Learning Outcome Based Curriculum of

2 Years-Master in Statistics (M.A./M.Sc)

अधिगम प्रतिफल आधारित

2. वर्षीय मास्टर इन सांख्यिकी

सांख्यिकी विभाग
विज्ञान संकाय
इलाहाबाद विश्वविद्यालय, प्रायागराज.211002

Department of Statistics
Faculty of Science
University of Allahabad, Prayagraj-211002
(Session 2025-2026 onwards)
(Revised in the light of NEP- 2020)

Prologue

About the Department:

The Statistics wing, under the Department of Mathematics, was started in 1968 with the initiative of Prof. H. C. Khare, then acting Head of the Department of Mathematics. The wing became operational immediately, with the first batch of B.Sc. I and M.Sc. previous students enrolling simultaneously. Soon, Mr. C. C. Pant was appointed in the Wing, and he primarily taught Statistics courses; other courses such as Numerical analysis and Econometrics were taught by faculty members from the Departments of Mathematics and Economics. In 1974, Dr. Samir K. Bhattacharya, Dr. Ram C. Tiwari and Dr. G. S. Pandey were appointed as regular teachers of Statistics and the Statistics wing continued to operate under the leadership of Prof. Bhattacharya. Professor Bhattacharya was the first Professor of Statistics, who became the Head of combined Department of Mathematics and Statistics in the year 1989 and continued till 1995 (till his death). In 1981, Dr. Anoop Chaturvedi joined as a faculty member of the Department, followed by Dr. S. Lalitha in 1986. Mr. Vishnu Kumar Srivastava, Mr. R.N. Verma, Mr. Raj Kumar Srivastava, Dr. Ajit Chaturvedi and Dr. Anup Kumar also served the department for a shorter period.

On August 21, 2000, the Department was established as a separate entity from Mathematics, with Dr. G. S. Pandey appointed as the first Head of Department, alongside Dr. Anoop Chaturvedi and Dr. S. Lalitha as faculty members. The department also attained the research calibre by the effort of Prof. Anoop Chaturvedi.
Many research projects including
(i) Control Charts for Auto correlated Observations on Grid
(ii) Sustainable Development: A Mathematical Model
(iii) Joint Big Data project could be successfully completed.

The Department offers courses for Undergraduate, Postgraduate degrees, and a Doctoral program leading to a Ph.D. degree. The UG and PG students are provided a broad foundation and training in statistical theory and applications. Such training has enabled them to earn jobs in government including civil services, Indian Statistical Services etc., all branches of industry, and academia. The Department has a well-equipped computer laboratory with modern computing facilities. The students are provided rigorous training of statistical computing and various statistical software. Presently, this is the first time in history of the Department when there are six faculty members with specialization in different areas of statistics such as Sampling Techniques, Reliability Theory, Bayesian Statistics, Order Statistics, Quality Control, Sequential Analysis, Survival Analysis, Mathematical Demography, Biostatistics and Stochastic Modelling. The Department also provides statistical support to various interdisciplinary branches as per their research needs.

Learning Outcomes Descriptors and Programme Outcomes for a Master's degree (e.g. M.A., M.Sc., etc.) (Level 6.5)

Master's degree (1 year/2 semesters of study): The Master's degree qualifies students who can apply an advanced body of knowledge in a range of contexts for professional practice, research, and scholarship and as a pathway for further learning. Graduates at this level are expected to possess and demonstrate specialized knowledge and skills for research, and/or professional practice and/or for further learning.

Master's Degree (2 years /4 semesters of study): The Master's degree qualifies students who can apply an advanced body of knowledge in a range of contexts for professional practice, research, and scholarship and as a pathway for further learning. Graduates at this level are expected to possess and demonstrate specialized knowledge and skills for research, and/or professional practice and/or for further learning. Master's degree holders are expected to demonstrate the ability to apply the established principles and theories to a body of knowledge or an area of professional practice.

Descriptor	**Knowledge and understanding:** The graduates should be able to demonstrate the acquisition of:
P.O. 1.	advanced knowledge about a specialized field of enquiry with a critical understanding of the emerging developments and issues relating to one or more fields of learning,
P.O. 2.	advanced knowledge and understanding of the research principles, methods, and techniques applicable to the chosen field(s) of learning or professional practice,
P.O. 3.	procedural knowledge required for performing and accomplishing complex and specialized and professional tasks relating to teaching, and research and development.
Descriptor	**General, technical and professional skills required to perform and accomplish tasks:** The graduates should be able to demonstrate the acquisition of:
P.O. 4.	advanced cognitive and technical skills required for performing and accomplishing complex tasks related to the chosen fields of learning.
P.O. 5.	advanced cognitive and technical skills required for evaluating research findings and designing and conducting relevant research that contributes to the generation of new knowledge.
P.O. 6.	specialized cognitive and technical skills relating to a body of knowledge and practice to analyze and synthesize complex information and problems.
Descriptor	**Application of knowledge and skills:** The graduates should be able to demonstrate the ability to:
P.O. 7.	apply the acquired advanced theoretical and/or technical knowledge about a specialized field of enquiry or professional practice and a range of cognitive and practical skills to identify and analyze problems and issues, including real-life problems, associated with the chosen fields of

	learning.
P.O. 8.	apply advanced knowledge relating to research methods to carry out research and investigations to formulate evidence-based solutions to complex and unpredictable problems.
Descriptor	**Generic learning outcomes:** The graduates should be able to demonstrate the ability to:
P.O. 9.	listen carefully, read texts and research papers analytically and present complex information in a clear and concise manner to different groups/audiences,
P.O. 10.	communicate, in a well-structured manner, technical information and explanations, and the findings/results of the research studies undertaken in the chosen field of study,
P.O. 11.	present in a concise manner view on the relevance and applications of the findings of recent research and evaluation studies in the context of emerging developments and issues.
P.O. 12.	evaluate the reliability and relevance of evidence; identify logical flaws and holes in the arguments of others; analyze and synthesize data from a variety of sources; draw valid conclusions and support them with evidence and examples, and addressing opposing viewpoints.
P.O. 13.	meet one's own learning needs relating to the chosen fields of learning, work/vocation, and an area of professional practice,
P.O. 14.	pursue self-paced and self-directed learning to upgrade knowledge and skills, including research-related skills, required to pursue a higher level of education and research.
P.O. 15.	problematize, synthesize, and articulate issues and design research proposals,
P.O. 16.	define problems, formulate appropriate and relevant research questions, formulate hypotheses, test hypotheses using quantitative and qualitative data, establish hypotheses, make inferences based on the analysis and interpretation of data, and predict cause-and-effect relationships,
P.O. 17.	develop appropriate tools for data collection for research,
P.O. 18.	the ability to use appropriate statistical and other analytical tools and techniques for the analysis of data collected for research and evaluation studies,
P.O. 19.	plan, execute, and report the results of an investigation,
P.O. 20.	follow basic research ethics and skills in practicing/doing ethics in thefield/ in one's own research work.
P.O. 21.	make judgements and take decisions regarding the adoption of approaches to solving problems, including real-life problems, based on the analysis and evaluation of information and empirical evidence collected.
P.O. 22.	make judgement across a range of functions requiring the exercise of full responsibility and accountability for personal and/or group actions to

	generate solutions to specific problems associated with the chosen fields/subfields of study, work, or professional practice.
Descriptor	**Constitutional, humanistic, ethical, and moral values:** The graduates should be able to demonstrate the willingness and ability to:
P.O. 23.	embrace and practice constitutional, humanistic, ethical, and moral values in one's life,
P.O. 24.	adopt objective and unbiased actions in all aspects of work related tothe chosen fields/subfields of study and professional practice,
P.O. 25.	participate in actions to address environmental protection and sustainable development issues,
P.O. 26.	support relevant ethical and moral issues by formulating and presenting coherent arguments,
P.O. 27.	follow ethical principles and practices in all aspects of research and development, including inducements for enrolling participants, avoiding unethical practices such as fabrication, falsification or misrepresentation of data or committing plagiarism.
Descriptor	**Employability and job-ready skills, and entrepreneurship skills and capabilities/ qualities and mindset:** The graduates should be able to demonstrate the acquisition of knowledge and skill sets required for:
P.O. 28.	adapting to the future of work and responding to the demands ofthe fast pace of technological developments and innovations that drive the shift in employers' demands for skills, particularly with respect to t h e transition towards more technology-assisted work involving the creation of new forms of work and rapidly changing work and production processes.
P.O. 29.	exercising full personal responsibility for the output of own work as well as for group/team outputs and for managing work that is complex and unpredictable requiring new strategic approaches.

