ST. JOSEPH'S UNIVERSITY

BENGALURU-27

School of Physical Sciences

DEPARTMENT OF STATISTICS

SYLLABUS FOR POST GRADUATE PROGRAMME

For Batch 2024 onwards

<table>
<tr><td colspan="4">Part A</td></tr>
<tr><td>1</td><td>Title of the Academic Program</td><td colspan="2">M.Sc. Statistics</td></tr>
<tr><td>2</td><td>Program Code</td><td colspan="2">SJC M.Sc. Statistics (To be given by Examination Section)</td></tr>
<tr><td>3</td><td>Name of the College</td><td colspan="2">St. Joseph’s College (Autonomous)</td></tr>
<tr><td>4</td><td>Objective of the College</td><td colspan="2">Academic Excellance
Character Formation
Social Concern</td></tr>
<tr><td>5</td><td>Vision of the College</td><td colspan="2">“Striving for a just, secular, democratic and economically sound society, which cares for the poor, the oppressed and the marginalized”</td></tr>
<tr><td rowspan="4">6</td><td rowspan="4">Mission of the College</td><td>M1</td><td>St. Joseph's College (Autonomous) seeks to form men and women who will be agents of change, committed to the creation of a society that is just, secular and democratic.</td></tr>
<tr><td>M2</td><td>The education offered is oriented towards enabling students to strive for both academic and human excellence.</td></tr>
<tr><td>M3</td><td>The college pursues academic excellence by providing a learning environment that constantly challenges the students and supports the ethical pursuit of intellectual curiosity and ceaseless enquiry.</td></tr>
<tr><td>M4</td><td>Human excellence is promoted through courses and activities that help students achieve personal integrity and conscientize them to the injustice prevalent in society.</td></tr>
<tr><td>7</td><td>Name of the Degree</td><td colspan="2">Master of Science (M.Sc.,) in STATISTICS</td></tr>
<tr><td>8</td><td>Name of the Department offering the program</td><td colspan="2">STATISTICS</td></tr>
<tr><td>9</td><td>Vision of the Program</td><td colspan="2">To enhance the quality of life for individuals and societies through the intelligent and ethical use of statistics.</td></tr>
<tr><td>10</td><td>Mission of the Program</td><td colspan="2">Department aims to instil scientific temper, analytical skills and intellectual vigour among students, that they may contribute to the needs of society.

Department aims to provide students with life-oriented education through projects of social relevance.</td></tr>
<tr><td>11</td><td>Duration of the Program</td><td colspan="2">2 years (FOUR semesters)</td></tr>
<tr><td>12</td><td>Total No. of Credits</td><td colspan="2">90</td></tr>
<tr><td rowspan="3">13</td><td rowspan="3">Program Educational Objectives (PEOs)</td><td>PEO1</td><td>To enable students about the methods of describing the data through various multivariate measures such as moments.</td></tr>
<tr><td>PEO2</td><td>Designing a model for a real time problem, testing the adequacy of the model and its implementation that will be helpful to the society.</td></tr>
<tr><td>PEO3</td><td>To develop skills like leadership quality and professional ethics in analysis of data that will enhance human excellence.</td></tr>
<tr><td>14</td><td>Graduation Attributes</td><td colspan="2">The Following graduate attributes reflect the particular quality and feature or characteristics of an individual, that are expected to be acquired by a graduate through studies at St. Joseph’s College.
• Disciplinary knowledge
• Communication Skills
• Critical thinking
• Problem solving
• Analytical reasoning
• Research-related skills
• Cooperation/Team work
• Reflective thinking</td></tr>
</table>

			• Information/digital literacy • Self-directed learning and Lifelong learner • Multicultural competence • Moral and ethical awareness/reasoning • Leadership readiness/qualities • International Outlook
15	Program Outcomes (POs)	PO1	Make significant contributions to the field of statistics through Research
		PO2	Collaborate with organizations for inter-disciplinary research.
		PO3	Inculcate lifelong learning and develop professional ethics.
		PO4	Empowering their knowledge with various software tools for analysing big data.
		PO5	To be employable in different sectors demanding for statistical analysis and interpretation.

Eligibility: A candidate who has passed B.Sc. Degree Examination with Statistics or Mathematics as major/optional subjects, securing a minimum of 50% in the Statistics and a minimum of 40% in the aggregate of this University or any other University equivalent thereto is eligible.

Part B

M.Sc. Statistics Curriculum

Courses and course completion requirements	No. of credits
STATISTICS	88
Open elective courses (non-professional)	2
Outreach activity	0

SUMMARY OF CREDITS

SUMMARY OF CREDITS

DEPARTMENT OF STATISTICS (PG) (2023-25) SEMESTER I									
	Code Number	**Title**	**No. of Hours of Instructions**	**Number of Hours of teaching per week**	**Number of credits**	**Continuous Internal Assessment (CIA) Marks**	**End Semester Marks**	**Total marks**	**Duration of Exam (in hrs)**
Theory	ST7121	Probability Theory	52	4	4	50	50	100	2
Theory	ST7221	Theory of Point estimation	52	4	4	50	50	100	2
Theory	ST7323	Sampling theory and Statistics for National Development	52	4	4	50	50	100	2
Theory	ST7421	Mathematical Analysis and Linear Algebra	52	4	4	50	50	100	2
Theory	ST7521	Statistical Computing (Soft Core)	39	3	2	25	25	50	1
Theory	ST 7622	Online course	60		2	-	50	50	-
Practical	ST7P1	Practical – I (Based on ST 7221, ST 7321)	44	4	2	15	35	50	3
Practical	ST7P2	Practical – II (Based on ST 7421, ST 7521)	44	4	2	15	35	50	3
					24				

DEPARTMENT OF STATISTICS (PG) (2023-2025) SEMESTER II									
	Code Number	**Title**	**No. of Hours of Instructions**	**Number of Hours of teaching per week**	**Number of credits**	**Continuous Internal Assessment (CIA) Marks**	**End Semester Marks**	**Total marks**	**Duration of Exam (in hrs)**
Theory	ST8121	Distribution Theory	52	4	04	50	50	100	2
Theory	ST8221	Testing of Hypothesis and Interval Estimation	52	4	04	50	50	100	2
Theory	ST8321	Multivariate Analysis	52	4	04	50	50	100	2
Theory	ST8421	Linear Models and Regression Analysis	52	4	04	50	50	100	2
Theory	ST8521	Introduction to Data science (Soft Core)	39	3	02	25	25	50	1
Theory	ST 8622	Mini Project			02	15	35	50	
Practical	ST8P1	Practical III (based on ST 8121, ST 8221 and ST 8321)	44	4	02	15	35	50	3
Practical	ST8P2	Practical IV (based on ST 8421, ST 8521)	44	4	02	15	35	50	3
					24				

DEPARTMENT OF STATISTICS (PG) (2023-2025) SEMESTER III									
	Code Number	**Title**	**No. of Hours of Instructions**	**Number of Hours of teaching per week**	**Number of credits**	**Continuous Internal Assessment (CIA) Marks**	**End Semester Marks**	**Total marks**	**Duration of Exam (in hrs)**
Theory	ST9120	Stochastic Processes	52	4	04	50	50	100	2
Theory	ST9220	Data Mining and Machine Learning	52	4	04	50	50	100	2
Theory	ST9323	Statistical Quality Control	52	4	04	50	50	100	2
Theory	STDE 9420 STDE 9520	Optimization Techniques (Elective-I) Operations Research (Elective-I)	52	4	04	50	50	100	2
Theory	ST 9623	Reliability Theory	52	4	04	50	50	100	2
Theory	ST9720	Research paper presentation	39	4	02	25	25	-	-
Practical	ST9P1	Practical V (based on ST 9120, and ST 9320)	44	4	02	15	35	50	3
Practical	ST9P2	Practical VI (based on ST9220, STDE 9420/STDE 9520)	44	4	02	15	35	50	3
					26				

DEPARTMENT OF STATISTICS (PG) (2023-2025) SEMESTER IV									
	Code Number	**Title**	**No. of Hours of Instructions**	**Number of Hours of teaching per week**	**Number of credits**	**Continuous Internal Assessment (CIA) Marks**	**End Semester Marks**	**Total marks**	**Duration of Exam (in hrs)**
Theory	ST 0123	Advanced Statistical Inference	52	4	04	50	50	100	2
Theory	ST0220	Design and Analysis of Experiments	52	4	04	50	50	100	2
Theory	STDE 0323 STDE 0420	Bio Statistics (Elective II) Survival analysis (Elective II)	52	4	04	50	50	100	2
Theory	STDE 0520 STDE 0620	Time series analysis (Elective III) Actuarial Statistics (Elective III)	52	4	04	50	50	100	2
Theory	ST 0720	Project Work	39	4	04	15	35		
Practical	ST0P1	Practical III (based on ST 0120, STDE 0320/STDE 0420 and STDE0520/STDE 0620)	44	4	02	15	35	50	3
Practical	ST0P2	Practical VIII (based on ST 0220)	44	4	02	15	35	50	3
					24				

CORE COURSES (CC)	
Course Title	Code Number
Probability Theory	ST7121
Theory of Point estimation	ST7221
Sampling theory and Statistics for National Development	ST 7323
Mathematical Analysis and Linear Algebra	ST7421
Distribution Theory	ST8121
Testing of hypothesis and Interval estimation	ST8221
Multivariate Analysis	ST8321
Linear Models and Regression Analysis	ST8421
Stochastic Processes	ST9120
Data Mining and Machine Learning	ST9220
Statistical Quality Control	ST9323
Reliability Theory	ST 9623
Advanced Statistical Inference	ST0120
Design and Analysis of Experiments	ST0220

DISCIPLINE-SPECIFIC ELECTIVE COURSES (DSE)	
Course Title	Code Number
Optimization Techniques	STDE 9420
Operations Research	STDE 9520
Bio-Statistics	STDE 0323
Survival Analysis	STDE 0420
Time Series Analysis	STDE 0520
Actuarial Statistics	STDE 0620

SKILL ENHANCEMENT COURSE (SEC) – Any practical oriented and software-based courses offered by departments to be listed below	
Course Title	Code Number
Statistical Computing (Soft Core)	ST7521
Introduction to Data science (Soft Core)	ST8521

VALUE ADDED COURSES (VAC) Certificate courses that add value to the core papers can be listed.	
Course Title	Code Number
Data science and analytics	
Data analysis using SPSS and MS Excel	

ONLINE COURSES (OLC) Certificate courses offered or recommended by the department to be listed	
Course Title	Code Number
Online courses related to SAS, MATLAB, etc.. (Swayam, MOOC, NPTEL, etc..)	ST 7622
Online courses related to SAS, MATLAB, etc.. (Swayam, MOOC, NPTEL, etc..)	ST 9620

Course Outcomes and Course Content

M.Sc. program in Statistics

First Semester

ST 7121: Probability Theory

Semester	First
Paper Code	ST 7121
Paper Title	Probability Theory
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(12L+2T) hrs**

Classes of sets: Field, σ-field, minimal σ – field, Borel σ-field on $\mathbf{R}^n$. Sequences of setsand their limits. Measure of a set and its properties. σ-finite measure. Counting measure. Lebesgue measure and Lebesgue- Stieltjes measure. Probability space and its properties. Probability measure induced by a random variable.

