understand
8.	obtain asymptotic tests for various testing problems	Evaluate

Unit I

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

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

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

Unit II

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

Test function, MP tests, Neyman- Pearson lemma, UMP tests, nonexistence of UMP tests MLR property.

Unit III

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

Maximum likelihood estimation, restricted parameter space, inconsistent MLEs, MLEs in irregular cases

Asymptotic distribution of MLE in special class of distributions: Cramer regularity conditions, Cramer- Huzurbazar theorem, extension to vector-valued parameters

Unit IV

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

Asymptotic confidence intervals: construction and examples.

Applications to categorical data analysis, tests for independence for three dimensional contingency tables

Books Recommended

1. Casella, G. and Berger, R. L. (2002). *Statistical Inference*. Duxbury Advanced Series, Second Edition.
2. Deshmukh S. R. and Kulkarni M. G. (2021). *Asymptotic Statistical Inference - A Basic Course Using R*, Springer
3. Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*. Cambridge University Press
4. Ferguson, T.S. (1996). *A Course in Large Sample Theory*, Chapman & Hall, London
5. Gupta Anirban Das (2008). *Asymptotic Theory of Statistics and Probability*, Springer
6. Kale, B.K. & Muralidharan, K. (2015) *Parametric Inference: An Introduction*, Alpha Science International Ltd.
7. Lehmann, E. L. and Romano, J. (2005). *Testing Statistical Hypotheses*, Springer
8. Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*. Springer, New York
9. Rao, C. R. (1995). *Linear Statistical Inference and its Applications*, Wiley
10. Rohatgi, V. K. and Saleh, A.K. Md. E. (2001). *Introduction to Probability and Statistics,* John Wiley & Sons, New York.
11. Shao, J. (2003). *Mathematical Statistics*, Springer-Verlag, New York,

ST 204: Regression Analysis

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, 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 II

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

Unit III

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

Unit IV

Non linear regression (NLS): linearization transforms, their uses, iterative procedures for NLS, grid search, Newton-Raphson, steepest descent, Marquardt's methods.

Semiparametric regression models, nonparametric regression methods, Introduction to additive regression models and generalized additive models (GAM)

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. Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models*, CRC Press
4. Hosmer, D. W. and Lemeshow, S. (1989). *Applied Logistic Regression*, Wiley.
5. Kleinbaum, D. G. & Klein, M. (2002). *Logistic Regression: A Self-Learning Text*, Springer
6. McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, Chapman& Hall.
7. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). *Introduction toLinear Regression Analysis*, Wiley.
8. Neter, J., W., and Kutner, M. H. (1985). *Applied Linear Statistical Models*, Wiley.
9. Ratkowsky, D. A. (1983). *Nonlinear Regression Modelling*, Marcel Dekker, London.
10. Ruppert, D., Wand, M. P. and Carroll, R. J. (2003) *Semi parametric Regression*, Cambridge University Press.
11. Seber, G. E. F. and Wild, C. J. (1989). *Nonlinear Regression*, Wiley.
12. Weisberg, S. (2005). *Applied Liner Regression*, Wiley.
13. Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R*, CRC Press
14. Yan, X. and Su, X. G. (2009). *Linear Regression Analysis: Theory & Computing*, World Scientific.

ST 205: Practical I

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. carry out regression analysis given the data	Analyze
2. carry out binary and multiple logistic regression	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 Linear Regression
2. Multiple Regression
3. Variable Selection Problem
4. Multicollinearity and Ridge Regression
5. Regularized Methods (LASSO)

Unit II

1. Nonlinear regression
2. Nonparametric regression
3. Logistic regression (binary and multiple)
4. Poisson/Negative binomial regression
5. GLM and GAM

Unit III

1. Graphical representation of multivariate data
2. Principal Component Analysis, Correspondence analysis
3. Factor Analysis
4. Cluster Analysis
5. Canonical Correlations

Unit IV

1. BIAS and MSE of estimators, Power and Size of tests and Coverage of confidence interval comparison in univariate inference problems
2. Verification of CAN property of estimators
3. Model Sampling from multivariate normal distribution
4. Likelihood ratio tests (multivariate)
5. Applications of Hotelling's T^2
6. MANOVA
7. Discriminant Analysis

Books Recommended:

1. Khattree, R. & Naik D. N. (2003), Applied Multivariate Statistics with SAS Software, Second Edition (SAS Institute and Wiley), Chapter 2
2. Recommended books in ST 202, ST 203 and ST 204courses

ST 301: Stochastic Processes

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

Unit I

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 ruin problem and one dimensional random walk. Concept of absorption probabilities, use of these to compute 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, 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 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 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*, 4^{th}Edn. Academic Press.
7. Ross, S. (2014). *Introduction to Probability Models*, 11th Edn. Academic Press.
8. Serfozo, R. (2009). *Basics of Applied Stochastic Processes*, Springer.

ST – 302: Design and Analysis of Experiments

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 and Evaluate
2. understand the concepts related to different factorial designs solve problems related to them	Understand and Evaluate
3. understand the concepts related various advanced designs and solve problems related them	Understand and Evaluate
4. understand the concepts related to response surface methodology and solve problems related to them	Understand and Evaluate
5. understand the concepts related to Taguchi methods and 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 wayclassification 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. 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

Analysis of 3^2 designs: contrasts for linear and quadratic effects, analysis of 3^2 design, confounding and fractional experiments in 3^2design. Response surface methodology (RSM): linear and quadratic model, stationary point, central composite designs(CCD), ridge systems, multiple responses, concept of rotatable designs, Box-Behnken design, 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

ST 303: Practical II

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

1. One way classification. Multiple comparison tests
2. 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)
3. Analysis of LSD and BIBD.
4. Analysis of covariance in one way and two way model.

Unit III

1. 2^k factorial experiments, Analysis of single replicate of 2^k.
2. Total and partial confounding in 2^k factorial experiments.
3. Analysis of 2^kfractional factorial experiments
4. Analysis of 3^2 factorial experiments

Unit IV

1. Random effect and mixed models
2. Analysis of first and second order response surface model.
3. Central composite design. Contour and surface plots, Box-Behnken design, Small composite designs, optimality of designs, simplex lattice designs, simplex centroid designs.
4. Taguchi methods: S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays.
5. Split-plot designs

ST 401: Sampling Methods and Applications

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related various standard sampling designsand solve problems related to them	Understand and 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 and 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).

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, New Delhi.
11. Thmpson, S. K. (2012). Samplig, 3rdEdn., Wiley

ST 402 Project

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

Project Guidelines:

1. ST 402 project is a **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 be two presentations (each graded out of 25 points by the Supervisor) for the continuous internal assessment (CIA). In the first 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 TWO 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.

Optional Courses

The syllabus of all the listed optional courses is given below. These optional courses will be offered subject to the availability of the expert faculty in that area. The available optional courses will be declared before the commencement of the semester. All these courses are of **4 credits**.

