UNIVERSITY OF MYSORE

Department of Studies in Statistics

Manasagangotri, Mysuru 570006

SCHEME of INSTRUCTION and SYLLABI

This document has Syllabi of M.Sc. Statistics program and Syllabi for the Ph.D. Statistics Course Work program given in the last three pages.

M.Sc. STATISTICS

under

CHOICE BASED CREDIT SYSTEM (CBCS)

and

FLEXIBLE CHOICE BASED CREDIT SYSTEM (FCBCS from 2019 admissions onwards)

1. Scheme of Admission: Based on Merit-cum-Roster system of Govt. of Karnataka and through University of Mysore Entrance Test

2. Eligibility: B.Sc. degree or equivalent with E grade (50%-60% marks) and above in Statistics and Mathematics as one of the core / optional subjects OR B.Sc. degree or equivalent with D grade (60% - 70% marks) and above in Mathematics OR B.E. Degree or equivalent with D grade (60% - 70%) in aggregate in Mathematics papers with level of Syllabi as prescribed by the Vishveshwaraya Technological University (VTU), Karnataka.

3. Evaluation: The Evaluation will be totally internal.

4. Scheme of Examination: As prescribed by University regulations.

5. Minimum Credits for getting the M.Sc. Statistics degree: 76 credits from 4 semesters.

6. Grades in each Course: As per University regulations.

7. Hard Core courses are compulsory courses and Soft Core courses are program specific courses which are available as choice course. A particular Soft Core will be offered depending upon the availability of qualified teachers. Open Elective Courses are courses designed for students studying programs in other subjects who wish to gain basic knowledge of Statistics.

M.Sc. Statistics Program:

Learning Objectives: The M.Sc. Statistics program is a rigorous program in Probability Theory and Statistical Inference and is designed to give a sound foundation in fundamentals and training in practical Statistics leading to statistical data analysis. The four semester 76 credit program has a variety of soft-core courses to choose from including enough courses on statistical software.

Program specific Outcomes: A person successfully completing the program will have enough knowledge and expertise to statistically analyze small and large data sets, pursue advanced courses in Statistics or a Ph.D. in Statistics, work in software industry as domain expert, independently consult for statistical data analysis and whatnot. The program has proved to be one of the best in traditional Indian Universities and has demand from students within and outside the State.

Syllabi for Semester I to IV:

Semester I :

Hard Core : Course 1: Real Analysis (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This course is intended as a foundational course and includes advanced real analysis needed to understand and apply Probability Theory and Mathematical Statistics. The course is designed as a prerequisite to understand subsequent courses in Probability Theory and Mathematical Statistics.

Course Outcomes : A person successfully completing the Course will have enough knowledge of Real Analysis including standard techniques used in proofs of results in Real Analysis. Standard skills to solve problems in Analysis are learnt in the Course and these are useful to understand topics in Probability Theory and Mathematical Statistics and to apply to obtain results and solve problems in these subjects.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in problem solving techniques in Real Analysis is part of the Prcatical Course in the Semester. Students are encouraged to use resources available on the web.

Assessment : The assessment is done through periodic written tests, oral quizzes, assignments and student seminars.

Unit 1: Elements of set theory, Sets in Euclidean space of k-dimensional R^k rectangles, neighbourhood, interior point and limit point, open and closed sets, Bolzano-Weierstrass theorem in R^2, Real valued functions continuity and uniform continuity.

Unit 2: Sequences and Series of constants- Limit superior, limit inferior and limit - properties. Cesaro sequences. Series of positive terms - Tests for convergence, divergence. Integral and Order tests and Kummers' test (statement only of all the tests)- Ratio and Raabe's tests as special cases of Kummers' test. Series of arbitrary terms - absolute and conditional convergence.

Unit 3: Sequences of functions-Uniform convergence and point wise convergence, Series of functions-uniform convergence-Weierstrass' M test. Power series and radius of convergence. Riemann-Stieltjes integration-continuous integrand and monotonic /differentiable integrator.

Unit 4: Functions of two variables-partial and directional derivatives. Maxima and minima of functions, maxima-minima under constraints(Lagrange's multipliers).

Unit 5: Parametric functions. Uniform convergence of improper integrals, Differentiation under integrals. Double integrals and repeated integrals. Change of variables under double integration-statement of the theorems without proof and solution of problems.

Books for reference:

Apostol, T.M. (1985): Mathematical Analysis, Narosa India Ltd.
Courant, R. and John, F. (1965): Introduction to Calculus and Analysis, Wiley.
Goldberg, R.R.(1970): Methods of Real Analysis, Oxford Publishing Co.
Khuri, A.T. (1993): Advanced Calculus with Applications in Statistics, John Wiley.
Rudin, W. (1976): Principles of Mathematical Analysis, Mc Graw Hill.
Shantinarayan (1950) : A course of Mathematical analysis, Sultan Chand and Co.

Hard Core : Course 2: Linear Algebra (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This course is intended as a foundational course and includes advanced topics in Linear Algebra needed to understand and apply Probability Theory and Mathematical Statistics dealing with two or more than two random variables. The course is designed as a prerequisite to understand subsequent courses in Probability Theory and Mathematical Statistics.

Course Outcomes : A person successfully completing the Course will have enough knowledge of Linear Algebra including well known techniques used in proofs of results in Linear Algebra. Skills to solve problems in Linear Algebra are learnt in the Course and these are useful to understand topics in Probability Theory and Mathematical Statistics in general and Linear Models and Multivariate analysis in particular and to apply to obtain results and solve problems in these subjects.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in Linear Algebra is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Fields, vector spaces, subspaces; linear dependence and independence; basis and dimension of a vector space, finite dimensional vector spaces completion theorem. Examples of vector spaces over real and complex fields. Linear equations. Vector spaces with an inner product, Gram-Schmidt orthogonalization process. Orthonormal basis and orthogonal projection of a vector.

Unit 2: Linear transformations, algebra of matrices, row and column spaces of a matrix. Elementary matrices, determinants, rank and inverse of a matrix. null space and nullity; partitioned matrices; Kronecker product. Hermite canonical form, generalized inverse, Moore- Penrose Inverse, Idempotent matrices. Solutions of matrix equations.

Unit 3: Triangular reduction of a positive definite matrix. Characteristic roots and vectors, Cayley-Hamilton theorem, minimal polynomial, similar matrices. Algebraic and geometric multiplicity of characteristic roots, spectral decomposition of a real symmetric matrix, reduction of a pair of real symmetric matrices, Hermitian matrices.

Unit 4: Real quadratic forms, reduction and classification of quadratic forms, index and signature. Singular values and singular decomposition, Jordan decomposition, extrema of quadratic forms. Vector and matrix differentiation.

Books for Reference:

Bellman, R. (1970): Introduction to Matrix Analysis, Second Edition, McGraw Hill.
Biswas, S. (1984): Topics in Algebra of Matrices, Academic Publications.
Graybill, F. A. (1983): Matrices with Applications in Statistics, Second Edition, Wadsworth..
Hadley, G. (1987): Linear algebra, Narosa.
Halmos, P. R. (1958): Finite Dimensional Vector Spaces, Second Edition, D. Van Nostrand Company.
Hoffman, K. and Kunze, R. (1971): Linear Algebra, Second Edition, Prentice Hall.
Rao, A. R. and Bhimasankaram, P.(1992): Linear Algebra, Hindustan Book Agency.
Rao, C. R (1973): Linear Statistical Inference and its Applications, Second Edition, Wiley.
Rao, C. R. and Mitra, S. K (1971): Genera1ized Inverse of Matrices and its Applications, Wiley.
Searle, S. R (1982): Matrix Algebra Useful for Statistics, Wiley

Hard Core : Course 3: Probability Theory and Distributions – I (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This is the first of two foundational courses in Probability Theory and Distributions and gives a good foundation in measure theoretic Probability and standard univariate and bivariate distributions. This is a prerequisite for courses on Mathematical Statistics.

Course Outcomes : A person successfully completing the Course will acquire basic knowledge of axiomatic Probability Theory and learn standard distributions and their properties. This basic course is a prerequisite to an advanced course as well as to understand topics in Mathematical Statistics.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in the subject is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: 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 on the real line. Events, Sample space, Probability measure, Additive property,properties related to sequences of events, Independent events, Conditional probability and Bayes' theorem.

Unit 2: Measurable functions, random variables, sums, product and functions of random variables, sequence of random variables. Induced Probability measure, Distribution function Jordan decomposition theorem. Bivariate distributions-joint marginal and conditional distributions. Expectations and conditional Expectations.

Unit 3: Standard discrete and continuous univariate distributions and their properties, Probability generating function and moment generating function. Bivariate normal and Multinomial distributions

Unit 4: Transformation technique. Sampling distributions, Chi-square,Students t, F and Non-central chi-square and their properties. Bivariate Negative Binomial, Beta and Gamma distributions.

Unit 5: Markov, Chebyshev, Hoelder, Minkowski, Jensen and Liapunov inequalities. Relationship between tail of distributions and moments.

Books for Reference:

Cramer, H. (1946): Mathematical Methods of Statistics, Princeton.
Johnson, S. and Kotz. (1972); Distributions in Statistics, Vols.I, II and III, Houghton and Miffin.
Mukhopadhyaya, P. (1996): Mathematical Statistics, Calcutta Publishing House.
Pitman, J. (1993): Probability, Narosa.
Rao, C. R (1973): Linear Statistical Inference and its Applications, Second Edition, Wiley Eastern.
Rohatgi, V. K (1984): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern
Lukacs, C (1970)' Characteristic functions, Griffin Publications.