Unit 2 **(10L+2T) hrs**

Random variable, cumulative distribution function (c.d.f.), decomposition of a c.d.f. into discrete and continuous c.d.f.'s, quantile function, sequences of random variables, Modes of convergence- Convergence in distribution / law, convergence in probability, Convergence in r^{th} mean, almost sure convergence. Limit theorems.

Unit 3 **(10L+4T) hrs**

Expectation of a random variable. Properties. Monotone convergence theorem. Dominated convergence theorem. Markov, Chebycheff, Jensen, Minkowski, and Holder inequalities.

Unit 4 **(10L+2T) hrs**

Generating function: Moment generating function, Characteristic function. Properties of MGF and CF. Generating moments. Inversion theorem and its applications. Uniqueness theorem.

References

1. Ash, R.B. and Doleans-Dade, C.A. (2000). *Probability and Measure Theory*, Academic Press, NewYork.
2. Bhat, B.R. (1999). *Modern Probability Theory,* 2/e, New Age International, NewDelhi.
3. Billingsley, P. (1995). *Probability and Measure*, 3/e, John Wiley, NewYork.
4. Burril, C. (1972). *Measure, Integration, and Probability,* McGrawHill International, New York.
5. Chung, K.L.(2001). *A Course in Probability, 3/e,* Academic Press, NewYork.
6. Clarke, L.E. (1975). *Random Variables,* Longman Mathematical Texts, London.

7. Khosnevisan, D. (2013). *Probability*, American Mathematical Society, Indian Edition, Universities Press, Hyderabad.
8. Rao, C.R. (1973). *Linear Statistical Inference and Its Applications,* John Wiley, New York.

BLUEPRINT

Code number: ST7121

Title of the paper: Probability Theory

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
19	14	I
16	12	II
19	14	III
16	12	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the student should be able to

CO1	Understand the meaning of measure and probability.
CO2	Learn to develop complex mathematical reasoning.
CO3	Identify application of inequalities in probability theory.
CO4	Possess techniques of proving theorems and thinking out counter examples.
CO5	Student should be able to apply the probability concepts for real life uncertainty problems.

First Semester

ST 7221: Theory of Point Estimation

Semester	First
Paper Code	ST 7221
Paper Title	Theory of Point Estimation
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(10L+2T) hrs**

Families of distributions: location and scale families, k-parameter exponential, Pitman and Cramer families.
Properties of estimators: Unbiasedness, convex combination of unbiased estimators, consistent estimators, sufficient condition for consistent estimators with examples. Mean square error.

Unit 2 **(10L+2T)hrs**

Sufficiency and completeness: Sufficiency, Fisher-Neymann factorization theorem, minimal sufficiency, likelihood equivalence, completeness, bounded completeness. Statement of the theorem on complete sufficient statistic in the k-parameter exponential family.

Unit 3 **(14L+2T) hrs**

Minimum variance of estimators. Fisher information function and Fisher information matrix. Cramer-Rao inequality, Rao-Blackwell, and Lehmann-Scheffe theorems. Uniformly minimum variance unbiased (UMVU) estimation. Simultaneous estimation of parameters of multinomial and normal distributions. Ancillary statistics, Basu's theorem, and its application in UMVU estimation. Efficiency of estimators.

Unit 4 **(10L+2T) hrs**

Methods of estimation: Maximum likelihood (ML) and moment estimation. Properties of ML estimators. Computation of ML estimates using Newton-Raphson, method of scoring, least-square estimators and minimum chi-square estimators.

References

1. Casella, G. and Berger, R.L. (2002). *Statistical Inference, 2/e,* Duxbury Press, Belmont, California, USA.
2. Dudewicz, E.J. and Mishra, S.N. (1980). *Modern Mathematical Statistics,* John Wiley, NewYork.
3. Kale, B.K. and Muralidharan, K. (2015). *Parametric Inference: An Introduction*, Narosa, New Delhi.
4. Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation*, Springer, New York.
5. Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses,* 2/e, John Wiley, NewYork.

6. Rohatgi, V.K.and Saleh, A.K.Md.E. (2002): *An Introduction to Probability and Statistics*, John Wiley, NewYork.
7. Zacks, S. (1981). *Parametric Statistical Inference*, John Wiley, NewYork

BLUEPRINT

Code number: ST7221
Title of the paper: Theory of Point Estimation

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
16	12	I
16	12	II
22	16	III
16	12	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the student should be able to

CO1	Understand the concepts and importance of properties of estimators.
CO2	To compare the efficiency of various estimation techniques for real life problems.
CO3	To understand Fisher Information, Lower bounds to variance of estimators, MVUE and apply them in practical situations.
CO4	To apply various estimation procedures to deal with real life problems.

First Semester

ST 7323: Sampling Theory and Statistics for National Development

Semester	First
Paper Code	ST 7323
Paper Title	Sampling Theory and Statistics for National Development
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(12L + 3T) Hrs**

Random sampling methods: Simple random sampling, stratified random sampling, and systematic sampling. Determination of sample size. PPS sampling: Probability proportional to size (PPS) with replacement: Cumulative total method and Lahiri's scheme. Hansen-Hurwitz estimator. PPS without replacement: Des Raj estimator for a general sample size and Murthy's estimator for a sample of size2.Horvitz-Thompson estimator, Midzuno-Sen estimator

Unit 2 **(11L+2T) Hrs**

Cluster sampling with equal and unequal cluster sizes. Two-stage sampling with equal number of second stage units. SRS at both the stages. Non-Sampling errors: Non response: Hansen-Hurwitz technique, Deming's technique.Randomized response: Warner's model.

Unit 3 **(10L+2T)Hrs**

Two phase sampling: SRSWOR in both phases. Ratio and regression estimators based on SRSWOR sampling: Bias, mean squared error, and variance estimation. Double sampling for ratio and regression estimation.

Unit 4 **(10L+2T) Hrs**

Population growth models-exponential, logistic. Population projection using Leslie matrix.
Statistics for national development: Estimation of national income- product approach, income approach, and expenditure approach. GNP and GDP of India, HPI, CPI, Inflation and unemployment in relation to GDP. Measuring inequality in incomes, Gini coefficient, and Theil's measure. Poverty measurement: measures of incidence and intensity, combined measures, Kakwani and Sen Indices.

References

1. Chaudhuri, A. and Mukherjee, R. (1988). *Randomized Response: Theory and Techniques*, Marcel Dekker, NewYork.
2. Cochran, W. G. (1977). *Sampling Techniques,* 3/e; John Wiley, NewYork.
3. Des Raj and Chandok, P. (1998). *Sample Survey Theory*, Narosa Publishing House, NewDelhi.
4. Mukhupadhyay, P. (2009). *Theory and Methods of Survey Sampling*, 2/e, Prentice Hall, NewDelhi.
5. Murthy, M. N. (1977). *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta.
6. Sampath, S. (2006). *Sampling Theory and Methods*, 2/e, Narosa Publishing House, New Delhi.
7. Singh, D. and Chaudhary, F.S. (1986). *Theory and Analysis of Sample SurveyDesigns,* New Age International Publishers, New Delhi.
8. Sukhatme, P.V., Sukhatme, B.V., Sukhatme, Sand Ashok C. (1984). *Sampling Theory of Surveys with Applications,* Iowa State University Press,USA.

BLUEPRINT

Code number: ST 7323

Title of the paper: Sampling Theory and Statistics for National Development

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
20	15	I
12	09	II
16	12	III
22	16	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the student should be able to

CO1	Understand need and apply conceptually sampling
CO2	Able to use and apply various sampling methods
CO3	Able to compare various sampling techniques
CO4	Able to understand the importance of survey methods and sampling methods in the process of data collection.
CO5	Able to design a sampling plan for a statistical project.

First Semester

ST 7421: Mathematical Analysis and Linear Algebra

Semester	First
Paper Code	ST 7421
Paper Title	Mathematical Analysis and Linear Algebra
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(12L+2T) Hrs**

Sets, sequence and series of real numbers, Cauchy criteria for convergence, supremum and infimum, their properties, Interior points and limit points of subsets of R. Open and closed subsets of R. Bolzanao- Weierstrass theorem. Sequences and series of functions. Pointwise and uniform convergence. Cauchy criteria for convergence, Weierstrass-M test.

Unit 2 **(12+2T) Hrs**

Riemann-Stieltjes (R-S) integral of a bounded real valued function. Necessary and sufficient condition for R-S integrability. Properties of R-S integrals. Integration by parts. Change of variables in R-S integrals. Mean value theorems for R-S integrals.

Improper Riemann integrals. Convergence and absolute convergence of improper integrals.
Cauchy criteria for convergence of improper integrals.

Unit 3 **(8L+2T) Hrs**

Beta and Gamma integrals and their properties. Legendre's duplication formula. Integrals involving parameters. Improper integrals involving parameters and their uniform convergence. Differentiation under the integral sign: Leibnitz rule for integral with constant and variable limits. Maxima and minima of functions of several variables. Lagrangian multipliers. Double integrals.

Unit 4 **(12L+ 2T) Hrs**

Vector spaces, subspaces, linear dependence and independence, basis and dimension of a vector space, inner product and orthogonality of vectors, Gram-Schmidt orthogonalization process, orthonormal basis and its properties. Partitioned matrices: computation of determinant and inverse. Characteristic roots, characteristic vectors. Diagonalization, Singular value decomposition.

Systems of linear equations: consistency, existence of solutions, number of solutions, and solving the system of equations. Generalized inverse of a matrix and its properties.

