FAROOK COLLEGE (AUTONOMOUS)

Farook College PO, Kozhikode-673632

P.G Programme in Statistics

Under

Choice Based Credit Semester System

SYLLABUS

(2022 Admission Onwards)

Prepared By:

Board of Studies in Statistics

Farook College (Autonomous)

CERTIFICATE

I hereby certify that the documents attached are the bona fide copies of the syllabus of M.Sc. Statistics programme to be effective from 2022 admission onwards.

Date: Principal

Place: Farook College

CONTENTS

M. Sc. Statistics Programme under
Choice Based Credit Semester System
FCCBCSS PG-2022
Programme Structure & Syllabi
(With effect from the academic year 2022-2023 onwards)

Programme Duration: **Two years,** divided into **four semesters** of not less than 90 working days each.

Course Code	Type	Course Title	Credits	Class hours	Ratio Internal: External
		I SEMESTER (Total Credits: 20)			
MST1C01	Core	Analytical Tools for Statistics – I	4	5	1:4
MST1C02	Core	Analytical Tools for Statistics – II	4	5	1:4
MST1C03	Core	Probability Theory-I	4	5	1:4
MST1C04	Core	Distribution Theory	4	5	1:4
MST1C05	Core	Sampling Theory	4	5	1:4
MST1A01	Audit	Ability Enhancement Course	4 Credits (Not included in CGPA)		
		II SEMESTER (Total Credits: 20)			
MST2C06	Core	Probability Theory-II	4	5	1:4
MST2C07	Core	Applied Regression Analysis	4	5	1:4
MST2C08	Core	Estimation Theory	4	5	1:4
MST2C09	Core	Stochastic Processes	4	5	1:4
MST2L01	Core	Statistical Computing-I	4	5	1:4
MST2A02	Audit	Professional Competency Course	4 Credits (Not included in CGPA)		
		III SEMESTER (Total Credits:20)			
MST3C10	Core	Time Series Analysis	4	5	1:4
MST3C11	Core	Design and Analysis of Experiments	4	5	1:4
MST3C12	Core	Testing of Statistical Hypotheses	4	5	1:4
MST3E---	Elective	Elective-I	4	5	1:4
MST3E---	Elective	Elective-II	4	5	1:4
		IV SEMESTER (Total Credits: 20)			
MST4C13	Core	Multivariate Analysis	4	5	1:4
MST4E---	Elective	Elective-III	4	5	1:4
MST4L02	Core	Statistical Computing-II	4	5	1:4
MST4P01	Core	Project/Dissertation	5	10	1:4
MST4V01	Core	Comprehensive Viva-Voce	3	5	1:4
		Total	**80**		----

Total credits: 80 (Core -60, Elective-12, Project & Comprehensive Viva-8)

The courses Elective –I, Elective –II and Elective –III shall be chosen from the following list.

Course Code	Course Title	Credits
01	Operations Research-I	4
02	Operations Research – II	4
03	Queueing Theory	4
04	Lifetime Data Analysis	4
05	Advanced Distribution Theory	4
06	Statistical Decision Theory	4
07	Reliability Modelling	4
08	Actuarial Statistics	4
09	Statistical Quality Control	4
10	Advanced Probability Theory	4
11	Official Statistics	4
12	Biostatistics	4
13	Econometric Models	4
14	Demographic Techniques	4
15	Stochastic Finance	4
16	Longitudinal Data Analysis	4
17	Data Mining Techniques	4
18	Statistical Machine Learning –I	4
19	Statistical Machine Learning –II	4
20	Advanced Statistical Machine Learning Techniques	4
21	Non-Parametric Statistical Methods	4
22	Statistical Modelling and Data Mining Techniques	4
23	Applied Algorithms and Analysis of Multi type and Big Data	4
24	Risk Modelling and Survival Analysis-I	4
25	Risk Modelling and Survival Analysis-II	4

EVALUATION AND GRADING:

Evaluation: The evaluation scheme for each course shall contain two parts; (a) Internal/ Continuous Assessment (CA) and (b) External / End Semester Evaluation (ESE). Of the total, 20% weightage shall be given to internal evaluation / continuous assessment and the remaining 80% to external/ESE and the ratio and weightage between Internal and External is **1:4.** Primary evaluation for Internal and External shall be based on 6 letter grades **(A+, A, B, C, D and E)** with numerical values (Grade Points) of **5, 4, 3, 2, 1 & 0** respectively.
The criteria and percentage of weightage assigned to various components for evaluation are as follows:

(A)Theory and Practical:

<u>Internal Evaluation</u>

(a) Theory :			
Sl.No	**Component**	**Percentage**	**Weightage**
1	Examination /Test	40%	2
2	Seminars / Presentation	20%	1
3	Assignment	20%	1
4	Attendance	20%	1
(b) Practical :			
1	Lab Skill	40%	4
2	Records/viva	30%	3
3	Practical Test	30%	3

<u>External Evaluation</u>

a) The semester-end examinations in theory courses shall be conducted with question papers set by external experts. The duration shall be **3 hours** and the total weightage should be **30**.

The question paper pattern is as follows:

b) Theory:				
Sl. No.	**Type of Questions**	**Individual weightage**	**Total Weightage**	**Number of questions to be answered**
1	Short Answer type questions	2	2 x 4 = 8	4 out of 7
2	Short essay/ problem solving type	3	3 x 4 = 12	4 out of 7
3	Long Essay type questions	5	5 x 2 = 10	2 out of 4
Total			**30**	**10/18**

c) **Practical:** The end semester evaluation in practical course shall be conducted by both internal and external examiners as per the stipulations in the syllabus. The duration shall be **3 hours** and the total weightage should be **30**.

