MANGALORE UNIVERSITY

Scheme of Examination and Syllabus for

Master of Science in Statistics Degree Programme

DEPARTMENT OF POST-GRADUATE STUDIES AND RESEARCH IN STATISTICS

MANGALAGANGOTHRI-574 199

2014

A. Preamble:

It is three years since the last revision of the syllabus took place. Based on the feedback from students and teachers, the faculty of this department felt there is a need to reorganize the syllabus of M.Sc. (Statistics) course. The internal members of the PG board of studies in Statistics deliberated upon this and prepared a draft syllabus and placed before the PG board of studies. PG board of Studies in Statistics thoroughly discussed and modified the draft syllabus. The syllabus is prepared according to existing Choice Based Credit System (CBCS), by keeping the same pattern and credits.

PROGRAMME OUTCOMES (POs)

The curriculum leading to M.Sc-Statistics degree prepares the students for the positions as Data scientists, Data Analyst, and Academicians in Business Intelligence, Information Technology, Software Industry and Government sectors. The curriculum's main objectives are to impart students with an understanding of the various techniques of data analysis, problem solving skills through algorithmic approaches and to prepare them for continued professional development.
Upon completion of M.Sc. Statistics degree, students will be able:

PO 1: To cultivate a statistical attitude and nurture interests in mathematical statistics.
PO 2: To provide theoretical foundations that will motivate and prepare the students to take up theoretical and applied research in statistics.
PO 3: To focus on algorithms, designs and advanced softwares to give statistical solutions to real life problems.
PO 4: To provide first hand practical experience in handling modern statistical software in data analysis
PO 5: To provide training for a career as a statisticians
PO 6: To train statisticians who can work on challenging problems in various industries.

PROGRAMME SPECIFIC OUTCOMES (PSO)

On completion of the M.Sc.-Statistics Degree programme the graduates of the M.Sc (Statistics) program will be expected to have to
PSO 1: Professionally inclined Statistics knowledge.
PSO 2: Deeper knowledge of statistical inference and be able to discuss and analyse its possibilities and limitations
PSO 3: Contributions as researchers in theoretical and applied fields of Statistics.
PSO 4: Evince an ability critically, independently, and creatively to identify and formulate problems of significance for statistical scienc

B. Course Pattern for M.Sc (Statistics) programme from 2014-15:

I Semester

Course Code	Title of Paper	Hrs./Week	Credits
ST 401	Real Analysis	4	4
ST 402	Matrix Theory and R programming	4	4
ST 403	Probability Theory	4	4
ST 404	Theory of Sampling	4	4
ST 405	Practical I (R Programming & EXCEL)	8	4
ST 406	Practical II (Based on ST404)	8	4
ST 407	Seminar/Mini Project*	1	1
		Total	25

II Semester

Course Code	Title of Paper	Hrs./Week	Credits
ST 451	Distribution theory	4	4
ST 452	Theory of Point Estimation	4	4
ST 453	Stochastic Processes	4	4
ST 454	Design & Analysis of Experiments	4	4
ST 455	Practical III (Based on St 454)	8	4
ST 456	Practical IV (Based on ST 451, ST 452 & ST 453)	8	4
ST 457	Seminar/Mini Project*	1	1
		Total	25

III Semester

Course Code	Title of Paper	Hrs./Week	Credits
ST 501	Statistical Methods (Choice paper as per CBCS)	4	4
ST 502	Testing of Hypothesis	4	4
ST 503	Econometrics	4	4
ST 504	Elective I	4	4
ST 505	Practical V(Based on ST 502)	8	4
ST 506	Practical VI(Based on ST 503)	8	4
ST 507	Seminar/Mini Project*	1	1
		Total	25

IV Semester

Course Code	Title of Paper	Hrs./Week	Credits
ST 551	Multivariate Analysis	4	4
ST 552	Operations Research	4	4
ST 553	Elective II	4	4
ST 554	Elective III	4	4
ST 555	Practical VII (Based on Theory papers)	8	4
ST 556	Project Work	8	4
ST 557	Seminar/Field Work*	1	1
		Total	25

(* All seminars/mini projects/field works shall be assessed internally)

(**The electives are offered from the list of elective courses given below)

C. Elective papers:

ST 504 Elective I is from:

a) Actuarial Statistics.
b) Survival Analysis.

ST 553 Elective II is from:

a) Time Series Analysis.
b) Data Mining Techniques.
c) Nonparametric Regression.

ST 554 Elective III are from:

a) Stochastic_Finance.
b) Bayesian Inference.

D. <u>Scheme of Internal Assessment Evaluation:</u>

The scheme of evaluation for internal assessment marks shall be as follows:

(i) Two tests each of 2 hrs. duration: 10x2 = 20 marks

(ii) Seminar/Assignment/Quizzes etc.: = 10 marks

Total: 30 marks

E. Question Paper Pattern:

The pattern of question papers in theory examinations shall be as follows:

1. There shall be totally 8 questions of which the Q.No. 1 is compulsory. Students have to answer any 4 questions from the remaining 7 questions.

2. Q.No.1 will contain 8 questions of short answer type, each question carrying 3 marks. Students will have to answer any 6 questions. Thus Q.No.1 carries 18 marks.

3. Q.No.2 to Q.No.8 will be of long answer type, each carrying 13 marks.

The distribution of marks will be as follows:

Q.1 -	3 x 6 = 18
Any four questions out of remaining 7	13 x 4 = 52
Total:	70

I SEMESTER	ST-401 REAL ANALYSIS	No. of hrs./week:4

Course Outcomes:

CO1: Explain the basics of Real analysis and to apply the acquired knowledge in their subsequent course work.

CO2: Describe the fundamental properties of the real numbers that lead to the formal development of real analysis.

CO3: understand of limits and how they are used in sequences, series, differentiation and integration

CO4: Solve the problems of convergence and divergence of sequences and series.

CO5: Classify and explain open and closed sets, limit points, convergent and Cauchy convergent sequences, complete spaces, compactness, connectedness, and uniform continuity etc. in a metric space.

CO6: Illustrate the effect of uniform convergence on the limit function with respect to continuity, differentiability and integrability.

Unit-I

Introduction to n-dimensional Euclidean space; open and closed sets, Bolzano-Weir strass theorem (statement only), Heine-Borel theorem (statement only), compact set. (8 hrs)

Unit-II

Sequences and their convergence- bounded sequences, monotone sequences, limit superior and limit inferior, Cauchy sequences. Series- convergence and divergence, tests for convergence. (10 hrs)

Unit-III

Real valued functions, continuous functions, monotonic functions, discontinuities of real valued functions, uniform continuity, sequences and series of functions, uniform convergence- uniform convergence of sequences, definition and examples, Cauchy criterion for uniform convergence, uniform convergence and integration, uniform convergence and differentiation, uniform convergence of infinite series of function, Weirstrass M-test, Dirichlet’s test. Power series-definition, radius of convergence and examples.

(12 hrs)

Unit –IV

The Reimann Stieltjes’ integration, properties. Integration by parts, change of variables, step functions as integrators, monotonically increasing integrators. (11 hrs)

Unit-V

Improper integrals – Beta and gamma integrals. Extrema of real valued functions- one variable & several variables, stationary point, saddle point, local and global extrema, extremum problems with restrictions-Lagrange’s method. (9 hrs)

References:

1. R.R. Goldberg (1978): Methods of Real Analysis, 4^{th} Ed., Oxford & I BH Publishing Co.
2. S.C. Malik and Savitha Arora (1993): Mathematical Analysis, Wiley Eastern.
3. Robert G. Bartle (1975): The elements of Real Analysis, 2 nd Ed., John Wiley & Sons.
4. W. Rudin (1976): Principles of Mathematical Analysis, 3^{rd} Ed., McGraw-Hill, New York.
5. Tom. M. Apostol (1977): Mathematical Analysis, 2^{nd} Ed., Addison-Wesley Publishing Co.

I SEMESTER	ST-402: MATRIX THEORY AND R -PROGRAMMING	No. of Credits:4

Course Outcomes:

CO1: Carry out matrix operations, including inverses and determinants.
CO2: Demonstrate understanding of the concepts of vector space and subspace.
CO3: Demonstrate understanding of linear independence, span, and basis.
CO4: Determine eigen values and eigenvectors and solve eigen value problems.
CO5: Apply principles of matrix algebra to linear transformations.
CO6: Demonstrate understanding of inner products and associated norms.
CO7: Make use of R loops and functions for computations.
CO8: Deal with vectors, matrices, data frames, lists, factors, tables in R environment.
CO9: identify basic R data structures relevant to modern data analysis (atomic vectors and data frames)
CO10: write syntactically correct R expressions that involve variables, variable assignment, operators and functions
CO11: Write and read documents from different environments into R.
CO12: Use several graphical tools for data interpretation.
CO13: Correct errors using debugging tools.
CO14: Find, install and use R packages from repositories.