References

1. Apostol, T.M. (1986). Mathematical Analysis, 2/e, Narosa Publishing House, New Delhi.
2. Bartle, R.G. (1975). The Elements of Real Analysis, 2/e, John Wiley.
3. Bilodeau, G.G., Thie, P.R., and Keough, G.E. (2010). An Introduction to Analysis, 2/e, Jones

and Bartlett (Indian Edition), New Delhi.
4. Goldberg, R.R. (1970). Methods of Real Analysis, Oxford and IBH Publishing Company, New Delhi.
5. Malik, S.C. and Arora, S. (1998). Mathematical Analysis, New Age, New Delhi.
6. Rudin, W. (2013). Principles of Mathematical Analysis, 3/e, Indian Print, Tata McGrawhill, New Delhi.
7. Bapat, R.B. (2011). Linear Algebra and Linear Models. Springer and Hindustan Book Agency.
8. Beezer, R. A. (2004). A First Course in Linear Algebra, Congruent Press, Washington
9. Hohn, F. E. (1973). Elements of Matrix Algebra, Macmillan
10. Kollo, T. and Rosen, D. von (2005). Advanced Multivariate Statistics with Matrices, Springer, New York.
11. Kumaresan, S. (2000). Linear Algebra: A Geometric Approach, Prentice Hall

BLUEPRINT

Code number: ST7421

Title of the paper: Mathematical Analysis and Linear Algebra

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
19	14	I
19	14	II
13	10	III
19	14	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the student should be able to

CO1	Understand various concepts of Riemann-Stieltjes integrals and their properties
CO2	Get an insight into improper integrals, their convergence and applications
CO3	Find extremes of functions of several variables under constraints
CO4	Understanding the basic concepts of matrix theory.
CO5	Apply principles of matrix algebra to linear transformations.

First Semester

ST 7521: Statistical Computing

Semester	First
Paper Code	ST 7521
Paper Title	Statistical Computing
Number of teaching hrs per week	2+1
Total number of teaching hrs per semester	39
Number of credits	2

Unit 1 **(8L+ 2T) Hrs**

(a) Theory of inverse transformation method (ITM) for random variate generation- definition of quantile function, its properties. Quantile function as a random variable and its distribution function. ITM based algorithms to generate random variates from standard discrete and continuous distributions.

(b) Random variate generation from bivariate and conditional distributions.

(c) Theory of random number generation -linear, multiplicative and mixed random number generators. Selection of a random number generator

Unit 2 **(8L+ 3T) Hrs**

Numerical algorithms as grid search, interpolation search, gradient search, Bisection and Newton-Raphson methods, Aitkens extrapolation, Simple applications of the above methods.

Unit 3 **(6L+ 2T) Hrs**

Probability and Distributions: Random sampling and combinatory; obtaining density, cumulative density and quantile values for discrete and continuous distributions; generating samples from discrete and continuous distributions; Plotting density and cumulative density curves; Q-Q plot.

Unit 4 **(8L+ 2T) Hrs**

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. Verification of WLLN, CLT and other approximations through simulation.

Reference

1. Introductory Statistics with R by Peter Dalgaard, Springer, 2ndedition, 2008.

2. The R Book by Michael J. Crawley, John Wiley and Sons, Ltd., 2007.Lab

BLUEPRINT

Code number: ST7521

Title of the paper: Statistical Computing

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
7	10	I
7	11	II
5	08	III
6	10	IV
25	**39**	**TOTAL**
Maximum marks for the paper (Including bonus question): 2**5**		

Course Outcomes: At the end of the Course, the student should be able to

CO1	Generate random numbers through various algorithms
CO2	Compare various random number generators and use them to develop simulation models
CO3	Construct a probability model for a given set of data and carryout goodness of fit procedures.
CO4	Do essential computing to apply numerical algorithms using excel and R to solve real life industry problems.
CO5	Design simulation study to solve real life problems.

ST 7P1: Practical I

Semester	First
Paper Code	ST 7P1
Paper Title	Practical I (Based on ST 7221, ST 7321)
Number of teaching hrs per week	4
Total number of teaching hrs per semester	44
Number of credits	2

List of Assignments

1. Maximum likelihood estimation - 1 (for standard probability models)
2. Maximum likelihood estimation - 2 (Implementing Likelihood functions in R-software and use of optimize () and optim(), Newton Raphson Method , Expectation-Maximization Algorithm)
3. Method of moments.
4. Method of scoring.
5. Least square estimation.
6. Simple, Stratified and Cluster Sampling
7. PPSWR
8. PPSWOR
9. Ratio and regression methods of estimation
10. Statistics for National Development - 1: Estimation of National Income, Income Inequality
11. Statistics for National Development - 2: Population Growth Models

ST 7P2: Practical – II

Semester	First
Paper Code	ST 7P2
Paper Title	Practical - II (Based on ST7421 and ST7521)
Number of teaching hrs per week	4
Total number of teaching hrs per semester	44
Number of credits	2

List of Assignments

1. Calculation of determinant and rank of a matrix.
2. Calculation of inverse and Moore-Penrose inverse
3. Calculation of eigen values, eigen vectors and g-inverse
4. Systems of linear equations.
5. Random number generation: Inverse transformation method
6. Random number generation: bivariate and conditional distributions
7. Numerical Algorithms: Applications of Gradient search, Newton-Raphson, Mullers Method
8. Generating samples from standard probability distribution
9. Plotting density and cumulative density curves
10. Monte Carlo integration and its application to compute expected values and probabilities
11. Verification of WLLN, CLT and other approximations through simulation.

Second Semester
ST 8121: Distribution Theory

Semester	Second
Paper Code	ST 8121
Paper Title	Distribution Theory
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit I **(12L + 2T) Hrs**

Recapitulation of random variables and probability distributions, Mixtures of probability distributions, Functions of random variables, symmetry of distribution, standard discrete and continuous distributions: Binomial, Poisson, Negative Binomial, Normal, Gamma, Pareto, Weibull. Truncated distributions (Binomial, Poisson, Normal), Applications of truncated distribution.

Unit 2 **(12L + 2T) Hrs**

Random vectors, distribution function of random vectors and its properties, joint m.g.f., joint c.f., mixed moments, variance-covariance matrix, multinomial distribution, multivariate normal distribution: pdf, mean vector, dispersion matrix, mgf. Bivariate exponential distributions- Gumbel type I, Gumbel type II, Marshall Olkin. Independence of random variables, convolutions of random variables, Jacobian transformation, conditional expectation and variances.

Unit 3 **(10L + 2T) Hrs**

Sampling distributions: central chi-square, t and F distributions: pdf, mgf, moments. Non-central chi-square, non-central t and non-central F distributions: their pdf.
Distribution of quadratic forms. Fisher Cochran theorem (Statement only) of statistics from univariate normal random samples, distributions of linear and quadratic forms involving normal random variables.

Unit 4 **(10L + 2T) Hrs**

Order statistics, Joint distribution of order statistics, distribution of r-th order statistic, joint distribution of r^{th} and s^{th} order statistics (r < s). Distribution of sample median and sample range, Illustration of independence of $_{(r)}\ and\ X_{(s)} - X_{(r)}\ for$ r < s for exponential, distribution of spacings, normalized spacings with illustration to exponential case.

Reference

1. Berger, R. and Casella G. (2002). Statistical Inference, Duxbury Resource Center, Second Edition.
2. Dasgupta, A. (2010) Fundamentals of Probability: A First Course, Springer, New York.
3. Rohatgi, V. K. & A. K. M. E Saleh (2001). Introduction to Probability and Statistics, Wiley.
4. Hogg, R. V. McKean, J. W. and Craig, T. T. (2005). Introduction to Mathematical Statistics, Sixth Edition, Pearson Prentice Hall, New Jersey.

5. Rao, C. R. (2002). Linear Statistical Inference and Its Applications, Wiley

BLUEPRINT

Code number: ST8121

Title of the paper: Distribution Theory

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
19	14	I
19	14	II
16	12	III
16	12	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able to

CO1	Understand the various probability distributions and their applications in real life.
CO2	Define and explain the different statistical distributions and the typical phenomena that each distribution often describes.
CO3	Apply key concepts of mean, variance, independence and conditional expectations and variances.
CO4	Apply problem solving techniques to solve real life problems involving uncertainty.

Second Semester

ST 8221: Testing of Hypothesis and Interval Estimation

Semester	First
Paper Code	ST 8221
Paper Title	Testing of Hypothesis and Interval Estimation
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(10L+2T)**

Formation of null hypothesis, simple and composite hypothesis, One tail and two-tail test, critical region, test functions, two types of errors, size and power of a test, level of significance, power function, expressions for size and power of a test, P-value of a test. Randomized and non-randomized test. Most Powerful (MP) test for testing simple hypothesis against simple alternative, Neyman-Pearson Lemma, illustrations.

Unit 2 **(10L+2T)**

Distributions with Monotone Likelihood Ratio (MLR) property, Uniformly Most Powerful (UMP) test for testing one tailed null hypothesis against one tailed alternative. Existence of UMP test and Illustrations. Extension of these results in Pitman family when only upper or lower end points depend on the parameter.

Non-existence of UMP test for testing null hypothesis against two-sided alternative. Unbiased test. Existence of Uniformly Most Powerful Unbiased (UMPU) test for testing null hypothesis against two-sided alternative. Illustrations. Neyman-Pearson generalized lemma.

Unit 3 **(14L+2T)**

Likelihood Ratio Test Procedure (LRTP), Asymptotic properties of LR test statistic, Pearson's chi-square test for goodness of fit, Assumptions for the validity of chi-square test, Bartlett's Test for homogeneity of variances. Wald’s Sequential Probability Ratio Test (SPRT), Wald’s equation, Score tests. Illustrations.

Unit 4 **(10L+ 2T)**

Confidence level and confidence coefficient, Interval estimation-confidence sets, Relation between confidence sets and testing of hypothesis, Shortest expected length confidence interval. Evaluating interval estimators using size and coverage probability and test-related optimality. Uniformly Most Accurate (UMA) confidence interval (One-sided interval) and its relation with UMP test.

References:

1. Casella G. and Berger R.L. (2002): Statistical Inference, Wadsworth Grou.
2. Gibbons J.D. (1971): Nonparametric Inference, McGraw-Hill.
3. Kale B.K. (1999): A First Course on Parametric Inference, Narosa Publishing House.

4. Kendall M.G. and Stuart A. (1968): The Advanced Theory of Statistics, Vol.II,Charles Griffin and Co.
5. Lehmann E.L. (1986): Testing Statistical Hypotheses, John Wiley.
6. Pratt T.W. and Gibbons, J.D. (1981): Concepts of Nonparametric Theory, Springer.
7. Rao C.R. (1973): Linear Statistical Inference and Its Applications, Wiley Eastern.
8. Silvey S.D. (1970): Statistical Inference, Chapman & Hall.