Hard Core : Course 4: Practicals based on Papers I – III (4 Credits – 8 hours of Practicals per week)

Learning Objectives : This is a Practical training course and exposes students to practical problems in the hard core courses Real Analysis, Linear Algebra, Probability Theory and Distributions I. The course is carried out with traditional problem solving techniques and in R software.

Course Outcomes : A person successfully completing the Course will acquire practical knowledge of solving problems and proof techniques in the subjects. Writing R codes, visualization and running R packages are part of the Course. .

Pedagogy : The course is taught using traditional chalk-and-talk method and on Computers using R software. Students are encouraged to use R resources available on the web.

Assessment : The assessment is done through periodic tests involving R software wherein the students are asked to write R codes and run them in R software and also perform data analysis on data available online and from standard text books. The students are also assessed for problem solving through oral quizzes, assignments and student seminars.

Soft Core : Course 5: Statistical Computing (4 Credits – 8 hours of Practicals per week)

Learning Objectives : As the name suggests, this course on Statistical Computing uses R software for implementation and students are trained on several practical problems in this course to equip them to think independently and for analysis of statistical data. After finishing this course, one can write R codes for statistical methods and implement R codes for the various methods learnt in the M.Sc. Program.

Course Outcomes : A person successfully completing the Course will acquire practical knowledge of solving problems and proof techniques in several topics in Statistical Computing. Writing R codes, visualization and running R packages are part of the Course. .

Pedagogy : The course is taught for advance data analysis using R software. Students are encouraged carryout data analysis using R.
Assessment : The assessment is done through periodic tests involving R software wherein the students are asked to write R codes and run them in R software and also perform data analysis on data available online and from standard text books. The students are also assessed for problem solving through oral quizzes, assignments and student seminars.

Unit 1: A general overview of R . Numeric/Character/logical; real /integer/complex string and the paste command matrices, dataframes, lists, setwd,read.table,read.csv.write. matrix, write.csv, creation of new variables, catergorization cut, factor,round, apply, creation of pattern variables, saving output files; source;print saving work space/ history. R-help command; help. search(), r mailing list, contributed documentation on cran. Descriptive statistics and Graphics in R: summary statistics for single group, the plot-command, histogram, box plot bar plot, lines, points, segments, arrows, paste, inserting math. symbols in a plot, pie diagram, customization of plot- setting graphical parameters text and m-text the pairs command. Graphical parameters such as par/main/mfrow/xlab/ylab/las/xaxp/xlim/ylim/cex/axis/tck/srt/main/title/legend/locator/ identity.

Unit 2: One and two sample t-tests, chi-squared tests, F- test for equality of variance, nonparametric tests, regression analysis, checking the assumptions of normality, Q-Q plots, P-P plots. ANOVA. Matrix operations, addition, subtraction, multiplication, linear equation and eigenvalues, finding rank, inverse, g-inverse, determinant. R –functions; some useful built in r functions, attach, detach, sort, order, rank, ceiling, floor, round, trunc, signif, apply, lapply, by. Programming in R;fro/while/loops, functions, the source command.

Unit 3: Numerical analysis and statistical applications. Numerical integration, root extraction, random number generation, Monte Carlo integration, matrix computations, drawing random samples from known univariate probability distributions -both discrete and continuous and bivariate normal distribution - the inverse method, the accept- rejection method, decomposition of discrete mixtures, Classical Monte Carlo integration. R-functions for generating random variables and simulations; rnorm, rbinom, rpiosson, runif,rchisq,rt, etc.,; sample, set.seed.

Unit 4: Writing / performing programs using MATLAB /MINITAB/ SPSS/ Excel on problems from the following topics: Descriptive statistics and Graphs, One and two sample parametric and nonparametric tests, Chi-square tests, Regression and correlation analysis, Analysis of Variance and Kruskal-Wallis Test.

References:

1. Dalgaard, P. (2002). Introductory Statistics with R. Springer Verlag, New York.
2. Kerns, G.J. (2010). Introduction to Probability and Statistics Using R. Free Software Foundation.
3. Kunte, Sudhakar (1999). Statistical Computing: 1. Understanding Randomness and Random Numbers,Resonance, Vol.4, No.10, pp.16-21.
4. Kunte. Sudhakar (2000). Statistical Computing: 2. Technique of Statistical *Simulation,Resonance,* Vo1.5, No.4, pp.18-27.
5. Robert, C. and Casella, G. (2010). Introducing Monte Carlo Methods with R. Springer Verlag, New York.
6. Sudha G Purohit, Sharad D Gore, Shailaja R Deshmukh (2010) Statistics Using R, Alpha Science Intl. Publ.
7. Verzani, J. (2005). Using R for Introductory Statistics. Taylor & Francis

Soft Core : Course 6: Sample Surveys and Statistics for National Development (4 Credits – 3 hours of Theory + 2 hours of Practicals teaching per week)

Learning Objectives : This course gives an overview of sampling design and methodology and also discusses important statistical indicators used in national development. Several sampling techniques are part of this course and a comprehensive introduction to statistics for national development is also included in the course.

Course Outcomes : A person successfully completing the Course will acquire a very good knowledge of standard sampling designs and a comprehensive knowledge of Statistics used in study of National Development and the Course also has Practical problem solving and data analysis techniques.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in the subjects is given as part of the Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

A. **SAMPLE SURVEYS**

Unit 1: Basic finite population sampling techniques (SRS WR/ WoR, stratified, systematic), related problems of population mean estimation, allocation problems in stratified sampling.

Unit 2: Unequal probability sampling: PPS WR / WoR methods (including Lahiri's scheme) and related estimators of a finite population mean (Hansen-Hurvitz and Desraj estimators for a general sample size and Murthy's estimator for a sample of size 2).

Unit 3: Ratio and regression estimators based on SRS WoR method of sampling, two-stage sampling with equal. number of second stage units, double sampling, cluster sampling.

B. STATISTICS FOR NATIONAL DEVELOPMENT

Unit 4: Economic development: growth in per capita income distributive justice. Indices of development, Human Development Index. Estimation of National Income - product approach, income approach and expenditure approach. Population growth in developing and developed countries. Population projection using Leslie matrix. Labour force projection.

Unit 5: Measuring inequality of incomes, Gini coefficient, Theil's measure. Poverty measurement-different issues, measures of incidence and intensity, combined measures, eg. Indices due to Kakwani, Sen. etc.

Books for Reference:

Choudhary, A and Mukherjee, R (1989): Randomized Response techniques, Marcel Decker.
Cochran, W. G. (1977): Sampling techniques, Third Edition, Wiley.
Des Raj and Chandok (1998): Sampling Theory, Narosa.
Murthy, M. N. (1977): Sampling Theory and Methods, Statistical Publishing Society, Calcutta.
Singh,D. and Choudhary, F. S. (1986): Theory and Analysis of Sample Survey Designs, New Age International.
Sukhatme et al. (1984): Sampling Theory of Surveys with Applications, Iowa State University Press.
C.S.O. (1980): National Accounts Statistics- Sources and Health.
Keyfitz, N. Mathematical Demography.
UNESCO: Principles of Vital Statistics Systems, Series M-12.
Sen, A(I997): Poverty and Inequality.

Second Semester:

Hard Core : Course 7: Probability Theory and Distributions – II (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This is the second course on the topic and includes advanced topics in Probability theory including integration with respect to a measure and convergence of random sequences. Order statistics and limit theorems are part of the course designed to expose students to advanced topics. This is a prerequisite for courses on Mathematical Statistics.

Course Outcomes : A person successfully completing the Course will acquire knowledge of many advanced topics in axiomatic Probability Theory and distributions and their properties. This advanced course is a prerequisite to many topics in Mathematical Statistics.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in the subject is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Lebesgue and Lebesgue Stieltjes measure on the real line. Integration of measurable functions with respect to measures. Monotone convergence theorem, Fatou's lemma and dominated convergence theorem.

Unit 2: Convergence in distribution, in Probability and with probability 1 and their implications. Slutsky's theorem. Weak law of large numbers- Ko1mogorov's generalized WLLN (proof of sufficient condition only), Khintchine's WLLN as special case, Chebyshev's WLLN.

Unit 3: Borel-Cantelli lemma, Kolmogorov's inequality. Strong law of large numbers - Kolmogorov's SLLN's for independent sequences and deduction for the ii.d. Case. Definitions and examples of Markov dependent, exchangeable and Stationary sequences. Characteristic function - properties, Inversion theorem (statement only and proof for density version), Uniqueness theorem, Continuity theorem (statement only). Central limit theorem- lindberg-Feller form (statement only). Deductions of Levy-Lindberg and Liapunov's forms

Unit 4: Order Statistics- their distributions and properties, Joint and marginal distributions. Extreme value distributions and their properties. Extreme value distributions as limit laws for the case of exponential, Normal, Uniform and Pareto.

Books for Reference:

Cramer, H. (1946): Mathematical Methods of Statistics, Princeton.
Johnson, S. and Kotz, (1972): Distributions in Statistics, Vols.I, II and III, Houghton and Miffin.
Mukhopadhyaya, P. (1996): Mathematical Statistics, Calcutta Publishing House.
Pitman, J. (1993): Probability ;Narosa.
Rao, C. R. (1973): Linear Statistical Inference and its Applications, 2 nd Ed., Wiley Eastern.
Rohatgi, V. K (1984): An Introduction to Probability Theory and Mathematical Statistics, Wiley Eastern.
Lukacs, C (1970): Characteristic functions, Griffin Publications.