(B) Project work/Dissertation and External Viva-Voce:

Project work				
Sl. No	**Criteria**	**% of weightage**	**Weightage External**	**Weightage Internal**
1	Review of literature, formulation of the problem and defining clearly the objective:	10%	4	1
2	Methodology and description of the techniques used	10%	4	1
3	Analysis, programming/simulation and discussion of results	20%	8	2
4	Presentation of the report, organization, linguistic style, reference etc	20%	8	2
5	Viva-voce examinations based on project/dissertation	40%	16	4
	Total Weightage	**100 %**	**40**	**10**

There shall be a comprehensive Viva Voce examination based on all courses of the programme with **3 credits,** internal and external being in the ratio 1:4. The Viva-Voce shall be conducted by a board of examiners consisting of at least one external expert and internal examiners.

Grading: Direct Grading System based on a 10 – Point scale is used to evaluate the performance (External and Internal Examination of students)

For all courses (Theory & Practical)/Semester/Overall Programme, Letter grades and **GPA/SGPA/CGPA** are given on the following way:

a) First Stage Evaluation for both Internal and External done by the Teachers concerned in the following Scale:

Grade	Grade Points
A+	5
A	4
B	3
C	2
D	1
E	0

The Grade Range for both Internal & External shall be:

Letter Grade	Grade Range	Range of Percentage (%)	Merit / Indicator
O	4.25 – 5.00	85.00 – 100.00	Outstanding
A+	3.75 – 4.24	75.00 – 84.99	Excellent
A	3.25 – 3.74	65.00 – 74.99	Very Good
B+	2.75 – 3.24	55.00 – 64.99	Good
B	2.50 – 2.74	50.00 – 54.99	Above Average
C	2.25 – 2.49	45.00 – 49.99	Average
P	2.00 -2.24	40.00 – 44.99	Pass
F	< 2.00	Below 40	Fail
I	0	-	Incomplete
Ab	0	-	Absent

No separate minimum is required for internal evaluation for a pass, but a minimum **P** Grade is required for a pass in the external evaluation. However, a minimum **P grade** is required for pass in a course.

PROGRAMME SPECIFIC OUTCOMES

PSO1: Learn how to solve practical statistical problems and develop personal skills including presentation and time management.

PSO2: Master the theories behind a variety of statistical techniques, and how to apply them in scenarios that professional statisticians face every day.

PSO3: Develop a detailed working knowledge of collection of data, important statistical techniques and concepts, including statistical modelling, inference and time series.

PSO4: Learn how to analyse and draw meaningful conclusions from data, and develop programming skills using the statistical computing software R.

SYLLABI OF CORE COURSES

SEMESTER- I

MST1C01: ANALYTICAL TOOLS FOR STATISTICS – I (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: State the fundamental theorems in Riemann Stieltjes integral.
CO 2: Determine the integrals using Riemann Stieltjes method.
CO 3: Distinguish between point wise and uniform convergence of sequences and series.
CO 4: Illustrate the consequence of uniform convergence on continuity, integrability and differentiability.
CO 5: Develop skills in generalizing the concepts in univariate calculus to multivariate setup.
CO 6: Calculate Laplace transforms and inverses.
CO 7: Apply Laplace transforms to solution of differential equations.

Unit-I: Riemann Stieltjes integral: Riemann-Stieltjes Integral:- Definition, linear properties, integration by parts, change of variable in a Riemann-Stieltjes integral, reduction to a Riemann integral, step functions as integrators, reduction of a Riemann-Stieltjes integral to a finite sum, Euler's summation formula, monotonically increasing integrators, Riemann's condition, comparison theorems, Functions of bounded variation (Definition and properties without proof), sufficient conditions for existence of Riemann-Stieltjes integrals, necessary conditions for existence of Riemann-Stieltjes integrals, Mean value theorems for Riemann-Stieltjes integrals.
25 hours

Unit-II: Convergence of sequences and series: Point wise convergence of sequence of functions, Examples of sequences of real valued functions, Definition of Uniform convergence, Uniform convergence and continuity, The Cauchy condition for uniform convergence, Uniform convergence of infinite series of functions, Uniform convergence and Riemann-Stieltjes Integration, Uniform convergence and differentiation. 25 hours

Unit-III: Multivariable functions: Limits and continuity of multivariable functions. Derivatives, directional derivatives and continuity. Total derivative in terms of partial derivatives, Taylor's theorem. Inverse and implicit functions. Optima of multivariable functions. Method of Lagrangian multipliers, Riemann integral of a multivariable function (Definition and related problems). 25 hours

Unit-IV: Laplace transform: Definition, Laplace transforms of some elementary functions, Some important properties of Laplace transforms: Linearity property, First translation or shifting property, Second translation or shifting property, Change of scale property, Laplace transform of derivatives, Laplace transform of integrals, Laplace transforms of functions when they are multiplied by t^n. Inverse Laplace transform: Definition and problems, Application of Laplace transforms in the solution of ordinary differential equations with constant coefficients and variable coefficients. 15 hours

Text Books

1. **Andre's I. Khuri (1993).** Advanced Calculus with applications in statistics. Wiley & Sons.
2. **Murray R Spiegel (1965).** Schaum's Outline of Theory and Problems of Laplace ransforms, Schaum's Outline series, McGraw- Hill, New York.
3. **Apostol, T.M. (2002).** Mathematical Analysis, Narosa Publshing House, New Delhi, Second Edition.

References

1. **Walter Rudin (1976)**. Principles of Mathematical Analysis. McGraw Hill.
2. **Malik, S.C and Arora. S (2006).** Mathematical Analysis, second edition, New age international
3. **Kasana H S (2001).** Complex Variables: Theory and Applications. Prentice-Hall of India.
4. **Churchill Ruel.V. (1975).** Complex variables and applications .McGraw Hill.
5. **Ajit kumar and S. Kumaresan (2014).** A basic course in real analysis. CRC Press, A Chapman and Hall Book.