<u>PSO for M.Sc. Statistics</u>

PSO1: Advanced Theoretical Foundation
Build deep understanding of advanced statistical theory — including measure-theoretic probability, estimation theory, hypothesis testing, multivariate methods, time series, nonparametric inference — to a level sufficient for research or teaching.

PSO2: Mathematical & Algebraic Skills
Use matrix algebra, linear models, optimization, and related mathematical tools effectively in statistical modeling and inference.

PSO3: Applied Statistical Methodology
Design and apply statistical methods (regression, ANOVA, sampling, survey techniques,

design of experiments) in real-world settings across domains (e.g. social sciences, biology, economics).

PSO4: Computational & Software Proficiency
Use statistical software (e.g. R, Python, SPSS, etc.) and programming skills to implement algorithms, carry out simulations, and analyze large datasets.

PSO5: Project & Research Competence
Conduct independent research (minor and major project work), including problem formulation, data collection, analysis, interpretation, and reporting of results.

PSO6: Critical Thinking & Problem Solving
Critically evaluate assumptions underlying statistical models, identify model limitations, propose improvements or alternate approaches.

PSO7: Communication & Collaboration
Communicate statistical results effectively (both orally and in writing) to varied audiences (statisticians, domain experts, laypersons), and work collaboratively in multidisciplinary teams.

PSO8: Ethical & Responsible Data Use
Appreciate issues of ethics, privacy, and responsibility in data collection, analysis, and reporting; ensure reproducibility, transparency, and integrity in statistical practice.

PSO9: Lifelong Learning & Adaptability
Acquire the capacity to keep up with evolving statistical/machine learning techniques and adapt to new trends, technologies, and domains.

PSO10: Readiness for Further Studies & Careers
Prepare students for doctoral work (PhD in Statistics or allied disciplines), or for professional roles as statisticians, data scientists, analysts in industry, government, academic or research organizations.

DEPARTMENT OF STATISTICS
FACULTY OF SCIENCE
UNIVERSITY OF ALLAHABAD, PRAYAGRAJ

(w.e.f. 2025-26 onwards)

(Revised in the light of NEP-2020)

PROGRAMME STRUCTURE FOR MA/M.Sc.

1st Semester July to December

Paper	Duration L-T-P-C	Maximum Marks
STA501: Matrix Algebra and Linear Estimation	4-1-0-5	100
STA503: Survey Sampling	4-1-0-5	100
STA504: Measure and Probability Theory	4-1-0-5	100
Elective (Chose any one)		
Elective I: STA551: R Programming and Statistical Computing	4-1-0-5	100
Elective II: STA669: Statistical Computing using Python	4-1-0-5	100
STA531: Lab	0-1-5-5	100

2nd Semester January to May

Paper	Duration L-T-P-C	Maximum Marks
STA502: Multivariate Analysis	4-1-0-5	100
STA562: ADVANCED ESTIMATION THEORY	4-1-0-5	100
STA 507: Regression Analysis	4-1-0-5	100
Elective (Chose any one)		
Elective I: STA506: Time series analysis	4-1-0-5	100
Elective II: STA561: Population Studies and Demography	4-1-0-5	100
STA532: LAB	0-1-5-5	100

3rd Semester July to December

Paper	Duration L-T-P-C	Maximum Marks
STA552: Knowledge discovery and data mining	4-1-0-5	100
STA601: Testing of hypothesis, Sequential Analysis AND NON-PARAMETRIC INFERENCE	4-1-0-5	100
STA670: Survival analysis and Clinical Trials	4-1-0-5	100
Elective (Chose any one)		
Elective I: STA653: Official and National	4-1-0-5	100

Development Statistics		
Elective II: STA654: Advanced Multivariate Analysis	4-1-0-5	100
ELECTIVE III: STA655: ECONOMETRICS	4-1-0-5	100
STA631: LAB	0-1-5-5	100

4th Semester January to May

Paper	Duration L-T-P-C	Maximum Marks
STA602: Analysis of Variance and Design of Experiments	4-1-0-5	100
STA656: Stochastic Process	4-1-0-5	100
Elective (Chose any one)	4-1-0-5	100
Elective I: STA652: Statistical Decision Theory and Bayesian theory	4-1-0-5	100
Elective II: STA671: Advanced Operation Research	4-1-0-5	100
Elective II: STA672: Biostatistics	4-1-0-5	100
Elective II: STA673: Forestry and Environmental Statistics	4-1-0-5	100
STA633: LAB	0-1-5-5	100
STA634: Major project	5 credit	100

DEPARTMENT OF STATISTICS
FACULTY OF SCIENCE
UNIVERSITY OF ALLAHABAD, PRAYAGRAJ
2 Years-Master in Statistics (M.A./M.Sc.)

Course Structure at Glance

	Semester I	Semester II	Semester III	Semester IV	Total Credits
Core Course	STA501: MATRIX ALGEBRA AND LINEAR ESTIMATION	STA502: MULTIVARIATE ANALYSIS	STA552: KNOWLEDGE DISCOVERY AND DATA MINING	STA602: ANALYSIS OF VARIANCE AND DESIGN OF EXPERIMENTS	
	STA503: SURVEY SAMPLING	STA562: ADVANCED ESTIMATION THEORY	STA601: TESTING OF HYPOTHESIS, SEQUENTIAL ANALYSIS AND NON-PARAMETRIC INFERENCE	STA656: STOCHASTIC PROCESS	
	STA504: MEASURE AND PROBABILITY THEORY	STA 507: REGRESSION ANALYSIS	STA670: SURVIVAL ANALYSIS AND CLINICAL TRIALS		
Core Elective	STA551: R PROGRAMMING AND STATISTICAL COMPUTING	STA506: TIME SERIES ANALYSIS	STA653: OFFICIAL AND NATIONAL DEVELOPMENT STATISTICS	STA652: STATISTICAL DECISION THEORY AND BAYESIAN THEORY	
	STA669: STATISTICAL COMPUTING USING PYTHON	POPULATION STUDIES AND DEMOGRAPHY	STA654: ADVANCED MULTIVARIATE ANALYSIS	STA671: ADVANCED OPERATION RESEARCH	
			STA655: ECONOMETRICS	STA672: BIOSTATISTICS	
				STA673: FORESTRY AND ENVIRONMENTAL STATISTICS	

First Semester

Core Courses

Paper	Duration L-T-P-C	Maximum Marks
STA501: MATRIX ALGEBRA AND LINEAR ESTIMATION	4-1-0-5	100
STA503: SURVEY SAMPLING	4-1-0-5	100
STA504: MEASURE AND PROBABILITY THEORY	4-1-0-5	100

(STA501) Matrix Algebra and Linear Estimation

Course Objectives:

1. To develop a strong foundation in matrix operations and algebraic properties essential for understanding advanced topics in statistics and data analysis.
2. To familiarize students with different types of matrices such as symmetric, orthogonal, idempotent, and their roles in statistical modeling.
3. To understand the computation and application of matrix inverses, partitioned inverses, and generalized inverses (g-inverses) in solving linear systems and statistical estimation problems.
4. To study eigenvalues, eigenvectors, and the Cayley–Hamilton theorem and their applications in multivariate analysis, principal component analysis, and canonical correlations.
5. To apply matrix theory in formulating and solving linear models, regression equations, and analysis of variance (ANOVA) problems.