ST O1 Advanced Probability

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to decomposition of distribution, Radon-Nikodym derivatives and solve related problems	Understand and Evaluate
2. understand the concept of Kolmogorov’s consistency theorem	Understand
3. understand the concept of convergence and solve related problems	Evaluate
4. understand the concept of conditional probability and expectation and solve related problems	Understand and Evaluate
5. understand the concept of martingales and solve problems involving martingales	Understand Evaluate
6. understand the central limit theorem for martingale and mixing sequences	Understand

Unit I
Measure and Integration, Integral of a measurable function with respect to a measure, its properties, L-p spaces, Hahn – Jordan decomposition, Lebesgue decomposition, Radon – Nikodym derivative, Product measure, Fubini's theorem, Convolutions.

Unit II
Probability spaces, Kolmogorov’s consistency theorem
Distribution Functions and their basic properties, Helly–Bray type results

Unit III
Convergence in measure, Almost everywhere convergence, Kolmogorov inequality, Kolmogorov three series criterion and strong law of large numbers, Introduction to weak convergence

The Paul Lévy Continuity Theorem, Convergence in distribution in the multidimensional case - The Cramér-Wold device with examples

Unit IV
Conditional probability and conditional expectations, their simple properties, discrete parameter martingales,

Limit theorems for dependent sequences of random variables, Martingale convergence theorems, central limit theoremfor martingales, Mixing sequences, Mixing coefficients, Coupling and covariance inequalities, Central limit theorems for mixing sequences

Books Recommended

1. Ash, R. B. (2000). *Probability & Measure Theory*. Academic Press.
2. Ash, R.B. (1972). *Real Analysis and Probability*. Academic Press
3. Athreya, K.B. and Lahiri, S.N. (2006). *Measure Theory and Probability Theory*. Springer.
4. Athreya, S. and Sunder, V. S. (2008). *Measure and Probability*, CRC Press
5. Billingsley, P. (1986). *Probability and Measure*. John Wiley
6. Dudley, R. M. (2004). *Real Analysis and Probability*. Cambridge University Press.
7. Rosenthal, J. (2006). *A First Look at Rigorous Probability Theory*, 2^{nd} Ed. World Scientific.
8. Roussas, G. G. (2014) *An Introduction to Measure-Theoretic Probability*, 2nd Edition, Academic Press
9. Taylor, J. C. (1997). *Introduction to Measure and Probability*. Springer.
10. Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.

ST O2: Advanced Stochastic Processes

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related advanced theory of Markov chain and solve problems related to this topic	Understand and Evaluate
2. understand the general theory of Markov sequences and solve related problems	Understand and Evaluate
3. understand various advanced concepts related to stochastic processes	Understand
4. understand the concepts related to diffusion and related processes and solve related problems	Understand and Evaluate

Unit I
Markov Chains: Taboo probabilities and ratio limit theorems. Invariant measures:

Unit II
Markov Sequences: Definitions, transition densities, stationary distribution, normal Markov sequences, Markov pure jumpprocesses

Unit III
Stochastic processes (general Theory): Probability spaces appropriate for stochastic processes, Kolmogorov's extension theorem. (only sketch of the proof), separability, progressive measurability and strong Markov property of stochastic processes

Unit IV

Diffusion processes: Definition, elementary properties, infinitesimal parameters, standard process and Dynkin's theorem, continuity and non-differentiability of diffusion processes. Modeling based on diffusion processes, standard Brownian motion, Ornstein-Ulhenbeck process and other processes, transformation of processes, distribution of hitting times and related problems, scale function and speed density, Kolmogorov's backward differential equations, forward differential equations (without proof), transition density and stationary distribution of a diffusion process.

Books Recommended

1. Adke, S.R. and Manjunath, S. M. (1984). *Finite Markov Processes*, Wiley.
2. Ash, R.B. and Gardner, M. F. (1975). *Topics in Stochastic Processes*. Academic Press.
3. Athreya, K.B. and Lahiri, S.N. (2006). *Measure Theory and Probability Theory*. Springer.
4. Feller, W. (1969). *An Introduction to Probability Theory*. (Vol. II)
5. Karlin, S. and Taylor H.M. (1981). *A Second course in Stochastic Processes*. Academic Press.
6. Resnick, S. (2002). *Adventures in Stochastic Processes*, Springer
7. van Casteran, J. A. (2013). *Advanced Stochastic Processes*, University of Waterloo

ST O3: Inference for Stochastic Processes

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
5. understand the concepts related to inference associated with Markov chain model and solve related problems	Understand and Evaluate
6. understand the concepts related to inference associated with Poisson and other pure jump processes and solve problems	Understand and Evaluate
7. understand the concepts related to inference associated with diffusion and branching processes and solve related problems	Understand and Evaluate
8. understand the concepts related to semi, nonparametric and Bayesian inference related to some stochastic process models	Understand and Evaluate

Unit I

Inference in Markov chains, estimation of transition probabilities, testing for order of a Markov chain, estimation of functions of transition probabilities, parametric models and their goodness of fit, Markov sequences, estimation of parameters based on likelihood and conditional least squares, auto-regressive time series. Models for higher order Markov chains, (Raftery's long memory model), Statement of martingale strong law of large numbers and CLT for martingales, CAN property of the MLEs from a general sequence of dependent random variables, Fisher information, applications to Markov chains and sequences

Unit II

Inference for Poisson process, Likelihood of Poisson and other pure Jump Markov processes from first principles, CAN property of MLEs, testing for a Poisson process, Inference for non-homogeneous Poisson process. Inference for parametric pure jump processes, such as birth process, birth-death process, birth-death-immigration processes.

Unit III

Diffusion processes and their likelihood, properties of estimators (without proof) Branching processes, ergodic and non-ergodic processes, inconsistency of MLE/moment estimators, properties of estimators on the non-extinction path, estimation of asymptotic distribution theory

Unit IV

Elements of semi-parametric and non-parametric analysis, theory and applications of optimal estimating functions, estimation of transition and stationary density, intensity function of a counting process. Methods based on estimating functions, inference on panel data models.

Introduction to Bayesian inference for stochastic processes

Books Recommended

1. Adke, S.R. and Manjunath. S. M. (1984). *An introduction to Finite Markov Processes*, Wiley Eastern.
2. Basawa, I.V. and Prakasa Rao, B.L.S.(1980). *Statistical Inference for Stochastic Processes*, Academic Press.
3. Bhat, B. R. (2000). *Stochastic Models: Analysis and Applications*. New Age International.
4. Billingsley, P. (1962). *Statistical Inference for Markov Chains*, Chicago University Press.
5. Broemeling, L. D. (2017). *Bayesian Inference for Stochastic Processes*, CRC Press
6. Guttorp, P. (1991). *Statistical Inference for Branching Processes*, Wiley.
7. Guttorp, P.(1995). *Stochastic Modelling for Scientific Data*, Springer.
8. Prakasa Rao, B.L.S. and Bhat, B.R.(1996). *Stochastic Processes and Statistical Inference*, New Age International.
9. Rajarshi M.B. (2013). *Inference for Discrete Parameter Stochastic Processes*, Springer India.
10. Rao, M. M. (2000). Stochastic Process Inference Theory, Springer