BLUEPRINT

Code number: ST8221
Title of the paper: Testing of Hypothesis and Interval Estimation

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
16	12	I
16	12	II
22	16	III
16	12	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able to

CO1	Draw Inference about unknown population parameters based on random samples.
CO2	Learn the basics of testing of hypothesis and understand MP and UMP tests.
CO3	Learn to construct confidence intervals for population parameters based on various statistical methods.
CO4	Apply LRT test and large sample tests for real life problems.
CO5	Identification of appropriate test for a given scenario for real life problems.

Second Semester

ST 8321: Multivariate Analysis

Semester	Second
Paper Code	ST 8321
Paper Title	Multivariate Analysis
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 (10L+2T)hrs

Introduction to Multivariate analysis, need and applications. Multivariate Normal Distribution (MVN). Sampling from MVN distribution – sample mean vector, sample variance covariance matrix. Distribution of the sample mean vector, independence of sample mean vector and sample variance covariance matrix. Assessing multivariate normality. Q-Q and chi-square plots. Multiple and Partial Correlation coefficients and their distributions (statements only).

Unit 2 (10L+2T)hrs

MLE's of the parameters of multivariate normal distribution and their sampling distributions. Likelihood ratio tests: Tests of hypotheses about the mean vectors and covariance matrices for multivariate normal populations.

Classification and discriminant analysis: Bayes, minimax, and Fisher's criteria for discrimination between two multivariate normal populations. Sample discriminant function. Tests associated with discriminant functions. Probabilities of misclassification and their estimation. Discrimination for several multivariate normal populations. Canonical discriminant function.

Unit 3 (12L+2H) hrs

Principal component analysis-sample principal components asymptotic properties. Canonical variables and canonical correlations: definition, estimation, computations. Test for significance of canonical correlations.

Multivariate regression model: Estimation of parameters, testing of linear hypotheses about regression coefficients. Multivariate analysis of variance (MANOVA) for one-way and two- way classified data.

Unit 4 (12L+2H) hrs

Cluster Analysis: distances and similarity measures, hierarchical clustering methods, K – means method. Multidimensional scaling: nature of the problem, classical solution.

Factor analysis: Orthogonal factor model, factor loadings, estimation of factor loadings, factor scores. Applications.

References

1. Anderson, T. W. (2004). *An Introduction to Multivariate Statistical analysis,* 3/e, John Wiley, NewYork.
2. Giri, N. C. (1977). *Multivariate Statistical Inference,* Academic Press, NewYork.
3. Johnson, R. A. and Wichern, D.W. (2003). *An Introduction to Multivariate Statistical Analysis*, 5/e, PearsonEducation.
4. Kshirsagar, A. M. (1972). *Multivariate Analysis,* Marcel Dekker, NewYork.
5. Morrison, D. F. (2005). *Multivariate Statistical Methods,* 4/e, McGrawhill, NewYork.
6. Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory,* John Wiley, New York.
7. Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2/e, John Wiley, New York.
8. Seber, G. A. F (1984). *Multivariate Observations*, John Wiley, NewYork.
9. Sharma, S. (1996). *Applied Multivariate Techniques*, John Wiley, NewYork.
10. Srivastava, M. S. (1979). *An Introduction to Multivariate Statistics*, NorthHolland.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press, NewYork.

Code number: ST8321

Title of the paper: Multivariate analysis

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
16	12	I
16	12	II
19	14	III
19	14	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able to

CO1	Understand multivariate normal distribution and their real-life applications
CO2	Understand Wishart distribution, Hotelling T^2 and Mahalanobis D^2 statistic.
CO3	Implement dimension reduction techniques using software on real life problems.
CO4	Demonstrate knowledge and understanding of the basic ideas behind discriminant and clustering analysis techniques with applications.

Second Semester

ST 8421: Linear models And Regression Analysis

Semester	Second
Paper Code	ST 8421
Paper Title	Linear Models and Regression analysis
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(10L+4T)hrs**

Linear model. Estimability of linear parametric functions in Gauss-Markov model, least squares estimation, BLUE, Gauss-Markov theorem. Distributional properties of least squares estimators, Independence of BLUE's and residual sum of square. Estimation under linear restrictions involving estimable functions.

Confidence intervals for BLUE's, General linear hypotheses, testable hypotheses. Likelihood ratio test procedure.

Unit 2 **(10L+2T) hrs**

Multiple linear regression, estimation, and properties. Prediction of new observations and prediction interval. Hidden extrapolation. Use of dummy variables. Generalized linear models: link function, binary logistic regression, and Poisson regression.

Unit 3 **(10L+2T) hrs**

Measures of model adequacy, coefficient of determination R^2, lack of fit test, residuals, scaling residuals, and residual analysis: residual plots as tests for departure from assumptions of homoscedasticity, normality (Q-Q plot), non-linearity, and detection of outliers. Detecting influential observations. Transformations: Box-Cox transformation and transforming the predictors. Subset selection of regressors: Mallow's C_p statistic, all possible, stepwise, forward and backward regressions.

Unit 4 **(12L+2T)hrs**

Heteroscedasticity and autocorrelation: sources, consequences, detection, and remedial procedures.

Multicollinearity: sources, consequences, detection, and remedial procedures. Ridge regression and generalized least squares. Validation of regression models. Analysis of model coefficients and predicted values. Collecting fresh data. Data splitting.

References

1. Cook, R. D. and Weisberg, S. (1982). *Residual and Influence in Regression,* Chapman and Hall,London,UK.
2. Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis, 3/e,* John Wiley, NewYork.
3. Gunst, R. F. and Mason, R. L. (1980). *Regression Analysis and Its Applications - A Data Oriented Approach*, Marcel and Dekker, NewYork.
4. Montgomery, D. C. and Peck, E. A., and Vining, G. G. (2012). *Introduction to Linear Regression*, 5/e, John Wiley, NewYork.
5. Ryan, T. P. (1997). *Modern Regression Methods*, John Wiley, NewYork.

6. Searle, S. R. (1971). *Linear Models,* John Wiley, NewYork.
Weisberg, S. (1985). *Applied Linear Regression*, John Wiley, NewYork.

BLUEPRINT

Code number: ST8421
Title of the paper: Linear Models and Regression analysis

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
19	14	I
16	12	II
16	12	III
19	14	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student will able to

CO1	Apply simple linear regression model to real life examples.
CO2	Understand multiple linear regression models with applications and concept of Multicollinearity and autocorrelation.
CO3	Compute multiple and partial correlation and checking residual diagnostic to validate model.
CO4	Apply Logistic and Poisson regression models and its implementation in real life situation

Second Semester

ST 8521: Introduction to Data Science

Semester	Second
Paper Code	ST 8521
Paper Title	Introduction to Data Science
Number of teaching hrs per week	2+1
Total number of teaching hrs per semester	39
Number of credits	2

Unit 1 **(10L+ 2T) Hrs**

a) Data processing using MS-Excel. Graphical and diagrammatic representation, exploring various functions, Data Analysis and Solve add-on packages in MS-Excel
b) R language Essentials: Expressions and objects, Assignments, creating vectors, vectorized arithmetic, creating matrices, operations on matrices, lists, data frames - creation, indexing, sorting and conditional selection; examples.
c) R Programming: conditional statements -if and if else; loops -for, while, do-while; functions -built-in and user defined; Data entry -reading from text file, data editor; examples. Data and file handling in R. Packages in R, installation, loading, accessing inbuild functions, use of help function. User defined functions- Exercises
d) Introduction to Python: Python Essentials; Data Management using dplyr package
e) Graphical Representation using R/Python and Tableau

Unit 2 **(7L+ 3T) Hrs**

Introduction to Big Data, structured and unstructured data, Data Management: Data cleaning; Missing value imputation techniques/methods, Outlier detection techniques: Graphical and analytical methods; Exploratory Data Analysis, Descriptive statistics, importance of EDA through case studies.

Unit 3 **(12L+ 5T) Hrs**

Model Building: Multiple linear regression, model adequacy-residual analysis, model selection and evaluation. Classification problems: decision trees, support vector machine; Random forest model. Clustering: k-means

Note: Above concepts/topics to be covered only through case studies

Reference

1. Jiawei Han, Micheline Kamber (2002), “Data Mining: Concepts and Techniques”, Morgan Kaufmann Publishers.
2. Margaret H.Dunham (2003), “Data Mining Introductory and Advanced Topics”, Pearson Education.
3. Alex Berson,Stephen J. Smith (2004), “Data Warehousing, Data Mining,& OLAP”, Tata McGrawHill.
4. Ralph Kimball (2007), “The Data Warehouse Life Cycle Toolkit”, John Wiley & Sons Inc.
5. O'Neil and Schutt (2013) “Doing Data Science” Shroff Publishers
6. Provost and Fawcett (2013) “Data Science for Business”, Shroff Publishers
7. Fung (2013) “Numbersense: how to use big data to your advantage” McGraw-Hill.
8. James, G., Witten, D., Hastie, T., Tibshirani, R. An introduction to statistical learning with applications in R. Springer, 2013.
9. Han, J., Kamber, M., Pei, J. Data mining concepts and techniques. Morgan Kaufmann, 2011.
10. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning, 2nd edition. – Springer, 2009.
11. Practical Data Science with R”. Nina Zumel, John Mount. Manning, 2014
12. “Data Science for business”, F. Provost, T Fawcett, 2013

BLUEPRINT

Code number: ST8521

Title of the paper: Introduction to Data science

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
4	6	I
5	8	II
8	12	III
8	13	IV
25	**39**	**TOTAL**
Maximum marks for the paper (Including bonus question): 2**5**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To develop practical data analysis skills, which can be applied to practical problems
CO2	To develop practical skills needed in modern analytics
CO3	To give a hands-on experience with real-world data analysis
CO4	To develop applied experience with data science software, programming and applications

Second Semester

ST 8P1: Practical – III

Semester	Second
Paper Code	ST8P1
Paper Title	Practical- III (Based on ST 8121, ST 8221 and ST 8321)
Number of teaching hrs per week	4
Total number of teaching hrs per semester	44
Number of credits	2

List of Assignments

ST 8121: Distribution Theory
ST 8221: Testing of Hypothesis and Interval Estimation
ST 8321: Multivariate Analysis

1. Sampling distribution
2. Order Statistics
3. Formulation of hypothesis, Size of the test, power of the test and plotting power function.
4. Most powerful tests
5. UMP one sided test including plotting of power function
6. UMPU test based on one parameter exponential family.
7. Interval estimation
8. Likelihood ratio test for finite sample based on one and two independent sample from normal distribution and exponential distribution.
9. Bartlett test for homogeneity of variances.
10. Discriminant Analysis
11. Principal component analysis and Factor Analysis

Second Semester

ST 8P2: Practical – IV

Semester	Second
Paper Code	ST8P2
Paper Title	Practical- IV (Based on ST 8421 and ST 8521)
Number of teaching hrs per week	4
Total number of teaching hrs per semester	44
Number of credits	2

List of Assignments

ST 8421: Linear Models and Regression Analysis
ST 8521: Introduction to Data Science

1. Computation of mean vector, covariance matrix, partial and multiple correlations. from a multivariate data.
2. Fitting multiple linear regression models - Stepwise regression
3. Multicollinearity diagnostics.
4. Residual analysis.
5. Tests for autocorrelation.
6. Fitting a ridge regression model.
7. Fitting logistic and Poisson regression models
8. Exploratory data analysis
9. Data Modeling
10. Graphic displays
11. Mining Databases

Third Semester

ST 9120: Stochastic Processes

Semester	Third
Paper Code	ST 9120
Paper Title	Stochastic Processes
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(12L+3T) Hrs**

Introduction to Stochastic Processes, Markov chain, one-step transition probabilities, Chapman-Kolmogorov equations, calculation of n-step transition probability and its limit, first passage probabilities. Stationary distribution.