Course Learning Outcome:

After completion of this course the students will be able to

CO1. Students will be able **to** apply matrix algebra concepts including partitioned inverses, generalized inverses, orthogonal and idempotent matrices, and characteristic roots

CO2. able to analyze and classify quadratic forms as definite, semi-definite, or indefinite, perform simultaneous reduction of two quadratic forms, and apply the properties of similar matrices in simplifying and interpreting multivariate statistical models.

CO3. Discriminate between diagonalizable and non-diagonalizable matrices; orthogonally diagonalizable symmetric matrices and quadratic forms

CO4. Combine methods of matrix algebra to compose the change-of basis matrix with respect to two bases of a vector space, identify linear transformations of finite dimensional vector spaces and compose their matrices in specific bases

CO5. They will get the knowledge of building and fitting linear regression models with software.

Mapping of Course outcomes:

Course Outcomes	PO1	PO4	PO7	PO16	PO29	PSO1	PSO2	PSO3	PSO4
CO1			✓	✓	✓	✓	✓	✓	✓
CO2		✓	✓			✓	✓	✓	✓
CO3	✓	✓				✓	✓	✓	✓
CO4	✓					✓	✓	✓	✓
CO5	✓								

Course Content

STA 501C: Matrix Algebra and Linear Estimation

UNIT I Matrix Theory- Inverse of partitioned matrices, g-inverse, orthogonal matrices, properties of idempotent matrices, characteristic roots and vectors, Cayley-Hamilton theorem.

UNIT II Quadratic forms, definite, semi-definite and indefinite forms, simultaneous reduction of two quadratic forms, properties of similar matrices.

UNIT III General linear model, assumptions, estimation of parameters by least squares, estimable functions, error and estimation space, Gauss-Markov theorem, use of g-inverse.

UNIT IV Distribution of quadratic form and its application in analysis of variance model, Estimable linear hypothesis, generalized F and t tests.

UNIT V Linear regression model, Random and mixed effect models, estimation of variance component in one way and two-way random effects model.

Books Recommended:

1. Draper, N.R. and Smith H. (1998); Applied Regression Analysis, 3rd Ed., Wiley.

2. Graybill, F.A. (1983). Matrices with Applications in Statistics, 2nd Ed., Wadsworth.

3. Joshi, D.D. (1987). Linear Estimation and Design of Experiments, Wiley Eastern.

4. Rao, C.R. and Mitra, S.K. (1971). Generalized Inverse of Matrices and its Application, John Wiley and Sons Inc.

5. Searle, S.R. (1982). Matrix Algebra for Statistical Applications, John Wiley and Sons inc.

6. Seber , George A. F. and Lee Alan J. (2003). Linear Regression Analysis, Wiley.

7. Shanti Narain. A text book of matrices, S. Chand and Company (Pvt.) Ltd.

(STA501) SURVEY SAMPLING

Course Objectives:

1. To provide an in-depth understanding of varying probability sampling techniques with and without replacement.
2. To familiarize students with key estimators such as the Horvitz-Thompson estimator and their properties.
3. To explore advanced estimation strategies including Desraj and unordered estimators.
4. To introduce complex sampling schemes such as those by Midzuno-Sen and Narain.
5. To understand stratification techniques including deep and post-stratification, and the use of double and sub-sampling in estimation.
6. To examine the impact of non-sampling errors and introduce students to methods like randomized response and ranked set sampling for sensitive surveys.

Syllabus

Unit	Course Content	Hours of Teaching
I	Varying probability sampling with and without replacement, cumulative total and Lahiri's methods of selection, Estimation of population mean, Desraj ordered estimates, unordered estimators.	
II	Horvitz-Thompson estimator, its variance and unbiased estimator of variance and Yates- Grundy modification, schemes of sampling due to Midzuno-Sen, Narain.	
III	Deep stratification, Post Stratification, Double sampling in ratio estimation.	
IV	Double sampling in regression estimation, sub sampling.	
V	Non-sampling errors, Randomized Response techniques (Warner's model: related and unrelated questionnaire methods), Ranked set sampling.	

Texts / Reference Books

1. Chaudhuri, A. (2010). Essentials of Survey Sampling, Prentice Hall of India.
2. Chaudhuri, A. and Vos, J.W.E. (1988). Unified Theory of Strategies of Survey Sampling, North Holland, Amsterdam.

3. Hedayat, A. S. and Sinha, B.K. (1991). Design and Inference in Finite Sampling, Wiley.

4. Murthy, M.N. (1967). Sampling Theory and Methods. Statistical Publishing Society, Kolkata.

5. Mukhopadhyay, P. (1996). Inferential Problems in Survey Sampling, .New Age International.

6. Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C. (1984). Sampling Theory of Surveys with Applications, Iowa State University Press and Indian Society of Agricultural Statistics.

7. Cochran, W.G. (1977). Sampling Techniques, Third Edition, Wiley.

8. Sarndal, C.E., Swensson and Wretman (1977). Model Assisted Survey Sampling. Springer.

9. Singh, D. and Chuddar, F. S. (1986). Theory and Analysis of Sample Survey Designs. New Age International.

10. Sarjinder Singh (2003). Advanced Sampling Theory with Application. Kluwer Academic Publishers.

Learning Outcomes:

Upon successful completion of this course, students will be able to:

1. Apply varying probability sampling methods both with and without replacement using cumulative total and Lahiri's selection methods.
2. Compute and evaluate population mean estimates using Desraj ordered estimators, unordered estimators and Horvitz-Thompson estimator.
3. Understand and implement complex sampling schemes including those due to Midzuno-Sen and Narain.
4. Distinguish and apply strategies for deep stratification, post-stratification, and double sampling in both ratio and regression estimation contexts.
5. Analyze subsampling techniques and assess their suitability for various survey contexts.

6. Identify and address non-sampling errors in surveys, and effectively implement randomized response techniques such as Warner's models (related and unrelated questions).

Course Outcomes (COs)	PO1	PO3	PO4	PO7	PO16	PSO1	PSO2	PSO3	PSO4	PSO5
CO1	✓		✓	✓		✓	✓	✓		
CO2		✓	✓	✓	✓	✓	✓	✓	✓	
CO3	✓		✓	✓		✓		✓	✓	
CO4		✓	✓	✓	✓	✓	✓	✓		✓
CO5			✓	✓	✓	✓		✓	✓	✓
CO6	✓			✓				✓		✓

(STA504) Measure and Probability Theory

Couse Objective:

1. Establish a rigorous measure-theoretic foundation for probability and convergence concepts.
2. Apply mathematical inequalities and limit theorems to derive key probabilistic results.
3. Utilize computational reasoning for probabilistic simulations and verifying convergence.
4. Enhance critical thinking for analyzing random variables, convergence modes, and distributional behavior.
5. Foster adaptability and lifelong learning through mastery of advanced theoretical tools.

Course Learning Outcomes:

By the end of this course, students will be able to:

1. **Understand the foundational structures of sets, rings, fields, and σ-fields** to formalize probability spaces and decision-theoretic models.

2. **Apply measure-theoretic concepts** including properties of measures and measurable functions to represent random variables rigorously.

3. **Differentiate types of convergence** (almost everywhere, in probability, in measure, and in mean) and use them to evaluate the stability and reliability of decision rules.