ST O4: Spatial Processes and their Applications

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to spatial models and and their applications in geostatistics	Understand and Understand
2. understand the concept of nonstationary and non Gaussian spatial process models and solve related problems	Understand and Evaluate

3.	understand the concepts related to random fields	Understand
4.	understand the concept related to spatial point processes solve problems related to such processes	Understand and Evaluate
5.	understand the concepts related to spatio-temporal models and solve related problems	Understand and Evaluate
6.	analyze the spatial and spatio-temporal data using the models discussed in the syllabus	Apply and Analyze

Unit I

Spatial sampling, Smoothing and Interpolation

Spatial models and geo-statistics, Classical Geostatistical Methods, kriging, Variogram and covariance models and estimation,

Nonstationary Spatial Processes, Non-Gaussian and Nonparametric Models for Continuous Spatial Data

Unit II

Autocorrelation on spatial network

Random fields, Markov Random fields, Statistical inference in random fields

Unit III

Spatial Point Process, Spatial Point Process Models (Poisson, Cox, Markov etc.) , Parametric and Nonparametric Methods, Modeling strategies, Multivariate and Marked Point Processes,

Point Process Models and Methods in Spatial Epidemiology, Isotropy for spatial point patterns

Unit IV

Space-time data, Space-time models, space-time symmetry, parametric and nonparametric methods for assessing space-time symmetry, Spatio-Temporal Processes, Continuous Parameter Spatio-temporal Processes, Dynamic Spatial Models Including Spatial Time, Spatio-temporal Point Processes.

Books Recommended

1. Blangiardo, M. and Cameletti, M. (2015) *Spatial and Spatio-Temporal Bayesian Models with R- INLA*. Wiley
2. Cressie, N. and Wikie, C. K. (2011). *Statistics for Spatio-Temporal Data*, Wiley
3. Gaetan, C. and Guyon, X. (2010). *Spatial Statistics and Modeling*, Springer
4. Gelfand, A. E., Diggle, P., Fuentes, M. and Guttorp, P. (2010). *Handbook of Spatial Statistics*, CRC Press
5. Haining, R. P. and Li, G. (2021). *Modeling Spatial and Spatial-Temporal Data: A Bayesian Approach*, CRC Press
6. Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*, CRC Press.
7. Oyana, T. J. (2021). *Spatial Analysis with R: Statistics, Visualization and Computational Methods*, CRC Press.
8. Ripley, B. D. (2004). *Spatial Statistics*, Wiley

9. Schinazi, R. B. (2010). *Classical and Spatial Stochastic Processes With Applications to Biology*, Birkhauser
10. Sherman, M. (2011). *Spatial Statistics and Spatio-Temporal Data: Covariance Functions and Directional Properties* , Wiley
11. Tautu, P. (1984). *Stochastic Spatial Processes: Mathematical Theories and Biological Applications*, Springer-Verlag
12. vanLieshout, M. N. M. (2019). *Theory of Spatial Statistics: A Concise Introduction*, CRC Press.
13. Wikle, C. K., Mangion, A. Z. and Cressie, N. (2019). *Spatio-Temporal Statistics with R*, CRS Press

ST O5: Stochastic Models in Finance

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to different financial derivatives and solve related problems	Understand and Evaluate
2. understand different types of options and the concepts related to binomial model and solve problems	Understand Evaluate
3. understand and apply the concepts related to Ito calculus	Understand
4. understand the BSM model and evaluate the price of an option using BSM model	Understand and Evaluate
5. understand the concepts related to market risk and evaluation of risk measures	Understand and Evaluate
6. understand the concept of Interest rate models and solve related problems	Understand and Evaluate

Unit I

Derivatives, hedging, forward and future contracts, Markets, prices, arbitrage and hedging Complete market, market risk and credit risks in the use of derivatives.

Options markets, properties of stock option prices, American and European options, Binomial model: One-step and two-step models, Binomial trees, Risk neutral valuation

Unit II

Behaviour of stock price, Conditional expectation, martingales, Brownian motion and Geometric Brownian motion, Markov property, Ito integral, Ito/diffusion and mean reverting processes process, Ito Lemma.

Unit III

Black Scholes model: Distribution of returns, volatility, risk neutral pricing, equivalent martingale measure, Black-Scholes-Merton differential equation. Estimating volatility (historical data, implied volatility)

Capital Asset Pricing Models

Options on stock indices, currencies and futures

Unit IV

Some exotic equity and foreign exchange derivatives, Greek Letters and hedging,

Market risk, Value-at-risk as a measure of risk, Expected Shortfall, Coherent Risk Measure

Interest rate derivatives, Black model, Models of the term structure of interest rates: one factor diffusion model, Vasicek, Cox-Ingersoll-Ross and Hull white models

Books Recommended

1. Baxter, M. and Rennie,.A. (1996). *Financial Calculus*, Cambridge University Press.
2. Bingham, N. and Keisel, R. (1998). *Risk-Neutral Valuation*, Springer.
3. Bodie Z., Kane A., Marcus A. and Mohanty P. (2009). *Investments*, 8th Edn., McGraw Hill.
4. David, R. (2004). *Statistics and Finance: An Introduction*. Springer.
5. Hull, J. (2008). *Options, futures and other derivatives*, International 7th Edn, Pearson Prentice Hall.
6. Joshi, M. S. (2002). *The Concepts of Mathematical Finance*, Lecture Notes
7. Lindstorm, E., Madsen, H. and Nielsen, J. N. (2015). *Statistics for Finance*, CRC Press
8. Ross.S. (2003). *Introduction to Mathematical Finance*, Cambridge University Press.
9. Ruppert, D. (2015). *Statistics and Data Analysis for Financial Engineering: with R examples*, 2^{nd} Ed., Springer
10. Shreve, S. E. (2004). *Stochastic Calculus for Finance* I, Springer.
11. Shreve, S. E. (2004). *Stochastic Calculus for Finance II*, Springer.

ST O6: Asymptotic Theory of Statistical Inference

Course Outcome (CO)	Cognitive level
After completion of this course the students will be able to	
1. understand the concepts related to CAN and BAN estimators and solve related problems	Understand and Evaluate
2. understand the concepts related to asymptotic efficiency in testing and evaluate the efficiency of tests	Understand and Evaluate
3. understand the concept of M estimators and the related theory	Understand
4. understand the concepts related to asymptotic expansions of estimators (Edgeworth, Fisher-Cornish etc.) and derive them	Understand Evaluate
5. understand the concepts related to local asymptotic normality and examine the LAN for various cases	Understand Evaluate
6. understand the asymptotic theory related to bootstrap and examine them empirically	Understand Evaluate

Unit I

Classical CAN theory: consistency (strong, weak and uniform), CAN estimators, best asymptotic normal (BAN) estimators, Asymptotic efficiency, Hodge's estimator, delta method, maximum likelihood estimators (MLE), CAN-ness of MLE in Cramer-regular families, scoring procedures, inconsistent MLEs (Neyman-Scott problems)