MST1C02: ANALYTICAL TOOLS FOR STATISTICS – II (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Illustrate vector space, subspaces, independence of vectors, basis and dimension, direct sum, complement and orthogonality with examples.
CO 2: Examine linear independence and to construct orthogonal and orthonormal vectors.
CO 3: Acquire the concept of linear transformation.
CO 4: Determine eigen values and eigen vectors of a given matrix.
CO 5: Execute the decomposition of a matrix.
CO 6: Develop skills in finding solution of homogeneous equations and their applications in real life situations .
CO 7: Acquire the concept of g-inverse.
CO 8: Classify the quadratic forms.

Unit-I: Basics of Linear Algebra-Definition of vector space, Subspaces, Linear dependence and independence, Basis and dimensions, Direct sum and complement of a subspace, Quotient space, Inner product and orthogonality.

15 hours

Unit-II: Algebra of Matrices-Linear transformations and matrices, Operations on matrices, Properties of matrix operations, Matrices with special structures-Triangular matrix, Idempotent matrix, Nilpotent matrix, Symmetric, Hermitian and Skew- hermitian matrices, Unitary matrix. Row and column space of a matrix, Inverse of a matrix, Rank of real and complex matrices, Rank- nullity theorem , Rank of product of matrices, Rank factorization of a matrix, Rank of a sum and projectors, Inverse of a partitioned matrix.

25 hours

Unit-III: Eigen Values, Spectral Representation and Singular value Decomposition-Eigen values, eigen vectors and eigen spaces, Cayley-Hamilton theorem, Minimal polynomial, Spectral representation of a semi simple matrix, Algebraic and geometric multiplicities, Jordan canonical form, Spectral representation of a real symmetric matrix, Concepts of Hermitian and normal matrices, Singular value decomposition.

25 hours

Unit- IV: **Linear Equations, Generalized Inverses and Quadratic forms**-Homogenous system, General linear system, Generalized inverse, Properties of g-inverse, Moore-Penrose inverse, Computation of g-inverse, Definition of quadratic forms, Classification of quadratic forms, Rank and signature, Extrema of quadratic forms, Simultaneous diagonalization of matrices.

25 hours

Text Books

1. **Rao A.R. & Bhimasankaram P. (1992)**.Linear Algebra. Hindustan Book Agency. New Delhi.
2. **Lewis, D.W. (1995).** Matrix Theory. Allied publishers, Bangalore.
3. **Rudin W. (1976).** Principles of Mathematical Analysis- Third Edition. McGraw Hill, New York

References

1. **Biswas S. (1997).** A text book of Linear Algebra. New Age International, New Delhi.
2. **Rao C.R. (2002).** Linear Statistical Inference and Its Applications- Second Edition. John Wiley & Sons, New York.
3. **Graybill F.A. (1983).** Matrices with Applications in Statistics. Wadsworth Publishing Company, Belmont, California.

MST1C03: PROBABILITY THEORY - I (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Explain probability space.
CO 2: Describe the basic concepts of Random variable from measure point of view.
CO 3: Explain the concept of distribution function.
CO 4: Discuss Product Spaces and Fubini's theorem.
CO 5: Distinguish between different types of convergences.

UNIT- I: Probability : Sets and classes of events, Sequences of sets and their limits, Field, Sigma field, Borel field, Measurable functions, Random variables, Probability space, Induced sigma field, Induced Probability Space, Limits of sequences of random variables, Generalized Probability Measure, Conditional Probability Measure, Counting Measure, Lebesgue Measure, Lebesgue-Stieltjes Measure, Signed Measure.

25 hours

UNIT- II: Distribution function: Distribution function of a random variable, Decomposition of distribution functions, Distribution function of random vectors, Correspondence theorem.

15 hours

UNIT-III: Integration with respect to measure (Introduction only), Expectation and moments: Definition and properties, Moment generating functions, Moment inequalities: Cr-, Holder's, Jenson's and basic inequalities, Product spaces and Fubini's theorem (idea and statement only), Independence of two events, Independence of classes, Independence of random variables, Borel 0-1 law, Kolmogorov 0-1 law.

25 hours

UNIT- IV: Convergence: Modes of convergence, Convergence in probability, in distribution, in rth mean, almost sure convergence and their inter-relationships, weak convergence of distribution functions, Slutsky's theorem, Helly Bray theorems.

25 hours

Text Books

1. **Bhat, B.R. (2011).** Modern Probability Theory, Second edition, Wiley Eastern.
2. **Billingsley, P. (1986).** Probability and Measure, Second Edition, John Wiley.
3. **Laha & Rohatgi (1979).** Probability theory, Wiley New York.

References

1. **Feller, W. (1966).** An Introduction to Probability Theory and Its Applications, Volume II, Wiley Eastern.
2. **Rao, C.R. (1973)**. Linear Statistical Inference and Its Applications, Wiley.
3. **Rohatgi, V.K. and A.K.E. Salah (2001).** Introduction to Probability and Statistics, John Wiley and Sons.
4. **Basu, A.K. (1999)**. Measure Theory and Probability, Prentice-Hall.

MST1C04: DISTRIBUTION THEORY (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe different types of discrete probability distributions.
CO 2: Explain the properties and applications of continuous distributions.
CO 3: Derive probability distributions of the different functions of discrete and continuous random variables.
CO 4: Acquire the concept of order statistics and derive their distributions.
CO 5: Describe different Sampling distributions and their interrelations
CO 6: Illustrate real data modeling using probability distributions.

Unit-1: **Discrete distributions:** Random variables, Moments and Moment generating functions, Probability generating functions, Discrete uniform, Binomial, Poisson, Geometric, Negative binomial, Hyper geometric and Multinomial distributions, Power series distributions. 20 hours

Unit-2: **Continuous distributions**: Uniform, Normal, Exponential, Weibull, Pareto, Beta, Gamma, Laplace, Cauchy and Log-normal distributions, Pearsonian system of distributions, location and scale families.