Hard Core : Course 8: Inference – I (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This course on Mathematical Statistics introduces point and interval estimation, testing of statistical hypotheses and Bayesian inference. The course is designed as a prerequisite to advanced statistical topics like linear models, multivariate analysis and design and analysis of experiments. A good foundation of this course gives the ability to study further advanced topics in Statistics.

Course Outcomes : A person successfully completing the Course will acquire knowledge of many topics in basics of Mathematical Statistics which is a prerequisite to advanced topics in Mathematical Statistics.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in the subject is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Sufficiency, completeness, Uniformly minimum variance unbiased estimators, C-R inequalities, exponential class of densities and its properties, some special classes of distributions admitting complete sufficient statistics, extensions of these results to multi-parameter situation.

Unit 2: Test function, Neyman- Pearson lemma for test functions. Uniformly most powerful tests for one sided alternative for one parameter exponential class of densities and extension to the distributions having monotone likelihood ratio property.

Unit 3: Confidence Intervals, shortest expected length confidence intervals, relations with testing of hypotheses, uniformly most accurate confidence intervals.

Unit 4: Bayesian estimation, prior distributions, posterior distribution, loss function, principle of minimum expected posterior loss, quadratic and other common loss functions, conjugate prior distributions. Common examples. Bayesian HPD confidence intervals.

Books for Reference:

1. Kale, B.K. (2005). A First Course on Parametric Inference. Second Edition. Narosa.
2. Casella, G. and Berger, R. L. (2002). Statistical Inference. 2nd Edition, Duxbury Advanced series.
3. Dudewicz, E. J. and Mishra, S.N.(1988). Modern Mathematical Statistics, John Wiley.
4. Roussas, G. G. (1973). First Course in Mathematical Statistics, Addison Wesley.
5. Silvey, S. D. (1975). Statistical Inference, Chapman and Hall.
6. Wilks, S. S. (1962). Mathematical Statistics, John Wiley.
7. Lehmann, E. L. (1986). Testing of Statistical hypothesis, John Wiley.
8. Lehmann, E. L. (1988). Theory of Point Estimation, John Wiley.
9. Rohatgi, V. K. (1976). Introduction to theory of probability and Mathematical Statistics, John Wiley and Sons.
10. Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, Second Edition, Springer - Verlag.
11. Ferguson, T.S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press.

Hard Core : Course 9: Linear Models and Regression Analysis (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This is the first of two courses on the subject introducing students to the fundamentals of linear models and regression analysis in detail. A good understanding of the topics is necessary and useful in solving a wide variety of problems in Science, Technology, Social Science, Data Science and almost all forms of human endeavour.

Course Outcomes : A person successfully completing the Course will acquire a good foundation on linear models and regression analysis and can independently carry out statistical modelling of several types of data.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in the subject is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Gauss-Markov setup, normal equations and least squares estimates, error and estimation space, variance and covariance of least squares estimates, estimation of error variance, estimation with correlated observations, least squares estimates with restriction and parameters. simultaneous estimates of linear parametric functions.

Unit 2: Test of hypothesis for one and more than one parametric functions. Confidence intervals and regions, analysis of variance table, power of F-test, multiple comparison test like Tukey and Scheffe, simultaneous confidence interval.

Unit 3: One way linear models when parameters are random and estimation of variance components. Simple linear regression, multiple regression - estimation, testing linear hypotheses, confidence interval, confidence region, prediction of new observations, prediction interval, fit of polynomials and use of orthogonal polynomial Introduction to non-linear models.

Unit 4: Model adequacy - residuals and their plot for examining the departure from assumptions such as fitness of the model, normality, homogeneity of variances and detection of outliers and remedies. Hat-matrix, leverages and detection of influential observations.

Unit 5: Multicollinearity, ridge and principal component regression. Validation of regression model: analysis of estimated coefficients and predicted values, collecting fresh data, data splitting. Subset selection of explanatory variables and Mallow's C_p statistic, all possible regressions, stepwise, forward and backward regressions.

Books for Reference:

Cook, RD. and Weisberg, S. (1982); Residual 'and Influence in Regression. Chapman and Hall, London.
Draper, N.R and Smith, H (1998): Applied Regression Analysis. Third Edition, WiIey, New York.
Gunst, R.F. and Mason, R.L. (1980) ; Regression Analysis and its Application - A Data Oriented Approach, Marcel-Dekker.
Montgomery,D.C, Peck, E.A. and Vining,G.G.(2003). Introduction to Linear Regression, John Wiley.
Rao, C. R (1913): Linear statistical Inference, Wiley Eastern.
Ryan,T.P.(1997). *Modern Regression Methods*, John Wiley, NY
Searle,S.R.(1971). *Linear Models*, John Wiley, NY.
Seber, G.A.F.(1997): *Linear Regression Analysis*. John Wiley, NY.
Seber, G.A.F. and Lee (2003): *Linear Regression Analysis*. 2/e John Wiley, NY
Weisberg, S. (1985): Applied Linear Regression, Wiley.

Hard Core : Course 10: Practicals based on Papers VI – VIII (4 Credits – 8 hours of Practicals per week)

Learning Objectives : This is a Practical training course and exposes students to practical problems in the hard core courses on Inference-I, Linear Models and Regression Analysis and Probability Theory and Distributions – II. Problems are solved traditionally with calculators and also using R software. The course is mainly intended to train students in solving statistical problems in the three areas.

Course Outcomes : A person successfully completing the Course will acquire a good foundation on linear models and regression analysis in R software and can independently carry out statistical modelling of several types of data using R.

Pedagogy : The course is taught using R software and practical training in solving problems by fitting appropriate models for bivariate and multivariate data is given.

Assessment : The assessment is done through periodic tests involving R software wherein the students are asked to write R codes and run them in R software and also perform data analysis on data available online and from standard text books. The students are also assessed for problem solving through oral quizzes, assignments and student seminars.

Soft Core : Course 11: Project Work (4 Credits consisting of analysis of a Primary data and a Secondary data)

Learning Objectives : Under the supervision of a guide, students learn the methodology of collecting primary data and associated problems and learn to analyse data using statistical methods, both primary and secondary. The course results in a dissertation on the data analysed.

Course Outcomes : A person successfully completing the Course will acquire experience in carrying out statistical projects including design of studies and data analysis using R software and writing reports in LaTeX.

Pedagogy : The course is taught using R software and LaTeX typesetting software. Practical training in data analysis is given including designing a statistical study and writing reports.

Assessment : The assessment is done through seminars and a dissertation.

Third Semester:

Hard Core : Course 12: Inference – II (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This is the second course in statistical inference and includes advanced topics on this very important subject. Students learn to obtain distribution of estimators and test statistics fro large sample sizes. Students also learn to test advanced statistical hypotheses inlcuding nonparametric tests.

Course Outcomes : A person successfully completing the Course will acquire knowledge of many advanced topics in basics of Mathematical Statistics including tests of hypotheses and nonparametric tests.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in the subject is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Consistency and asymptotic normality (CAN) of real and vector parameters. Invariance of consistency under continuous transformation. Invariance of CAN estimators under differentiable transformations, generation of CAN estimators using central limit theorem.

Unit 2: Method of moments, method of maximum likelihood, Special cases such as exponential class of densities and multinomial distribution, Cramer-Huzurbazar theorem, method of scoring.

Unit 3: Tests based on MLEs. Likelihood ratio tests, asymptotic distribution of log likelihood ratio, Wald Test, Score Test, locally most powerful tests. Applications to categorical data analysis, three dimensional contingency tables, Pearson's chi-square test and LR test.

Unit 4: Asymptotic comparison of tests. Asymptotic Relative Efficiency (Pitman's). Introduction to Nonparametric Methods, one sample tests; Kolmogorov-Smirnov test, sign test, Wilcoxon –signed rank test. Wilcoxon rank sum test for two sample problem**.**

Books for Reference:

1. Casella, G. and Berger, R. L. (1990). *Statistical Inference*. Pacific Grove, CA: Wadsworth/Brooks Cole.
2. Cramer, H.(1974). Mathematical Methods in Statistics, Princeton Univ. Press.
3. Ferguson, T.S. (1996). A Course in Large Sample Theory, Chapman and Hall.
4. Gibbons,J.D., Chakraborti,S (2003). Nonparametric Statistical Inference, Fourth edition, CRC press.
5. Kale B.K. (2005). A First Course on Parametric Inference. Second Edition, Narosa.
6. Rao, C. R.(1995). Linear Statistical Inference and its Applications, Wiley Eastern Ltd.
7. Silvey, S. D.(1975). Statistical Inference, Chapman- Hall.

Hard Core : Course 13: Design and Analysis of Experiments (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This is a sequel to the course on linear models and exposes students to several standard designs including factorial experiments. Several advanced topics are introduced in this course and helps students to do advanced data analysis.

Course Outcomes : A person successfully completing the Course will acquire a good foundation on designing and analysing statistical experiments and can independently carry out advanced statistical modelling of several types of data using designs.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in the subject is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Introduction to designed experiments; General block design and its information matrix, criteria for connectedness, balance and orthogonality; Intrablock analysis (estimability, best point estimates/interval estimates of estimable linear parametric functions and testing of linear hypotheses) of BIBD- recovery of interblock information.