Matrix Theory

Unit 1: Vector spaces, linear dependence and independence; basis and dimension of a vector space. Orthonormal basis and orthogonal projections. Gram-Schmidt orthogonalization process. (8 hrs)

Unit 2: Types of matrices, determinant, row and column spaces of a matrix, rank and inverse of a matrix. Null space and nullity; partitioned matrices; Kronecker product. Generalised inverse, Moore- Penrose Inverse. Linear equation-homogenous and non-homogenous systems, solution spaces. (12 hours)

Unit 3: Characteristic roots and vectors, Cayley-Hamilton theorem, algebraic and geometric multiplicity of characteristic roots. Determinant, rank and trace of a matrix in terms of characteristic roots. Real quadratic forms, classification of quadratic forms, reduction of quadratic forms, index and signature. Sylvester's law of Inertia. Vector and matrix differentiation. (12 hrs.)

R Programming

Unit 4

Introduction to R. Storing data. Starting R, setting directories. Regular expressions in R and their evaluation. Vectors and matrix. Operations with matrix, submatrices, subsetting, missing values. Subscripting, rbind() and cbind(). Functions, Data frames, names, attach, detach, expanding data frames. Libraries of R. Script editor in R. Syntax in R programming. Using logical expressions. If functions. Loops in R – for, while, repeat.

(10 hrs.)

Unit 5

Handling data files: Reading data from text file, from excel file and web pages. Scan() function, Saving output in a file, printing outputs and files. Graphics, designing graphs. User defined function programs. Packages, loading packages.

Functions for statistical distributions, some statistical tools in R. Some program examples using these features. Debugging. (8 hrs.)

Reference:

[1] Franz E Hohn(1971): Elementary Matrix Algebra, Second edition, Amerind Publishing Co. Pvt.Ltd.
[2] Hadley, G.(1987): Linear Algebra, Narosa
[3] Rao, C.R. (1973): Linear Statistical Inference and its Applications, second edition, Wiley.
[4] Searle S.R.(1982): Matrix Algebra Useful for Statistics, John Wiley& Sons.
[5] John Verzani(2005), "Using R for Introductory Statistics", Chapman & Hall/CRC
[6] Alain F. Zuur et. al. (2009) "A Beginner's Guide to R. Use R! Series" Springer.
[7] Phil Spector (2008), "Data Manipulation with R. Use R! Series", Springer.
[8] A. Ramachandra Rao, and P. Bhimasankaram (2000), "Linear Algebra", Hindustan Book Agency.

I SEMESTER	ST 403: PROBABILITY THEORY	No. of credits: 4

Course Outcomes:

CO1: Learn to develop complex mathematical reasoning.

CO2: Know the main tools to describe a random variable, such as the probability density function, the cumulative distribution function, and the characteristic function

CO3: Recognize the importance of the central limit theorem and understand when it is appropriate to use normal approximations for the distribution of a statistic.

CO4: Possess techniques of proving theorems and thinking out counter-examples.

Unit-I

Classes of sets, sequence of sets, limit superior and limit inferior of a sequence of sets. Fields, sigma-fields, minimal sigma-field, Borel sigma-field in R. (10 hrs)

Unit-II

Measure, probability measure, properties of probability measure, Independence. General distribution functions, Lebesgue and Lebesgue-Steiltjes measures on R. Measurable functions, random variables, induced probability measure and distribution function and properties. (10 hrs)

Unit-III

Integration with respect to a measurable function. Expectation of a random variable. Monotone convergence theorem, Fatou's lemma, Dominated convergence theorem. Sequence of random variables. Convergence in distribution, convergence in probability. Properties and examples. Slutsky's Theorem. (10 hrs)

Unit-IV

Almost sure convergence, convergence in r^{th} mean. Borel-Cantelli Lemma. Khintchine and Chebychev's weak law of large numbers. Kolmogorov's strong law of large numbers for sequence of independent and sequence of iid random variables. (10 hrs)

Unit-V

Characteristic functions: Definition and simple properties, inversion formula (density function), uniqueness theorem, Levy's continuity theorem. Central limit theorem, Lindeberg-Levy and Liapounov central limit theorems. Statement of Lindeberg-Feller central limit theorem. Applications of these theorems. (10 hrs)

References:

1) Bhat B.R. (1999):Modem Probability Theory, 3rd Ed., New Age Publishers.
2) Basu, A.K. (1999): Measure Theory and Probability, Prentice-Hall of India.
3) Chow Y.S. and Teicher H. (1979): Probability Theory, Narosa Publishing House.
4) Kingman JFC and Taylor S.J. (1966): Introduction to Measure and Probability, Cambridge University Press.
5) Laha R.G. and Rohatgi V.K. (1979): Probability Theory, John Wiley.
6) Robert B.Ash (2000): Probability and Measure Theory, A Harcourt Science and Technology Company
7) Rohatgi V.K. and A.K.E. Saleh (2001): Introduction to Probability and Statistics, John Wiley & Sons.
8) David Stirzaker (1994) "Elementary Probability", Cambridge University Press.
9) Geoffrey Grimmett and Dominic Welsh (2003) "Probability- An Introduction" Oxford Science Publishers.
10) Kai Lai Chung (2001) A Course in Probability Theory, Academic Press.

I SEMESTER	ST-404: THEORY OF SAMPLING	No. of Credits:4

Course Outcomes:

CO1: Understand the principles underlying sampling as a means of making inferences about a population

CO2: Learn the different sampling techniques and able to apply the same in different area and able to analyze data from multi-stage surveys

CO3: Apply the statistical method to get response in case sensitive matter and able to analyse the data to get accurate results

CO4: Learn the concepts of ratio and regression method of estimation, estimate the bias and sampling variability and straggles for reducing these

CO5: Compute the HDI from data ,interpretation of results and understand need, importance of measuring inequality in income and concept of NSS and CSO Understand and solve the practical issues arising in sampling studies

Unit-I

Basic Concepts: Sampling design, sampling scheme, sampling strategy, interpenetrating subsampling.

Probability Proportion to Size with Replacement (PPSWR) Sampling: Selection of PPSWR sample. Estimation of population mean, total and their sampling variances – Hansen-Hurwitz strategy. Estimation of sampling variance. Comparison with SRSWR, Estimation of Gain due to PPSWR sampling.

(12 hrs)

Unit-II

Varying Probability Without Replacement (PPSWOR) Sampling: Some properties of sampling design, Horwitz-Thompson estimator, sampling variance of population total and its unbiased estimator. Sen-Midzuno Sampling Scheme, Des-Raj's Ordered estimator (general case), Murthy's unordering principle (sample of size two), Rao-Hartley-Cochran sampling strategy.

(10 hrs)

Unit-III

Single stage cluster sampling: Concepts, estimation of efficiency of cluster sampling, clusters of varying sizes.

Two stage sampling: Notions, estimation of population total and its variance, when SRSWR is used at first stage and SRSWOR at the second stage, SRSWOR at both stages and PPSWR at the first stage and SRSWOR at the second stage. Efficiency of two-stage sampling relative to cluster and SRS sampling.

(10 hrs)

Unit-IV

Ratio and regression estimators based on SRSWOR, method of sampling, bias and mean square errors, comparison with mean per unit estimator.

Two phase sampling: notion, double sampling for ratio estimation, double sampling for regression estimation. (10 hrs)

Unit-V

Randomized response techniques: Warner's model, related and unrelated questionnaire methods, Nonsampling errors.

Statistics for National Development: NSO, CSO, Human Development Index, measuring inequality in income: Lorenz Curve, Gini coefficient. (8 hrs)

References:

1. Cochran W.G. (1977): Sampling Techniques, 3rd Ed., Wiley.
2. Des Raj and Chandok (1998): Sampling Theory, Narosa Publication.
3. Mukhopadhyay P. (1998): Theory and Methods of Survey Sampling, Prentice-Hall of India.
4. Murthy M.N. (1977): Sampling Theory and Methods, Statistical Publishing Society, Calcutta.
5. S. Sampath (2001): Sampling Theory and Methods, Narosa Publishers.
6. Sen A. (1997): Poverty and Inequality.
7. Singh D. and Chaudhary F.S. (1986): Theory and Analysis of Sample Survey Designs, New Age International Publishers.
8. Sukhatme P.V., Sukhatme B.V, Sukhatme S. and Ashok(1984): Sampling Theory of

Surveys with Applications, ICAR publication.
9. Vic Barnett (2002): Sample Survey – Methods and Principles, Arnold Publishers.