4. **Use Lebesgue Dominated Convergence, Monotone Convergence, Fatou, and Fubini Theorems** to handle expectations, integrals, and conditional decisions under uncertainty.
5. **Decompose distribution functions** into discrete, absolutely continuous, and singular components for accurate probabilistic modeling in decision problems.
6. **Apply fundamental inequalities** (Hölder, Minkowski, Lyapunov, Kolmogorov) to bound risks, errors, and expectations in decision-making scenarios.
7. **Demonstrate Weak and Strong Laws of Large Numbers (WLLN, SLLN)** to justify long-run average outcomes in sequential decisions.
8. **Apply Khintchin's Theorem, Kolmogorov Strong Law, and Borel-Cantelli Lemma** to determine almost sure events and formulate reliable strategies in uncertain environments.
9. **Understand Borel zero-one laws** for events with certainty in the long run, guiding probabilistic decision-making.
10. **Analyze weak and complete convergence of distribution functions** to study limiting behavior of sequences of random variables in decision contexts.
11. **Use Helly-Bray Lemma and Weak Compactness Theorem** to evaluate convergence properties for optimal statistical inference.
12. **Employ characteristic functions, inversion and continuity theorems** to uniquely determine distributions and guide decisions based on limit laws.
13. **Apply one-dimensional Central Limit Theorems** (Lindeberg-Levy, Lyapunov, Lindeberg & Feller) to approximate distributions of sums of independent random variables.
14. **Use CLTs to assess the variability and risk** in aggregated decisions and to justify asymptotic approximations for large-sample decision-making.

UNIT I

Basics of ring, field and set function, Monotone collection of sets.

Measure, Properties of measure, Random variable as a measurable function. Almost everywhere convergence, Convergence in Probability and in measure, Mean convergence, Almost sure convergence, σ-field induced by sequence of random variables.
Lebesgue Dominated Convergence Theorem, Monotone Convergence Theorem, Fatou Lemma, Fubini Theorem (Without Proof).

UNIT II

Decomposition of distribution functions in purely discrete, absolutely continuous and singular components. Holder's inequality, Minkowski inequality, Lyapunov inequality, Kolmogorov's inequality.

UNIT III

Weak (WLLN) and Strong (SLLN) Law of Large Numbers, Khintchin's Theorem and Kolmogorov Strong Law of Large Numbers, Borel zero- one Law, Borel – Cantelli lemma.

UNIT IV

Weak and complete convergence of sequence of distribution functions, Weak compactness Theorem, Helly Bray Lemma & Theorem, Characteristic function, Inversion Theorem, Continuity Theorem.

UNIT V

One dimensional Central Limit Theorems: Lindeberg- Levy for i.i.d. random variables, Lyapunov (without proof), Lindeberg & Feller Theorem (without proof) for independent random variables.

Books recommended:

1. Bhat, B.R. (1999). Modern Probability Theory, 3rdedn. New Age International.
2. Chow Y.S. and Teicher, H. (2008) Probability theory, 3rdedn. Springer.
3. Chung, K.L. (2001). A course in probability theory, Academic Press.
4. Eisen, M. M. (1969) Introduction to mathematical probability theory Prentice-Hall.
5. Feller, W. (1968 and 1971) .An introduction to probability theory and its applications, vol I and II, Wiley.
6. Gnedenko, B.V. (1978). The theory of probability MIR Publishers.
7. Loève, M. (1977 & 1978) Probability theory, vol I & II Springer.
8. Rohatgi, V.K. and Saleh, A.K. Md. E. (2001). An Introduction to Probability and Statistics. Wiley.
9. Tucker, H.G. (1967). A graduate course in probability theory Academic Press Inc.
10. Ash, Robert. (1972). Real Analysis and Probability. Academic Press.
11. Bhat, B.R. (1999). Modern Probability Theory, 3rdedn. New Age International.
12. Bauer, H. (1991): Probability theory and elements of measure theory, Academic Press.
13. Burrill C. W. (1972). Measure, Integration, and Probability, McGraw-Hill
14. Basu, A.K. (2004). Measure Theory and Probability, Prentice Hall of India.
15. Halmos, P.R. (1974). Measure Theory, Springer.

CO- PO & PSO Mapping Matrix:-

CO No.	PO1	PO4	PO7	PO16	PO28	PSO1	PSO3	PSO4	PSO5	PSO6
CO1	✓	✓	✓	✓		✓	✓			✓
CO2	✓	✓	✓	✓		✓	✓			✓
CO4		✓	✓	✓		✓	✓	✓	✓	✓
CO6		✓	✓	✓		✓	✓		✓	✓
CO9	✓		✓	✓	✓	✓	✓			✓

SECOND SEMESTER

Paper	Duration L-T-P-C	Maximum Marks
STA502: MULTIVARIATE ANALYSIS	4-1-0-5	100
STA562: ADVANCED ESTIMATION THEORY	4-1-0-5	100
STA 507: REGRESSION ANALYSIS	4-1-0-5	100

(STA502) MULTIVARIATE ANALYSIS

Course Objectives:

The objective of this course is to introduce students to the theory and applications of multivariate statistical methods. Specifically, the course aims to:

1. Develop understanding of multivariate distributions.
2. Build skills in parameter estimation and distribution theory.
3. Understand sampling distributions in multivariate settings.
4. Apply multivariate hypothesis testing procedures.
5. Develop classification and dimensionality-reduction competencies.
6. Promote analytical and computational proficiency.

Course Learning Outcomes:

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

CO1: Understand and apply the concepts of the multivariate normal distribution, including its moment generating and characteristic functions, as well as derive marginal and conditional **distributions.**

CO2: Compute and interpret multiple and partial correlation coefficients, and understand their roles in modeling interrelationships among several variables.

CO3: Derive and apply maximum likelihood estimators (MLEs) of the mean vector and covariance matrix, and understand the distribution of the sample mean vector and the Wishart distribution with its key properties.

CO4: Explain and use the sampling distributions of sample correlation, multiple correlation, and regression coefficients under null hypotheses, and apply them in hypothesis testing contexts.

CO5: Develop and apply classification and discrimination procedures between two or more multivariate normal populations, construct sample discriminant functions, and evaluate misclassification probabilities.
CO6: Apply Principal Component Analysis (PCA) for dimensionality reduction and data summarization, and perform Canonical Correlation Analysis (CCA) to explore relationships between variable sets.
CO7: Integrate theoretical understanding with practical data analysis skills, demonstrating the ability to interpret multivariate relationships, make statistical inferences, and apply techniques to real-world multivariate datasets.

Mapping of Course outcomes:

Course Outcomes (COs)	PO1	PO3	PO4	PO7	PO16	PSO1	PSO2	PSO3	PSO4	PS6
CO1	✓		✓	✓		✓	✓			✓
CO2	✓		✓	✓		✓	✓	✓		✓
CO3	✓	✓	✓	✓	✓	✓	✓	✓		✓
CO4	✓	✓	✓	✓	✓	✓	✓	✓		✓
CO5	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CO6	✓		✓	✓		✓	✓	✓	✓	✓
CO7	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Course Contents:

UNIT I

Multivariate normal distribution, moment generating function and Characteristic function, marginal and conditional distributions, multiple and partial correlation coefficients.

UNIT II

Maximum likelihood estimators of the mean vector and covariance matrix, Distribution of sample mean vector. Wishart distribution and its properties.

UNIT III

Null distribution of sample correlation coefficient, sample multiple and partial correlation coefficients and their null sampling distributions, distribution of sample regression coefficient.

UNIT IV

Null distribution and non-null distribution of Hotelling's T^2 statistic, Applications in tests forthe mean vector of one and more multivariate normal populations, equality of the components of a mean vector in a multivariate normal population and their applications, Mahalanobis' D^2,

UNIT V

Classification and discrimination procedures for discrimination between two multivariate normal populations, sample discriminant function, tests associated with discriminant functions, probabilities of misclassification and their estimation, classification into more than two multivariate normal populations, Fisher-Behren Problem, Principal component analysis, Canonical correlations and variables.