Unit 2 **(10L + 3T) Hrs**

Classification of states, transient Markov chain, absorption probabilities, absorption and recurrence times, random walk and gambler's ruin problem. Estimation of transition probabilities of a Markov chain, applications.

Unit 3 **(12L+2T) Hrs**

Discrete state space continuous time MC: Kolmogorov-Feller differential equations; Poisson process, pure birth process, birth and death process, applications to queuing theory. Renewal process, renewal function. Renewal equation. Elementary renewal theorem and its applications.

Unit 4 **(8L+2T) Hrs**

Branching process: Galton-Watson branching process, probability of ultimate extinction, distribution of population size, Martingale in discrete time.

BLUEPRINT

Code number: ST 9120

Title of the paper: Stochastic Processes

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
20	15	I
18	13	II
19	14	III
13	10	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To construct transition matrices for Markov dependent behaviour and summarize process information.
CO2	To understand the principles and objectives of the model building based on Markov chains and Poisson processes.
CO3	To learn random walk and gambler's ruin problem
CO4	To use notions of long-time behaviour including transience, recurrence, and equilibrium in applied situations such as branching processes

References

1. Karlin, S and Taylor, H. M. (1975). *A First Course in Stochastic Processes*, Academic Press, New York.
2. Bhat, B. R. (2000). *Stochastic Models*, New Age International, New Delhi.
3. Medhi, J. (2017). *Stochastic Processes*, 4/e. New Age International, New Delhi.
4. Ross, S.M. (1996). *Stochastic Processes*, 2/e, John Wiley, New York.

5. A.K. Basu (2003): *Introduction to Stochastic Processes*, Narosa Publishers.

6. Cox, D.R. and Miller, H.D. (2001): *The theory of Stochastic Processes,* Chapman & Hall.

7. Gallager, R.G. (2013): *Stochastic Processes: Theory for applications*, Cambridge University press.

8. Jones, P.W. and Smith, P. (2017): *Stochastic processes: An introduction,* Chapman & Hall.

Third Semester

ST 9220: Machine Learning Techniques

Semester	Third
Paper Code	ST 9220
Paper Title	Machine Learning Techniques
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

UNIT 1 **(6L+1T) hrs**

Introduction to Machine Learning: Definitions of Machine Learning, Paradigms of Machine Learning- Supervised, Unsupervised and Reinforcement Learning with examples, Regression versus Classification Problems with examples, Batch Learning and Online Learning, Instance Based Versus Model Based Learning, Training, Testing and Validation

Challenges of Machine Learning- Prediction Accuracy and Model Interpretability Trade-off, Bias-Variance Trade off (Underfitting and Overfitting), Insufficient Quantity of Training Data, Non-representative Training data, Definition of Feature Engineering.

UNIT 2 **(4L+2T) Hrs**

Computational Foundations: Basic Python Using Jupiter Notebooks, Scientific Computing using Numpy, SciPy and Matplotlib, Data Preprocessing techniques- Handle Missing Values, Label Encoding and One-hot Encoding, Standardization etc., Data Wrangling and Preprocessing using Pandas, Machine Learning with Scikit Learn.

UNIT 3 **(16L+4T) hrs**

Multiple Linear Regression- Interaction, Transformations, Categorical Predictors.

Logistic Regression.

Performance Metrics for Regression and Classification- MSE, Confusion Matrix, ROC Curve, AIC, BIC

Resampling Methods- Cross Validation and Bootstrap

Polynomial Regression, Regression Splines, Generalized Additive Methods

Dimensionality Reduction

Curse of Dimensionality, Subset Selection and Regularization- Ridge, Lasso, Elastic Net

Principal Component Analysis, Kernel PCA, Considerations in High Dimensions.

UNIT4 **(15L + 4T) hrs**

Tree Based Models:

Decision Trees- Regression Trees, Classification Trees, Trees vs Linear Models

Ensemble Models- Bagging, Random Forests, Feature Importance, Boosting-Adaboost, Gradient Boosting.

Support Vector Machines:

Linear SVM Classification, Non-Linear SVM Classification, SVM Regression

Unsupervised learning

Challenge of Unsupervised Learning, K Means, PCA, Hierarchical Clustering, Gaussian Mixture Models

BLUEPRINT

Code number: ST 9220

Title of the paper: Machine Learning Techniques

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
9	07	I
8	06	II
27	20	III
26	19	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To understand the basic theory underlying machine learning.
CO2	To be able to formulate machine learning problems corresponding to different applications.
CO3	To understand a range of machine learning algorithms along with their strengths and weaknesses.
CO4	To apply the algorithms to a real-world problem, optimize the models learned and report on the expected accuracy that can be achieved by applying the models.

References

1. Alpaydin, E. (2014). *Introduction to Machine Learning*, 3e., MIT Press, USA.
2. Bishop, C. (2007). *Pattern Recognition and Machine Learning*, Springer, New York.
3. Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. CRC Press, London.
4. Jiawei Han, Micheline Kamber. (2002). *Data Mining-Concepts and Techniques*, Morgan Kaufman Publishers, U.S.A
5. Raschka, S., Mirjalili,V. (2019). *Python Machine learning,* Packt Publishing, UK.
6. Gareth, J., Daniela Witten, Trevor, H., Robert, T. (2013). *An Introduction to Statistical Learning: With Applications in R,* Springer, New York.
7. Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The Elements of Statistical Learning,* Springer series in statistic, New York, U.S.A.

Third Semester

ST 9323: STATISTICAL QUALITY CONTROL

Semester	Third
Paper Code	ST 9323
Paper Title	Statistical Quality Control
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(14L + 2T) Hrs**

Concept of quality. Quality function and quality characteristics. Quality assurance - its evolution and modern trends. Chance and assignable causes. Principles of a process control chart and associated decision rules. The seven QC tools. Design quality and conformance quality. Quality costs. Quality and productivity. Analysis of OC, ARL, and other measures. Techniques for improving sensitivity of a chart. Process capability and its measures. Design and implementation of SPC and six-sigma programmes, BIS and ISO certification.

Unit 2 (10L+2T) hrs

Shewhart control charts for monitoring process level and process dispersion. Rational subgroups. Pre-control and analysis of patterns on a control chart, CUSUM and EWMA charts, Process control with autocorrelated observations. Modifications of Shewhart control chart. Multivariate control charts.

Unit 3 (10L+2T) hrs

Acceptance sampling plan: Principle of acceptance sampling plans. Single and Double sampling plan their OC, AQL, LTPD, AOQ, AOQL, ASN, ATI functions with graphical interpretation, use and Dodge-Romig and MIL-STD systems- interpretation of Dodge and Romig's sampling inspection plan tables.

Unit 4 (14L+2T) hrs

Sequential Probability Ratio Test; concept and operational definition. Determination of stopping bounds A and B, OC and ASN functions of SPRT for testing the mean of a normal distribution with known variance. Statement of the optimal property of SPRT.

References

1. Alwan, L.C. (2000). *Statistical Process Analysis,* McGraw Hill, New York.
2. Grant, E. L. and Leavenworth, R. S. (1996). *Statistical Quality Control*. 7th edition, McGrawHill, New York.
3. Mittage, H.J. and Rinne, H. (1993). *Statistical Methods of Quality Assurance,* Chapman and Hall, London, UK.

4. Montgomery, D.C. (2012). *Introduction to Statistical Quality Control*, 7/e, John Wiley, New York.
5. Smith, G.M. (1991). *Statistical Process Control and Quality Improvement*, 3/e, Prentice Hall, New York.
6. Wetherill, G.B. and Brown, D.W. (1991). *Statistical Process Control: Theory and Practice*, Chapman and Hall, London, UK

Code number: ST 9323

Title of the paper: Statistical Quality Control

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
22	16	I
19	14	II
16	12	III
13	10	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question): **70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	Introduce the principles and techniques of Statistical Quality Control and their practical uses in product and/or process design and monitoring
CO2	Demonstrate the approaches and techniques to assess and improve the process and/or product quality.
CO3	Illustrate the basic concepts and techniques of reliability theory
CO4	To apply the techniques of statistical quality control and reliability to solve real life problems.

Third Semester

STDE 9420: OPTIMIZATION TECHNIQUES (ELECTIVE- I)

Semester	Third
Paper Code	STDE 9420
Paper Title	Optimization Techniques
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit I **(14L+2T) hrs**

Linear Programming: Review of basic concepts, Computational complexity of LPP, Ellipsoid method, Polynomial time algorithm, Karmarkar's polynomial time algorithm, Convergence and complexity,
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, Dynamic programming, sensitivity analysis, Bellman's optimality principle.

Unit II **(10L+2T) hrs**

General Transportation problem, Transportation matrix, L.P formulation of a Transportation problem. Finding an initial Basic feasible solution: NWCR, LCM and VAM. Test for optimality: Stepping stone algorithm and modified distribution method. Problem of degeneracy and its resolving degeneracy. Unbalanced and maximization transportation problem, Assignment problem: Mathematical formulation, Hungarian algorithm. Travelling salesman problem.