I SEMESTER	STH-405: Practical using EXCEL and R- Programming	No. of Credits:4

Course Outcomes:

CO1: Learn the Use graphical tools for data interpretation in R and Excel
CO2: Understand the R - programming language
CO3: Develop Macro programming for iterative computing using Excel
CO4: Develop User defined function using visual basic for different distributions
CO5: List the probabilities for various continuous and discrete distributions
CO6: Write and read documents from different environments into R.
CO7: Understand the use of loops and functions for computations
CO8: Understand R coding and implementation.
CO9: Know about debugging tools.
CO10: Learn to install and use R packages from repositories.

EXCEL Exercises

1. Reading data and creating data, certain computations using data. Descriptive Statistics and construction of frequency distribution. (At least two practicals).
2. Listing probabilities for standard distributions and plotting its probabilities and distribution functions.
3. Plotting density functions and distribution functions for standard continuous distribution functions.
4. Finding probabilities of certain sets in case of discrete and continuous distribution functions and Finding probabilities and critical values.
5. User defined function using visual basic (VB) – Plotting some general distribution function and finding certain probabilities.
6. Computation of annual salary of a randomly drawn employee (create problem so as to use LOOKUP function) and finding her net payable tax according to that year's Income tax.
7. Using macro programming for certain iterative computing (at least two practicals.)

R Practicals

1. Simple R exercises, using – scan function, reading data from EXCEL and exercises, vectors, matrices, rbind and cbind
2. Exercises on Matrices.
3. Reading data from text file. Data frames , names etc., Exercises based on these data and exercises on graphics
4. Exercises using iterative computations.
5. Functions in R Exercises - 1
6. Functions in R Exercises - 2
7. Generating a sample from general discrete distribution
8. Generating a sample from general continuous distribution
9. Verification law of large numbers and central limit theorem.

I SEMESTER	STH-406: Practical based on ST 404	No. of Credits:4

Course Outcomes:

CO1: Learn to determine sample size in various sampling schemes

CO2: Able to apply the unequal probability sampling such as PPSWR and PPSWOR for real life situations.

CO3: Learn to select samples and estimate population parameters using cluster sampling and multistage sampling

CO4: Apply the regression and ratio method estimation for various sampling scheme

CO5: Learn to select samples using Two Phase sampling scheme and estimate the population parameters

Practical's on Theory of Sampling

1. Determination of sample size.
2. PPSWR sample selection by (i) Cumulative Total method and (ii) Lahiri's method.
3. PPSWR sampling: Hansen-Hurwitz estimator and its sampling variance and Comparison of PPSWR sampling with SRSWR sampling based on PPS sample.
4. PPSWOR sampling: Horvitz-Thompson sampling strategy.
5. PPSWOR sampling: Sen- Midzuno sampling strategy.
6. PPSWOR sampling: Desraj's ordered estimator and Murthy's unordered estimator and their sampling variance and PPSWOR sampling : Rao-Hartley-Cochran strategy.
7. Cluster sampling with clusters of equal size.
8. Cluster sampling with clusters of unequal size.
9. Two stage sampling with SRSWOR at both the stages.
10. Two stage sampling with SRSWR at the first stage and SRSWOR at the second stage.
11. Ratio Method of estimation.
12. Regression method of estimation.
13. Two Phase sampling.
14. Two stage sampling with PPSWR at the first stage and SRSWOR at the second stage.

II SEMESTER	ST-451 DISTRIBUTION THEORY	No. of credits: 4

Course Outcome

CO1: Apply problem-solving techniques to solving real-world events.

CO2: Apply selected probability distributions to solve problems.

CO3: Apply key concepts of probability, including discrete and continuous random variables, probability distributions, conditioning, independence, expectations, and variances.

CO4: Define and explain the different statistical distributions (e.g., Normal, Binomial, Poisson) and the typical phenomena that each distribution often describes.

Unit-I

Random experiments and its sample spaces, random variables, cdf, pdf and pmf, absolutely continuous and discrete distributions. Jordan decomposition theorem. Mixture of probability distribution.

Discrete Distributions: Power series family and its properties, bionomial, negative binomial, Poisson logarithmic series as special cases. Generating functions- probability generating function, moment generating functions. Truncated distributions. (12 hrs)

Unit II

Continuous univariate distributions : Weibull, lognormal, Pareto, Laplace, Cauchy, Logistic, inverse Gaussian distributions, extreme value distributions. – Properties and applications. (8 hrs)

Unit III

Functions of random variables and their distributions using Jacobian of transformation. Probability integral transformation. Independence, sum of independent random variables, convolutions, conditional expectation. Independence of mean and variance of a random sample from normal population. (12 hrs)

Unit IV

Order statistics –their distributions and properties, joint and marginal distributions of order statistics. Distributions of range and median. (8 hrs)

Unit V

Sampling distributions, Non-central chi-square, mgf of non central chi-square distribution, reproductive property. Non-central t and Non-central F (statements only). Distributions of quadratic forms under normality and related distributions.
(10 hrs)

References:

1) Arnold B.C., Nagaraja H.N. and Balakrishna N. (2008): First Course in Order Statistics, John Wiley.
2) Johnson N.L., Kotz S and Balakrishna N (1994): Continuous Univariate Distributions-1, John Wiley.
3) Johnson N.L., Kotz S and Balakrishna N (1994): Continuous Univariate Distributions-2, John Wiley.
4) Johson N.L and Kemp (1992): Univariate Discrete Distributions, John Wiley.
5) Kendal M.G. and Stuart A. (1977): The advance Theory of Statistical (Vol.1) Charles Griffin and Company Limited.
6) Rohatgi V.K. V. K. Rohatgi, A. K. Md. Ehsanes Saleh (2011): Introduction to Probability Theory and Statistics, Wiley Eastern.
7) Rao CR. (1995): Linear Statistical Inference and Its Applications (Wiley Eastern) 2nd Ed.
8) Parimal Mukhopadhyay (2012) Theory of Probability,World Scientific.

II SEMESTER	ST-452: THEORY OF POINT ESTIMATION	No. of credits: 4

Course Outcomes:

CO1: To apply various estimation and testing procedures to deal with real life problems.

CO2: To understand Fisher Information, Lower bounds to variance of estimators, MVUE and apply them in practical situations.

CO3: To understand consistency, sufficiency, unbiasedness, CAN and BAN estimators

Unit-I

Parametric models, likelihood function; examples from standard discrete and continuous models.

Information in data. About the parameters as variation in 1ikelihood function, concept of no information, sufficiency, Neyman factorizability criterion, likelihood equivalence. Fisher information for single and several parameters. (10 hrs)

Unit-II

Minimal sufficient statistic, Exponential families and Pitman families.

Minimum Variance Unbiased Estimation, unbiasedness, locally unbiased estimators, minimum variance, locally minimum variance, mean squared error, Cramer-Rao lower bound approach. (10 hrs)

Unit-III

Minimum variance unbiased estimators(MVUE), Rao-Blackwell theorem, completeness, Lehman-Scheffe theorem, necessary and sufficient condition for MVUE.

(10 hrs)

Unit-IV

Consistent estimation of real and vector valued parameter, invariance of consistent estimator under continuous transformation: Consistency of estimators by method of moments and method of percentiles, mean squared error criterion, Asymptotic relative efficiency. Consistent asymptotic normal (CAN) estimator.

(10 hrs)

Unit-V

Method of Maximum Likelihood: notion, MLE in exponential family, Cramer Family, Cramer-Huzurbazar Theorem, Multinomial with all probabilities depending on a parameter, solutions to likelihood equations, method of scoring, Newton-Raphson and other iterative procedures. Fisher lower bound to asymptotic variance, extension to multiparameter case (without proof). (10 hrs)

References:

1. Casella G. and Berge R.L. (2002): Statistical Inference, 2nd Ed., Thomson- Duxbury, Singapore.
2. Kale B.K.(1999): A First Course on Parametric Inference, Narosa Publishing House.
3. Kendall M.G. and Stuart A. (1968): The Advanced Theory of Statistics, Vol.II, Charles Griffin and Co.
4. Lehman E.L. (1986): Theory of Point Estimation, John Wiley.
5. Rao C.R. (1973): Linear Statistical Inference and Its Applications. Wiley Eastern.
6. Rohatgi V.K. and A.K.L. Salah (2001): An Introduction to Probability and Mathematical Statistics. Wiley Eastern.
7. Silvey S.D. (1970): Statistical Inference. Chapman and Hall.
8. Zacks S. (1981): Parametric Statistical Inference, Pergamon Press.

II SEMESTER	ST- 453: STOCHASTIC PROCESSES	No. of credits: 4

Course Outcomes:

- CO1: To construct transition matrices for Markov dependent behavior and summarize process information.
- CO2: To learn random walk and gambler's ruin problem; Statistical inference in MC
- CO3: To understand the principles and objectives of model building based on Markov chains, Poisson processes and Brownian motion.
- CO4: To learn Renewal theory and Stationary process
- CO5: To use notions of long-time behavior including transience, recurrence, and equilibrium in applied situations such as branching processes.