20 hours

Unit-3: Functions of random variables: Joint and marginal distributions, Conditional distributions and independence, Bivariate transformations, Covariance and Correlations, Bivariate normal distributions, Multivariate Normal distribution: Definitions, Mean, Variance, Characteristic function. Mixture distributions(Introduction only), Cauchy-Schwarz and Jenson Inequalities, Order statistics.

20 hours

Unit-4: **Sampling distributions:** Basic concept of random sampling, Sampling from normal distributions, Properties of sample mean and variance. Chi-square distribution and its applications, t-distribution and its applications. F-distribution- properties and applications. Non-central Chi-square, t, and F- distributions. 30 hours

Text Books

1. **Rohatgi, V.K. (1976).** Introduction to probability theory and mathematical statistics. John Wiley and sons.
2. **Alexnder Mood, Graybill and Bose (1973)**. Introduction to the Theory of Statistics- McGraw Hill
3. **Parimal Mukhopadhyay (2018).** Mathematical Statistics, Book and Allied Publishers,(Ltd.), Calcutta.

References

1. **Johnson ,N.L.,Kotz.S. and Balakrishnan, N.(1995).** Continuous univariate distributions, Vol.I&Vol.II, John Wiley and Sons, New York.
2. **Johnson ,N.L.,Kotz.S. and Kemp.A.W.(1992).**Univariate Discrete distributions, John Wiley and Sons, New York
3. **Kendall, M. and Stuart, A. (1977).** The Advanced Theory of Statistics Vol I: Distribution Theory, 4th Edition.

MST1C05: SAMPLING THEORY (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Distinguish between Probability and Non-Probability Sampling.
CO 2: Apply the sampling methods: simple random sampling, systematic sampling, stratified sampling and cluster sampling.
CO 3: Estimate the population parameters for variables and attributes under the above procedures.
CO 4: Estimate the population parameters concerning the study variables under auxiliary information (Ratio and regression methods).
CO 5: Discuss probability proportional to size (PPS) sampling strategies.
CO 6: Explain the concepts of ordered and unordered estimators and its properties.
CO 7: Discuss the multi stage and multiphase sampling.
CO 8: Describe non-sampling errors.

Unit-I: Census and Sampling-Basic concepts, probability sampling and non-probability sampling, simple random sampling with and without replacement- estimation of population mean and total-estimation of sample size- estimation of proportions. Systematic sampling- linear and circular systematic sampling-estimation of mean and its variance- estimation of mean in populations with linear and periodic trends. 25 hours

Unit-II: Stratification and stratified random sampling. Optimum allocations, comparisons of variance under various allocations. Auxiliary variable techniques. Ratio method of estimation-estimation of ratio, mean and total. Bias and relative bias of ratio estimator. Mean square error of ratio estimator. Unbiased ratio type estimator. Regression methods of estimation. Comparison of ratio and regression estimators with simple mean per unit method. Ratio and regression method of estimation in stratified population.

25 hours

Unit-III: Varying probability sampling-pps sampling with and without replacements. Des- Raj ordered estimators, Murthy's unordered estimator, Horvitz-Thompson estimators, Yates and Grundy forms of variance and its estimators, Zen-Midzuno scheme of sampling, πPS sampling.

20 hours

Unit-IV: Cluster sampling with equal and unequal clusters. Estimation of mean and variance, relative efficiency, optimum cluster size, varying probability cluster sampling. Multi stage and multiphase sampling. Non-sampling errors.

20 hours

Text Books

1. **Cochran W.G. (1992):** Sampling Techniques, Wiley Eastern, New York.
2. **D. Singh and F.S. Chowdhary (1986):** Theory and Analysis of Sample Survey Design**,** Wiley Eastern (New Age International), New Delhi.

References

1. **P.V.Sukhatme**et.al. (1984): Sampling Theory of Surveys with Applications. IOWA State University Press, USA.
2. **Des Raj (1976):** Sampling Theory. McGraw Hill
3. **Mukhopadhyay. P. (1999).** Theory and Methods of Survey Sampling. Prentice-Hall India, New-Delhi.

SEMESTER II

MST2C06: PROBABILITY THEORY – II (Credits: 4, Hours per week: 5)

Course Outcomes: After successful completion of the course, students will be able to:

CO 1: Apply inversion formula.
CO 2: Acquire knowledge in some of the very important theorems like WLLN, CLT and their applications.
CO 3: Identify infinitely divisible distributions.
CO 4: Explain Radon Nikodym Theorem.
CO 5: Know the concept of Martingales

.

UNIT-I: Characteristic functions: Definition and simple properties, Inversion Theorem, Inversion formula, Uniqueness theorem, Characteristic function and moments, Bochner's Theorem (Statement only), Levy's continuity theorem. 15 hours

UNIT-II: Laws of Large Numbers: Convergence of series of independent random variables, Kolmogorov's inequality, Three series theorem, Weak law of large numbers (Khintchine's and Kolmogorov's), Kolmogorov's strong law of large numbers, Glivenko-Cantelli theorem, Kolmogorov's law of iterated logarithms (without proof).

25 hours

UNIT-III: Central Limit Theorem (CLT): Lindeberg-Levy theorem, Liapounov form of CLT. Lindeberg-Feller CLT (no proof required). Association between Liapounov's condition and Lindeberg conditions; Simple applications of CLT, Infinitely divisible distributions--definition, elementary properties and examples, Canonical representation : Levy-Khintchine representation and Kolmogorov representation (both without proof). 25 hours

UNIT-IV: Conditioning: Conditional expectation and its properties, Conditional probabilities, Radon-Nikodym Theorem (Statement only) and its applications. Martingales, Sub martingales, Martingale convergence theorem, Decomposition of sub martingales.