Unit 2: Youden design - intrablock analysis. Analysis of covariance in a general Gauss-Markov model, applications to standard designs. Missing plot technique - general theory and applications.

Unit 3: Fixed, mixed and random effects models; Variance components estimation - study of various methods; General factorial experiments, factorial effects: best estimates and testing the significance of factorial effects; study of 2^M and 3^M factorial experiments in randomized blocks.

Unit 4: Complete and partial confounding. Fractional replication for symmetric factorials. Response surface experiments.

Books for Reference:

Aloke Dey (1986): Theory of Block Designs, Wiley Eastern.
Angela Dean and Daniel Voss (1999): Design and Analysis of Experiments, Springer.
Chakrabarti,M.C.(1962): Mathematics of Design and Analysis of Experiments, Asia.
Cocharan and Cox,D.R.(1957): Experimental Designs, John Wiley.
Das, M.N. and Giri, N.(1979): Design and Analysis of Experiments, Wiley Eastern.
Giri, N.(1986): Analysis of Variance, South Asian Publishers.
John, P.W.M.(1911): Statistical Design and Analysis of Experiments, Macmillan.
Joshi, D.D.(1987): Linear Estimation and Design of Experiments, Wiley Eastern.
Kempthorne,O. (1952); Design Analysis of Experiments, Wiley Eastern.
Montgomery,C.D.(1976): Design and Analysis of Experiments, Wiley, New York.
Mukhopadhyay,P.(1998): Applied Statistics, Books and Allied (P) Ltd.
Myers, R. H.(1971): Response Surface Methodology, Allyn and Bacon.
Pearce,.S.C.(1984): Design of Experiments, Wiley, New York .
Rao,C.R..andKleffu, J.(1988): Estimation of Variance Components and Applications, North Holland.
Searle, S. R., Casella, G. and McCullugh, C. E. (1992): Variance Components, Wiley.

Hard Core : Course 14: Multivariate Analysis (4 Credits – 3 hours of Theory teaching and 2 hours of Practical per week)

Learning Objectives : This is a standard course on multivariate analysis analysing vector observations. The course will train students on several multivariate techniques useful to analyze multivariate data.

Course Outcomes : A person successfully completing the Course will acquire knowledge in analysing multivariate data and learn special techniques that are used to analyse multivariate data.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems in the subject is given as part of the Practical Course. Multivariate data analysis is carried out using R software. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Random sampling : multivariate normal distribution, maximum likelihood estimators of parameters, distribution of sample mean vector Wishart distribution (statement only) and its properties; distribution of sample generalized variance. Null distributions of sample correlation coefficient, partial. And multiple Correlation coefficients, distribution of sample regression coefficients. Application in testing and interval estimation.

Unit 2: Hotellings- T^2, Null distribution of Hotelling's T^2 statistic, Application in test on mean vectors for single and several multivariate normal populations.

Unit 3: Multivariate linear regression model, estimation of parameters, testing linear hypothesis about regression coefficients. Likelihood ratio test criterion. Multivariate analysis of variance of one and two way classified data.

Unit 4: Classification and discrimination procedures for discrimination into one of two multivariate normal populations. Sample discriminant function, tests associated ,with discriminant function, probabilities of misclassification and their estimation, classification into more than two mul1ivariate normal populations.

Unit 5: Principal components. Dimension reduction, canonical correlations and canonical variable - definition, use, estimation and computation. . Factor Analysis , Cluster analysis and Multi-domensional Scaling .

Books for Reference:

Anderson, T.W. (1983): An. Introduction to Multivariate Statistical Analysis, Second Edition, Wiley.
Giri, N. C. (1977): Multivariate Statistical Inference, Academic Press.
Johnson and Wichern (1986) : Applied Multivariate Analysis, Wiley.
Kshirsagar,. A.M. (1972): .Multivariate Analysis, Marcel-Dekker. .
Morrison, D.F. (1976): Multivariate Statistical Methods, Second Edition, McGraw Hill.
Muirhead, R.J. (1982): Aspects of Multivariate Stalistica1 Theory, Wiley.
Rao, C. R. (1973). Linear Statistical Inference and its Applications, Second Edition, Wiley Eastern.

Seber, G.A.F. (1984) : Multivariate Observations, Wiley.
Sharma, S. (1996). Applied Multivariate Techniques, Wiley.
Srivastava, M.S. and Khattree, C. G. (1979). An Introduction to Multivariate Statistics, North Holland.

Hard Core : Course 15: Practicals based on Papers XII – XIII (4 Credits – 8 hours of Theory teaching per week)

Learning Objectives : This is a Practical training course based on Inference – II and Design and Analysis of Experiments, in which practical problems on these topics are discussed and exposes students to practical problems in these two areas.

Course Outcomes : A person successfully completing the Course will acquire a practical knowledge of solving problems and data analysis in the subjects using R software and can independently carry out statistical modelling of several types of data using R.

Pedagogy : The course is taught using R software and practical training in solving problems in fitting appropriate models for bivariate and multivariate data.

Assessment : The assessment is done through periodic tests involving R software wherein the students are asked to write R codes and run them in R software and also perform data analysis on data available online and from standard text books. The students are also assessed for problem solving through oral quizzes, assignments and student seminars.

Soft Core : Course 16: Reliability Analysis (4 Credits – 3 hours of Theory teaching + 2 hours of Practical teaching per week)

Learning Objectives : As the name suggests, this course is an introduction to reliability analysis and several methods used to carry out reliability analysis are discussed in this course. Many of the topics overlap with methods in survival analysis. This course discusses statistics of reliability along with distributional aspects of reliability analysis.

Course Outcomes : A person successfully completing the Course will acquire knowledge in the theory of statistical reliability analysis along with learning special techniques to analyse positive valued data.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems on reliability data is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Life distributions : reliability function; hazard rate; common life distributions exponential, Weibull, Gamma, etc. Estimation of parameters and test in these models.

Unit 2: Notions of ageing : IFR, IFRA, NBU, DMRL, and NBUE classes and their duals; loss of memory property of the exponential distribution; closures of these classes under formation of coherent systems, convolutions and mixtures.

Unit 3: Univariate shock models and life distributions arising out of them; bivariate shock models; common bivariate exponential distributions and their properties.

Unit 4: Reliability estimation based on failure times in various censored life tests and tests with replacement of failed items; stress-strength reliability and its estimation.

Unit 5: Maintenance and replacement policies; availability of repairable systems; modelling of a repairable system by a non-homogeneous Poisson process. Reliability growth models; probability plotting techniques; Hollander-Proschan' and Deshpande tests for exponentiality; test for HPP vs. NHPP with repairable systems. Basic ideas of accelerated life testing.

Books for Reference:

Barlow, R.E. and Proschan, F. (1985). Statistical Theory of Reliability.
Lawless, J.R (1982). StaJistica1 Models and Methods of Life Time Data.
Bain, L.J. and Engelhardt (1991). Statistical Analysis of Reliability and Life Testing Data.
Zacks, S. Reliability Theory.

Soft Core : Course 17: Nonparametric and Semiparametric Methods (4 Credits – 3 hours of Theory + 2 hours of Practicals teaching per week)

Learning Objectives : This is a soft core course intended to expose students to nonparametric and semiparametric methods in Statistics. These methods are very powerful when parametric models are not appropriate. Along with standard material some advanced topics are also discussed in this course.

Course Outcomes : A person successfully completing the Course will acquire knowledge in using nonparametric and semiparametric methods to analyse data.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in solving problems using nonparametric and semiparametric methods is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Empirical distribution function, Glivenko-Cantelli theorem, Kolmogorov goodness of fit test. One sample U-statistics, Kernel and symmetric kernel, two sample U-statistics. asymptotic distribution of U-statistics. UMVUE property of U-statistics, asymptotic distribution of linear function of order statistics.

Unit 2: Rank tests, locally most powerful rank test, linear rank statistics and their distributional properties under null hypothesis, Pitman's asymptotic relative efficiency.

Unit 3: One sample location problem, sign test and signed rank test, two sample Kolmogorov Smirnov tests. Two sample location and scale problems. Wilcoxon-Mann-Whitney test, normal score test, ARE of various test based linear rank statistics. Kruskal-Wallis K sample test.

Unit 4: Cox's proportional hazards model, rank test (partial likelihood) for regression coefficients, concepts of jackknifing method of Quenouille for reducing bias, bootstrap methods, confidence intervals.

Books for Reference:

1. Cox, D.R. and Oakes, D. (1983). Survival Analysis, Chapman and Hall.
2. Davison, A.C. and Hinkley, D.V. (1991). Boo-:strap methods and their application, Cambridge University Press.
3. Fraser, D.A.S. (1957). Nonparametric me1hods in Statistics, John Wiley.
4. Gibbons, J.D. (1985). Nonparametric Statistical Inference. Second Edition, Marcel-Dekker.
5. Hajek, J. and Sidak, Z. (1961): Theory of Rank Tests, Academic Press.
6. Puri, M.L and Sen, P.K (1971). Nonparametric methods in multivariate analysis, Wiley.
7. Randles, R.H and Wolfe, D.A. (1979). Introduction to the theory of nonparametric statistics, Wiley.

Soft Core : Course 18: Operations Research (4 Credits – 3 hours of Theory + 2 hours of Practical teaching per week)

Learning Objectives : Operations research techniques are discussed in this course and these have wide applications. The course also considers advanced topics like queueing models and is designed to give a full view of most of the topics on this important subject.