Unit I

Introduction to Stochastic Processes; classification according to state space and time domain. Countable state Markov Chains(MC's), Chapman-Kolmogorov equations; calculation of n-step transition probability and its limit. (8 hrs)

Unit II

Stationary distribution, classification of states, transient MC; random walk and gambler's ruin problem; Statistical inference in MC. (8 hrs)

Unit III

Discrete state space continuous time MC: Kolmogorov-Fel1er differential equations; Poisson process, birth and death process; Applications to queues and storage problems. Brownian motion process and its properties, Wiener process as a limit of random walk; first-passage time and other problems. (12 hrs)

Unit-IV

Renewal theory: Elementary renewal theorem and applications. Statement and uses of key renewal theorem; study of residual 1ife time process. Stationary process: weakly stationary and strongly stationary processes; Moving average and auto regressive processes. (12 hrs)

Unit-V

Branching process: Galton-Watson branching process, probability of ultimate extinction, distribution of population size, Martingale in discrete time, inequality, convergence. (10 hrs)

References:

1) A.K. Basu (2003): Introduction to Stochastic Processes, Narosa Publishers.
2) Bhat B.R. (2000): Stochastic Models: Analysis and Applications, New Age International.
3) Cinlar E.(1975):Introduction to Stochastic Processes, Prentice Hall.
4) Karlin S. and Taylor, H.M. (1975): A First Course in Stochastic Processes, Vol.I, Academic Press.
5) Medhi J. (1982): Stochastic Processes, Wiley Eastern.
6) Ross S.M.(1983): Stochastic Processes, John Wiley & Sons.

II SEMESTER	ST-454: DESIGN AND ANALYSIS OF EXPERIMENTS	No. of credits: 4

Course Outcomes:

CO1: Construct and analyse incomplete block designs, Latin square designs and Youden square designs.
CO2: Identify the effects of different factors and their interactions and analysis of factorial experiments.
CO3: Construct complete and partially confounded factorial designs and their analysis.
CO4: Able to analyse the experimental designs with missing values.

Unit-I

Gauss-Markov set-up, normal equations and least squares estimators, error and estimation spaces, Variances and Covariances of least squares estimates, estimation of error variance, least squares estimators, simultaneous estimates of linear parametric functions.

Tests of hypothesis for one and more than one linear parametric functions, Confidence intervals and regions, multiple comparison tests, simultaneous confidence intervals.

(12 hrs)

Unit-II

Introduction to designed experiments, General block design information matrix (C-matrix) and its properties; connectedness, balance and orthogonality, Intra block analysis (contrast estimators, multiple comparisons and testing of linear hypothesis).

(10 hrs)

Unit-III

Balanced incomplete block design (BIBD) – Definition and relations among the parameters, Intrablock analysis. Youden square design. (8 hrs)

Unit-IV

General factorial experiments, factorial effects - best estimators and testing the significance of factorial effects, study of 2 and 3 level factorial experiments in randomized blocks; complete and partial confounding of 2 and 3 level symmetric factorial experiments;

notion of fractional factorial experiments for factors with 2 levels.
(10 hrs)

Unit-V

Split-plot designs analysis of split-plot design with complete blocks. Analysis of covariance for CRD and RBD designs. Missing plot techniques for RBD. (10 hrs)

References:

1) Aloke Dey (1986): Theory of Block Designs, Wiley Eastern.
2) Angela Dean and Daniel VOSS (1999): Design and Analysis of Experiments. Springer.
3) Das M.N. and Giri N.C. (1979): Design and Analysis of Experiments, 2nd Ed., Wiley.
4) Giri N.C. (1986): Analysis of Variance. South Asian Publishers.
5) Hinkleman and Kempthrone C. (1994): Design and Analysis of Experiments, Vol.I, John Wiley.
6) Joshi D.D. (1987): Linear Estimation and Design of Experiments,Wiley Eastern.
7) Montgomery D.C. (2001): Design and Analysis of Experiments, John Wiley.
8) Rao C.R. (1973) Linear Statistical Inference and its Applications. Wiley Eastern.
9) R B Bapat (2011), “Linear Algebra and Linear Models”, Hindustan Book Agency.
10) Parimal Mukhopadhyay (1999) “Applied Statistics”, Books and Applied.

II SEMESTER	ST-455: Practicals based on ST 454	No. of credits: 4

Course Outcomes:

CO1: Construct complete and partially confounded factorial designs and their analysis.

CO2: Able to analyse the experimental designs with missing values.

CO3: Learn to analyse IBD, BIBD and YSD layouts

ST455: Practical's on Design and Analysis of Experiments

1) Linear estimation: Estimability of linear parametric function, Least squares estimators.
2) Testing Linear hypothesis. Analysis of one way and two way classified data.
3) Incomplete Block Design-1: computations of Incidence matrix, C-matrix, Q-matrix, estimability of contrasts, Determining estimable and non estimable treatment contrasts. Best estimates and testing linear restrictions
4) Incomplete Block Design-2 : Intra block Analysis.
5) Balanced Incomplete block design (BIBD): Verifying the relationship between the parameters of the design, computation of C-matrix of the design, best estimates.
6) BIBD: Intra block Analysis
7) Analysis of Youden square Design.
8) Analysis of covariance ANCOVA.
9) Analysis of 2^3 /2^4 Factorial Experiment : Yates table, estimation of main effect and interaction effect , testing the significance of factorial effects.
10) Analysis of 3^2 Factorial Experiment.
11) Complete Confounding $2^3/2^4$ and $3^2/3^3$ Factorial Experiment.
12) 2^3 Partial Confounding
13) 3^2 Partial Confounding
14) Split Plot experiment.
15) Missing plot techniques.

II SEMESTER	ST- 456: Practicals based on ST 451, ST 452 and ST 453.	No. of credits: 4

Course Outcomes:

CO1: To carry out method of scoring and to fit the truncated distributions

CO2: To compute the benefits of insurance, reserves, premiums.

CO3: To generate the random observations from different distributions.

CO4: To familiarize the students with the stochastic processes.

CO5: To familiarize the students with the applications of stochastic methods in practical situations.

1. Generating observations from mixture distributions.
2. Non-central sampling distributions – generating and properties –I
3. Non-central sampling distributions – generating and properties –II
4. Fitting truncated distributions -1
5. Fitting truncated distributions -2
6. Newton-Rapson method to determine maximum likelihood estimate.
7. Method of scoring -1
8. Method of scoring -2
9. Maximum likelihood estimator (when closed form solution does not exist)
10. Sample path of a Markov Chain.
11. Stationary probabilities of a Markov Chain.
12. Poisson process Homogeneous and non-homogeneous.
13. Weiner process, hitting time.
14. Branching process.
15. Simulation of AR(p), MA(q) and plotting the sample path.

III SEMESTER	ST 501: STATISTICAL METHODS	No. of credits: 4

(Choice Paper)

Course Outcomes:

CO1: Demonstrate their knowledge of the basics of inferential statistics by making valid generalizations from sample data

CO2: Perform Test of Hypothesis as well as calculate confidence interval for a population parameter for single sample and two sample cases. Understand the concept of p-values.

CO3: Learn non-parametric test such as the Chi-Square test for Independence as well as Goodness of Fit.

CO4: Compute and interpret the results of Bivariate and Multivariate Regression and Correlation Analysis

Unit I:

Statistics: meaning and role as a decision making science, Data-types and scales of measurement.

Descriptive Statistics - measures of central tendency, positional averages, measures of dispersion, skewness and kurtosis. Presentation-tables, diagrammatic and graphical methods. Exploratory Data Analysis using descriptive measures and graphical tools.

(10 hrs.)

Unit II:

Probability theory: random experiment, simple events, sample space - types of events, probability of an event, rules of probability, conditional probability, Bayes' theorem.

Probability distributions: random variables - discrete and continuous type, Bernoulli, Binomial, Poisson and normal distributions - applications. (10 hrs.)

Unit III:

Sampling methods - population and sample, parameter and statistic, concept of a random sample, simple random sampling, stratified sampling, systematic sampling, sample size determination. (8 hrs.)

Unit IV:

Testing of hypothesis: null hypothesis, alternate hypothesis test statistic, level of significance, p-value, Testing hypothesis about population mean, tests for proportions. Confidence intervals. Contingency tables - Chi-square test for independence of attributes.

(12 hrs.)

Unit V:

Correlation: bivariate data, correlation, scatterplot, correlation coefficient and its properties, testing for correlation coefficient, rank correlation.

Regression: linear relationship, linear regression model, simple linear regression, fitting the regression model, coefficient of determination, standard error of the estimated model. Testing regression coefficients.

(10 hrs.)