Unit II

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

Concept of M estimators, influence curve, asymptotic linearity of estimators, introduction to functional calculus

Edgeworth expansions and cumulants, expansion for means, sample percentiles, t-statistics, Cornish-Fisher expansions

Unit III

Contiguity of probability measures, local asymptotic normality (LAN), efficiency of estimators, Le Cam's lemmas, limitations and redundancy of Cramer's regularity conditions, a class of non-differentiable location models which yield efficient estimators, quadratic mean differentiable (QMD) families, applications of LAN theory of QMD families,

Unit IV

Bootstrap distribution and the meaning of consistency, delta theorem for the bootstrap, failure of the bootstrap, bootstrap confidence intervals, bootstrap confidence intervals for quantiles, Bootstrap in regression, residual bootstrap, problems with bootstrapping in the case of dependent observations

Books Recommended

1. Das Gupta, A. (2008), *Asymptotic Theory of Statistics & Probability*, Springer, New York
2. Ferguson, T. S. (1996), *A Course in Large Sample Theory*, Chapman & Hall, London
3. Fernholz, (1983). Von Mises Calculus for Statistical Functionals, Spribger
4. Le Cam, L. M. and Yang, G. (1990), *Asymptotics in Statistics: Some Basic Concepts*, Springer, New York
5. Lehmann, E. L. (1999), *Elements of Large Sample Theory*, Springer, New York
6. Rajarshi, M. B. (2012), *Statistical Inference for Discrete Time Stochastic Processes*, Springer, New Delhi
7. Roussas, G. G. (1972), *Contiguity of Probability Measures: Some Applications in Statistics*, Cambridge University Press, London
8. van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press, London

ST O7: Time Series Analysis

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.

ST O8: 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, ordinal responses: 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*, 2^{nd} Ed. Chapman & Hall
3. Jiang, J. (2007). *Linear and Generalized Linear Mixed Models and their Applications*, 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

ST O9Astrostatistics

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand different types of astronomical data and their measurements, sources etc.	Understand
2. understand and apply the concepts related to smoothing of astronomical data	Understand and Apply
3. understand and apply the concepts related to the analysis of	Understand and

	astronomical data using regression techniques	Apply
4.	understand and apply the concepts related to the analysis of astronomical data using multivariate analysis methods	Understand and Apply
5.	understand and apply the concepts related to the analysis of astronomical data using time series techniques	Understand and Apply

Unit I

Basics of astronomy: observing the sky (naked eye, telescopes and instruments), distance scales, units

Basics of astrophysics: blackbody radiation, relation between magnitude and luminosity, stellar parallax and stellar distances, populations, clusters stars, galaxies and their distribution, scaling relations

Historical data driven discoveries in astronomy, sources and nature of astronomical data; syndication of data (archive, virtual observatory) Sloan Digital Sky Survey, vizier service, data on eclipsing binary stars, extra Galactic Distance Data Base (EDD), data on pulsars, gamma ray bursts etc

Nature of data gathering with discussion on the underlying statistical assumptions; sources of

errors in measurement and calibration

Unit II

Data smoothing: density estimation, histogram method (choosing optimal bin-width), kernel density estimation (univariate and multivariate kernel density estimation, choosing band widths by cross-validation), adaptive smoothing (adaptive kernel estimators, nearest-neighbour estimators),

Unit III

Regression techniques: least-squares regression (symmetric least-squares regression, bootstrap error analysis, robust regression, quantile regression), non-linear regression non-parametric regression (Nadaraya-Watson estimator, local regression), model validation and selection, introduction to penalized regression

Unit IV

Multivariate methods: outlier detection, quadratic discriminant analysis, model based clustering (Gaussian mixture models, estimation via EM), model validation and selection

Time series methods: basic time series methods, review of time-domain analysis, spectral analysis of evenly spaced data (Periodogram, Fourier power spectrum), spectral analysis of unevenly spaced data (non-Fourier periodograms, statistical significance of periodogram peaks, spectral analysis of event data)

Note: The concepts from above topics may be illustrated using case studies related to cosmology, big bang, LSS, gravitational wave, LIGO, TMT, SKA, stellar spectra and their classification, solar corona, helioseismology, neutron stars, pulsars etc.

Recommended Books

1. Andreon, S. and Weaver, B. (2015). *Bayesian Methods for the Physical Sciences: Learning from Examples in Astronomy and Physics*, Springer
2. Babu, G. J. and Fiegelson, E. D. (1996). *Astrostatistics*, CRC Press, London
3. Chattopadhyay, A. K. and Chattopadhyay, T. (2014). *Statistical Methods for Astronomical Data Analysis*, Springer
4. Feigelson, E. D. and Babu, G. J. (2012). *Modern Statistical Methods for Astronomy with R Applications*, Cambridge University Press
5. Feigelson, E. D. and Babu, G. J. (2012). *Statistical Challenges in Modern Astronomy V*, Springer.
6. Hilbe, J. M. (Ed.), (2013). *Astrostatistical Challenges for the New Astronomy*, Springer
7. Jenkins, C. R and Wall, J. V.(2012) *Practical Statistics for Astronomers*, 2nd Edition, Cambridge
8. Sarro, L.M., Eyer, L., O'Mullane, W., De Ridder, J. (Eds.) (2012). *Astrostatistics and Data Mining*, Springer

ST O10 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

ST O11 Reliability and 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 characteristic	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 independentcomponents, 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, poperties 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.

ST O12 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.

ST O13 Statistical Learning and 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

ST O14 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 under a semi-parametric regression set up (Cox PH model)	Understand and 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	Understand and

	of survival function estimation and analysis	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-dependent covariates, 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.

ST O15 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 solve associated problems	understand and 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.*

ST O16 Medical and Health Statistics

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand different measures of disease occurrence, accuracy of findings and diagnostic tests and evaluate them	Understand and Evaluate
2. understand different ethical issues associated with medical and health studies	Understand
3. understand the theory associated with longitudinal analysis of binary data from medical and health domain	Understand and Analyze
4. analyze the data from matched case-control studies	Analyze
5. estimate the survival function using K-M estimator from survival data	Apply/Analyze
6. carry out meta analysis of given data	Analyze

Unit I
Study designs in epidemiology. Measures of disease occurrence and association, variation and bias, identifying non-causal association and confounding, communicating results of epidemiological studies, ethical issues in epidemiology

Unit II
Defining and assessing heterogeneity of effects, interaction. Sensitivity and specificity of diagnostic test, cohort study designs, statistical power and sample size computations.