25 hours

Text Books

1. **Bhat, B.R. (2011).** Modern Probability Theory, Second edition, Wiley Eastern.
2. **Laha. R.G. and Rohatgi V.K.** (1979). Probability Theory, John Wiley.

References

1. **Billingsley, P. (1986)**. Probability and Measure, Second edition, John Wiley.
2. **Feller, W. (1976)**. An Introduction to Probability Theory and its Applications, Volume II Wiley Eastern.
3. **Hoffmann - Jorgensen J. (1994)**. Probability with a view towards Statistics, Chapman & Hall.
4. **Loeve M. (1977)**. Probability Theory, Volume I, Fourth edition, Springer-Verlag.
5. **Rohatgi, V.K. and Salah, A.K.E. (2001).** An Introduction to Probability and Statistics, John Wiley & Sons.
6. **Sidney I. Resnick (1999)**. A Probability Path, Bikhanser.

MST2C07: APPLIED REGRESSION ANALYSIS (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Illustrate the concept of linear regression model.
CO 2: Estimate and test the significance of regression parameters and explain properties of estimators.
CO 3: Check the model adequacy of regression models using residual analysis.
CO 4: Discuss polynomial, step-wise regression models.
CO 5: Explain non-linear and non-parametric regression models.
CO 6: Explain logistic and Poisson regression models for binary and count data and estimate their parameters.
CO 7: Discuss generalized linear models and estimation of its parameters.

Unit-I: Linear Regression Model, Least squares estimation, Gauss Markov Theorem, Properties of the estimates, Distribution Theory, Maximum likelihood estimation, Estimation with linear restrictions, Generalised least squares; Hypothesis testing - likelihood ratio test, F-test; Confidence intervals.

25 hours

Unit-II: Residual analysis, Departures from underlying assumptions, Effect of outliers, Collinearity, Non-constant variance and serial correlation, Departures from normality, Diagnostics and remedies.

20 hours

Unit-III: Polynomial regression in one and several variables, Orthogonal polynomials, Indicator variables, Subset selection of explanatory variables, stepwise regression and Mallows Cp - statistics, Introduction to non-parametric regression.

25 hours

Unit-IV: Introduction to nonlinear regression, Least squares in the nonlinear case and estimation of parameters, Models for binary response variables, estimation and diagnosis methods for logistic and Poisson regressions. Prediction and residual analysis, Generalized Linear Models – estimation and diagnostics.

20 hours

Text Books

1. **Seber, A.F. and Lee, A.J. (2003).** Linear Regression Analysis, John Wiley, Relevant sections.
2. **Montegomery, D.C., Peck, E.A. and Vining, G.G. (2001).** Introduction to Regression Analysis, Third edition. Wiley.
3. **B. Abraham and Ledotter, J. (1983).** Statistical Methods for Forecasting, John Wiley & Sons.

References

1. **Searle, S.R. (1971).** Linear models, John Wiley & Sons, Inc.
2. **N. Draper and H. Smith (1986).** Applied Regression Analysis – John Wiley & Sons.
3. **Fox, J. (1984)**. Linear Statistical Models and Related methods, John Wiley,
4. **Christensen, R. (2001).** Advanced Linear Modelling

MST2C08: ESTIMATION THEORY (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe the properties of estimators: unbiasedness, consistency and sufficiency.
CO 2: Explain exponential family and Pitman family of distributions, with illustrations.
CO 3: Describe the method of finding sufficient statistics, minimum variance unbiased estimators, consistent estimators and consistent and asymptotically normal estimators.
CO 4: Relate sufficient statistic and ancillary statistic using Basu's thorem.
CO 5: Determine UMVUE using complete sufficient statistic using Rao- Blackwell, and Lehmann-Scheffe theorems.
CO6: Determine the estimators using method of moments, method of percentiles, maximum likelihood method and Bayesian method.
CO7: Explain the concept of interval estimation- SELCI, Bayesian and Fiducial Intervals.

Unit-I: Sufficient statistics and minimum variance unbiased estimators: Sufficient statistics, Factorisation theorem for sufficiency, Joint sufficient statistics, Exponential family, Pitman family, Minimal sufficient statistics (MSS). Criteria to find the MSS, Ancillary statistics, Complete statistics, Basu's theorem, Unbiasedness, Best Linear Unbiased estimator(BLUE),Minimum variance unbiased estimator(MVUE),Rao-Blackwell theorem, Lehman-Scheffe theorem, Necessary and sufficient condition for MVUE, Fisher Information, Cramer Rao inequality and its applications. 30 hours

Unit-II: Consistent estimator and Consistent asymptotically normal estimators: Consistent estimator, Invariance property of consistent estimator, Method of moments-method of percentiles to determine consistent estimators, choosing between Consistent estimators. CAN estimators.

20 hours

Unit-III: Methods of estimation: Method of moments-method of percentiles-method of maximum likelihood-MLE in exponential family- Cramer family, Cramer Huzurbazar Theorem, Solution of likelihood equations-Bayesian method of estimation-Prior information-Loss functions (squared error absolute error and zero-one loss functions) – Posterior distribution-estimators under the above loss functions. 25 hours

Unit-IV: Interval estimation: Definition, Shortest expected length confidence interval-large sample confidence intervals-unbiased confidence intervals-examples-Bayesian and Fiducial intervals.

15 hours

Text Books

1. **Kale,B.K** .(2005)..A first course in parametric inference, Second Edition, NarosaPublishing House, New Delhi.
2. **George Casella and Roger L Berger** (2002). Statistical inference, Second Edition, Duxbury, Australia.