References:

1. R.C. Campbell.(1974) : Statistics for Biologists, Cambridge University Press
2. Christopher Chatfield (1981) : Statistics for Technology, Chapman and Hall
3. Harry Frank and Steven C. Athoen (1997) : Statistics: Concepts & Applications, Cambridge University Press.
4. J.Medhi (1992): Statistical Methods : An Introductory Text, Wiley Eastern Limited.
5. Douglas A. Lind, William C. Marchal, Samuel A. Wathen (2012), "Basic Statistics for Business & Economics" McGraw-Hill Education

III SEMESTER	ST-502: TESTING OF HYPOTHESES	No. of credits: 4

Course Outcomes:

CO1: To learn the basics of testing of hypothesis and understand MP and UMP tests.

CO2: To learn Non-existence of UMP test and unbiased test

CO3: To learn to construct confidence intervals for population parameters based on various statistical methods.

CO4: To understand LRT test and large sample tests.

CO5: To familiarize the concepts of non-parametric tests

Unit-I

Framing of null hypothesis, critical region, test functions, two kinds of error, size of a test, p-value, power function, level of a test. Randomized and non-randomised tests, most powerful tests in class of size α - test, Neyman-Pearson lemma, MP test for simple null against simple alternative hypothesis. Distributions with monotone likelihood ratio, UMP tests for one sided null against one sided alternatives,. Extension of these results in Pitman family when only upper or lower end points depend on the parameter.

(12 hrs)

Unit-II

Non-existence of UMP test for simple null against two sided alternatives in one parameter exponential family. Neyman-Pearson generalized lemma. Unbiasedness for hypothesis testing – concept with application to one parameter exponential family. (8 hrs)

Unit-III

Interval estimation, confidence level, construction of confidence intervals by inverting a test statistic and using pivots. Shortest expected length confidence interval, evaluating interval estimators using size and coverage probability and test-related optimality. Uniformly most accurate one-sided confidence interval and its relation to UMP test for one sided null against one sided alternative hypothesis. (10 hrs)

Unit-IV

Likelihood Ratio Test (LRT), Asymptotic distribution of LRT statistic, Pearson's chi-square test for goodness of fit, Bartlett's Test for homogeneity of variances. Large Sample Tests – Wald and Score tests.

(10 hrs)

Unit-V

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. Two-sample tests: Sign test, Wilcoxon signed rank test, the median test, the Wilcoxon-Mann-Whetney test, the Kolmogorov Smirnov two sample test.

(10 hrs)

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.

III SEMESTER	ST-503: ECONOMETRICS	No. of credits: 4

Course Outcomes:

CO1: A broad knowledge of regression analysis relevant for analysing economic data.

CO2: Learn estimation of model parameters ,inference problems in case of simple and multiple linear regression model

CO3: estimate the model parameters of regression model when some of the basic ideal conditions are violated

CO4: regression model adapted to cross section and time-series data

CO5: Learn to select the best subsets of regressors for the model.

CO6: Learn how to use estimated regression models for prediction.

CO7: Understand the estimation techniques and inference procedure in case of simultaneous equations model

Unit-I

Introduction to Econometrics. Nature of econometric study. Simple linear regression, multiple linear regression, basic assumptions. Ordinary Least Squares (OLS) estimation and their properties. Use of prior information. Restricted least squares estimators Tests of hypothesis about regression coefficients and ANOVA. Mixed regression estimator.

(10 hrs)

Unit-II

Dummy variables. Prediction – best linear unbiased predictor. Regression diagnostics and specification tests: Residual analysis for identifying influential observations, recursive residuals and their applications, specification tests. Subset selection of explanatory variables, Mallows C_p-statistic.

(10 hrs)

Unit-III

Violation of basic ideal conditions: Disturbance with non-zero mean; asymptotically unco-operative regressors. Multicollinearity – its consequences and testing. Ridge estimator and its properties. Ridge regression. Stochastic regressors, autoregressive models, Instrumental

variables, Errors in variables.
(10 hrs)

Unit-IV

Heteroscedasticity, tests for heteroscedasticity. Generalised Least Squares (GLS) estimators and its properties. Feasible generalized least squares estimators. Grouping of observations. Sets of Regression Equations. Auto correlation, its consequences and testing for autocorrelation. Estimation. (10 hrs)

Unit-V

Simultaneous equation models. Identification problem. Identification using linear homogeneous restrictions on structural parameters, rank and order conditions.

Estimation in simultaneous equation models – Indirect Least Squares (ILS) estimators, Two State Least Squares (2SLS), Least Variance Ratio(LVR) estimators and their properties. Three stage least squares estimation. (10 hrs)

References:

1. Badi H. Battagi (2002): Econometrics, 3rd Ed., Springer.
2. Draper N.R. and Smith H. (1998): Applied Regression Analysis, 3rd Ed., John Wiley and Sons, Inc.
3. Fomby T.B., Hill C.R. and Johnson S.R. (1988): Advanced Econometric Methods, Springer-Verlag.
4. Greene W.H. (1993): Econometric Analysis, Macmillan, New York.
5. Johnston J. (1984): Econometric Methods, 3rd Ed., McGraw Hill.
6. Johnston J. and Dinardo J. (1997): Econometric Methods, 4th Ed., McGraw-Hill Companies.
7. G.S. Maddala (1977): Econometrics, McGraw-Hill Inc.
8. Peter Schmidt (1976): Econometrics, Marcel Dekker.
9. Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining (2013), Introduction to Linear Regression Analysis, John Wiley
10. Damodar N. Gujarati (2003) “Basic Econometrics”, McGraw Hill

III SEMESTER	ELECTIVE I: ST 504 a) ACTUARIAL STATISTICS	No. of credits: 4

Course Outcomes:

CO1: To understand how actuarial science is used in finance, investments, banking and insurance.

CO2: Explain the concept of survival models

CO3: Describe estimation procedures for lifetime distributions.

CO4: To understand the statistical behaviour of actuarial indicators.

To solve the problems related to the benefit amounts in insurance, annuities, premiums and reserves.

Unit 1

Utility theory, insurance and utility theory, elements of insurance, Individual risk models for a short term, models for individual claims and their sums, survival function, probability of age at death, time until death for a person age x, curtate future lifetime, force of mortality.

(8 hrs)

Unit 2

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

Life table and its relation with survival function, examples, assumptions for fractional ages, some analytical laws of mortality, select and ultimate tables.

(10 hrs)

Unit 3

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

Life annuities: Single payment, continuous life annuities, discrete life annuities, life annuities with monthly payments, commutation functions, varying annuities, recursions, complete annuities.

(10 hrs)

Unit 4

Collective Risk Models for a Single Period: Claim number distribution, distributions of individual and aggregate claims, Reinsurance, the effect of reinsurance, Recursive calculation of aggregate claims distributions, Extensions of the Panjer recursion formula, approximations to the distribution of aggregate claims.

(10 hrs)

Unit 5

Collective Risk Models over an Extended Period: Discrete and continuous time surplus models, compound Poission process as a model for aggregate claim. Ruin probability, Lunberg inequality. Approximations of aggregate claims. Surplus below the initial level, the maximal aggregate loss.

(12 hrs)

References:

1. N.L. Bowers, H.U. Gerber, J.C. Hickman, D.A. Jones and C.J. Nesbitt (1997), "Actuarial Mathematics", Second Edition, The Society of Actuarials.
2. David C. M. Dickson (2005) "Insurance Risk and Ruin" Cambridge University Press.
3. Thomas Mikosch (2006), "Non-Life Insurance Mathematics -An Introduction with Stochastic Processes", Springer.
4. Alexander J. McNeil, Rüdiger Frey, Paul Embrechts (2005), "Quantitative Risk Management: Concepts, Techniques, and Tools", Princeton University Press.
5. Paul Embrechts, Claudia Klüppelberg, Thomas Mikosch (1997), "Modeling Extremal Events: For Insurance and Finance", Springer
6. Tomasz Rolski, Hanspeter Schmidli, Volker Schmidt, and Jozef Teugels (2008) "Stochastic Processes for Insurance & Finance", Wiley Series in Probability and Statistics.

III SEMESTER	ELECTIVE I : 504 b) SURVIVAL ANALYSIS	No. of credits: 4

Course Outcomes:

CO1: To Identify characteristics of survival data and their implications for analysis
CO2: To Perform and interpret univariate analyses of survival data
CO3: To Compare groups using common statistical procedures
CO4: To Analyze survival data and interpret results using Cox proportional hazards model.
CO5: To Assess models for fulfillment of proportional hazards & other aspects of model adequacy

Unit-I

Complete and censored samples, Type I, Type II, and random censoring, Life distributions -Exponential, Gamma, Weibull, Lognormal, Pareto, Proportional Hazards family. Estimation of parameter for exponential and gamma distribution under various censoring situations.