Unit III

Log-linear models, 2xk and 2x2x2 contingency tables, logistic model, analysis of binary data. causal Inference, longitudinal data cross-control study designs, matched case-control studies

Unit IV

Survival data, product-limit estimator, proportional hazards model, multivariate survival data
Agreement and reliability
Meta analysis

Recommended books

1. Agresti, A. (2002). *Categorical Data Analysis*, Wiley
2. Armitage, P., Berry, G., Matthews, J. N. S. (20) *Statistical Methods in Medical Research*, 4th Ed., Blackwell
3. Bland, M. (2015). *An Introduction to Medical Statistics*, OUP
4. Brookemeyer, R. and Gail, M. H. (1994). *AIDS Epidemiology : A Quantitative Approach*,OUP
5. Clayton, D. and Hills, M. (2013). *Statistical Methods in Epidemiology*, OUP
6. Daniel, W. W. and Cross, C. L. (2012). Biostatistics: A Foundation for Analysis in the Health Sciences, 10th Edition, Wiley
7. Diggle, P. J., Heagerty, P., Liang, K-Y and Zeger, S. L. (2013). *Analysis of longitudinal data*, OUP
8. Harris, M. and Taylor, G. (2014). *Medical Statistics Made Easy*, Vol. 1-3, Scion
9. Matthews, D. E. and Farewell, V. T. (2015). *Using and Understanding Medical Statistics*, 5th Ed., Karger
10. McCullagh, P. and Nelder, J. A. (1999). *Generalized Linear Models*, Chapman & Hall
11. Pagano, M. and Gauvreau, K. (2018). *Principles of Biostatistics*, Taylor & Francis
12. Piantadosi, S. (2017), *Clinical Trials*, 3rd Ed., Wiley
13. Rao, P. S. R. S. (2017). *Statistical Methodologies with Medical Applications*, Wiley
14. Rosner, B. (2010). *Fundamentals of Biostatistics*, Harvard
15. Selvin, S. (2004). *Statistical Analysis of Epidemiological Data*, 3rd Ed., OUP.
16. Zhou, X. H., Obuchowski, N. A., and McClish, D. K. (2011). *Statistical Methods in Diagnostic Medicines*, 2nd Ed., Wiley

ST O17 Design and 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	Apply

Bioequivalence of more than two drugs

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 collection systems 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. ICH E9 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.

ST O18 Statistical Methods in Micro-array Data Analysis

Course Outcome (CO)	Cognitive level
After completion of this course the students will be able to	
1. understand the experimental set up of microarray and quantify the information available, including normalization	Understand Apply
2. choose appropriate inference procedures associated with microarray data and apply them for the analysis	Understand and Apply/Analyze
3. formulate and implement multiple testing problems (FDR etc.) associated with microarray data	Understand and Apply/Analyze
4. carry out concordance/discordance analysis for microarray data	Apply/Analyze
5. apply PCA methods for microarray data	Apply/Analyze
6. apply cluster analysis methods for clustering microarray data	Apply/Analyze

Unit I

Microarrays and normalization techniques, introduction to Biology relevant to microarray experiment, microarray experimental set up and quantification of information available from microarray experiments, data cleaning, transformation of data , Between array & within array normalization, quantile and LOWESS normalization, stage wise normalization, concordance coefficient and its role in normalization

Unit II

Statistical inference procedures in comparative experiments, inference procedures for single channel microarray data, application of two sample t–test, Tests for validating assumptions of two sample t-test. Application of Welch testand Wilcoxon rank sum test, Inference procedures for two channel microarray data. Paired t –test, Tests for validating assumptions of paired t-test. Wilcoxon signed rank test, Comparison of more than two types of mRNA samples in single channel or two channel microarray experiments. One way ANOVA F test, one way ANOVA Welch F test, Kruskal-Wallis test, pairwise t-test, pairwise Welch test and pairwise Wilcoxon rank sum test, Strip charts and its role to decide the profile of differentially expressed genes

Unit III

Multiple hypotheses testing and principal component analysis, adjustments for multiple hypotheses testing, adjusted p-values, false discovery rate and its application to microarray data analysis. Principal component analysis for microarray data, scree plot, plot of scores to rectangular matrix and the concept of ballot, its application to microarray

Unit IV

Cluster analysis and logistic regression, hierarchical cluster analysis of microarray data, K - means cluster analysis of microarray data, application of logistic regression for microarray data, concept of AIC and BIC and its role to identify marker genes

Books Recommended

1. Amartunga D. and Cabrera J. (2004). Exploration and Analysis of DNA Microarray and Protein Array Data. Wiley.
2. Deshmukh S.R. and Purohit S.G. (2007). *Microarray Data: Statistical Analysis Using R*, Narosa.
3. Draghici, S. (2003). *Data Analysis Tools for DNA Microarrays*, Chapman and Hall/CRC.
4. Dov, S. (2003). *Microarray Bioinformatics*, Cambridge University Press,
5. McLachlan, G.J.; Do, K.A. and Ambroise, C. (2004). *Analyzing Microarray Gene Expression Data*, Wiley.
6. Simon, R.M ;Korn, E.L. ; McShane, L.M. ; Radmacher, M.D. ; Wright, G.W. and Zhao, y. (2003). *Design and Analysis of DNA Microarray Investigations*. Springer.
7. Speed, T. (2003). *Statistical Analysis of Gene Expression Microarray Data*, Chapman and Hall/CRC.

ST O19 Statistical Methods for Bio-computing

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand different types of genetic data	Understand
2. understand the concept of entropy, different entropy measures and their applications in biological data analysis	Understand and Evaluate
3. align different biological sequences using various algorithms	Apply/Analyze
4. carry out molecular phylogeny analysis using different methods	Understand and Apply/Analyze
5. use of Markov chain algorithms fordemarcation of a region in Biological sequence analysis	Understand and Apply/Analyze
6. apply different algorithms such as Verterbi's algorithm, Forward and backward algorithm, Baum – Welch algorithm etc. for biological sequence analysis (use of HMM)	Apply Analyze

Unit I

Type of genetic data, differences and advantages of molecular data on morphological data, Character data and distance data, their relative merits and demerits, concept of entropy, entropy as a measure of uncertainty, entropy of single and combined scheme/s, measure of information content based on entropy, relative entropy its similarity with likelihood ratio. Applications of these to biological sequences

Unit II

Alignment of biological sequences: Pairwise and local alignment of biological Sequences (DNA/protein sequences). How biological sequences are different from mathematical sequences? The scoring matrices for alignment algorithms PAM and BLOSUM matrices, algorithm for global alignment (Needleman–Wunch algorithm), Local alignment algorithms (Smith - Waterman) Gap Model, dynamic programming algorithms for alignment with gaps such as linear gap model, affine gap model. Introduction to heuristic alignment algorithms such as BLAST, FASTA

Unit III

Molecular phylogeny Analysis: Tree of life, gene and species tree. Distance based methods for reconstruction of phylogenetic tree such as UPGMA, weighted UPGMA, transformed distance method, nearest – neighbor joining method. Comparison of trees generated using different distance function Requisites of a good distance function. Character based methods for molecular phylogeny, maximum likelihood method and maximum parsimony method. Assessing trees via bootstrap, probabilistic approach to phylogeny. Probabilistic models of evolution, Felsentein's algorithm for likelihood computation. Juke – Canter model and Kimura and other probabilistic models for evolution.