Unit III **(10L+2T) hrs**

Nonlinear programming: Karush-Kuhn-Tucker conditions, Convexity, Quadratic programming, Wolfes and Beales algorithms for solving quadratic programming problems.

Unit IV **(10L+2T) hrs**

Networking models: Network flows, maximal flow in the network, Transportation problems, transhipment problems and assignment problems as networking problems. Network scheduling by CPM and PERT. Critical path. Float and slack. Resource Analysis in network scheduling.

Code number: STDE 9420

Title of the paper: Optimization Techniques

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
22	16	I
16	12	II
16	12	III
16	12	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	Enumerate the fundamental knowledge of Linear Programming.
CO2	Learn classical optimization techniques and numerical methods of optimization
CO3	Explain Integer programming techniques and apply different optimization techniques to solve various models.
CO4	To apply the optimization techniques to solve real life problems.

References

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

Third Semester

STDE 9520: OPERATIONS RESEARCH (ELECTIVE- I)

Semester	Third
Paper Code	STDE 9520
Paper Title	Operations Research
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(12L+2T) hrs**

Linear Programming: Review on basics of LPP, Simplex algorithm, Big-M and two-phase method. Primal and dual LPP. Dual simplex method. Transportation problem and test for optimality. Assignment problem.

Unit 2 **(14L+2T) hrs**

Integer programming: Pure and mixed Integer programming problems. Cutting plane methods - Gomory’s algorithms. Branch and bound technique. Zero-one programming.

Nonlinear programming: Formulation of nonlinear programs. Unconstrained and constrained optimization problems. The Lagrangian method. Karush-Kuhn-Tucker optimality conditions. Quadratic programming. Wolfe’s modified simplex method.

Unit 3 **(10L+2T) hrs**

Queueing theory: General description and characteristics of a queueing system. M/M/1 and M/M/c queueing systems and their waiting time distributions. M/M/1/N and M / M / ∞ queues. Transient solution of M / M / ∞ queueing system. Introduction to Non-Markovian queues.

Unit 4 **(8L+2T) hrs**

Inventory models: Basic characteristics of inventory systems (models). ABC analysis. Deterministic inventory systems EOQ Models with quantity discounts, price breaks, and storage limitations. Multiperiod dynamic inventory models. Continuous review stochastic inventory systems. The (s, S) policy.

Code number: STDE 9520

Title of the paper: Operations Research

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
19	14	I
22	16	II
16	12	III
13	10	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question): **70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	Identify and develop operational research models from the verbal description of the real system.
CO2	Understand the mathematical tools that are needed to solve optimisation problems.
CO3	Use mathematical software to solve the proposed models.
CO4	Develop a report that describes the model and the solving technique, analyse the results in the decision-making processes

References

1. Gross, D and Harris, C. M. (1986). *Fundamentals of Queueing Theory*, 2/e, John Wiley.
2. Taha, H. A. (2002). *Operations Research*, 7/e; Macmillan.
3. Medhi. J. (1991). *Stochastic models in queueing theory*, Academic Press.
4. Bazaara, M. S. and Shetty, C. M. (1979). Nonlinear *Programming: Theory and Algorithms*, John Wiley, New York.
5. Hillier, F. S. and Liebermann, G. J. (1986). *Introduction to Operations Research*, Holden Day, New York.
6. Kambo, N. S. (1991). *Mathematical Programming Techniques*, Affiliated East-West Press, New Delhi.
7. Murthy, K. G. (1995). *Operations Research: Deterministic Optimization Models*, Prentice Hall, New Delhi.
8. Swarup, K. et. al. (1985). *Operations Research*, Sultan Chand and Co., New Delhi.

9. Wayne, L. W. (1996). *Introduction to Mathematical Programming*, 2/e, Duxbury Press, New York.

Third Semester

ST 9623: RELIABILITY THEORY

Semester	Third
Paper Code	ST 9623
Paper Title	RELIABILITY THEORY
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit I **(9L+2T Hrs)**
Introduction to Reliability and its needs; Structural properties of coherent system: components and systems, coherent structures, representation of coherent systems in terms of paths and cuts, relevant & irrelevant structure; Reliability of a coherent systems; Reliability importance of components.

Unit II **(8L+2T Hrs)**
Life Distributions: Concept of distribution function, hazard function, Reliability function, Mean Time to Failure, Bathtub failure rate; loss of memory property of Exponential distribution - parametric families of some common life distributions -life distributions and its characterization - Reliability estimation of parameters in these models.

Unit III **(13L+3T Hrs)**
Classes of lifetime distributions. Concept of ageing, positive and negative ageing, IFR, IFRA, NBU, NBUE, DMRL classes of distributions and their dual classes. Interrelations among the classes of life time distributions. Closures of these classes under formation of coherent systems, convolutions, and mixtures. System Reliability: Series and parallel systems. k-out- of -n system. Structure function and block diagrams of these systems. Bounds on System Reliability. System life as a function of component lives. Expected system lifetime.

Unit IV **(12L+3T Hrs)**
Reliability modelling: Introduction to shock models- Univariate shock models: cumulative damage model, shock models leading to univariate IFR, Successive shock model, stress-strength models- Concepts and its estimation for some life distribution and proportional hazard model, Maintenance policies: Age replacement and block replacement policies and their characteristics.

References:

1. Barlow, R.E. and Proschan F. (1985) Statistical Theory of Reliability and Life Testing; Rinehart and Winston.
2. Lawless, J.F. (2003): Statistical Models and Methods of Life Time Data; John Wiley.
3. Zacks, S (1992): Introduction to Reliability Analysis, Springer Verlag.
4. Deshpande, J. V. and Purohit, S. G. (2005). Life Time Data: Statistical Models and Methods, World Scientific.
5. Ross, S.M. (2010). *Introduction to Probability Models,* 10/e, Academic Press, New York.
6. Bhat, B. R. (2000). *Stochastic Models*, New Age International, New Delhi.

7. Sinha, S.K. (1986). Reliability and Life Testing. John Wiley and sons.

Code number: ST 9623

Title of the paper: Reliability Theory

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
15	11	I
13	10	II
22	16	III
20	15	IV
70	**35**	**TOTAL**
Maximum marks for the paper (Including bonus question): **70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	Describe and discuss the key terminology, concepts and techniques used in statistics
CO2	Discuss critically the uses and limitations of statistical analysis
CO3	To gain effective skills to perform data analysis using statistical tools
CO4	To apply the statistical tools to solve real-life problems

References:

1. Harry, F., and Steven C. (1997). *Statistics: Concepts & Applications*, Cambridge University Press.
2. Medhi, J. (1992). *Statistical Methods: An Introductory Text*, Wiley Eastern Limited.
3. D. A., Lind, W. C. Marchal, and S. A. Wathen (2012). *Basic Statistics for Business and Economics,* Mc Graw Hill, London.
4. R.C. Campbell. (1974). *Statistics for Biologists*, Cambridge University Press
5. Christopher Chatfield. (1981). *Statistics for Technology*, Chapman and Hall
6. Douglas A. Lind, William C. Marchal, Samuel A. Wathen (2012). *Basic Statistics for Business & Economics,* McGraw-Hill Education
7. Sheldon M Ross. 2007. *Introductory Statistics*. Elsevier.

Third Semester

ST 9P1: Practical V (based on ST 9120 and ST 9320)

Semester	Third
Paper Code	ST 9P1
Paper Title	Practical V (based on ST 9120 and ST 9320)
Number of teaching hrs per week	4
Total number of teaching hrs per semester	44
Number of credits	2

List of Assignments

ST 9120: Stochastic Processes
ST 932: Quality Assurance and reliability theory

1. Computing n-step and stationary probabilities
2. Sampling path of a Markov chain
3. Computation of first passage probabilities and time, mean recurrence time.
4. Stationary Probabilities of a Markov chain
5. Poisson Process
6. Branching Process
7. OC and ARL curves of X and R control charts
8. CUSUM and EWMA control Charts
9. Multivariate control charts.
10. Single and double attribute sampling plans.
11. Reliability Estimation.

Third Semester

ST 9P2: Practical VI (based on ST 9220 and STDE 9420/STDE 9520)

Semester	Third
Paper Code	ST 9P2
Paper Title	Practical VI (based on ST 9220 and STDE 9420/STDE 9520)
Number of teaching hrs per week	4
Total number of teaching hrs per semester	44
Number of credits	2

List of Assignments

ST 9220: Data mining and Machine Learning
STDE 9420/STDE 9520: Optimization Techniques/Operations Research

1. Support vector Machine
2. Bayesian classifier
3. Bootstrap sampling
4. EM algorithm
5. k-NN classifier
6. Decision tree for classification
7. Simplex Method /Integer linear Programming Problem
8. Big M Method/Dynamic programming
9. Integer Programming/Transportation problem
10. Nonlinear optimization/Quadratic program
11. Inventory Models/Assignment Problem

Fourth Semester

ST 0123: Advanced Statistical Inference

Semester	Fourth
Paper Code	ST 0123
Paper Title	Advanced Statistical Inference
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(15L + 2T) Hrs**

Consistency: Definition and problems for consistency. Weak and strong consistencies. Marginal and jointly consistent estimators. Invariance property. Comparison of consistent estimators. Consistent asymptotically normal (CAN) property: Definition and methods of obtaining CAN estimators. Example of consistent but not asymptotic normal in Pitman family. Invariance property. CAN property of MLE in Cramer's family. Best asymptotically normal (BAN) estimator. Asymptotic relative efficiency (ARE).

Unit 2 **(7L + 2T) Hrs**

Robust estimation: The influence curve and empirical influence curve. M-estimation: Median, Trimmed and winsorized mean. Influence curve for M-estimators. Limiting distribution of M-estimators. Resampling methods: Quenouille's Jackknife estimation, parametric and nonparametric bootstrap methods.

Unit 3 **(10L+2T) Hrs**

Decision Theory and Bayesian Analysis: Elements of a decision problem. Estimation and testing as decision problems. Bayes paradigm. Prior and posterior distributions. Conjugate and non-informative priors. Construction of Bayes estimators relative to squared error, weighted squared error, absolute error, Stein and LINEX loss functions. Minimax estimation.

Unit 4 **(12L + 2T) Hrs**

Non parametric Tests: One sample test: Test based on total number of runs, the ordinary sign test, the Wilcoxon signed - rank test, the Kolmogorov-Smirnov one sample goodness of fit test. Definition of U-statistic and properties. Hoeffding's one-sample U – statistic theorem. Two-sample tests: Sign test, Wilcoxon signed rank test, the median test, the Wilcoxon-Mann-Whetney test, Mood's test for two sample scale problem, the Kolmogorov Smirnov two sample test. Analysis of variance by ranks.