(10 hrs)

Unit-II

Life tables: Standard methods for uncensored and censored data; Asymptotic properties of estimates under a random censorship model. Failure rate, mean residual life and their elementary properties.

Estimation of survival function - Kaplan Meier Estimator, Greenwood's formula. Other life table estimators. (10 hrs)

Unit-III

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

(10 hrs)

Unit-IV

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. (10 hrs)

Unit-V

Two sample problem – Gehan test, log rank test, Mantel Haenszel test. Competing risks model, parametric and non-parametric inference for this model. (10 hrs)

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. Hosmer D.W., Lemeshow S. and May S.(2008): Applied Survival Analysis: Regression Modeling of Time-to-Event Data (2nd Edition), John Wiley& Sons, Inc.

III SEMESTER	ST-505: Practical V (based on ST-502)	No. of credits: 4

Course Outcomes:

CO1: To learn parametric, non-parametric and testing (simple, as well as, composite procedures.

CO2: To learn a strong theoretical background to develop test procedures for any type of populations

CO3: To demonstrate computational skills to implement various statistical inferential approaches.

1) Computing size of the test, power of the test and plotting power function: Discrete distribution.
2) Computing size of the test, power of the test and plotting power function: Continuous distribution.
3) Most powerful tests (Continuous as well as discrete distributions).
4) UMP one sided test including plotting of power function: Discrete distributions.
5) UMP one sided test including plotting of power function: Continuous distributions.
6) UMPU test based on one parameter exponential family (Examples based on Binomial, Poisson distribution).
7) UMPU test based on one parameter exponential family (Examples based on Normal, Lognormal and Exponential distribution).
8) Interval estimation: Pivotal approach and Interval estimation: Through the acceptance region of one sided UMP test and two sided UMPU tests for one parameter exponential family.
9) Likelihood ratio test for finite sample based on one and two independent sample from normal distribution and exponential distribution.
10) Likelihood ratio goodness of fit test and likelihood ratio test for contingency table.
11) Bartlett test for homogeneity of variances.
12) Wald and Score test for large samples.
13) Non parametric test 1: Tests and confidence intervals based on one and two sample sign and Wilcoxon signed rank test, Kolmogorov Smirnoff goodness of fit test, run test.
14) Non parametric test 2: Two and k-sample median test, two sample Wilcoxon-Mann Whitney test, two sample Kolmogorov Smirnoff test.

III SEMESTER	ST-506: Practical VI(based on ST-503)	No. of credits: 4

Practicals on Econometrics

Course Outcomes:

CO1: Learn estimation of model parameters ,inference problems in case of simple and multiple linear regression model

CO2: estimate the model parameters of regression model when some of the basic ideal conditions are violated

CO3: Learn to select the best subsets of regressors for the model.

CO4: Learn how to use estimated regression models for prediction.

CO5: Understand the estimation techniques and inference procedure in case of simultaneous equations model.

1. Simple linear regression.
2. Multiple linear regression.
3. Testing the significance of regressors and ANOVA.
4. Restricted least squares estimators and Testing linear restrictions.
5. Residual Analysis.
6. Best Linear Unbiased Prediction (BLUP) and confidence interval.
7. Testing for autocorrelation and fitting auto-correlated model.
8. Testing Heteroscedasticity in multiple linear regression model.
9. Recursive residuals and their applications.
10. Feasible generalised least squares estimation.
11. Specification Tests of model specification.
12. Multicollinearity.
13. Best subset selection based on MSE, R^2 and Mallows C_p-criterion.
14. Ridge regression.
15. Indirect Least squares(ILS)
16. Two stage least squares (2SLS) estimation.

IV SEMESTER	ST 551: MULTIVARIATE ANALYSIS	No. of credits: 4

Course Outcomes:

CO1: Use multivariate distributions and analysis techniques appropriately

CO2: Undertake multivariate hypothesis tests, and draw appropriate conclusions.

CO3: Analyze multivariate data , dimension reduction and the dependence structure of variables to extract the useful information from multivariate dataset

CO4: Apply suitable tools for classification to formulate and solve real-life problems.

CO5: Apply suitable tools for exploratory data analysis and clustering techniques to extract the useful information from real-life problems.

Unit I

Nature of a multivariate problem, main types of multivariate problems, objectives of multivariate analysis. Organisation of multivariate data, descriptive statistics, visualization techniques.

Multivariate normal distribution – properties, maximum likelihood estimators of the parameters, multiple and partial correlation coefficients, Independence of sample mean vector and sample covariance matrix. Assessing the assumptions of normality – Q-Q plot, Chi-square plot, transformations to near normality. Wishart matrix – statement of Wishart distribution – its properties and applications. (12 hrs)

Unit II

Inference problems in multivariate normal distribution – Hotelling's T^2 & Mahalanobis D^2 – statistics, likelihood ratio tests – one sample and two sample problems, q-sample problem, test for symmetry, confidence regions, simultaneous confidence statements. Independence of subvectors, sphericity test. Test for covariance matrices (Statistics and their distributions- Statements only and applications) (10 hrs)

Unit III

Principal Component Analysis (PCA) – definition and properties, graphing the principal components, sample principal components, interpretation of zero, small and repeated eigenvalues, component loadings and component correlations, the problem of scaling, tests of hypotheses.

Canonical Correlation Analysis – canonical variates and canonical variables, sample canonical variates, sample canonical correlations, inference problems. (10 hrs)

Unit IV

Classification and Discrimination problems – concepts of separation and classification, Bayes', minmax and Fisher's criteria, classification rules based on Expected Cost of Misclassification (ECM) and Total Probability of Misclassification (TPM), classification with two multivariate normal populations (equal and unequal covariance matrices), evaluating classification rules, classification with several populations, Fisher's linear discriminant function, tests associated with discriminant functions.
(10 hrs)

Unit V

Factor Analysis: orthogonal factor model, factor loadings, estimation of factor loadings, factor scores.

Cluster Analysis: distances and similarity measures, hierarchical clustering methods, K – means method.

Multidimensional scaling: nature of the problem, classical solution. (8 hrs)

References :

1. T.W.Anderson (1984): An Introduction to Multivariate Analysis, 2nd Ed., John Wiley.
2. Bernard Flury (1997): A First Course in Multivariate Statistics, Springer Texts in Statistics.
3. A.M.Kshirasagar (1972): Multivariate Analysis, Marcel Dekker.
4. K.V.Mardia, J.T. Kent and J.M. Bibby (1979): Multivariate Analysis, Academic Press.
5. C.R. Rao (1973): Linear Statistical Inference and its Applications, Wiley Eastern.

6. Richard Arnold Johnson, Dean W. Wichern (2007) Applied Multivariate Statistical Analysis, Prentice Hall.

7. Alvin C. Rencher, William F. Christensen (2012), "Methods of Multivariate Analysis" John Wiely.

IV SEMESTER	ST-552: OPERATIONS RESEARCH	No. of credits: 4

Course Outcomes:

CO1: To apply the theorems on duality to problems appropriately.

CO2: To explain the concept of complementary slackness and its role in solving primal / dual problem pairs.

CO3: To be able to modify a Primal Problem, and use the Fundamental Insight of Linear Programming to identify the new solution, or use the Dual Simplex Method to restore feasibility.

CO4: To solve the problems related to linear programming

Unit I

Linear Programming Problem (LPP) – definition, formulation, simplex method – canonical form, improving nonoptimal basic feasible solution (b.f.s), conditions for optimality, conditions for unboundedness. Two phase method, Big M method. Convex sets, geometry of simplex method- extreme point and b.f.s., existence of b.f.s., existence of optimal b.f.s.

(12 hrs)

Unit II

Duality theory of LPP – weak duality theorem and its properties, the fundamental duality theorem, complementary slackness theorem. Dual simplex method. Sensitivity analysis. Integer programming-cutting plane technique, Gomory's algorithm for pure integer program.

(10 hrs)

Unit III

Dynamic Programming - Multistage decision making problems, Bellman's principle of optimality, recursive nature of computation, applications of dynamic programming, probabilistic dynamic programming.