Course Outcomes : A person successfully completing the Course will acquire knowledge in several techniques of problem solving using Operations research.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in applying Operations research in solving specialized problems is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Definition and scope of Operational research; phases in Operational research; models and their solutions; decision-making under uncertainty and risk, use of different criteria; sensitivity analysis.

Unit 2: LP Problem - Simplex method, Karmarkar's algorithm, duality theorem; transportation and assignment Problems; Wolfe's and Beale's algorithms for solving quadratic programming problems. Dynamic programming, Bellman's principle of optimality, general formulation, computational methods and applications of dynamic programming.

Unit 3: Analytical structure of inventory problems; EOQ formula of Ranis and its sensitivity analysis and extensions allowing quantity discounts and shortages. Multi-item inventory subject to constraints. Models with random demand, the static risk model. P and Q -systems with constant and random lead times, (t-S) po1icy.

Unit 4: Queueing .models - specifications; and effectiveness measures, steady-state solutions of M/M/I and M/M/C models with associated distributions of queue-length and waiting time. Queues with truncation Machine interference problem. M/G/I queue and Pollaczek-Khinchine result Steady - state solution of. M/E_k/I.

Unit 5: PERT and CPM Probability of project completion; PERT - crashing.

Books for Reference:

Churchman, C.W., Ackoff, R.L. and Amoff, E.L. (1957). Introduction to Operations Research.
Hillier, F.R and Leibennan, G (1962). Introduction to Operations Research
Kanti Swarup and Gupta, M.M. (1985); Operations Research.
Philips, D.T., Ravindran, A. and Solberg, J. Operations Research, Principles and Practice.
Taha, HA. (1982). Operational Research.

Soft Core : Course 19: Statistical Process Control and Total Quality Management (4 Credits – 3 hours of Theory + 2 hours of Practical teaching per week)

Learning Objectives : This is a soft core course intended to expose students to statistical process control and total quality management, two applied fields of statistics. The applied flavour of the course having practical component will help prospective students to apply the topics of this course in practical situations.

Course Outcomes : A person successfully completing the Course will acquire knowledge in applying techniques of statistical quality control and statistical process control to data on quality for statistical quality assurance and management.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in applying techniques in statistical quality control and statistical process control in solving specialized problems in statistical quality management is given as part of the Practical Course. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Concept of quality. Quality characteristics. Control charts for variables and attributes. OC and ARL of control charts. Moving average (MA) and EWMA and CUSUM charts. Process control with auto correlated observations. Multivariate control charts.

Unit 2: Capability indices Cp, Cpk, and Cpm; estimation. confidence intervals and test of hypotheses.

Unit 3: Acceptance sampling plans for attribute inspection Single, Double. Multiple and Sequential sampling plans and their properties; Plants for inspection by variable for one-sided and two-sided specifications; continuous sampling plans - chain sampling plans TQM

Unit 4: TQM the seven QC tools. Quality costs. Six-sigma programme. Quality Systems - ISO 9000 standards series. Quality circles.

Books for Reference:

Johnson, N.L. and Kotz,S. (1993): Process capability indices, Chapman and Hall.
Montgomery; D. C. (1985): Statistical PrOi:eSS Control,Wiley.
Shridhara Bhat K. (2002): Total Quality Management Himalaya Publishing House.

Fourth Semester:

Hard Core : Course 20: Stochastic Processes (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This hard core course is intended to expose students to a variety of stochastic processes. Since these are studied as part of a course in Probability, this course will help students understand some stochastic processes and techniques used in understanding properties of some stochastic processes.

Course Outcomes : A person successfully completing the Course will acquire fundamental and advanced knowledge in stochastic processes which should help apply these models to modelling random processes.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Introduction to stochastic processes (SP's); classification of SP's 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, Stationary distribution, classification of states; transient MC; Random walk., and gambler's ruin problem. Applications from social, biological and physical sciences.

Unit 2: Discrete state space continuous time MC: Kolmogorov- Feller differential equations; Poisson process, birth and death process; Applications to queues and storage problems. Wiener process as a limit of random walk first -passage time and other problems.

Unit 3: Renewal theory: Elementary renewal theorem and applications. Statement and uses of key renewal theorem; study of residual life time process. Stationary process: weakly stationary and strongly stationary. processes: Moving average and auto regressive processes.

Unit 4: Branching processes: Galton-Watson branching process, probability of ultimate extinction, distribution of population size. Martingale in discrete time, inequality, Convergence and smoothing properties. Statistical inference in MC and Markov processes.

Books for Refernece:

Adke S. R. and Manjunath, S. M. (1984): An Introduction to Finite, Markov Processes, Wiley Eastern.
Bhat,B. R (2000): Stochastic Models: Analysis and Applications, New Age International, India.
Cinlar, E. (1975): Introduction to Stochastic Processes, Prentice Hall.
Feller, W. (1968): Introduction to Probability and its Applications, Vol.1, Wiley Eastern.
Harris T.E (1963): The Theory of Branching Processes, Springer Verlag.
Hoel, P.G, Port S.C. and Stone, C.J. (1972): Introduction to Stochastic Processes, Houghton Miffin and Co.
Jagers, P. (1974): Branching Processes with Biological Applications, Wiley.
Karlin, S. and Taylor,H.M. (1975): A First course in Stochastic Processes, Vol.1,Academic Press.
Medhi, J. (1982): Stochastic Processes, Wiley Eastern.
Parzen, E. (1962): Stochastic Processes, Holden- Day.

Soft Core : Course 21: Probability Measures on Metric Spaces (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This is an advanced soft core course and is intended to expose students to advanced topics in Probability theory. This will help students pursue more abstract concepts in Probability.

Course Outcomes : A person successfully completing the Course will become aware of the abstract concepts of probability measures on metric spaces and this should be of help to do further research in the field and study further advanced concepts in probability.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Probability measures on metric spaces: Measures and integrals, tightness, (measure) determining class; Weak convergence of measures: properties of weak convergence and Portmanteau theorem, Convergence detennin1ng class.

Unit 2: Some special cases: Convergence of probability measures and convergence determining class in the Euclidean space, the circle, the space R^\\infty, the product spaces etc.

Unit 3: Random clement of a metric space and its distribution, convergence in distribution of a sequence of random elements - various equivalent criteria for convergence; convergence in probability; weak convergence and mappings.

Unit 4: Relative compactness and Prohorov's theorem. Weak convergence and tightness in C[0, l]; Wiener measure. Weak convergence and tightness in D[O, 1]; Donsker's theorem.

Books for Reference:

1. Billingsley; P. (1968): Convergence of Probability Measures; Wiley.
2. Parthasarathy, K. R (1967): Probability Measures on Metric Spaces, Academic Press.

Soft Core : Course 22: Time Series Analysis (4 Credits – 3 hours of Theory + 2 hours of Practical teaching per week)

Learning Objectives : This soft core course introduces students to the important applied field of time series. The practical training of the course prepares students to analyze various time series data arising in real life situations such as economic data and stock market data etc.

Course Outcomes : A person successfully completing the Course will be exposed to specialized techniques to analyse data on time series and the practical component aids in understanding fitting of suitable time series models to time series data.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in modelling time series data is given. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Time series as discrete parameter stochastic process, auto-covariance and auto-correlation functions and ,their properties.

Unit 2: Detailed study of the stationary processes: (i) moving average (MA), (ii) auto-regressive (AR), (iii) ARMA, and, (iv) AR integrated MA (ARIMA) models. Box-Jenkins models. Discussion (without proof) of estimation of mean, auto-covariance and auto-correlation functions under large sample theory.

Unit 3: Choice of AR and MA orders. Estimation of ARIMA model parameters. Forecasting. Residual analysis and diagnostic checking.

Unit 4: Spectral analysis of weakly stationary process, periodogram and correlogram analysis, computation based on Fourier transforms, Spectral decomposition of weakly AR process and representation as a one-sided MA process -necessary and sufficient conditions.

Unit 5: Implication of spectral decomposition in prediction problems. State space representation of time series. Kalman filter techniques.

Books for Reference:

1. Anderson. T.W. (1971). The Statistical Analysis of Time Series. Wiley.
2. Bloomfield, P. (2000). Fourier Analysis of Time Series: An Introduction. Second Edition, Wiley.
3. Box, G.E.P., Jenkins, G. W. and Reinsel, G.C. (1994). Time Series Analysis:Forecasting and Control Prentice Hall.
4. Box, G.E.P. and Jenkins, G.M (1976). Time Series Analysis - Forecasting and Control Holden-day, San Francisco.
5. Chatfield, C. Analysis of Time Series - Theory and Practice, Chapman and Hall.
6. Chow, C.G. (1985). Econometrics.. Mc Graw Hill.
7. Findley, D.F..ed., (1981). Applied Time Series Analysis II. Academic. Press.
8. Fuller, W.A. (1976). Introduction to Statistical Time series. Wiley.
9. Granger, C W.J. and Newbold (1984). Forecasting Econometric Time Series, Third Edition, Academic Press.
10. Granger, C.W.J. and Hatanka, M. (1964). Spectral Analysis of Economic Time Series, Princeton University Press.
11. Hannan. E.J. (1960). Time Series Analysis, Metheun, London.
12. Kendall, MG. (1974). Time Series, CnMles Griffin, London.
13. Kendall, MG> and Sroan, A. (1966). The Advanced Theory of Statistics, Vol. 3, Charles Griffin, London.
14. Koopmans, L.H. (1974). The Spectral Analysis of Time Series, Academic Press.
15. Montgomery, D.C. and Johnson, L.A (1977) Forecasting and Time Series Analysis, McGraw Hill.
16. Nelson, C.R (1973). Applied Time Series for managerial forecasting. Holden-day.
17. Priestly, MB. (1981). Spectral Analysis and Time Series. Griffin, London

Soft Core : Course 23: Biostatistics (4 Credits – 3 hours of Theory + 2 hours of Practicals teaching per week)

Learning Objectives : This is a soft core course intended to expose students to the applied field of biostatistics and discusses some topics in biostatistics. Clinical trials are introduced as well as clinical epidemiology. The objective is to expose students to the topics whereby interested can pursue these topics further.