Unit IV

Applications of HMM to biological sequence analysis, Markov chain as a classifier, use of Markov chain Model for demarcation of a region in Biological sequence analysis, Application of these in genetic sequence analysis such as detection of CPG Island. Testing whether given stretch of sequence is coming from CPG Island (use of Markov model for discrimination) Markov model based classification & clusterization, testing order of a Markov model, testing homogeneity of two Markov models, Use of these test to design clustering algorithm. Hidden Markov/chains, Difference between these and simple Markov chains, Analysis of Hidden Markov Models/chains, Verterbi's algorithm, forward and backward algorithm for hidden Markov model, parameter estimation in hidden Markov model when path is known as well as unknown, Baum – Welch algorithm.

Recommended Books

1. Durbin R., Eddy S. Krogh A. Michelson G. (1998). *Biological Sequence Analysis*, Cambridge University Press.
2. Haubold, B. and Thomas, W. (2006). *Introduction to Computational Biology*, Birkhauser
3. Isaev, A. (2006). *Introduction to Mathematical Methods Bioinformatics*. Springer.
4. Kelly, S. T. and Didulo,D. (2018). *Computational Biology: A Hyper Textbook*, Wiley
5. Robin S., Rudolph F, Schboth S. (2003) *DNA Words and Models Statistics of Exceptional Words*, Cambridge University Press
6. Satpathy, R., Choudhury, T., Satpathy, S., Mohanty, S. N. and Zhang, X. (Editor), (2021).*Data Analytics in Bioinformatics: A Machine Learning Perspective*, Wiley

ST O20 Bayesian Inference

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 and Evaluate
2. understand the concepts related to loss functions, posterior loss and 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 and solve related problem	Understand and Evaluate
5. understand the asymptotics related to posterior distribution	Understand
6. compute estimators using Bayesian computing techniques such as 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 HPD confidence intervals, testing, credible intervals, prediction of a future observation

Unit II

Bayesian analysis with subjective prior, robustness and sensitivity, classes of priors, conjugate class, neighborhood class, density ratio class, different methods of construction of objective 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

ST O21 Expert Systems with Applications

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts of fuzzy theory and learn how it can be used for various decision making problems	Understand and Evaluate
2. understand the concepts of GA and apply them for various optimization problems	Understand and Apply
3. understand various concepts related to neural network analysis and apply them for different data sets	Understand and Apply

Unit I

Introduction of expert system, fuzzy sets, operations on fuzzy sets, fuzzy arithmetic and relations, fuzzy logic and controllers, multi criteria decision making, multi attribute decision making, fuzzy decision making, applications of fuzzy set theory in reliability

Unit II

Introduction to genetic algorithm (GA), Binary GA, Real coded GA, parameters of GA, different types of crossover operators, multi objective GA, Pareto solutions, Elitist and non-elitist solutions, non-dominated sorting GA

Unit III

Constrained multi-objective GA, Modified GA, Convergence of GA, Elitist and non-elitist GA under constrained optimization

Unit IV

Introduction to Neural network, problem of network learning, multi-layer networks and back-propogation, Hebbian learning, Statistical aspects of learning, VC dimensions, Gauss-Newton Method. Hybridization of fuzzy, GA and NN with real world case studies.

Recommended books

1. Buckley, W. S. J. (2004). *Fuzzy Expert Systems and Fuzzy Reasoning*, Wiley.
2. Deb, K. (2001). *Multi-objective Optimization Using Evolutionary Algorithms*, Wiley
3. Donald Waterman. (1986). *A Guide to Expert Systems*. Pearson India.
4. Goldberg, D. E. (1989). *Genetic Algorithms*, Addison-Wesley Professional
5. Klir, G. J. and Yuan, Bo. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Application*, Prentice Hall
6. Langley, P. (1995). *Elements of Machine Learning, Morgan Kaufmann Series in Machine Learning*
7. Mitchell, T. (1997). *Machine Learning*, McGraw-Hill
8. Nikolopoulose, C. (2019). *Expert Systems: Introduction to First and Second Generation and Hybrid Knowledge Based Systems*., CRC Press

ST O22 Financial Econometrics

(This course cannot be opted simultaneously with ST O23)

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to spectral analysis of time seriesand solve related problems	Understand and Evaluate/Analyze
2. model and analyze the vector financial time series data	Analyze
3. understand the concepts related to unit root testing in time series and apply them for financial time series data	Understand and Apply/Analyze
4. understand the concepts related to Granger causality and cointegration and apply them for vector financial time series data	Understand and Apply/Analyze
5. understand the concepts related to state-space modeling (dlm) and apply them for given financial time series data	Understand and Apply/Analyze
6. use nonlinear and Markov switching models for modeling financial time series data	Understand and Apply/Analyze
7. understand the theory related to SV models and apply them for financial time series	Understand and Apply/Analyze

Unit I

Frequency domain analysis of time series, periodicity, spectral density, periodogram and DFT, spectral representation, inference, examples & data analysis

Vector time series, cross correlations, VAR and VARMA models, stability condition, impulse response function, causality & invertibility, estimation – least squares and MLE (Gaussian), order determination, multivariate Portmanteau tests

Granger causality, cointegration and ECM, cointegrating VAR, cointegration tests, applications to the PPP (purchasing power parity). Applications to the net present value model of stock prices, market microstructure and the efficient market hypothesis, examples & data analysis

Unit II

State-space models, State-space representations, local-trend model, The basic structural model, State-space representation of ARIMA models, Filtering and smoothing, The Kalman recursions, estimation for State-space models, generalized state-space models, parameter & observation-driven models, non-Gaussian state-space models, APM with time varying parameters, examples & data analysis

Unit III

Nonlinear time series models, modeling regime shifts in time series, Markov-switching models, estimation, Bayesian methods, diagnostic checking, Forecasting, examples & data analysis

Unit IV

Volatility definition and estimation, volatility forecast evaluation, stochastic volatility models, MCMC approach, option pricing with stochastic volatility

Recommended Books

1. Ait-Sahalia, Y. & Hansen, L. P. (Ed.) (2010). *Handbook of Financial Econometrics: Tools and Techniques*, Vol.1& 2, Elsevier, Amsterdam.
2. Campbell J., Lo A. & McKinley C. (1997). *The Econometrics of Financial Markets*. PrincetonUniversity press. Princeton.
3. Chan, N. H. (2002). *Time Series: Applications to Finance*, Wiley
4. Fan J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York.
5. Hamilton (1994). *Time Series Analysis*. Princeton University press. Princeton
6. Knight, J. and Satchell, S. (2007). *Forecasting Volatility in the Financial Markets*, Third Edition, Elsevier, Amsterdam.
7. Lutkepohl, H. and Kratzing, M. (Ed.) (2004). *Applied Time Series Econometrics*, Cambridge University Press, Cambridge
8. Poon Ser-Huang (2005). *A Practical Guide to Forecasting Financial Market Volatility*, Wiley, New York.
9. Rachev. S. T., Mittnik, S., Fabozzi, F. J. ,Focardi, S. M. and Jasic, T. (2007). *Financial Econometrics :From Basics to Advanced Modeling Techniques*, Wiley, New York.
10. Ruppert, D. (2004). *Statistics and Finance: An Introduction*, Springer, New York.
11. Shephard, N. (2004). *Stochastic Volatility: Selected Readings*, Oxford University Press, Oxford.
12. Soderlind, P. (2010). *Lecture Notes in Financial Econometrics*, University of St. Gallen, Switzerland.
13. Tsay, (2014). *Multivariate Time Series Analysis: With R and Financial Applications*, Wiley

14. Tsay, R. S. (2010). *Analysis of Financial Time Series*, Third Edition, Wiley, New York.
15. Wang, P. (2003). *Financial Econometrics: Methods and Models*, Routledge, London.