References

1. **Lehmann, E.L(1983).** Theory of point estimation, John Wiley and sons, New York.
2. **Rohatgi, V.K(1976).** An introduction to Probability Theory and Mathematical Statistics, John Wiley and sons, New York.
3. **Rohatgi, V.K (1984).** Statistical Inference, John Wiley and sons, New York.
4. **Rao, C.R (2002).** Linear Statistical Inference and its applications, Second Edition, JohnWiley and sons, New York.

MST2C09: STOCHASTIC PROCESSES (4 Credits, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Recollect the basic concepts of random variables and conditional probabilities.
CO 2: Explain Markov Chain with illustrations.
CO 3: Classify the States of a Given Markov Chain.
CO 4: Describe inter arrival time and waiting time distributions and their properties.
CO 5: Explain generalized Poisson process and their properties.
CO 6: Describe the concept and applications of renewal process.
CO 7: Explain the basic characteristics of queues and the properties of Brownian motion.

Unit-I: Introduction to stochastic processes - Time and state space, classification of stochastic processes, second order processes, stationarity, Gaussian process, Martingales, Markov process, random walk and Wiener process (examples). Markov chains- Chapman Kolmogorov equations – classification of states – limiting probabilities, first passage time distribution, stationary distribution, ergodic theorem, Gamblers ruin problem, Hidden Markov chains. 30 hours

Unit-II: Poisson processes, postulates for Poisson process, properties of Poisson process, Poisson process and related distributions of arrival times. Generalization of Poisson processes - compound Poisson process, pure birth process, birth immigration process, non-homogeneous Poisson process, Birth and death processes. 20 hours

Unit-III: Renewal processes-Renewal function and renewal density, renewal equation, stopping time, Wald's equation, limit theorems and their applications. Renewal reward process. Regenerative processes, Semi-Markov process, Brownian motion Process – hitting time – Maximum variable – variations on Brownian motion.

20 hours

Unit-IV: Basic characteristics of queues – Little's formula, steady state probabilities, M/M/1, M/M/s, M/M/∞, network of queues-open system, closed system. Non Markovian queueing models-M/G/1, G/M/1. 20 hours

Text Books

1. **Ross, S.M. (2007).** Introduction to Probability Models. IXth Edition, Academic Press.
2. **Medhi, J. (1996).** Stochastic Processes. Second Editions. Wiley Eastern, New-Delhi.

References

1. **Karlin, S. and Taylor, H.M. (1975).** A First Course in Stochastic Processes. Second Edition Academic Press. New-York.
2. **Cinlar, E. (1975).** Introduction to Stochastic Processes. Prentice Hall. New Jersey.
3. **Basu, A.K. (2003).** Introduction to Stochastic Processes. Narosa, New-Delhi.

MST2L01: STATISTICAL COMPUTING-I (Credits: 4, Hours per week: 5)

Teaching scheme: 5 hours practical per week.

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Develop scientific and experimental skills.
CO 2: Apply the principles of Analytical Tools for Statistics- II, Distribution Theory, Sampling Theory, and Estimation Theory using real data sets.
CO 3: Know the formulas to be applied for the analysis.
CO 4: To install and load the packages required to run the R codes
CO 5: Write the R codes for the analysis of the given data.
CO 6: Enter the data given for the analysis.
CO 7: Explain how to make conclusions and write the inference for the data analysis based on the output obtained.

Statistical Computing-I is a practical course. Its objectives are to develop scientific and experimental skills of the students and to correlate the theoretical principles with application based studies. The practical is based on the following FIVE courses of the first and second semester.

1.MST1C02: Analytical Tools for Statistics –II
2.MST1C04: Distribution Theory
3.MST1C05: Sampling Theory
4.MST2C07: Applied Regression Analysis
5.MST2C08: Estimation Theory

Practical is to be done by using R or Python. At least five statistical data oriented/supported problems should be done from each course. Practical Record shall be maintained by each student and the same shall be submitted for verification at the time of external examination.

The Board of Examiners (BoE) shall decide the pattern of question paper and the duration of the external examination. The external examination at the centre shall be conducted and evaluated on the same day jointly by two examiners – one external and one internal, appointed by the College on the recommendation of the Chairman, BoE. The question paper for the external examination at the centre will be set by the external examiner in consultation with the Chairman, BoE and the HoDs of the centre. The questions are to be evenly distributed over the entire syllabus. Evaluation shall be done by assessing each candidate on the scientific and experimental skills, the efficiency of the algorithm/program implemented, the presentation and interpretation of the results. The valuation shall be done by the direct grading system and grades will be finalized on the same day.

SEMESTER III

MST3C10: TIME SERIES ANALYSIS (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe the basics of time series data, its auto-covariance, auto-correlation and stationarity.
CO 2: Illustrate test for trend and seasonality.
CO 3: Explain the smoothing methods for determining trend of the data.
CO 4: Describe the properties of linear time series models.
CO 5: Fit linear models for time series data sets.
CO 6: Describe the maximum likelihood, Yule-Walker and least square estimation methods.
CO 7: Learn to validate a model using residual analysis.
CO 8: Define ARCH and GARCH models and derive their properties.
CO 9: Analyse spectral density and periodogram.