Code number: ST 0123

Title of the paper: Advance Statistical Inference

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
23	17	I
12	9	II
16	12	III
19	14	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To understand the importance of CAN and BAN estimators
CO2	To work on several standard examples to understand the various inherent concepts.
CO3	To understand why nonparametric techniques are needed and useful.
CO4	To apply the general theory of statistical inference to specific problems

References

1. Casella G. and Berger R.L. (2002): Statistical Inference, 2nd Ed., Thomson- Duxbury, Singapore.
2. Kale B.K. and Muralidharan (2015): Parametric Inference, An Introduction, Alpha Science International Limited.
3. Dudewicz, E. J. and Mishra, S. N. (1980). *Modern Mathematical Statistics*, John Wiley, New York.
4. Lehmann, E. L. and Cassella, G. (1998). *Theory of Point Estimation,* 2/e, Springer Verlag, New York.
5. Rohatgi, V. K. and Saleh, A. K. Md. E. (2002). *An Introduction to Probability and Statistics*, 2/e, John Wiley, New York.

6. Zacks, S. (1981). *Parametric Statistical Inference,* John Wiley, New York.
7. Rohatgi, V.K. (2003). *Statistical Inference*, Dover Publications.
8. Wald, A. (1973). *Sequential Analysis*, Dover Publications.
9. Gibbons, J.D., Chakraborti. (201). *Nonparametric Statistical Inference,* Taylor and Francis.
10. Berger, J. O. (1980). *Decision Theory and Bayesian Analysis*, John Wiley, New York.
11. Sinha, S.K. (1988). *Bayesian Estimation,* New Age International, New Delhi.

Fourth Semester

ST 0220: Design and Analysis of Experiments

Semester	Fourth
Paper Code	ST 0220
Paper Title	Design and Analysis of Experiments
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(12L+3T) Hrs**

Introduction to design of experiments. Fixed, random and mixed effects models. One way ANOVA, Completely randomized design. General block design: complete block and incomplete block design. C matrix and its properties, concepts of connectedness, orthogonality, variance balance. Intra block analysis of general block design: Estimability, BLUEs, interval estimates of estimable linear parametric functions and testing of linear hypotheses.

Unit 2 **(12L + 3T) Hrs**

Two-way ANOVA, Randomized block design. Balanced incomplete block design (BIBD) – Definition and relations among the parameters, Intra block analysis. PBIBD. Multiple comparison test: Tukey, Scheffe, Duncan and Dunnett's procedures.

Unit 3 **(10L+2T) Hrs**

Three-way ANOVA, LSD, Youden square design (YSD). Intra block Analysis of YSD. Analysis of covariance for CRD and RBD designs. Missing plot techniques for RBD and LSD.

Unit 4 **(8L+2T) Hrs**

Factorial experiments: concepts, symmetric factorial experiments. Analysis of 2^n and 3^n factorial experiments in randomized blocks. Complete and partial confounding, Layout and analysis of confounded 2^n and 3^n factorials. Fractional replication for 2^n factorials.

Code number: ST0220

Title of the paper: Design and analysis of experiments

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
21	15	I
20	15	II
16	12	III
13	10	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To understand the potential practical problems in its implementation.
CO2	To appreciate the advantages and disadvantages of a design for a particular experiment.
CO3	To construct optimal or good designs for a range of practical experiments.
CO4	To describe how the analysis of the data from the experiment should be carried out.

References

1. Das M.N. and Giri N.C. (1979): *Design and Analysis of Experiments*, 2nd Ed., Wiley.
2. Giri N.C. (1986): *Analysis of Variance.* South Asian Publishers.
3. Hinkleman and Kempthrone C. (1994): *Design and Analysis of Experiments,* Vol.I, John Wiley.
4. Joshi D.D. (1987): *Linear Estimation and Design of Experiments,* Wiley Eastern.
5. Montgomery D.C. (2001): *Design and Analysis of Experiments*, John Wiley.
6. Chakrabarti, M.C. (1962). *Mathematics of Design and Analysis of Experiments*, Asia Publishing House, New Delhi.
7. Kempthorne, O. (1952). *Design and Analysis of Experiments*, Wiley Eastern, New Delhi.
8. Cochran, W.G. and Cox, G. M. (1957). *Experimental Designs,* 2/e, John Wiley, New York.

9. Dean, A. and Voss, D. (2006). *Design and analysis of experiments,* Springer.
10. Cox, D.R. and Reid, N. (2000). *The Theory of the Design of Experiments*, Chapmann and Hall.
11. Toutenburg, H. and Shalabh. (2009). Statistical Analysis of Designed experiments.
12. Parimala Mukhopadhyay. (1999). *Applied Statistics*, Books and Applied Publisher.
13. Mohan Madhyastha, Ravi, S., Praveena, A.S. (2020). *A First Course in Linear Models and Design of experiments,* Springer.

Fourth Semester – Elective II

STDE 0323: Biostatistics

Semester	Fourth
Paper Code	STDE 0323
Paper Title	Biostatistics
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(13L + 2T) Hrs**

Probability density function, survival function, hazard rate function, relationship between three types of functions, Life time distributions - exponential, gamma, Weibull, lognormal, Pareto, Rayleigh. Tests of goodness of fit for survival distributions (WE test for exponential distribution, W-test for lognormal distribution, Chi-square test for uncensored observations).

Type I, Type II and progressive or random censoring with biological examples, Estimation of mean survival time and variance of the estimator for type I and type II censored data with numerical examples. Parametric methods for comparing two survival distributions viz. L.R test, Cox's F-test. P-value.

Unit 2 **(10L + 3T) Hrs**

Life tables: Standard methods for uncensored and censored data;

Non-parametric methods for estimating survival function and variance of the estimator viz. Actuarial and Kaplan -Meier methods. Properties of Kaplan Meier Estimator- Self consistency, Generalized MLE, Statement of Asymptotic properties of KM estimators, Nelson-Aalen estimator, treatment of ties (Peto's method)

Unit 3 ***(*10L+2T)hrs**

Analysis of Epidemiologic and Clinical Data: Planning and design of clinical trials, Phase I, II, and III trials. Consideration in planning a clinical trial, designs for comparative trials. Sample size determination in fixed sample designs. Studying association between a disease and a characteristic: (a) Types of studies in Epidemiology and Clinical Research (i) Prospective study (ii) Retrospective study (iii) Cross-sectional data, (b) Dichotomous Response and Dichotomous Risk Factor: 2X2 Tables (c) Expressing relationship between a risk factor and a disease (d) Inference for relative risk and odds ratio for 2X2 table, Sensitivity, specificity and predictivities, Cox proportional hazard model.

Unit 4 **(10L+2T) hrs**

Competing risk theory, Indices for measurement of probability of death under competing risks and their inter-relations. Estimation of probabilities of death under competing risks by maximum likelihood and modified minimum Chi-square methods. Theory of independent and dependent risks. Bivariate normal dependent risk model. Conditional death density functions. Stochastic epidemic models: Simple and general epidemic models (by use of random variable technique).

Code number: STDE 0323

Title of the paper: Bio Statistics

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
20	15	I
17	13	II
16	12	III
16	12	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To describe various application areas of biostatistics
CO2	To Describe the roles biostatistics serves in the discipline of medical sciences.
CO3	To discuss the principal ethical issues that arise in clinical trials
CO4	To explain how statistical techniques are incorporated in the analysis of medical research data and its presentation.

References

1. Biswas, S. (1995). Applied Stochastic Processes. A Biostatistical and Population Oriented Approach, Wiley Eastern Ltd.
2. Collett, D. (2003). Modelling Survival Data in Medical Research, Chapman & Hall/CRC.

3. Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data, Chapman and Hall.
4. Elandt Johnson R.C. (1971). Probability Models and Statistical Methods in Genetics, John Wiley & Sons.
5. Ewens, W. J. (1979). Mathematics of Population Genetics, Springer Verlag.
6. Ewens, W. J. and Grant, G.R. (2001). Statistical methods in Bio informatics: An Introduction, Springer.
7. Friedman, L.M., Furburg, C. and DeMets, D.L. (1998). Fundamentals of Clinical Trials, Springer Verlag.
8. Gross, A. J. And Clark V.A. (1975). Survival Distribution; Reliability Applications in Biomedical Sciences, John Wiley & Sons.
9. Indrayan, A. (2008). Medical Biostatistics, Second Edition, Chapman & Hall/CRC.
10. Lee, Elisa, T. (1992). Statistical Methods for Survival Data Analysis, John Wiley & Sons.
11. Li, C.C. (1976). First Course of Population Genetics, Boxwood Press.
12. Miller, R.G. (1981). Survival Analysis, John Wiley & Sons.
13. Robert F. Woolson (1987). Statistical Methods for the analysis of biomedical data, John Wiley & Sons.

Fourth Semester

STDE 0420: Survival analysis (Elective -II)

Semester	Fourth
Paper Code	STDE 0420
Paper Title	Survival Analysis
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(13L+3T) Hrs**

Complete and censored samples, Type I, Type II, and random censoring, Survival function, Failure rate, mean residual life and their elementary properties. Life time distributions - Exponential, Gamma, Weibull, Lognormal, Pareto, Proportional Hazards family. Estimation of parameters of exponential and gamma distributions under various censoring situations.

Unit 2 **(10L + 3T) Hrs**

Life tables: Standard methods for uncensored and censored data;
Estimation of survival function – Actuarial Estimator, Kaplan Meier Estimator, Greenwood's formula. Properties of Kaplan Meier Estimator- Self consistency, Generalized MLE, Statement of Asymptotic properties of KM estimators, Nelson-Aalen estimator, treatment of ties (Peto's method)

Unit 3 **(8L+2T) Hrs**

Fully parametric analysis of dependency - accelerated life model - simple form, log logistic accelerated life model, proportional hazards model - relation with accelerated life model.

Unit 4 **(10L+3T) Hrs**

Semi-parametric regression for failure rate – Cox's proportional hazards model with one and several covariates, log likelihood function, log linear hazards, test for regression coefficients, Discrete failure time: ties. Competing risk models-Estimation of cumulative hazard function.