Course Outcomes : A person successfully completing the Course will be exposed to a variety of methods used in biostatistics and the practical component helps in understanding and solving problems in biostatistics.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Practical training in problems in biostatistics is given. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Definition, agent, host and environment, mode of transmission; incubation period, spectrum of disease, herd immunity, classification of cause, of death, measures of mortality, studies of mortality. Measures of morbidity, I1lOIbidity surveys, issues and problems. Risk, cause and bias. Observational studies: retrospective, cross sectional and prospective studies.

Unit 2: Clinical trials: Methods of randomization, ethical issues, cross over trials. Sequential and group sequential trials. Interim analysis, multiple testing and stopping rules. Equivalence trials.

Unit 3: Clinical Epidemiology: Definition, reliability, validity, sensitivity, specificity, predictive values, likelihood ration test, selection and interpretation of diagnostic test. Deciding on the .best therapy.

Unit 4: ROC curves, multiple and parallel test. Screening for disease, critical appraisal, Meta analysis. Epidemiologic Models - Epidemometric studies- Deterministic epidemic models: Simple, General, Recurrent.

Unit 5: Time Series (Epidemic or others) Applications of Time series analysis in epidemiology - Simple descriptive techniques for detecting seasonal, Cyclical, secular and random variations

Books for Reference:

1. Lilienfeld, A.M. and Lilienfeld, D.C. Foundations of epidemiology, Second Edition, Oxford Univ. Press, New York, 1980.
2. Fletcher, R.H., Fletcher, S. W. and Wagner, E.H. Clinical Epidemiology - the essentials, II ed., 1982.
3. Harold A Hahn, Christopher T. Sempos. Statistical Methods in Epidemiology, Oxford Univ. Press, New York, 1989.
4. David G.Kleinbaum, Lawrence L.Kupper and Hall Morgenstem. Epidemiologic Research, Van Nostrand, USA, 1982.
5. Chatfield, C. The Analysis of Time Series - An Introduction. III Ed. Chapman and Hall, London, 1984.
6. Bailey N.T.J. The Mathematical Approach to Biology and Medicine. - Chapters 1, 2, and 9, John Wiley, 1967.

Soft Core : Course 24: Analysis of Categorical Data (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : Categorical data are plenty and require special methods for their analysis. This course discusses the analysis of categorical data and methods used to analyse categorical data. The methods are useful for applications and the student completing this course will have enough knowledge to apply these practically.

Course Outcomes : A person successfully completing the Course will be exposed to specialized statistical methods applicable for analysing categorical data.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Categorical or count data, contingency table investigations associated with contingency table hypotheses, of independence in two-and three-way tables; analysis by fitting marginals log linear representation, internal' constraints problem, analysis of information, fitting strategies, algorithms to calculate quadratic approximations,confidence intervals.

Unit 2: Application, general linear hypotheses; external constraints problem; general formulation of MDI analysis, - single and K-sample cases, minimum modified chi-square estimations.

Unit 3: No linear interaction; general formularize in three-way table 2x2x2 table 4x2x2 table as examples.

Unit 4: Discrete classification models, full multinomial model and its variations.

Unit 5: Models for multivariate dichotomous responses- Bahadur model, loglinear and logit methods, Martin-Bradley model, KronmaI- Ott Tarter model, procedure leased on distributional distance. Variable -selection problem -m based on difference in discriminate score, Raiffas methods, Lachin's procedure, Kullback's divergence ,statistics procedure, Mixture of variables - Krazanowski's model and its analysis

Books for Reference:

1. Gokha1e, D.V. and Kullback,S. The information in contingency Table,Marcel Dekker, 1979.
2. Goldstein, M. and Dillon, W.R. Discrete Discriminant Analysis John Wiley, New York 1978.

Soft Core : Course 25: Demography (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This is a soft core course intended to expose students to demographic methods which deal with study of data on human populations. Several methods and measures used in such studies are discussed and the aim is to expose students to these topics in detail.

Course Outcomes : A person successfully completing the Course will be exposed to specialized statistical methods used to analyse demographic data.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Sources of Demographic data, coverage and content errors. Use of balancing equation and

Chandrasekharan Deming formula, to check completeness of registration data. Use of Whipple's Myres Baachi's and UN indices.

Unit 2: Measures of fertility period and cohort measures. Use of birth order Statistics and child - Woman ratio. Brass technique to estimate current-fertility levels Estimation of TFR age patten! of fertility, differential fertility. Measures of reproduction and replacement.

Unit 3: Measures of mortality - standard death rates, neo-natal, perinatal death rates, maternal and infant mortality rates standardization of mortality rates. Construction of complete and abridged life tables. Expectation of life uses and application of life tables.

Unit 4: Rates and ratios of Migration Methods of estimating the volume of migration - Vital statistics method and survival ratio methods. Measurement of population growth. Population estimates and projections.

Unit 5: Stable and quasi stable population analysis. Statistical models for population growth. Mathematical models in mortality, fertility and human reproductive process.

Books for Reference:

1. Barclay. G. W. : Techniques of population analysis, John Wiley and Sons.
2. Spiegelman, H : Introduction to demography, Harvard University press.
3. Keyfitz, N. : Introduction to the mathematics of population; Addison -Wesley Publishing Co.,
4. Pollard, J.H. : Mathematical Models for the growth or human population, Cambridge University press.
5. Biswas, S.: Stochastic processes in Demography and applications, Wiley Eastern Ltd.
6. Chiang, C.L. : Introduction to Stochastic Processes m Biostatistics, John Wiley and Sons.
7. Ravikumar, R. Technical demography, Wiley Eastern Ltd.
8. Wolfenden, H.B. : Population Statistics and their compi1ation, American Actuarial Society.
United Nations Manuals: II, III, IV, VII.

Soft Core : Course 26: Survival Analysis (4 Credits – 3 hours of Theory + 2 hours of Practical teaching per week)

Learning Objectives : This is a soft core course dealing with methods of survival analysis. Survival analysis deal with life time data on human populations and the course gives a sound foundation to the various techniques applied in survival analysis with some advanced topics.

Course Outcomes : A person successfully completing the Course will be exposed to specialized statistical methods used to analyse life time data and to model life time data practically.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. The practical component is to train students in analysing real life time data in survival analysis. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Concepts of Time, Order and Random Censoring. Life distributions - Exponential Gamma, Weibull, Lognormal, Pareto, Linear Failure rate. Parametric inference Point estimation, Confidence Intervals, Scores, tests based on LR , MLE

Unit 2: Life tables, Failure rate, mean residual life and their elementary properties. Ageing classes - IFR, IFRA, NBU, NBUE, HNBUE and their duals, Bathtub Failure rate.

Unit 3: Estimation of survival function - Actuarial Estimator, Kaplan - Meier Estimator, Estimation under the assumption of IFR/DFR.

Unit 4: Tests of exponentiality against non-parametric classes - Total time on test, Deshpande test. Two sample problem - Gehan Test, Log rank test. Mantel - Haenszel Test, Tarone - Ware tests.

Unit 5: Semi-parametric regression for failure rate - Cox's proportional hazards model with one and several covariates.

Books for Reference:

1. Cox, D.R. and Oakes, D. (1984). Analysis of Survival Data, Chapman and Hall, New York.
2. Gross, A. J. and Clark, V. A. (1975). Survival Distributions: Reliability Applications in the Biomedical Sciences, John Wiley and Sons.
3. Elandt - Johnson, R.E., Johnson, N.L. (1980). Survival models and Data Analysis, John Wiley and Sons.
4. Miller, R.G. (1981). Survival Analysis, Wiley.
5. Zacks, S. Reliability.

Soft Core : Course 27: Computational Statistics (4 Credits – 2 hours of Theory + 4 hours of Practical teaching per week)

Learning Objectives : This is a soft core course on computational methods in Statistics using R software for the computational part. The course contains a variety of statistical methods as well as practical training material in computation. The course will be useful for aspiring data scientists and is intended to train students in advanced data analysis.

Course Outcomes : A person successfully completing the Course will be exposed to specialized computational statistical tools including software.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises and the practical component taught in R software is to train students in using the computational statistics to analyse statistical data. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic tests involving R software wherein the students are asked to write R codes and run them in R software and also perform data analysis on data available online and from standard text books. The students are also assessed for problem solving through oral quizzes, assignments and student seminars.

Unit 1: Statistical Simulations.

Unit 2: Stochastic simulations: generating random variables, simulating normal, gamma and beta random variables.

Unit 3: Comparison of algorithms to generate random variables. Generating random variables from failure rates. Simulating multivariate distributions, MCMC methods and Gibbs sampler, Simulating random fields, simulating stochastic processes.

Unit 4: Variance reduction techniques : importance sampling for integration, control variates and antithetic variables.

Unit 5: Simulating a non-homogeneous Poisson process. Optimization using Monte Carlo methods, simulated annealing for optimization. Solving differential equations by Monte Carlo methods.