Unit-I: Motivation, Time series as a discrete parameter stochastic process, Auto – Covariance, Auto-Correlation and spectral density and their properties. Exploratory time series analysis, Test for trend and seasonality, Exponential and moving average smoothing, Holt – Winter smoothing, forecasting based on smoothing, Adaptive smoothing. 25 hours

Unit-II: Detailed study of the stationary process: Autoregressive, Moving Average, Autoregressive Moving Average and Autoregressive Integrated Moving Average Models. Choice of AR / MA periods. 25 hours

Unit-III: Estimation of ARMA models: Yule – Walker estimation for AR Processes, Maximum likelihood and least squares estimation for ARMA Processes, Discussion (without proof) of estimation of mean, Auto-covariance and auto-correlation function under large samples theory, Residual analysis and diagnostic checking. Forecasting using ARIMA models, Use of computer packages like R. 25 hours

Unit-IV: Spectral analysis of weakly stationary process. Herglotzic Theorem. Periodogram and correlogram analysis. Introduction to non-linear time Series: ARCH and GARCH models. 15 hours

Text Books

1. **Box G.E.P and Jenkins G.M. (1994).** Time Series Analysis, Forecasting and Control. Holden-Day
2. **BrockwellP.J.and Davis R.A. (2006).** Time Series: Theory and Methods, Springer – Verlag.
3. **Abraham B and Ledolter J.C. (1983).** Statistical Methods for Forecasting, Wiley
4. **Robert H Shumway and Davis S Stoffer(2016).** Time series analysis and its applications with R examples. Springer.

References

1. **Anderson T.W (1971).** Statistical Analysis of Time Series, Wiley.
2. **Fuller W.A. (1978).** Introduction to Statistical Time Series, John Wiley.
3. **William W S Wei (2006).** Time Series Analysis. Univariate and Multivariate Methods. Pearson. Addison Wesley

MST3C11: DESIGN AND ANALYSIS OF EXPERIMENTS (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe the concept of estimable functions and best estimates.
CO 2: Explain the Principles of planning of an experiment.
CO 3: Discuss and compare different complete block designs.
CO 4: Analyze experiments with and without missing values.
CO 5: Apply incomplete block designs and balanced incomplete block designs.
CO 6: Explain factorial experiments, total confounding and partial confounding.
CO 7: Differentiate between strip plot and split plot designs.

Unit-1: Linear model: Estimable functions and best estimates, Normal equations, Sum of squares, Distribution of sum of squares, Estimates and error sum of squares. Test of linear hypothesis. Basic Principles and Planning of Experiments, Experiments with Single Factor-ANOVA, Analysis of Fixed Effects Model, Model Adequacy Checking. 20 hours

Unit-II: Complete Block Designs: Completely Randomized Design, Randomized Block Design, Latin Square Design, Graeco Latin square design, Analysis with Missing Values, (RBD and LSD). Multiple comparison test. Least significant difference. Duncans's multiple range test, ANCOVA(one way and two way only). 25 hours

Unit-III: Incomplete Block Designs: Balanced Incomplete Block designs, Construction of BIB Designs, Analysis with recovery of inter-block information and intra-block information. Partially balanced incomplete block designs, Analysis of partially balanced incomplete block designs with two associate classes, Youden square design, Lattice designs. 25 hours

Unit-IV: Factorial Designs: Basic definitions and principles, Analysis of 2^n factorial experiments. Total confounding of 2^n designs in 2^n blocks. Partial confounding in 2^n blocks, 3^n factorial designs. Fractional factorial designs. Concepts of Split plot design and strip plot design.
20 hours

Text Books

1. **Joshi D.D. (1987).** Linear Estimation and Design of Experiments, Wiley Eastern.
2. **Montgomery D C (2001).** Design and Analysis of Experiments, John Wiley.
3. **Das M N and Giri N C (1979).** Design and Analysis of Experiments, second edition, Wiley.

References

1. **Chakrabarti, M.C. (1964).** Design of experiments, ISI, Calcutta.
2. **Hinkleman and Kempthrone C** (1994). Design and Analysis of Experiments Volume I, John Wiley.

MST3C12: TESTING OF STATISTICAL HYPOTHESES (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO1: Explain the problem of testing of hypotheses and the concept of p value.
CO2: Construct most powerful tests using Neyman-Pearson lemma, one-sided and two-sided UMP tests and UMP unbiased tests.
CO3: Describe the concept of α-similar tests and construct such tests.
CO4: Apply nonparametric tests for testing goodness of fit, homogeneity and independence.
CO5: Develop SPRT for different problems.

Unit-I: Tests of hypotheses & Most Powerful Tests: Simple versus simple hypothesis testing problem –Error probabilities, p-value and choice of level of significance – Most powerful tests – Neyman Pearson Lemma – Generalized Neyman–Pearson Lemma, One-sided UMP tests, two- sided UMP tests and UMP unbiased tests.
25 hours

Unit-II:UMP test for multi-parameter case: UMP unbiased test, α-similar tests and α-similar tests with Neyman structure, construction of α-similar tests with Neyman structure. Principle of invariance in testing of hypotheses, locally most powerful tests – Likelihood ratio tests – Bayesian tests.
20 hours

Unit-III: Non-parametric Tests: Single sample tests – testing goodness of fit, Chi-square tests- Kolmogorov– Smirnov test – sign test – Wilcoxon signed rank test. Two sample tests – the chi-square test for homogeneity – Kolmogorov – Smirnov test; the median test – Mann- Whitney-Wilcoxon test - Test for independence – Kendall's tau – Spearman's rank correlation coefficient – robustness.
25 hours

Unit-IV: Sequential Tests: Some fundamental ideas of sequential sampling – Sequential Probability Ratio Test (SPRT) – important properties, termination of SPRT – the fundamental identity of SPRT – Operating Characteristic (OC) function and Average Sample Number (ASN) of SPRT – Developing SPRT for different problems. 20 hours

Text Books

1. **Casella, G. and Berger, R.L. (2002).** Statistical Inference, Second Edition Duxbury, Australia.
2. **Rohatgi, V.K. (1976).** An Introduction to Probability Theory and Mathematical Statistics,John – Wiley Sons, New – York.
3. **Manojkumar Srivastava and Namita Srivstava(2009).** Statistical Inference: Testing of Hypothesis, Eastern Economy Edition, PHI Learning Pvt. Ltd., New Delhi.