Code number: STDE 0420

Title of the paper: Survival Analysis

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
22	16	I
17	13	II
13	10	III
18	13	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To Identify characteristics of survival data and their implications for analysis
CO2	To Perform and interpret univariate analyses of survival data
CO3	To describe the key features of survival data, the Kaplan Meier estimator, explain their theoretical basis, and be able to apply them to real data.
CO4	To Analyze survival data and interpret results using Cox proportional hazards Mode

References

1. Cox D.R. and Oakes D. (1984). *Analysis of Survival Data*, Chapman and Hall, New York.
2. Kalbfeisch J.D. and Prentice R.L. (2002). *The Statistical Analysis of Failure Time Data*, John Wiley& Sons, Inc. 2nd Edition.
3. Lawless J.F. (2002). *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, Inc.
4. Miller R.G. (1981). *Survival Analysis*, John Wiley& Sons, Inc.
5. Nelson. B. (2003). *Applied life Data Analysis,* Wiley Series.
6. Deshpande, J.V., Purohit, S.G. (2006). *Life time data: Statistical Models and Methods*, World Scientific.

Fourth Semester

STDE 0520: Time Series Analysis (ELECTIVE – III)

Semester	Fourth
Paper Code	STDE 0520
Paper Title	Time Series Analysis
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(6L+1T) hrs**

Exploratory Time Series Analysis, test for randomness, Tests for trend and seasonality. Estimation of trend by moving average, estimation of seasonal effect for additive and multiplicative models, de-seasonalising and detrending an observed time series.

Unit 2 **(16L+3T) hrs**

Time-series (t.s) as discrete parameter stochastic process, definition of strict and weak stationarity of a t.s., Gaussian t.s., ergodicity, autocovariance and autocorrelation functions (ACF) and their properties, partial autocorrelation function (PACF).

General linear processes (G.L.P), autocovariance generating function, stationarity and invertibility conditions of a G.L.P; autoregressive processes (AR(p)), stationarity condition, ACF, PACF, Yule-Walker equations, Moving average (MA(q)) processes, Invertibility condition, ACF, PACF, duality between AR(p) and MA(q) processes; ARMA(p,q) processes, stationarity, invertibility, ACF, PACF, particular cases of these processes.

Unit 3 **(11L+2T) hrs**

Linear Non-stationary time Series models: ARIMA(p,d,q) processes, general form, three explicit forms, IMA(0,1,1) process, seasonal ARIMA processes. Forecasting: minimum mean square error forecast, BLUP, three basic forms for the forecast, forecast error and its properties, examples; forecasting through exponential and Holt-Winter smoothing.

Unit 4 **(11L+2T) hrs**

Estimation: sample ACF, sample PACF, fitting AR(p), MA(q), ARMA(p,q) models; model identification: determination of p, d, q: method of differencing, unit root test, using sampleACF, sample PACF, Bartlett and Anderson bounds; diagnostics: residual analysis, Box-Pierce portmanteau statistic, Ljung-Box test; AIC and BIC criteria.

Introduction to Financial Time series. ARCH and GARCH models.

Code number: STDE 0520

Title of the paper: Time Series Analysis

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
9	07	I
26	19	II
18	13	III
17	13	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To Understand the concept of time series and its components.
CO2	To Understand the bases of different models of time series analysis including decomposition.
CO3	To learn proper model identification and its estimation.
CO4	To learn and apply different forecasting methods with the least forecasting error.

References

1. Anderson, T. W. (1971). *The Statistical Analysis of Time Series,* Wiley, New York.
2. Box, G. E. P, Jenkins, G. M, Reinsel, G. C. and Ljung, G. M. (2015). *Time Series Analysis - Forecasting and Control*, 5/e, Wiley.
3. Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*, 2/e, Indian Print, Springer, New Delhi.
4. Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, 2/e, Springer, New York.
5. Chatfield, C. (1996). *The Analysis of Time Series: Theory and Practice*, 5/e, Chapman and Hall, London.
6. Chatfield, C. (2003). *Analysis of Time Series: An Introduction*, CRC Press, New Delhi.
7. Nachane, D. M. (2006). *Econometrics: Theoretical Foundations and Empirical Perspectives,* Oxford University Press, London.
8. Cryer, J. D. and Chan, K. S. (2008). *Time Series Analysis with Application in R*, 2/e,

Springer, New York.
9. Kendall, M. G. and Ord, J. K. (1990). *Time Series*, 3/e, Edward Arnold, New York.
10. Montgemory, D. C. and Johnson, L. A. (1977). *Forecasting and Time Series Analysis*, McGrawHill, New York.

11. Kirchgassner, G and Walters, J. (2008). *Introduction to Modern Time series analysis*, springer.

12. Tsay, R.S. (2010). *Analysis of Financial Time series*. *3/e*. Wiley.

13. Stoffer, D.S. and Shumway, R.H. (2010). *Time Series Analysis and its applications: With R examples*. Springer.

Fourth Semester

STDE 0620: Actuarial Statistics (ELECTIVE – III)

Semester	Fourth
Paper Code	STDE 0620
Paper Title	Actuarial Statistics
Number of teaching hrs per week	4
Total number of teaching hrs per semester	52
Number of credits	4

Unit 1 **(10L+2T) hrs**

Utility theory, insurance and utility theory, models for individual claims and their sums, survival function, curtate future lifetime, force of mortality. Life table and its relation with survival function, examples, assumptions for fractional ages, some analytical laws of mortality, select and ultimate tables.

Unit 2 **(10L+2T) hrs**

Multiple life functions, joint life and last survivor status, insurance and annuity benefits through multiple life functions evaluation for special mortality laws. Multiple decrement models, deterministic and random survivorship groups, associated single decrement tables, central rates of multiple decrements, net single premiums and their numerical evaluations. Distribution of aggregate claims, compound Poisson distribution and its applications.

Unit 3 **(12L+2T) hrs**

Principles of compound interest. Nominal and effective rates of interest and discount, force of interest and discount, compound interest, accumulation factor, continuous compounding.

Life insurance: Insurance payable at the moment of death and at the end of the year of death-level benefit insurance, endowment insurance, deferred insurance and varying benefit insurance, recursions, commutation functions. Life annuities: Single payment, continuous life annuities, discrete life annuities, life annuities with monthly payments, commutation functions, varying annuities, recursions, complete annuities-immediate and apportionable annuities-due.

Unit 4 **(12L+2T) hrs**

Net premiums: Continuous and discrete premiums, true monthly payment premiums, apportionable premiums, commutation functions, accumulation type benefits.

Payment premiums, apportionable premiums, commutation functions, accumulation type benefits. Net premium reserves: Continuous and discrete net premium reserve, reserves on a semicontinuous basis, reserves based on true monthly premiums, reserves on an apportionable or discounted continuous basis, reserves at fractional durations, allocations of loss to policy years, recursive formulas and differential equations for reserves, commutation functions. Some practical considerations: Premiums that include expenses-general expenses types of expenses, per policy expenses. Claim amount distributions, approximating the individual model, stop-loss insurance.

Code number: STDE 0520

Title of the paper: Time Series Analysis

Total marks for which the questions are to be asked (including bonus questions)	Number of hours	Chapter/Unit number
16	12	I
16	12	II
19	14	III
19	14	IV
70	**52**	**TOTAL**
Maximum marks for the paper (Including bonus question):**70**		

Course Outcomes: At the end of the Course, the Student should be able

CO1	To understand how actuarial science is used in finance, investments, banking and insurance.
CO2	To understand the statistical behaviour of actuarial indicators
CO3	To solve the problems related to the benefit amounts in insurance, annuities, premiums and reserves.
CO4	To apply mathematical and statistical methods to assess risk in insurance, finance and other industries and professions.

References:

1. Atkinson, M.E. and Dickson, D.C.M. (2000). An Introduction to Actuarial Studies, Elgar Publishing.
2. Bedford, T. and Cooke, R. (2001). Probabilistic risk analysis, Cambridge.
3. Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones D.A. and Nesbitt, C. J. (1986). 'Actuarial Mathematics', Society of Actuaries, Ithaca, Illinois, U.S.A., Second Edition (1997).
4. Medina, P. K. and Merino, S. (2003). A discrete introduction: Mathematical finance and Probability, Birkhauser.
5. Neill, A. (1977). Life Contingencies, Heineman.
6. Philip, M. et. al (1999). Modern Actuarial Theory and Practice, Chapman and Hall.
7. Rolski, T., Schmidli, H., Schmidt, V. and Teugels, J. (1998). Stochastic Processes for Insurance and Finance, Wiley.
8. Spurgeon, E.T. (1972). Life Contingencies, Cambridge University Press.
9. Relevant Publications of the Actuarial Education Co., 31, Bath Street, Abingdon, Oxfordshire OX143FF (U.K.)

Fourth Semester

ST 0P1: Practical V II (based on ST 0120, STDE 0320/STDE 0420 and STDE 0520/STDE 0620)

Semester	Fourth
Paper Code	ST 0P1
Paper Title	Practical VII (based on ST 0123, STDE 0320/STDE 0420 and STDE 0520/STDE 0620)
Number of teaching hrs per week	4
Total number of teaching hrs per semester	44
Number of credits	2

List of Assignments

ST 0123: Advanced Statistical Inference

STDE 0320/STDE0420: Biostatistics/Survival analysis

STDE 0520/STDE 0620: Time series analysis/Actuarial Statistics

1. CAN estimators
2. Quenouille's and Jackknife estimation
3. Bayes estimators.
4. Minimax rule.
5. One and Paired sample tests: Wilcoxon signed - rank test and KS one sample goodness of fit test.
6. Two sample tests: Two sample run: Kolmogorov Smirnov two sample, Wilcoxon Mann Whitney
7. Life time distributions
8. Actuarial and Kaplan Meier Estimator
9. Estimation and elimination of trend and seasonal components /Future lifetime random variable and related measures
10. Examining Stationarity using sample ACF and PACF/ Computation of various measures using Gompertz and Makeham's Model.
11. Fitting MA and AR model/Cox proportional Hazard models

Fourth Semester

ST 0P2: Practical VIII (based on ST0220)

Semester	Fourth
Paper Code	ST 0P2
Paper Title	Practical VIII (based on ST0220)
Number of teaching hrs per week	4
Total number of teaching hrs per semester	44
Number of credits	2

List of Assignments

ST 0220: Design and Analysis of Experiments

1. Two -way ANOVA
2. Analysis of BIBD
3. Analysis of PBIBD
4. Analysis of Youden square design
5. Missing Plot Techniques for RBD and LSD
6. Analysis of Covariance for CRD
7. Analysis of Covariance for RBD
8. Analysis of 2^3 factorial experiments
9. Analysis of 3^2 factorial experiments
10. Confounding in 2^3 factorial experiments
11. Confounding in 3^2 factorial experiments