Markovian Decision Processes, finite stage dynamic programming model, infinite-stage model, policy iteration methods. (8 hrs)

Unit IV

Inventory theory – nature of inventory problem, motives for carrying inventory, deterministic inventory model with decay, finite horizon model with variable demand rate. Probabilistic inventory models – continuous review and periodic review systems, (s, S) policy, heuristic solution of lot size reorder point model [(Q, r) policy]. (10 hrs)

Unit V

Queuing theory – characteristics of queues, M/M/1 system, steady state solution, measures of effectiveness, waiting time distributions, Little's formula, M/M/1/K system, M/M/c system, machine interference problem, M/G/1 system- Pollaczek-Khintchine formula. (10 hrs)

References:

1. D.Gross and C.M.Harris (1985): Fundamentals of Queuing Theory, 2nd Ed., John Wiley.
2. G. Hadley (1975): Linear Programming, Addison Wesley.
3. Katta G. Murthy (1976): Linear and Combinatorial Programming, John Wiley & Sons.
4. N.S. Kambo (1991): Mathematical Programming Techniques, Affiliated East-West Press.
5. H. A. Taha (2001): Operations Research – An Introduction (6th Edition), Prentice-Hall, India.
6. B.D. Sivazlian and L.E. Stanfel (1975): Analysis of Systems in Operations Research, Prentice-Hall.
7. H.G.Daeallenbach & John A.George(1978): Introduction to Operations Research Techniques, Allyn & BaconInc.

IV SEMESTER	ELECTIVE II: ST 553 a) TIME SERIES ANALYSIS	No. of credits : 4

Course Outcomes:

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

Unit-I

Simple Descriptive Techniques: time series plots, trend, seasonal effect. Tests for trend and seasonality: estimation and elimination of trend and seasonal components. Exponential and moving average smoothing.

Time Series as discrete parameter stochastic process. Stationarity, autocovariance and autocorrelation function and their properties. Partial autocorrelation function.

(10 hrs)

Unit-II

Probability Models: White noise model, random walk, linear processes, Moving Average (MA), Autoregressive (AR), ARMA and ARIMA, seasonal ARIMA models. Invertibility. ACF and PACF of these processes. Sample ACF and PACF. Model identification.

(10 hrs)

Unit-III

Model Building: Estimation of mean, autocovariance function and autocorrelation function. Estimation of AR models – Yule-Walker equations, estimation of MA model and ARMA models. Order selection in AR and MA models. (10 hrs)

Unit-IV

Forecasting: Forecast mean square error (FMSE), Least squares prediction. BLUP. Box-Jenkins forecasting. Forecasting through exponential smoothing and Holt-Winters smoothing. Residual analysis and diagnostic checking. Nonstationary time series models and their identification. (10 hrs)

Unit-V

Spectral Analysis: Simple sinusoidal model, Spectral analysis of covariance stationary processes. Spectral distribution function, Spectral densities. Periodogram analysis: relation between periodogram and autocovariance. Estimation of spectral densities of AR, MA and ARMA models.

(10 hrs)

References:

1. Box GEP and Jenkins G.M. (1976): Time Series Analysis: Forecasting and Control, Holden-day, San Franscisco.
2. Brockwell P.J. and Davis R.S. (2002): Introduction to Time Series and Forecasting,2nd Ed., Springer.
3. Chatfield C. (1996): The Analysis of Time Series An Introduction, Chapman & Hall.
4. Janacek G. (2001): Practical Time Series Arnolds Texts in Statistics.
5. Kendall M.G. and Ord J.K. (1990): Time Series, 3rd Ed., Edward Arnold.
6. Montgomory D.C. and Johnson L.A. (1977): Forecasting and Time Series Analysis, McGraw Hill.
7. K. Tanaka (1996): Time Series Analysis, Wiley Series.
8. Dilip M. Nachane (2006) " Econometrics- Theoretical Foundations and Empirical Perspectives", OUP India

IV SEMESTER	ELECTIVE II : ST 553 b) DATA MINING TECHNIQUES	No. of credits : 4

Course Outcomes:

CO1: Design data warehouse with dimensional modelling and apply OLAP operations.

CO2: Gain knowledge about basic concepts of Machine Learning and Identify machine learning techniques suitable for a given problem

CO3: Compare and evaluate different data mining techniques like classification, prediction, clustering and association rule mining

CO4: Apply Dimensionality reduction techniques.

CO5: To assess the strengths and weaknesses of the different algorithms, identify the application area of algorithms and apply them.

CO6: Apply data mining techniques as well as methods in integrating and interpreting the data sets and improving effectiveness, efficiency and quality for data analysis.

Unit I

Data Mining – motivations and importance, Knowledge Discovery in Databases (KDD) process - search, induction, querying, approximation and compression. Kinds of data considered for data mining, basic data mining tasks, data mining issues, Data Mining models - predictive and descriptive, inter-connections between Statistics, Data Mining, Artificial Intelligence and Machine Learning. Applications tof data mining. (8 hrs)

Unit II

Data marts, databases and data warehouses - OLTP systems, multidimensional models – data cubes, OLAP operations on data cubes, multidimensional schemas.

Data pre-processing – data cleaning, data integration, data transformation and data reduction.

Visualisation techniques for multidimensional data - scatter plot matrix, star plots, Andrews plots, Chernoff faces, parallel axis plots.

(10 hrs)

Unit III

Supervised learning – classification and prediction, statistical classification-Linear Discriminants-Mahalanobis' linear discriminant, Fisher's linear discriminant; Bayesian classifier, Regression based classification, k-NN(nearest neighbour) classifier. Tree classifiers-decision trees, ID3 algorithm CART.

Artificial Neural Networks (ANN)-the learning problem, perceptron, the delta rule, multilayer feed forward neural network, back propagation learning algorithm

Support Vector Machines-Lagrangian formulation and solution, Measuring classifier accuracy. (12 hrs)

Unit IV

Unsupervised learning – Clustering problem, similarity and distance measures, Partitioning algorithms-k-means & k-medoids(PAM) algorithms. Density based clustering algorithms (DBSCAN).

Association Rule Mining - market basket analysis-frequent item sets, support and confidence of an association rule, a priori algorithm, Partition algorithm. (10 hrs)

Unit V

Computational methods useful in datamining: Expectation-Maximisation (EM) algorithm, Genetic algorithm, Markov Chain Monte Carlo(MCMC) method.

Resampling Techniques - Gibbs sampler, Bootstrap sampling, Cross-validation, Bagging, Arcing. (10 hrs)

References:

1. .Jiawei Han, Micheline Kamber: (2002): Data Mining-Concepts and Techniques, Morgan Kaufman Publishers, U.S.A
2. Margaret.H.Dunham (2005): Data Mining-Introductory and Advanced Topics, Pearson Education.
3. Trevor Hastie, Robert Tibshirani & Jerome Friedman (2001):The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer, New York,
4. Michael Berthold, David J. H and (Eds): (2003) Intelligent Data Analysis - An Introduction (2^{nd} Ed), Springer.
5. J.P. Marques de Sa: (2001):Pattern Recognition - Concepts, Methods and Applications, Springer 6.
6. Rajan Chattamvelli: (2009): Data Mining Methods, Narosa Publishing House

IV SEMESTER	ELECTIVE II: ST 553 c) NONPARAMETRIC REGRESSION	No. of credits: 4

Course Outcomes:

CO1: To understand nature and scope of nonparametric regression

CO2: To learn kernel approaches to density estimation and regression

CO3: Ability to apply spline and basis approaches and its computational issues

CO4: An insight into asymptotic properties different method.

CO5: To introduce generalised additive models

Unit-I

Nature and scope of nonparametric regression Basic idea of smoothing, Smoothing histograms and nonparametric probability density function. Random design and fixed design model, Bin smoothers, running mean and running line smoothers. Univariate Kernel density estimation. Local regression estimate. (12 hrs)

Unit-II

Kernel Regression: Kernel smoothing, Local polynomial Kernel estimators, Kernel estimators of derivatives, computational aspects of Kernel smoothing, K-nearest neighbor (KNN) estimates. Computational aspects of K-NN estimators. (10 hrs)

Unit-III

Spline Smoothing: Roughness penalties, quantifying roughness of a curve, cubic splines, computational aspects of natural cubic splines, Orthogonal series estimators. (8 hrs)

Unit-IV

Data-Driven Choice of Smoothing Parameters: Cross validation, Risk estimation. The Plug-in-method. Bandwidth choice for derivative estimation. Local adaption of the smoothing parameter. (10 hrs)

Unit-V

Lack of Fit Tests: Testing the fit of a linear model. Lack of fit tests based on linear smoothers: Smoothing residuals, comparing parametric and nonparametric models, Introduction to additive models. Semiparametric regression models. (10 hrs)

References:

1. Clive Loader (1999): Local Regression and Likelihood, Springer.
2. Hardle (1990): Applied Non-parametric Regression, Cambridge University Press.
3. Hart J.D. (1997): Non-parametric Smoothing and Lack of Fit Tests, Springer Verlag.
4. Hastie T.J. and Tibshirani R.J. (1990): Generalised Additive Models, Chapman & Hall.
5. John Fox (2000): Nonparametric Sample Regression, Sage Publications.
6. Takezawa K. (2005): Introduction to Non-parametric Regression - Wiley Series in Probability and Statistics, John Wiley and Sons.
7. Wand and Jone (1995): Kernel Smoothing, Chapman & Hall.