Books for Reference:

1. Fishman, G.S. (1996) Monte Carlo: Concepts, Algorithms, and Applications. Springer.
2. Rubinstein, R.Y. (1981); Simulation and the Monte Carlo Method. Wiley.
3. Ripley B.D. (1987) Stochastic Simulations. Wiley.
4. Ross, S.M.(2002) Simulation, Third Edition, Academic.

Soft Core : Course 28: Project Work (4 credits: One Primary data analysis and One Secondary data analysis OR One theoretical paper writing project in lieu of the latter)

Learning Objectives : This is a soft core course dealing with data analysis and / or theoretical project work for students resulting in a dissertation. Apart from collecting data, analysing data and project report writing, the course trains students to think independently while performing statistical data analysis.

Course Outcomes : A person successfully completing the Course will be exposed to advanced data analysis and report writing.

Pedagogy : The course is taught to collect primary and secondary data and to study research articles as well as in analysing data using R software and writing reports using LaTeX.

Assessment : The assessment is done through oral quizzes, assignments and seminars.

Soft Core Elective for Non-Statistics students: Course 29: Statistical Methods and Applications (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This soft core course is intended to introduce basic Statistics to students who do not have any background in statistics. The course has less stress on Mathematical Statistics and is oriented to expose students to the basic concepts of Statistics. .

Course Outcomes : A person successfully completing the Course will be exposed to basic statistical methods used to analyse data and enough applications of such methods.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

DESCRIPTIVE STATISTICS:

Unit 1: Collection of data, Tabular and graphical representation of data. Attribute and Variable discrete and continuous. Analysis of data - Frequency distribution, histogram and Ogive.

Unit 2: Measures of location, dispersion/scale and skewness. Bivariate data - scatter diagram. Product moment correlation and linear regression, Spearman's rank correlation.

PROBABILITY THEORY:

Unit 3: Concepts of Probability, Standard Probability Distributions - Binomial Poisson, Geometric, Exponential normal t, F and Chi-square and properties.
SAMPLING:

Unit 4: Population and sample - simple random sampling. drawing random samples using random tables. Concepts of stratified random sampling. Standard error of sample mean and sample proportion.

STATISTICAL INFERENCE:

Unit 5: Testing for means, proportions and variance in one sample and two sample problems. Chi-square test for attributes. Analysis of variance - principles - one way and two way classification models

Books for Reference:

1. Goon A.M., Gupta, M.K. and Das Gupta B: Fundamental of Statistics Vol.1 and II. World Press Pvt. ltd., Calcutta.
2. Bhattacharya, G.K. and Johnson, R.A. Statistical concepts and methods. Wiley Eastern. Calcutta, Bombay and Delhi.
3. Levin, R.I. : Statistics for Management, Prentice Hall of India, New Delhi.
4. Hines, W. W. and Montgomery, D.C: Probability and Statistics -In Engineering and Management Science. Royal Press, New York.
5. Medhi J: Statistical Methods. Wiley Eastern Limited, New Delhi.

Soft Core Elective for Non-Statistics students: Course 30 : Probability Theory and Mathematical Statistics (4 Credits – 4 hours of Theory teaching per week)

Learning Objectives : This soft core course is intended to introduce basic Mathematical Statistics to students who do not study Statistics as part of their program. Some advanced topics are also part of the course and students finishing this course will have enough knowledge to explore further advanced topics in Statistics.

Course Outcomes : A person successfully completing the Course will be exposed to basics of probability theory and mathematical statistics.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 0: Review of set theory and calculus.

PROBABILITY THEORY:

Unit 1: Axiomatic Probability, Random Variables and distributions. Expectations and moments.

Unit 2: Standard discrete and continuous probability models. Sampling distributions. Modes of Convergence , WLLN, SLLN and CLT with applications (SLLN and CLT - statements only)

STATISTICAL METHODS:

Unit 3: Measures of central tendency and dispersion. Correlation and regression - curve fitting.

STATISTICAL INFERENCE:

Unit 4: Parameters and estimates. Methods of estimation - moment, maximum Likelihood Properties of estimators - Unbiasedness, MSE and consistency Interval. Estimation.

Unit 5: Testing of hypotheses - Basic concepts of testing, Neyman - Pearson lemma (Statement only) and examples Likelihood Ratio Tests - One sample and two sample problems.

Books for Reference:

1. Rohatgi,V.K.: An introduction to Probability Theory and mathematical Statistics, Wiley Eastern.
2. Hogg, R V. and Tanis, E.A.: Probability and Statistical Inference. McMillan, New York
3. Hogg, R. V. and Craig: Introduction to Mathematical Statistics, McMillan, New York.
4. Feller,W.: Introduction to Probability Theory. Vol.1, Wiley.
5. Freund, J.E: Modern Elementary Statistics, Prentice Hall of India, New Delhi.

Soft Core Elective for Non-Statistics students: Course 31: An Introduction to Statistics and R (4 Credits – 2 hours of Theory teaching + 4 hours of Practical teaching per week)

Learning Objectives : This soft core course is intended to introduce basic Statistics to students using R software and does not assume any prior knowledge of Statistics. After finishing this course, a student will have the double advantage to explore Statistics as well as R program further.

Course Outcomes : A person successfully completing the Course will be exposed to basics of statistics and R programming.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises and the practical component is taught using R software. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit-1: Basic Statistics: Types of data: Discrete and continuous data, Different types of scales, Primary data (designing a questionnaire and schedule), Secondary data (major sources including some government publications).
Diagrammatic and graphical representations of data, frequency and cumulative frequency distribution and their applications, Histogram, frequency polygon, ogives, stem and leaf chart, box plot.
Concept of central tendency, Partitioned values, Dispersion and relative dispersion, Sample moments, skewness, kurtosis and their measures.
Correlation and Linear Regression.
Unit-2: Basic concepts of Probability and inference: Basic probability theory – definitions and consequences. Bayes' theorem (statement only) and applications. Conditional probability.
Discrete and Continuous distributions- Bernoulli, Binomial, Geometric, Poisson, Uniform, Exponential, Normal distributions. Reading normal probability table.
Basics of Point and Interval Estimation and Testing of hypotheses- Estimation methods, tests for mean, proportion, goodness-of-fit, equality of means (one-way and two-way ANOVA), independence of attributes.
Unit-3 : Computational techniques using R-Package: Introduction and preliminaries: The R environment, R commands, Simple manipulations; numbers and vectors, Objects, their modes and attributes, Arrays and matrices, Lists and data frames, Reading data from files.Probability distributions: Computation of probabilities for various discrete and continuous distributions.

Unit-4: Simple Programming using R: Programming using R: Grouping, loops and conditional execution.
Graphics : Different graphic commands like plot etc.,
Statistical models in R: Defining statistical models; formulae, Linear models, Analysis of variance and model comparison, ANOVA tables.

Books for reference:

1. Medhi, J. (1992): Statistical Methods - an introductory text, New Age International Publishers.
2. Hogg, Robert V., Tanis, Elliot, A.Tanis and Rao, M.Jagan Mohan (2010): Probability and Statistical Inference, Pearson education.
3. Purohit, Sudha G., Gore, Sharad D. and Deshmukh, Shailaja R.(2008): Statistics using R. Narosa publications.
4. Crawley, Michael (2007): The R book, John Wiley publications.

Soft Core : Course 32: Data Science (4 Credits – LTP 202 - 2 hours of Theory + 4 hours of Practical teaching per week)

Learning Objectives : This soft core course is intended to introduce data science to students of the M.Sc. Statistics program. This recent course exposes students to the basic concepts of data science and prepares students to explore further topics in data science.

Course Outcomes : A person successfully completing the Course will be exposed to basics of data science including practical aspects of data science.

Pedagogy : The course is taught using traditional chalk-and-talk method using problem solving through examples and exercises and the practical component is taught using R software. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Introduction to data science-Current landscape of perspectives and needed skill set. Data Science and Ethical Issues- Discussions on privacy, security, ethics.
Introduction to big data, necessity of big data. Big data sources: hunting for the data, setting the goal, diving deeper into big data sources, getting started with big data acquisition. The nuts and bolts of big data: the storage dilemma, bringing structure to unstructured data, processing power, choosing among in-house, outsourced or hybrid approaches. Analysis of scalability of algorithms to big data, Data warehouses and online analytical processing, Efficient storage of big data including data streams.

Unit 2: Introduction to data science in R. Applied data science with R.
Machine learning algorithms, three basic algorithms- Linear regression, k-Nearest Neighbors (k-NN), k-means. Basic machine learning algorithms and solutions. Motivating application: Filtering Spam, Why Linear Regression and k-NN are poor choices for Filtering Spam, Naive Bayes and why it works for Filtering Spam, Data Wrangling: APIs and other tools for scrapping the Web.

Unit 3: Logistic Regression- Estimation of the parameters, Newton's Method, Stochastic Gradient Descent, Implementation, Evaluation, Logistic Regression case study.

Unit 4: Building a User-Facing Data Product, Algorithmic ingredients of a Recommendation Engine, Dimensionality Reduction, Singular Value Decomposition (SVD), Properties of SVD, Principal Component Analysis (PCA), Cluster Analysis. Basic principles, ideas and tools for data visualization. Examples of inspiring (industry) projects, Exercise: create your own visualization of a complex dataset.

Unit 5: Data mining: methods for learning descriptive and predictive models from data, Distributed algorithms over very large graphs and matrices, Social media analysis, Visualization methods and interactive data exploration.