Books recommended:

Anderson, T.W. (2003). An Introduction to Multivariate Statistical Analysis, 3rd edn., Wiley.
Anderson, T.W. (2003). An Introduction to Multivariate Statistical Analysis, 3rd ed., John Wiley & Sons.
Giri, N. C. (1977). Multivariate Statistical Inference, Academic Press. 3.
Hardle, W. K. and Simar, L. (2015). Applied Multivariate Statistical Analysis, 4th Edn., Springer. 4.
Johnson, R. A. and Wichern, D. W. (2015). Applied Multivariate Statistical Analysis, 6th Edn., Pearson Education India.
Kshirsagar, A. M. (1996). Multivariate Analysis, 2nd ed., Marcel Dekker.
Lawley, D. N. and Maxwell, A. E. (1971). Factor Analysis as a Statistical Method, 2nd Edn., London Butterworths.
Muirhead, R. J. (1982). Aspects of Multivariate Statistical Theory, John Wiley & Sons.
Rao, C. R. (1973). Linear Statistical Inference and its Applications, 2nd ed., John Wiley & Sons.
Srivastava, M. S. and Khatri, C. G. (1979). An Introduction to Multivariate Statistics, North Holland.

(STA562) ADVANCED ESTIMATION THEORY

Course Objective

CO1- Build a strong theoretical understanding of estimation principles, sufficiency, and completeness.
CO2- Apply mathematical tools to derive bounds (Cramér–Rao, Bhattacharyya, CRK) and evaluate estimator efficiency.
CO3- Implement estimation techniques (MLE, Minimum Chi-square) for optimal decision-making.
CO6- Develop analytical insight into asymptotic properties and optimal inference under various criteria.
CO10- Prepare for research and higher studies by integrating theoretical, computational, and applied inference skills.

Course Learning Outcomes:

By the end of this course, students will be able to:

1. **Construct point estimators** using different criteria, ensuring unbiasedness and consistency for informed decision-making.
2. **Identify sufficient statistics** using the Fisher-Neyman-Halmos-Savage factorization criterion to maximize information extraction.
3. **Determine minimal sufficiency and completeness** of n-tuples of order statistics to guide optimal inference strategies.
4. **Analyze completeness and bounded completeness** to ensure uniqueness and optimality of estimators in decision problems.

5. **Apply Basu's theorem** to assess independence of statistics, enhancing efficiency of inference.
6. **Use properties of exponential family distributions** to develop optimal estimators and decision rules.
7. **Compute lower bounds on estimator variance** using Bhattacharyya, Chapman-Robbins-Kiefer (CRK), and generalized Rao-Cramér bounds for multi-parameter cases.
8. **Assess efficiency and risk of estimators** to select optimal solutions in statistical decision-making.
9. **Apply Maximum Likelihood Estimation (MLE)** for parameter estimation that maximizes likelihood and minimizes expected loss.
10. **Use Zehna's theorem (invariance) and Cramér's theorem (weak consistency)** to ensure robustness and reliability of estimators in practical decisions.
11. **Analyze asymptotic properties** of estimators, including asymptotic normality, BAN and CAN estimators, asymptotic efficiency, and equivariant estimation.
12. **Implement method of minimum chi-square and optimal decision rules** to minimize expected loss in complex decision problems.
13. **Determine existence and application of optimal stopping rules** to guide sequential decisions under uncertainty.

UNIT I

Point estimation- Different criterion of point estimation. Unbiasedness, Reducing bias. Consistent estimators, Multiparameter cases, Cramér theorem for weak consistency.

UNIT II

Sufficiency, Fisher-Neyman-Halmos-Savage factorization criterion, minimal sufficiency, Sufficiency and completeness of n-tuple of order statistic.

Completeness, Bounded completeness, Uses of completeness.

UNIT III

Ancillary statistics, Basu's theorem on independence of Statistics, Exponential family.

Bhattacharyya bound, Chapman Robbins and Kiefer (CRK) bound, Generalized RaoCramér bound for the multiparameter case.

UNIT IV

Different methods of estimation- method of moments. Maximum likelihood estimation, Zehna theorem for invariance. Method of minimum x? and method of least squares.

UNIT V

Concept of efficient estimation. Asymptotic normality, BAN and CAN estimators, asymptotic efficiency, equivariant estimation.

Concept of optimal decisions. Existence of optimal stopping rule.

Books Recommended:

1. Casella, G and Berger, R. L. (2002). Statistical Inference, 2nd edition, Duxbury Press.

2. DeGroot, Morris H. (2004). Optimal statistical decisions, Wiley-Interscience.

3. Lehmann, E.L. and Casella, G. (1998). Theory of Point Estimation, Springer.

CO- PO & PSO Mapping Matrix:-

CO No.	PO1	PO4	PO7	PO16	PO29	PSO1	PSO3	PSO4	PSO5	PSO6
CO1	✓	✓	✓	✓		✓	✓			✓
CO2	✓	✓	✓	✓		✓	✓	✓		✓
CO3		✓	✓	✓		✓	✓	✓	✓	✓
CO6		✓	✓	✓		✓	✓		✓	✓
CO10	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

(STA 507) REGRESSION ANALYSIS

Course Objectives:

1. To understand the fundamental concepts of regression analysis and its role in statistical modeling, prediction, and inference.
2. To develop the ability to formulate and estimate regression models, including simple and multiple linear regression models, using real-world data.
3. To interpret regression coefficients and assess model adequacy through diagnostic tools, goodness-of-fit measures, and residual analysis.
4. To understand the assumptions underlying regression models and learn methods to detect and address violations such as multicollinearity, heteroscedasticity, and autocorrelation.
5. To gain familiarity with model selection and validation techniques, including variable selection.
6. To introduce extensions of the classical linear regression model, such as polynomial regression, logistic regression, and nonlinear regression.
7. To enable students to apply regression techniques to real-life problems in various disciplines such as economics, agriculture, biology, engineering, and social sciences. Also use statistical software (such as R, SPSS, or Python) for implementing regression analysis and interpreting the output effectively.

Course Learning Outcome:

8. After completion of this course the students will be able to:
9. CO1. Learn how to apply linear regression models in practice: identify situation where linear regression is appropriate; build and fit linear regression models with software; interpret estimates and diagnostic statistics; produce exploratory graphs.
10. CO2. Model selection: Mallow's Cp, AIC, BIC, R-squared, subset selection of independent variables, transformation of dependent and independent variables, multicollinearity
11. CO3. Assessing goodness-of-fit, normality, homogeneity of variances, detection of outliers and influential observations; Diagnostic plots for linear regression models.
12. CO4. Estimation of the parameters in General linear regression; least squares, maximum likelihood, method of moments; Confidence Intervals for parameters in General linear regression.
13. CO5. Estimation of regression coefficients under exact and stochastic restrictions.

Mapping of Course outcomes:

Course Outcomes	PO1	PO3	PO7	PO16	PO29	PSO1	PSO2	PSO3	PSO4
CO1		✓	✓		✓		✓	✓	✓
CO2		✓	✓		✓		✓	✓	✓
CO3		✓	✓		✓	✓			
CO4	✓		✓		✓				
CO5	✓		✓		✓				

Course content

STA507: Regression Analysis

UNIT I Multiple linear regression model and assumptions, estimation of parameters, estimable functions, error and estimation space, Gauss-Markov theorem, use of g-inverse.

UNIT II Model in deviation form, ANOVA for linear model, R2, adjusted R2 and other model selection criterion, tests of linear hypothesis, forecasting.

UNIT III Model Adequacy Checking: checking of linear relationship, residual analysis and scaling of residuals, regression variable hull, PRESS residuals, R-student residuals, residual plots, partial residual plots, detection and treatment of outliers, Diagnostics for leverage and influence, measures of influence.

UNIT IV Model specification tests, tests for parameter constancy and structural change, use of dummy variables. Estimation of parameters by generalized least squares (GLS) in linear models with non-spherical disturbances, Gauss Markov theorem for GLS estimator, heteroscedasticity of disturbances, estimation under heteroscedasticity and tests of heteroscedasticity, tests for autocorrelation, estimation and forecasting under autocorrelated disturbances.