IV SEMESTER	ELECTIVE III: 554 a) STOCHASTIC FINANCE	No. of credits : 4

Course Outcomes:

CO1: The ability to model the returns.

CO2: The ability to understand the basic concepts of financial system.

CO3: The ability to explain portfolio theory.

CO4: The ability to understand Black- Scholes models.

CO5: The ability to deal with forward contracts and futures

Unit I

Basic concepts of financial markets and financial systems. Functions of financial markets, types of traders, stock markets.

Interest rates, continuous compounding, present value analysis - effective interest rate, present value and future value. Bond pricing, bond yield and par yield. Returns, gross returns, log returns. (10 hrs)

Unit II

Portfolio theory – mean variance portfolio theory. Risk and return, risk free interest rate. One risky asset and one risk free asset. Two risky assets. Sharpe's ratio, tangency portfolio, optimal mix of portfolio. Market portfolio, beta, security market line, and capital asset pricing model (CAPM) and their assumption.

Value at Risk (VaR) – Nonparametric and parametric estimation of VaR , VaR for a derivative and for a portfolio of assets – delta-normal method. Simulation of VaR models.

(10 hrs)

Unit III

Forward contracts and Futures. Call and put options, European option and American options, short and long positions. Financial derivatives, options, pricing via arbitrage, law of one price. Risk neutral valuation, arbitrage theorem. Convexity of cost of call option, Risk neutral probabilities- Binomial model, and multi-period model.

Modeling returns: lognormal model, random walk model, modeling through geometric Brownian motion process. (10 hrs)

Unit IV

Differential form of Weiner process, stock price process, Ito lemma (without proof). The Black-Scholes formula and assumptions. Properties of the Black-Scholes option cost. Delta, gamma and other Greeks.

Volatility and estimating the volatility parameter. Implied volatility. Pricing American options. Pricing of a European option using Monte-Carlo and pricing an American option using finite difference methods. Call options on dividend-paying securities. (10 hrs)

Unit V

Financial Time Series – Review of AR(p) and MA(q) processes, ARMA (p, q) models, the first and second order moments, estimation of parameters. Models for conditional heteroscedasticity - ARCH(1), ARCH(q), and GARCH(1,1) models elementary properties, moments and their estimation, diagnosis check in linear time series analysis. (10 hrs)

References:

1. Sheldon M. Ross (2003): "*An elementary introduction to Mathematical Finance*", Cambridge University Press.
2. David Ruppert (2004) *"Statistics and Finance an Introduction"* – Springer International Eddition.
3. John C. Hull (2008) "*Options, Futures and other derivatives*", Pearson Education India.
4. Cuthbertson K and Nitzsche D (2001): "*Financial Engineering - Derivatives and Risk Management*", John Wiley & Sons Ltd.
5. David G Luenberger(1998): "*Investment Science*", Oxford University Press.
6. Paul Wilmott (2000): "*Quantitative Finance*", John Wiley & Sons.
7. Ruey S. Tsay (2005) "*Analysis of Time Series*", John Wiley & Sons.

IV SEMESTER	ELECTIVE III: ST 554 b) BAYESIAN INFERENCE	No. of credits : 4

Course Outcomes:

CO1: Explain in detail the Bayesian framework for data analysis and its flexibility and be able to demonstrate when the Bayesian approach can be beneficial.

CO2: Develop, analytically describe, and implement both single and multi parameter probability models in the Bayesian framework.

CO3: Demonstrate the role of the prior distribution in Bayesian inference and be able to articulate the usage of non-informative priors and conjugate priors.

CO4: Show high level Interpretation of Bayesian Analysis Results and be able to readily perform Bayesian model evaluation and assessment.

CO5: Demonstrate the necessary skills to: fit hierarchical models, provide thorough technical specifications for these models.

CO6: Perform Bayesian computation using Markov chain Monte Carlo methods

Unit-I

Limitations of empirical and logical theories of probability, Subjective probability, determination of subjective probability, likelihood function, prior distribution, posterior distribution. Bayes' theorem, methods of construction of priors and computation of the posterior distribution.

Natural conjugate family of priors for a model. Hyper parameters of a prior from conjugate family. Conjugate families for (i) exponential family models, (ii) models admitting sufficient statistics of fixed dimension.
(10 hrs)

Unit-II

Enlarging the natural conjugate family by (i) enlarging hyper parameter space (ii) mixtures from conjugate family, choosing an appropriate member of conjugate prior family. Non informative, improper and invariant priors. Jeffrey's invariant prior. (10 hrs)

Unit-III

Bayesian point estimation: As a prediction problem from posterior distribution. Bayes estimators for (i) absolute error loss (ii) squared error loss (iii) 0-1 loss function. Generalization to convex loss functions. Evaluation of the estimate in terms of the posterior risk. (10 hrs)

Unit-IV

Bayesian interval estimation: Credible intervals. Highest posterior density regions. Interpretation of the confidence coefficient of an interval and its comparison with the interpretation of the confidence coefficient for a classical confidence interval.

Bayesian testing of hypothesis: Specification of appropriate form of the prior distribution for a Bayesian testing of hypothesis problem. Prior odds, Posterior odds, Bayes factor for various types of testing hypothesis problems depending upon whether
the null hypothesis and the alternative hypothesis are simple or composite. Specification of the Bayes tests in the above cases. (10 hrs)

Unit-V

Bayesian prediction problem. Large sample approximations for the posterior distribution. Monte Carlo Methods and Importance Sampling, Markov Chain Monte Carlo (MCMC) Methods: The Gibbs sampler. (10 hrs)

References:

1. Bernando J.M. and Smith A.F.M.: Bayesian Theory, John Wiley.
2. Berger J.O. (1988): Statistical Decision Theory and Bayesian Analysis, Springer- Verlag, New York Inc.
3. Degroot M.H.: Optimal Statistical Decisions, McGrow Hill.
4. Ghosh J.K., Delampady M. and Samanta T. (2006): An Introduction to Bayesian Analysis: Theory and Methods, Springer.
5. Leonard T. and Hsu J.S.J. (1999): Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers, Cambridge University Press.
6. Robert C.P. and Casella G.: Monte Carlo Statistical Methods, Springer-Verlag.

IV SEMESTER	**ST 555 : Practical VII (Based on Theory Papers)**	**No. of credits : 4**

Course Outcomes:

CO1: The ability to forecast future observations of the time series.

CO2: A running knowledge of R for applied time series analysis

CO3: To solve the problems related to linear programming

CO4: A running knowledge of R in the estimation techniques in the other applied areas.

CO5: Perform Bayesian computation using methods in R

Practicals on ST 551 Multivariate Analysis

1. Plotting of bivariate normal distribution.
2. Assessing normality of data – Q-Q plot and Chi-Square plot.
3. Hoteling's T^2 statistic – I (one sample and two sample problem)
4. Principal Component Analysis
5. Classification and discrimination -II

Practicals on ST 552 Operations Research

1. Simplex Method
2. Two phase method
3. Big M method
4. Dual LPP and Dual Simplex method.
5. Integer Programming

Practicals on ST 553 a) Time Series Analysis

1. Estimation and elimination of trend component. Variate difference method.
2. Sample ACF and PACF.
3. Identification of moving average (MA) and Auto regressive(AR) process and its order selection.
4. Yule-Walker estimation for AR(p) model. Fitting MA model using Least squares regression.
5. Residual Analysis and Diagnostic checking.

Practicals on ST 553 b) Data Mining Techniques

1. Decision Tree for classification and Classification using ANN

2. Bayesian classifier
3. k-NN classifier
4. Clustering techniques
5. Association Rule Mining

Exercises can be done using WEKA software, which is freely downloadable.

Practicals on ST 553 c) Nonparametric Regression

1. Density Estimation through histogram function for univariate data.
2. Kernel density estimation for univariate data
3. Nadaraya –Watson Kernel Regression estimator .
4. Local polynomial Kernel Regression.
5. Smoothing splines.

Practicals on ST 554 a) Stochastic Finance

1. Exercises on interest rates and net present values
2. Bond analysis, bond pricing bond yield, par yield
3. Returns, log returns, plots, value at risk
4. Option pricing - Binomial tree and multiperiod option pricing, applications of Black-Scholes formula
5. Option pricing when dividends are paid.

Practicals on ST 554 b) Bayesian Inference

1) Bayes estimation under conjugate family, hyperparameters of the conjugate family and mixtures of conjugate families.
2) Bayesian credible interval, HPD credible interval from exponential family.
3) Estimation of posterior density, HPD credible intervals using importance sampling from exponential family.
4) Posterior density estimation, HPD credible intervals using Gibbs sampler.
5) Posterior density estimation, HPD credible intervals and prediction of future observation using Metropolis Hasting algorithm.

~~~~~~~~~~~~~
~~~~~~~~~~~~~