ST O23 Advanced Time Series Analysis
(This course cannot be opted simultaneously with ST O22)

Course Outcome (CO)	Cognitive level
After completion of this course the students will be able to	
1. understand the theory related to estimating functions and bootstrapping In time series and apply them	Understand and Apply
2. understand the concepts of SETAR models and apply them for modeling and forecasting time series data	Understand Apply/Analyze
3. understand the concepts related to spectral analysis of time series and solve related problems	Understand and Evaluate/Analyze
4. understand the concepts related to unit root testing in time series and apply them for financial time series data	Understand and Apply/Analyze
5. understand the concepts related to Granger causality and cointegration and apply them for the analysis of time series data	Understand and Apply/Analyze
6. understand the concepts related to INAR models, apply them for integer time series data	Understand and Apply/Analyze

Unit I
Estimating functions in time series, bootstrapping in time series, long memory time series models, examples & data analysis

SETAR models, inference and applications.

Unit II
Spectral analysis of stationary time series: Fourier analysis, Fourier representation of periodic and non-periodic series, discrete Fourier transforms, spectral theory of stationary processes, spectrum of common processes, estimation of the spectral density and spectrum, examples & data analysis

Unit III
Non-stationarity, Granger causality and co-integration: test for stationarity and unit roots, DF, ADF, PP and KPSS tests, Granger causality, cointegration and ECM, VECM, cointegrating VAR, cointegration tests, applications to the purchasing power parity (PPP), applications to the net present value model of stock prices. examples & data analysis

Unit IV
Time series models for counts: thinning-based approach, POINAR models, INAR, INARMA models, categorical time-series models, estimation of parameters and forecasting,

mode of the predictive distribution, models which do not require thinning. examples & data analysis

Recommended Books:

1. Brockwell, P.J. and Davis, R. A. (2003). *Introduction to Time Series Analysis*, Springer
2. Commandeur, J. J. F. and Koopman, S. J. (2007). *An Introduction to State Space Time Series Analysis*, Oxford University Press
3. Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd Ed. Wiley.
4. Hamilton N. Y. (1994). *Time Series Analysis*, Princeton University press.
5. Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*, Springer
6. Palma, W. (2007). *Long-Memory Time Series: Theory & Methods,* Wiley.
7. Peters, G. W. (2018). *General Quantile Time Series Regressions for Applications in Population Demographics*. Risks 2018, 6, 97; doi:10.3390/risks6030097
8. Rajarshi, M. B. (2012). *Statistical Inference for Discrete Time Stochastic Processes*, Springer.
9. Shimizu, K. (2010). *Bootstrapping Stationary ARMA-GARCH Models*, Vieweg+Teubner
10. Shumway, R. H. and Stoffer, D. S. (2010). *Time Series Analysis & Its Applications*, Springer.
11. Tsay, R. S. (2010). *Analysis of Financial Time Series*, Third Edition, Wiley, New York.
12. Weiss, C., H. (2018). *An Introduction to Discrete-Valued Time Series*, Wiley.

ST O24 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

Recommended Books:

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 of Infectious 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 in Medical 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 Health Research : 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

ST O25 Statistical Foundations of Data Science

Course Outcome (CO)	**Cognitive level**
After completion of this course the students will be able to	
1. understand the concepts related to L2 penalized/Ridge regression and apply them	Understand and Apply/Analyze
2. understand the concepts related to penalized least squares in regression models and apply them	Understand and Apply/Analyze
3. understand the concepts related to penalized least squares in generalized linear models and apply them	Understand and Apply/Analyze
4. understand the theory related to high dimensional inference in linear regression apply them to high dimensional data	Understand and Apply/Analyze
5. understand the theory related to high dimensional inference in generalized linear models and apply them	Understand and Apply/Analyze

Unit I

Review of multiple and monparametric regression: multiple linear regression, weighted least-squares, ridge regression, bias-variancetradeoff, L2-Penalized least squares, Bayesian interpretation, Ridge regression solution path, kernel ridge regression

Regression in reproducing kernel Hilbert space, leave-one-out and generalized cross-validation

Unit II

Introduction to penalized least-squares: review of classical variable selection criteria, subset selection, relation with penalized regression, selection of regularization parameters

Folded-concave penalized least squares, Lasso and L1-regularization, nonnegative garrote, Lasso, adaptive Lasso, elastic net, Dantzig selector

Unit III
Introduction to generalized linear models and penalized likelihood, Sparest solution in, high confidence set, variable selection via penalized likelihood

Unit IV
High dimensional inference, Inference in linear regression, inference in generalized linear, Gaussian graphical models

Feature screening, correlation screening, generalized and rank correlation screening, feature screening for parametric models, model-free feature screening, refitted cross-validation,

Books Recommended

1. Blum, A., Hopcroft, J. and Kannan, R. (2020). *Foundations of Data Science*, CUP.
2. Fan, J, Li, R., Zhang, C-H and Zou, H. (2020). *Statistical Foundations of Data Science*, CRC Press
3. Hastie, T. and Efron, B. (2016*). Computer Age Statistical Inference: Algorithms, Evidence and Data Science*, CUP
4. Hastie, T., Tibshirani, R. and Friedman, J. (2017). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, 2ndEdn., Springer
5. Hastie, T., Tibshirani, R., Wainwright, M. (2016). *Statistical Learning with Sparsity: The Lasso and Generalizations*, CRC Press
6. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2017). *An Introduction to Statistical Learning: With Applications in R*, Springer

Project Work in lieu of an Optional Course
Rules & Regulations

1. Students with backlogs or less than B grade in any of the courses will not be eligible for such project work.
2. Whenever a project is planned, the concerned student and the guide shall submit a one page proposal about the project they are planning to do (including for a reading course project based on advanced probability, inference or related books) to the department Head and Departmental examination committee (DEC). The Head and DEC together will sanction the project. This will be in lieu of a 4 credit optional taught course.