UNIT-V Estimation of regression coefficients under exact and stochastic restrictions, restricted regression and mixed regression estimators and their properties. Model with stochastic regressors and errors in variable model, instrumental variable estimator.

Books recommended:

1. Draper, N.R. and Smith H. (1998), Applied Regression Analysis, 3rd Ed. Wiley.

2. Johnston, J. (1984). Econometric methods, Third edition, McGraw Hill.

3. Montgomery, D.C. Elizabeth A. P., Vining G.G. (2006), Introduction to Linear Regression Analysis, Wiley.

4. Rao, C. R., Toutenburg, H., Shalabh, Heumann, C (2008). Linear Models and Generalizations-Least squares and alternatives, Springer.

5. Monahan J.F. (2008). A Primer on Linear Models, CRC Press.

6. Khuri Andre I. (2010). Linear Model Methodology, CRC Press.

7. Seber , George A. F. and Lee Alan J. (2003), Linear Regression Analysis, Wiley.

Third Semester

Paper	Duration L-T-P-C	Maximum Marks
STA552: KNOWLEDGE DISCOVERY AND DATA MINING	4-1-0-5	100
STA601: TESTING OF HYPOTHESIS, SEQUENTIAL ANALYSIS AND NON-PARAMETRIC INFERENCE	4-1-0-5	100
STA670: SURVIVAL ANALYSIS AND CLINICAL TRIALS	4-1-0-5	100

(STA552) KNOWLEDGE DISCOVERY AND DATA MINING

Course Objectives (COBs):

The main objectives of this course are to:

- Provide a comprehensive understanding of database systems, data warehouses, and data management architectures.
- Introduce concepts of data mining, machine learning, and their applications in data-driven decision-making.
- Develop analytical skills for handling multidimensional and high-dimensional data.
- Equip students with practical knowledge of clustering, classification, and regression techniques.
- Familiarize students with neural network architectures and learning algorithms for predictive modeling.
- Introduce advanced machine learning models such as Support Vector Machines (SVM) and Independent Component Analysis (ICA).
- Integrate theoretical concepts with practical applications for solving real-world data analysis problems.

UNIT I

Database Systems Concepts and Architecture, concept of Dataware houses, DBMS, and RDBMS, online analytical data processing, Data mining and machine learning, supervised and unsupervised learning. Linear dimensionality reduction: principal component analysis for linear feature space, scree plot and its use for determining the number of principal components to retain, basic idea of non parametric kernel density estimation, non linear principal component analysis.

UNIT II

Clustering: Similarity and distance measures, Outliers, Minimum spanning tree, squared error clustering, K-means clustering, Hierarchical clustering, Block clustering and two way clustering: Hartigan's block clustering algorithm, Biclustering, Plaid models for biclustering.

UNIT III

Classification and Regession Trees (CART): Classification trees, node impurity function and entropy function, choosing the best split pruning algorithm for classification trees. Regression trees, terminal node value and splitting strategy, pruning the tree and best pruned subtree. Committee Machine: Bagging tree based classifiers and regression tree predictors, Boosting, ADABOOST algorithm for binary classification.

UNIT IV

Artificial Neural Network: and extensions of regression models, McCullon-Pitts Neuron (Threshold Logic Unit), Rosenblatt's Single layer perceptron, single unit perceptron gradient descent learning algorithm, Multilayer perceptron, feed forward and back propogation learning algorithm, Self organizing maps (SOM) or Kohonen neural network, on-line and batch versions of SOM algorithm, U matrix.

UNIT V

Support vector machine (SVM) with linear class boundaries, multiclass SVM, Latent variable models for blind source separation: Independent component analysis (ICA) and its applications, linear mixing and noiseless ICA, FastICA algorithm for determining single source component, deflation and parallel FastICA algorithm for extracting multiple independent source components.

Books Recommended:

1. Ramez Ellmasri and Shamkant B. Navathe (2010). Fundamentals of Database systems, sixth Edition, Tata McGraw Hill Publications.
2. Izenman, A.J., (2008), Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold learning, Springer
3. Han, J. and Kamber, M (2006). Data Mining: Concepts and Techniques, 2nd edition, Morgan Kaufmann.
4. Dunham, M. H. (2003). Data Mining: Introductory and Advanced Topics, Pearson Education.
5. Sheskin, D. J. (2004). The Handbook of Parametric and Nonparametric Statistical Procedures, 3rd Edition, Chapman and Hall/CRC.

Course Learning Outcomes:

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

CO1: Understand and explain the fundamental concepts of Database Systems, Data Warehouses, DBMS, and RDBMS, and distinguish between Online Analytical Processing (OLAP), Data Mining, and Machine Learning paradigms.

CO2: Apply supervised and unsupervised learning techniques, perform Principal Component Analysis (PCA) for linear dimensionality reduction, interpret scree plots, and explain the concepts of kernel density estimation and non-linear PCA.

CO3: Compute and interpret similarity and distance measures, identify outliers, and apply clustering algorithms such as K-means, Hierarchical, Block, and Biclustering, including Hartigan's and Plaid model approaches.

CO4: Construct and analyze Classification and Regression Trees (CART), compute impurity measures, implement pruning techniques, and apply ensemble methods such as Bagging, Boosting, and the AdaBoost algorithm for improved predictive performance.

CO5: Explain the architecture and functioning of Artificial Neural Networks (ANN), including McCulloch–Pitts neurons, perceptrons, and Multilayer Perceptrons (MLP), and implement learning algorithms such as gradient descent, backpropagation, and Self-Organizing Maps (SOM).

CO6: Apply Support Vector Machine (SVM) models for linear and multiclass classification, and explain latent variable models for Blind Source Separation using Independent Component Analysis (ICA) and the FastICA algorithm.

CO7: Integrate concepts from database systems, machine learning, and neural computation to analyze, classify, and interpret high-dimensional data, demonstrating the ability to apply advanced data analytics and pattern recognition techniques in real-world contexts.

Mapping of Course Outcomes:

Course Outcomes (COs)	PO1	PO3	PO4	PO7	PSO1	PSO2	PSO3	PSO4	PSO5	PSO6	PSO9
CO1	✓		✓		✓		✓	✓			✓
CO2	✓	✓		✓	✓	✓		✓		✓	✓
CO3		✓	✓	✓		✓	✓			✓	✓
CO4	✓	✓		✓	✓		✓	✓	✓		✓
CO5	✓			✓	✓		✓	✓		✓	✓
CO6	✓	✓	✓	✓	✓		✓		✓	✓	✓
CO7	✓	✓		✓	✓	✓		✓	✓		✓

(STA601) TESTING OF HYPOTHESIS, SEQUENTIAL ANALYSIS AND NON-PARAMETRIC INFERENCE

Course Objectives:

1. Develop a deep understanding of the theoretical relationship between interval estimation and hypothesis testing within the framework of classical statistical inference.
2. Introduce the principles of uniformly most powerful (UMP), uniformly most powerful unbiased (UMPU), and invariant tests, emphasizing their theoretical basis and practical implications.
3. Explain sequential methods of testing and estimation, including the Sequential Probability Ratio Test (SPRT), Wald's equation, and Stein's two-stage procedures, along with their asymptotic properties.
4. Provide theoretical and computational knowledge of order statistics, their moments, asymptotic distributions, and applications in nonparametric estimation.
5. Introduce nonparametric hypothesis testing procedures based on ranks and order statistics, such as Mann–Whitney U, Kolmogorov–Smirnov, Kruskal–Wallis, and Friedman tests, along with their efficiency measures.