References:

1. Alan Anderson and David Semmelroth (2015). Statistics for Big Data for Dummies, Wiley.
2. Aravind Sathi. (2012). Big Data Analytics: Disruptive Technologies for changing the Game, MC Press.

3. Deborah Nolan and Duncan Temple Lang (2015). Data Science in R- A case studies approach to computational reasoning and problem solving, CRC Press.
4. Kevin P. Murphy. (2013). Machine Learning-A Probabilistic Perspective. ISBN 0262018020.
5. Montgomery,D.C, Peck, E.A. and Vining,G.G.(2003). Introduction to Linear Regression, John Wiley.
6. Rachel Schutt and Cathy O'Neil. (2015). Doing Data Science-Straight Talk From The Frontline.
7. Rao, C. R . (1913). Linear statistical Inference, Wiley Eastern.
8. Searle, S. R .(1982). Matrix Algebra Useful for Statistics, Wiley.

Soft Core : Course 33: Data Science - Practical Multivariate Data Analysis using R (4 Credits – LTP 004 - 8 hours of Practical teaching per week)

Learning Objectives : This soft core course is intended to introduce practical data science to students of the M.Sc. Statistics program. This recent course exposes students to multivariate data analysis using R and prepares students to explore further multivariate topics in data science.

Course Outcomes : A person successfully completing the Course will be exposed to advanced multivariate techniques used in analysis of data.

Pedagogy : The course is taught entirely using Computers and R software through problem solving and analysing multivariate data. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic tests involving R software wherein the students are asked to write R codes and run them in R software and also perform data analysis on data available online and from standard text books. The students are also assessed for problem solving through oral quizzes, assignments and student seminars.

Unit 1: Introduction to the nature of the multivariate data. Matrix Algebra- Types of matrices, eigen values and eigen vectors, quadratic forms, vector and matrix differentiation, Lagrange multipliers. Basic multivariate statistics- population moments, multivariate distributions, statistical inference. Graphical representation of multivariate data.

Unit 2: Principal component analysis- Component scores, component correlations, Biplots. Correspondence analysis- Distance for contingency tables, multiple correspondence analysis. Cluster analysis- Hierarchical methods, K-means clustering, model based methods. Factor analysis- Principal components factor analysis, the factor analysis model, partial least squares regression.

Unit 3: Multidimensional scaling- classical scaling, nonmetric scaling, Procrustes analysis, individual differences scaling. Linear regression analysis- multiple linear regression, multivariate multiple linear regression. Multivariate analysis of variance- univariate one-factor model, multivariate one-factor model, profile analysis. Canonical correlation analysis, Discriminant analysis and canonical variates analysis-quadratic discrimination, non-parametric discrimination, support vector machines, CART. Loglinear modelling- loglinear models for Poisson means, contingency tables, higher dimensional tables.

Unit 4: Latent variable models- Latent class analysis, latent trait analysis, latent profile analysis.

Graphical modelling- conditional independence, graphs, fitting graph model to the data, categorical data. Data mining- KDD, Data mining, text analysis.

References:

1. Daniel Zelterman (2015). Applied Multivariate Statistics with R, Springer.
2. Shailaja R. Deshmukh and Sudha G. Purohit (2011). Microarray Data, Narosa.
3. Trevor F. Cox (2014). An Introduction to Multivariate Data Analysis, Wiley.

SYLLABUS FOR Ph.D. COURSE WORK
Course 1: Advanced Research Methodology

Duration : 55 hours

Learning Objectives : This course contains advanced topics from Probability Theory, Mathematical Statistics and Statistical Inference and is intended to expose every student pursuing Ph.D. in Statistics to some advanced topics which may help in the research work of the student. The units of the course consists of chapters from standard text and reference books in Probability and Statistics.

Course Outcomes : A person successfully completing the Course will be exposed to advanced concepts in probability theory and mathematical statistics useful for pursuing research in statistics.

Pedagogy : The course is taught using chalk-and-talk method and many topics are for self study by the students. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through periodic written tests involving problem solving, oral quizzes, assignments and student seminars.

Unit 1: Chapter 6 in Athreya and Lahiri (2006): Central Limit Theorems – Lindeberg-Feller theorems; Stable distributions; Infinitely divisible distributions, Refinements and extensions of the CLT – The Berry-Esseen theorem, Edgeworth expansions, Large deviations, The functional central limit theorem, Empirical process and Brownian bridge; Problems. Chapter 11 in Athreya and Lahiri (2006): Limit Theorems for Dependent Processes – Mixing sequences – Mixing coefficients, Coupling and covariance inequalities; Central limit theorems for mixing sequences; Problems. (12 hours)

Unit 2: Chapter 1 in Joe (1997): Conditions for multivariate distribution functions – Properties of a bivariate cdf F, Properties of a multivariate cdf F; Types of dependence; Copulas. Chapter 2 in Joe (1997): Basic Concepts of Dependence – Dependence properties and measures – Positive quadrant and orthant dependence, Stochastic increasing positive dependence, Right-tail increasing and left-tail decreasing, Associated random variables, Total positivity of order 2, Positive function dependence, Relationships among dependence properties, Max-infinite and min-infinite divisibility, Kendall's tau and Spearman's rho, Tail dependence, Examples; Dependence orderings – Concordance ordering, Axioms for a bivariate dependence ordering, Axioms for a multivariate dependence ordering, More SI bivariate orderings, More TP_2 bivariate orderings, Positive function dependence ordering, Examples: bivariate, Examples: multivariate; Exercises; Unsolved problems. (12 hours)

Unit 3: Chapter 10: Brownian Motion and Stationary Processes in Ross (2000): Brownian Motion; Hitting Times, Maximum Variable, and the Gambler's Ruin Problem; Variations on Brownian Motion – Brownian Motion with Drift, Geometric Brownian Motion; Pricing Stock Options – An Example in Options Pricing, The Arbitrage Theorem, The Black-Scholes Option Pricing Formula; White Noise; Gaussian Processes; Stationary and Weakly Stationary Processes; Harmonic Analysis of Weakly Stationary Processes; Exercises. (08 hours)

Unit 4: Chapter 3 Randles and Wolfe (1979): U-Statistics – One-sample U-Statistics; Some Convergence Results; The Projection Principle and the One-Sample U-Statistic Theorem; Two-Sample U-Statistics. Chapter 5 in Randles and Wolfe (1979): Asymptotic Relative Efficiency of Tests – Pitman Asymptotic Relative Efficiency; Methods for Evaluating ARE(S,T) – Noether's Theorem; Extended U-Statistics Theorems; Examples of Pitman's ARE for Translation Alternatives; Discussion – Deficiency and Bahadur Efficiency. Chapter 8 in Randles and Wolfe (1979): Linear Rank Statistics under the Null Hypothesis – Linear Rank Statistics; Distributional Properties; Some Preliminaries for Asymptotics; Asymptotic Normality under H_0. Chapter 9 in Randles and Wolfe (1979): Two-Sample Location and Scale Problems – The Two-Sample Location Problem; Asymptotic Properties in the Location Problem; The Two-Sample Scale Problem. (15 hours)

Unit 5: Chapter 12 in Athreya and Lahiri (2006): The Bootstrap – The bootstrap method for independent variables – A description of the bootstrap method, Validity of the bootstrap: Sample mean, Second order coorectness of the bootstrap, Bootstrap for lattice distributions, Bootstrap for heavy tailed random variables. Chapter 11 in Ross (2000): Simulation – General Techniques for Simulating Continuous Random Variables – The Inverse Transformation Method, The Rejection Method, The Hazard Rate Method; Special Techniques for Simulating Continuous Random Variables – The Normal Distribution, The Gamma Distribution, The Chi-Squared Distribution, The Beta (n, m) Distribution, The Exponential Distribution – The von Neumann Algorithm; Simulating from Discrete Distributions – The Alias Method; Stochastic Processes – Simulating a Nonhomogeneous Poisson Process, Simulating a Two-Dimensional Poisson Process; Variance Reduction Techniques – Use of Antithetic Variables, Variance Reduction by Conditioning, Control Variates, Importance Sampling; Determining the Number of Runs; Exercises. R software and LaTeX typesetting program. (08 hours)

References:

1. Athreya, K.B. and Lahiri, S.N. (2006). Probability Theory, TRIM 41, Hindustan Book Agency.
2. Joe, H. (1997). Multivariate Models and Dependence Concepts. Chapman and Hall.
3. Randles, R.H. And Wolfe, D.A. (1979). Introduction to the Theory of Nonparametric Statistics, John Wiley and Sons.
4. Ross, S.M. (2000). Introduction to Probability Models, Sixth Edition, Harcourt Asia Pvt. Ltd., Academic Press.

SYLLABUS FOR Ph.D. COURSE WORK
Course 2: Discipline centric Survey of Literature

Duration : 55 hours

Learning Objectives : This course shall contain literature survey on the topic in which a Ph.D. student will work for his / her thesis and shall be decided by the supervisor of the Ph.D. student. The course shall contain topics relevant to the work of the student and shall be evaluated based on seminars and dissertation containing the literature survey which is intended to result in the Ph.D. synopsis of the student.

Course Outcomes : A person successfully completing the Course will be exposed to advanced concepts in the area in which the student wishes to pursue research for Ph.D.

Pedagogy : The course is through self study and discussions with the supervisor. Students are encouraged to use text and video resources available on the web.

Assessment : The assessment is done through seminars and a dissertation.