3. No project involving only data analysis will be allowed, however good they are. Theory and problem solving should be part of a project.
4. In a project based on reading a book, i.e., reading course on advanced probability, inference or related books, the candidate is expected to read at least 5 to 6 chapters of the book proposed in the proposal and solve as many problems from the exercises of that book. The final report should consist of theory learned, problems and solutions worked out.
5. In a project based on reading research papers (but no data analysis), the candidate is expected to read at least two research articles (appeared in Scopus reviewed journals) and prepare the project report based on them.
6. There shall be two internal assessments for the project. The candidate shall submit a report about the work done until then, including the literature review from at least two Scopus reviewed journals (about 5 to 10 pages) for the first evaluation. For the second evaluation, the candidate is expected to present the analysis and preliminary findings of the project. For the second evaluation also the candidate should submit a write up of the work done (until then). A presentation has to be arranged, in which the guide should evaluate the candidate during both the occasions. The candidate may be evaluated out of 25 by the guide at both these occasions. (CIA = 50)
7. The candidate shall submit the final report of the project one week prior to the ETE. The project report will be subjected to the plagiarism test. If there is a similarity of more than 50 per cent, the project report will be rejected and no resubmission will be allowed. However, if the similarity is between 20 and 50 percent, it should be reworked and resubmitted by the candidate with a similarity less than or equal to 20 per cent.
8. It is the responsibility of the Head and DEC to conduct this test and take the decision. The student should submit a soft copy of the report to Head or DEC. The ETE of project will be in the form of a viva voce and presentation. This will be evaluated out of 50 by the external examiner (Report or Presentation-25, Viva-25).
9. The Head in consultation with DEC will appoint an external examiner to examine the project report. The examiner of the project should read the project before the viva and presentation and examine the details of the findings and assess the learning outcome of the project.

Service Courses

The syllabus of two service courses is given below. These service courses are meant for students from other departments in the campus and will be offered only subject to the availability of the faculty. These courses are of 2 credits each.

ST S1: Introduction to Statistics (Service Course) – 2 Credits

Course Outcome (CO)	Cognitive level
After completion of this course the students will be able to	
1. understand the basic concepts related to different types of data	Understand and
2. make graphical display of different types of using different visualization methods	Apply
3. model univariate data using simple distributions	Apply
4. carry out inference about the population using sample data with the help of simple statistical inference procedures	Apply
5. carry out regression analysis of the given data and interpret the results	Apply

Unit I: Data Structures and Data Summarization

Concepts of population and sample, methods of data collection, types of data (nominal, ordinal, interval, numeric, categorical etc)

Data summary and descriptive measures, measures of central tendency, measures of dispersion, measures of skewness and kurtosis, measures of association between two or more variables

Graphical display of data (visualization), different types of charts, histogram, Pie-charts, P-P and Q-Q plots, box plots, scatter plots.

Unit II: Modeling Univariate Data

Concept of randomness, random variables, data as a random sample, interrelations between summary measures of random variables and those of the observed data

Models for the data: Concept of probability and probability distribution of a random variable, discrete and continuous probability distributions, brief introduction to some well-known probability distributions (Discrete: Bernoulli, Binomial, Poisson; Continuous:

Uniform, Exponential, Gaussian), applications and discussions of normal (Gaussian) distribution
Generating data from the probability distributions

Unit III: Inferences about the Population Using Samples

Introduction to statistical inference, concept of parameter and parameter estimation, Estimation of parameters-distribution based and distribution free methods

Formal assessment of the proposed probability distributions for the data, goodness of fit tests, Chi-square tests

Concept of statistical hypothesis testing, test statistic, p-value, level of significance, false positives and false negatives in testing

The principle of approximation for large samples: Is everything "normal" in large samples?
Conclusions about the population characteristics based on sample attributes, some commonly used large sample and small sample tests for population descriptive measures such as mean and variance, Z-test, t-test, F-test, Introduction to Analysis of variance.

Unit IV: Regression Modeling and Model Selection

Causal associations among variables, Features of supervised statistical learning, evaluating the dependence of a variable on independent variables, principles and assumptions of regression modeling, philosophy and implementation of least-square estimation
Simple and multiple regressions, determining the subset of important variables, forward selection and backward elimination methods, model diagnostics, residual analysis, model evaluation

Recommended Books:

1. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer
2. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). *Introduction to Linear Regression Analysis*, Wiley.
3. Purohit, S.G, Gore, S.D. and Deshmukh, S.R. (2008). *Statistics Using R*. Narosa Publishing House, India. (2008)
4. Ross, S. M. (2014). *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press.
5. Shaefer, S. J. and Theodore, L. (2007). *Probability and Statistics Applications for Environmental Science*, CRC Press
6. Walpole, R. E., Myers, R. H., Myers, S. L. and Ye, K. (2012). *Probability and Statistics for Engineers and Scientists*, 9th Edition, Prentice Hall.
7. Wasserman, L. (2013). *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & Business Media.

ST S2: Advanced Statistical Analysis (Service Course) – 2 Credits

Course Outcome (CO)	Cognitive level
After completion of this course the students will be able to	
1. understand the concepts related to survey sampling and apply them for sampling problems	Understand and Apply
2. understand the concepts related to experimental design and apply them for designing scientific experiments	Apply
3. understand the analysis methods related to categorical and count data with covariates	Understand Apply
4. analyze multivariate data using different methods	Apply
5. analyze time series data using different methods	Apply

Unit I: Survey Sampling Techniques

Concepts of population and sample, sample survey prerequisites, probability sampling, simple random sampling, stratified sampling and multi-stage sampling, sample size determination, sampling and non-sampling errors, analysis of survey data

Unit II: Analysis of Variance & Covariance, Design of Experiments

One-way ANOVA, two-way ANOVA, introduction to analysis of covariance
Basic principles of experimental designs, completely randomized designs, randomized block designs, factorial designs

Unit III: Non normal Regression

Review of normal regression analysis (multiple),

Regression models for binary, categorical and count data, analysis and diagnostics

Unit IV: Multivariate Data Analysis

Multivariate data, exploratory analysis and multivariate plots, principal component analysis, factor analysis and cluster analysis.

Unit V: Introduction to Time Series Analysis

Time series, plots, decomposition of time series, time series regression, non-parametric smoothing, exponential smoothing, Holt-Winter seasonal forecasting

Introduction to time series models, AR, MA, ARMA and ARIMA models

Recommended Books:

1. Bryan F. J. Manly, Jorge A. Navarro Alberto (2016). *Multivariate Statistical Methods: A Primer* (4th Edition) , Chapman & Hall/CRC
2. Chatfield, C. (1995). *The Analysis of Time Series: An Introduction*, Chapman & Hall
3. James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*, Springer
4. Jonson, R. A. and Wichern, D. W.(2013). *Applied Multivariate Statistical Analysis*. PHI
5. Montgomery, D. C., (2010). *Design and Analysis of Experiments*, Fifth Edition, Wiley.
6. Montgomery, D. C., Peck, E. A. and Vining, G. G. (2003). *Introduction to Linear Regression Analysis*, Wiley.
7. Ross, S. M. (2014). *Introduction to Probability and Statistics for Engineers and Scientists*. Academic Press
8. Walpole, R. E., Myers, R. H., Myers, S. L. and Ye, K. (2012). *Probability & Statistics for Engineers & Scientists*, 9th Edition, Prentice Hall.
9. Wasserman, L. (2013). *All of Statistics: A Concise Course in Statistical Inference*. Springer Science & Business Media.
10. Wayne Daniel (1991). *Biostatistics: A foundation for analysis in Health Science*. Wiley.