Syllabus

Unit	Course Content	Hours of Teaching
I	Relationship between interval estimation and hypothesis testing, Generalized Neyman Pearson lemma, UMP tests for distributions with MLR tests. UMPU tests, similar regions, Neyman structure, Invariant tests, Properties of LR tests (only statement).	
II	SPRT, Fundamental identity, OC and ASN functions, Wald's equation. Wolfowitz generalization of FRC bound, Stein's two stage procedure, asymptotic theory of sequential estimation, sequential estimation of normal mean.	
III	Moments of Order Statistics, Asymptotic distribution of an order statistic, nonparametric estimation of distribution function and Glivenko-Cantelli fundamental theorem of statistics.	
IV	The Mann-Whitney U test, Application of U-statistic to rank tests, One sample and two sample Kolmogorov-Smirnov tests.	

V	The Kruskal-Wallis One-Way ANOVA Test, Friedman's Two-Way Analysis of Variance by ranks. Efficiency criteria, Theoretical basis for Calculating the ARE, Pitman ARE.	

Texts / Reference Books

1. David, H.A. and Nagaraja, H.N.. (1981), Order Statistics Wiley.
2. Gibbons, J.D. and Chakraborti,S. (2003). Nonparametric Statistical Inference, 4-th Edition, Marcel Dekker, Inc., New York.
3. Hájek, J. (1969) A course in Nonparametric statistics. Holden-Day, San Francisco
4. Lehmann, E.L. (2006), Nonparametrics. Statistical methods based on ranks Springer.
5. Randles, R.H. and Wolfe, D.A. (1979). Introduction to the Theory of Nonparametric Statistics, Wiley.
6. Ghosh, B.K.. (1970). Sequential tests of statistical hypothesis, Addison Wesley.
7. Ghosh, M., N. Mukhopadhyay, and P.K. Sen. (2011), Sequential Estimation John Wiley.
8. Lehmann, E L (1986). Testing Statistical Hypotheses, Springer.
9. Rohatgi, V. K. and Saleh, A. K. Md E. (2001). An Introduction to Probability and Statistics, Wiley.
10. Srivastava, M. and Srivastava, N. (2009). Statistical Inference -Testing of Hypotheses, Prentice Hall.

Learning Outcomes:

Upon successful completion of this course, students will be able to:

1. Explain the theoretical relationship between interval estimation and hypothesis testing, and apply the Generalized Neyman–Pearson Lemma to construct most powerful tests.

2. Derive and evaluate UMP, UMPU, and invariant tests for different statistical models and understand their optimality properties.

3. Describe the properties and structure of Likelihood Ratio Tests and discuss their role in hypothesis testing.

4. Apply sequential testing and estimation procedures such as SPRT, Wald's equations, Wolfowitz generalization, and Stein's two-stage method, and interpret their OC and ASN functions.

5. Compute moments and asymptotic distributions of order statistics and apply them to nonparametric estimation of distribution functions.

6. Perform and interpret nonparametric rank-based tests including Mann–Whitney U, Kolmogorov–Smirnov, Kruskal–Wallis, and Friedman tests.

7. Evaluate the efficiency of nonparametric tests using Pitman's Asymptotic Relative Efficiency (ARE) and related efficiency criteria.

Course Outcomes (COs)	PO1	PO3	PO4	PO7	PO16	PSO1	PSO2	PSO3	PSO 5	PSO 6
CO1	✓	✓	✓	✓	✓	✓	✓			
CO2	✓		✓	✓	✓	✓	✓			✓
CO3	✓		✓	✓		✓				✓
CO4		✓	✓	✓	✓	✓		✓	✓	
CO5			✓	✓	✓	✓	✓	✓		
CO6		✓	✓	✓	✓	✓		✓	✓	
CO7			✓	✓	✓	✓	✓	✓		✓

(STA670) SURVIVAL ANALYSIS AND CLINICAL TRIALS

Course Objective:

CO1 Build a strong theoretical foundation of survival analysis and reliability models using hazard functions, censoring, and truncation principles.

CO3 Apply statistical methods (Kaplan–Meier, Log-Rank Test, Cox Model) to analyze lifetime and medical survival data.

CO4 Use computational and decision-theoretic approaches for optimizing clinical trial design and analysis.

CO5 Design and evaluate fixed, sequential, and randomized trials ensuring ethical and statistical rigor.

CO6 Strengthen critical and analytical thinking to interpret risk, uncertainty, and model-based decisions in real-world medical research.

Course Learning Outcomes:

By the end of this course, learners will be able to:

1. **Understand and interpret survival data** through decision-oriented perspectives, recognizing the role of hazard functions and their properties in modeling lifetime uncertainty and reliability-based decisions.
2. **Formulate and evaluate parametric survival models** (Exponential, Weibull, Gamma, Normal, and Log-Normal) using estimation and hypothesis testing as decision rules that balance bias, variance, and model adequacy.
3. **Apply decision-theoretic principles to censored and truncated data**, selecting optimal methods (such as Kaplan–Meier estimation, life tables, and Log-Rank tests) for inference and comparison of survival functions under uncertainty.
4. **Analyze and extend Cox Proportional Hazards Models**, incorporating stratified and time-dependent covariates, while making optimal model-based decisions about risk factors and their proportional influence on hazard rates.
5. **Evaluate recurrent event and competing risks models** using frailty and dependency structures to make informed statistical decisions regarding event interdependence and risk prioritization.
6. **Integrate decision-theoretic reasoning into clinical trial design**, optimizing trial phases, sample sizes, and timing to minimize error probabilities and ethical risks while maximizing inferential power.
7. **Compare and select appropriate trial designs** (fixed, sequential, or randomized) using Bayesian or frequentist decision frameworks to determine optimal stopping, allocation, and blinding strategies.
8. **Develop a comprehensive understanding of uncertainty quantification**, decision loss functions, and risk-benefit trade-offs in survival modeling and clinical trial analysis.

UNIT I

Introduction, Outlines and objectives, Applications. Basic terms and their inter-relationships. Various properties of hazard function.

Parametric Survival Models- Exponential, Weibull, Gamma, Normal, Log-normal models. Estimation and testing procedures on these models.

Types of censoring and truncation, Uses of Life table, Kaplan–Meier Survival Curves and the Log–Rank Test, Log–Rank Statistic for Several Groups.

UNIT II

Proportional Hazard Models- Assumption. The Cox Proportional Hazards Model and its Characteristics. The Stratified Cox Procedure. Extension of the Cox Proportional Hazards Model (Time-Dependent).

UNIT III

Recurrent Event Survival Analysis- Introduction, Outline and objectives, Competing Risks Survival Analysis- Competing risk events and Frailty models**.**

UNIT IV

Introduction to clinical trials, Historical development. Uses of clinical trials, Problems in the timing of a clinical trials. Phases of a clinical trial.

UNIT V

Statistical designs-fixed sample trials: simple randomized design, stratified randomized crossover design; Sequential design - open and close sequential design. Randomization Dynamic randomization, Permuted block randomization; Blinding-Single, double and triple.

Books Recommended:

1. Allison , P. D. (2010). Survival Analysis Using SAS: A Practical Guide, SAS Publishing.
2. Kleinbaum, D. G. and Klein, M. (2012). Survival Analysis: A Self-Learning Text, Springer-Verlag New York.
3. Klein , J. P. and Moeschberger, M. L. (2005) Survival Analysis–Techniques for Censored and Truncated Data, Springer.
4. Hosmer, D. W., JR and Lemeshow , S. (2008) Applied survival Analysis: regression modeling of time to event data, Wiley.
5. Cleves , M., Gould, W. and Gutierrez, R. (2010) An introduction to survival analysis using STATA, Stata Press.
6. Friedman IM Furberg CD Demets DL. Fundamentals of clinical trials. 4th edition. Springer. 2010.
7. Meinert CL. Clinical trials: Design conduct and analysis. 2nd edition. New York: Oxford University Press. 2012.
8. Pocock S. Clinical trials – A practical approach. John Wiley & Sons. 2010.
9. Daniel WW. Biostatistics: A foundation for analysis in the health sciences. 10th edition. John Wiley & Sons. 2013.

CO- PO & PSO Mapping Matrix:

COs	PO1	PO4	PO7	PO15	PO16	PSO1	PSO3	PSO5	PSO6	PSO8
CO1	✓		✓			✓	✓		✓	
CO2		✓			✓	✓	✓		✓	
CO3		✓	✓			✓	✓	✓		
CO4		✓		✓		✓	✓		✓	
CO5			✓		✓	✓		✓	✓	
CO6				✓		✓	✓	✓		✓

Fourth Semester

Paper	Duration L-T-P-C	Maximum Marks
STA602: ANALYSIS OF VARIANCE AND DESIGN OF EXPERIMENTS	4-1-0-5	100
STA656: STOCHASTIC PROCESS	4-1-0-5	100

(STA602) ANALYSIS OF VARIANCE AND DESIGN OF EXPERIMENTS

Course Objectives:

1. Introduce the fundamental principles of experimental design including randomization, replication, and local control. Develop the ability to construct and analyze various experimental designs such as CRD, RBD, LSD, and factorial designs**.**
2. Promote critical thinking to select suitable experimental designs for specific research problems and ensure efficient resource utilization.
3. Provide practical skills to plan, conduct, and analyze experiments in fields like agriculture, industry, and life sciences**.**
4. Familiarize students with commonly used non-parametric tests such as the Sign test, Wilcoxon test, Mann–Whitney U test, Kruskal–Wallis, and Friedman test**.**
5. Encourage analytical reasoning and the use of statistical software for applying non-parametric tests to real-life data.

Course Learning Outcome:

After completion of the course the students will be able to: understand the design and conduct experiments, as well as to analyze and interpret data.

CO1. Carry out one way and two-way Analysis of Variance (ANOVA), Analyze covariance (ANCOVA) models and understand their applications in improving precision of experimental results.

CO2. Understand the basic terms used in Intra and inter block analysis of Incomplete block design.

CO3. Identify and apply appropriate incomplete block designs (e.g., BIBD, PBIBD and Lattice) to handle practical constraints in experimentation.

CO4. Understand the basic concepts Association schemes and partially balanced incomplete block designs.

CO5. Apply factorial experiment concepts to study the interaction effects among multiple factors and estimate main and interaction effects. Formulate and communicate scientifically valid conclusions from designed experiments relevant to agriculture, industry, and social sciences.

Mapping of Course outcomes:

Course Outcomes	PO1	PO4	PO7	PO16	PO29	PSO1	PSO2	PSO3	PSO4
CO1				✓			✓		
CO2		✓	✓	✓				✓	
CO3	✓		✓	✓	✓				
CO4					✓	✓			✓
CO5					✓	✓			

Course content

UNIT I

Linear regression model, Random and mixed effect models, estimation of variance component in one way and two-way random effects model. General two- way classification.

UNIT II

Tukey's test, general two-way classification. Intra and inter block analysis of Incomplete block design.

UNIT III

General block design and its information matrix (C). Criteria for connectedness, balanced and orthogonality: Balanced Incomplete Block Design (BIBD) – Intra and inter block analysis, Simple lattice designs.

UNIT IV

Association schemes and partially balanced incomplete block designs – construction and parameter identification, Analysis of covariance.

UNIT V

General factorial experiments, factorial effects, study of n 2 and n3 factorial experiments in randomized blocks, complete and partial confounding, construction of confounded factorial experiments, split plot experiment.

Books recommended:

1. Das, M.N. and Giri, N. (1979). Design and Analysis of Experiments, Wiley Eastern.
2. Dean, A. and Voss, D. (1999). Design and Analysis of Experiments, Springer.
3. Dey, A. (1986). Theory of Block Designs, Wiley Eastern.
4. Giri, N. (1986). Analysis of Variance, South Asian Publishers.
5. Joshi, D.D. (1987). Linear Estimation and Design of Experiments, Wiley Eastern.
6. Montgomery, C.D. (1976). Design and Analysis of Experiments, Wiley.
7. Toutenburg, H. and Shalabh (2009). Statistical Analysis of Designed Experiments, Springer.

(STA656) STOCHASTIC PROCESS

Course Objectives:

1. Introduce the fundamental concepts and classifications of discrete and continuous-time Markov processes and their long-term behavior.

2. Develop analytical skills to compute n-step transition probabilities, steady-state distributions, and understand properties like periodicity, recurrence, and ergodicity in Markov chains.

3. Explain key continuous-time stochastic models such as the Poisson process, random walks, Brownian motion, and their applications in various fields.

4. Familiarize students with branching processes, birth–death processes, renewal processes, and their relevance in population growth and queueing systems.

5. Explore advanced stochastic concepts including the Wiener process, martingales, stopping times, and the optional sampling theorem, emphasizing both theoretical and applied perspectives.

Syllabus

Unit	Course Content	Hours of Teaching
I	Two state Markov sequences, Markov chains, determination of n-step transition probabilities, Chapman-Kolmogorov equations, first return and first passage probabilities, classification of states, communicating states, periodicity, stationary probability distributions and limit theorems for ergodic chains.	
II	Continuous time Markov processes, Poisson (point) process, Inter arrival time distribution, Random walk and Brownian motion as a random walk, gambler's ruin problem.	

III	Branching processes of discrete type, average size and variance of the population in the n-th generation, fundamental theorem of extinction.	
IV	Birth and death processes, renewal processes, Queueing Theory: M/M/1, M/M/k and M/G/1 queueing processes.	
V	Wiener process, Arc-sine law, Martingales, stopping times, optional sampling theorem.	

Texts / Reference Books

1. Adke, S. R. and Manjunath, S. M. (1984). An Introduction to Finite Markov Processes, Wiley Eastern.
2. Cinlar, E. (1975). Introduction to Stochastic Processes, Prentice Hall.
3. Feller, W. (1968). Introduction to Probability and Applications, New Age India International.
4. Harris, T. E. (1963). The Theory of Branching Processes, Springer Verlag.
5. Hoel, P. G., Port, S. C. and Stone, C. J. (1991). Introduction to Stochastic Processes, University Book Stall.
6. Karlin, S. and Taylor, H. M. (1995). A First Course in Stochastic Processes, Academic Press.
7. Medhi, J. (2012). Stochastic Processes, 3rd edition, New Age India International.
8. Ross, S. M. (1996). Stochastic Processes, Wiley.

Learning Outcomes:

Upon successful completion of this course, students will be able to:

1. Explain the concepts of Markov chains, classify states, and determine n-step transition probabilities using Chapman–Kolmogorov equations.
2. Analyze the long-run behavior of Markov chains through stationary distributions, periodicity, and limit theorems for ergodic chains.
3. Describe continuous-time Markov processes and the Poisson process, including inter-arrival time distributions and their applications.
4. Model and analyze random walks, Brownian motion, and the gambler's ruin problem as stochastic processes.
5. Evaluate properties of branching processes and compute expected population size, variance, and extinction probabilities.
6. Apply the theory of birth–death and renewal processes to derive and interpret results for queueing models such as M/M/1, M/M/k, and M/G/1.
7. Understand and apply the concepts of Wiener processes, martingales, stopping times, and optional sampling theorem in stochastic modelling and inference.

Course Outcomes (COs)	**PO1**	**PO3**	**PO4**	**PO7**	**PO16**	**PSO1**	**PSO2**	**PSO3**	**PSO5**	**PSO6**
CO1	✓		✓	✓		✓	✓			✓
CO2	✓		✓	✓	✓	✓				✓
CO3	✓		✓	✓		✓	✓			
CO4	✓		✓	✓	✓	✓	✓			✓
CO5	✓		✓	✓	✓	✓	✓		✓	✓
CO6	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
CO7	✓	✓	✓	✓	✓	✓	✓		✓	✓