References

1. **Rohatgi,V.K. (1984)**. Statistical Inference, John-Wiley and Sons, New-York.
2. **Lehman, E.L. (1983).** Theory of Point Estimation, John-Wiley and Sons, New-York

3. **Kale, B.K. (2005).** A First Course on Parametric Inference. Second Edition, Narosa Publishing, New-Delhi.
4. **Lehman, E.L. and Romano, Joseph P.(2005).** Testing Statistical Hypotheses. Third Edition, Springer, New-York.

MST3 …: ELECTIVE-I (Credits: 4, Hours per week: 5)

(Electives are to be chosen from the list attached)

MST3E …: ELECTIVE-II (Credits: 4, Hours per week: 5)

(Electives are to be chosen from the list attached)

SEMESTER IV

MST4C13: MULTIVARIATE ANALYSIS (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Identify the marginal and conditional distributions of multivariate normal distribution.

CO 2: Recall the properties of multivariate normal distribution.

CO 3: Compute the MLE estimates of the parameters of multivariate normal distribution and determine their sampling distributions.

CO 4: Describe the independence and probability distributions of quadratic and linear forms.

CO 5: Discuss Wishart's distribution and its properties.

CO 6: Describe the testing problems in connection with multivariate normal distribution.

CO 7: Illustrate and apply the techniques of Classification, principal component analysis, factor analysis and cluster analysis.

Unit-I: Multivariate Normal Distribution- Properties and estimation: Conditional distribution, marginal distribution, maximum likelihood estimation of the mean vector and dispersion matrix, sufficient statistics for mean vector and dispersion matrix, independence of mean vector and sum of squares and product matrix, the distribution of sample mean vector, inference concerning the mean vector when the dispersion matrix is known for single and two populations. Distribution of simple, partial and multiple (null-case only) correlation coefficients.

20 hours

Unit-II: Quadratic forms and associated distributions: Test concerning the mean vector when dispersion matrix is known (one sample and two sample), confidence interval **(**one sample and two sample), distribution of quadratic forms. Independence of a linear form and quadratic form, independence of two quadratic forms, Cochran's theorem, matrix variate Gamma distributions (one and two parameters:- definition and characteristic function only), matrix variate Beta distributions (one and two parameters-definition and characteristic function only), Wishart distribution: properties–rectangular coordinates, generalized variance, distribution of sample correlation coefficient ($\rho=0$, $\rho\neq0$), sampling distribution of partial and multiple correlation coefficients, partial regression coefficient.

25 hours

Unit-III: Testing Problems: Likelihood ratio tests–Hotelling's T^2 Statistics-properties, applications (one sample and two sample), problem of testing symmetry, Mahalanobi's D^2 statistic- properties and applications, Fisher-Behrens test, test for independence of sets of variates, testing the hypothesis of equality of covariance matrices, test for identical populations, sphericity test. 20 hours

Unit-IV: Applied Multivariate Analysis: Classification problems, construction of classification rules, Classification of one of two multivariate normal population**.**

Dimension reduction problems**:** principal component analysis – Summarizing sample variation by principal components – Iterative procedure to calculate sample principal components, canonical correlation analysis, computation of canonical correlations and variables. Factor analysis-factor loading principal factor method, Factor Rotations. Cluster analysis- different types of clustering- basic concepts and applications.

25 hours

Text Books

1. **Anderson, T.W. (2003).** Multivariate Analysis. John – Wiley, New York.
2. **Mathai, A.M., Provost, S.B., Haubold, H.J. (2022).** Multivariate Statistical Analysis in the Real and Complex Domains, Springer, New York.
3. **Johnson, R.A. and Wichern, D.W. (2001).** Applied multivariate statistical analysis, 3rd Edn.,Prentice Hall of India, New Delhi.
4. **Härdle,W. K. and Simar, L.(2012).**Applied Multivariate Statistical Analysis, Third Edition, Springer, New York.
5. **Rao, C.R. (2002).** Linear Statistical Inference and Its Applications, Second Edition, John Wiley andSons, New York.

References

1. **Mardia,K.V,Kent,J.T. and Bibby, J.M.(1995)** . Multivariate Analysis, Accademic Press, New York
2. **Giri, N.C. (1996).** Multivariate Statistical Analysis. Marcel Dekker. Inc., New York.
3. **Kshirasagar, A.M. (1972).** Multivariate Analysis. Marcel Dekker. New-York
4. **Rencher, A.C. (1998).** Multivariate Statistical Analysis. Jon Wiley, New York.
5. **Morrison, D.F. (1976).** Multivariate statistical methods, McGraw Hill, New York.

MST4 …: ELECTIVE- III (Credits: 4, Hours per week: 5)

(Electives are to be chosen from the list attached)

MST4L02: STATISTICAL COMPUTING-II (Credits: 4, Hours per week: 5)

Teaching scheme: 5 hours practical per week.

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Develop scientific and experimental skills of the students.
CO 2: Apply the principles of the courses in third and fourth semesters using real data sets.
CO 3: Know the formulas to be applied for the analysis.
CO 4: To install and load the packages required to run the R codes
CO 5: Write the R codes for the analysis of the given data.
CO 6: Enter the data given for the analysis.
CO 7: Explain how to make conclusions and write the inference for the data analysis based on the output obtained.

Statistical Computing-II is a practical course. Its objectives are to develop scientific and experimental skills of the students and to correlate the theoretical principles with application based studies. The practical is based on the following courses of the third and fourth semesters.

1. MST3C10: Time Series Analysis
2. MST3C11: Design and Analysis of Experiments
3. MST3C12: Testing of Statistical Hypotheses
4. MST3E--: Elective-I
5. MST3E--: Elective -II
6: MST4C13: Multivariate Analysis
7. MST4E--: Elective -III

Practical is to be done by using R or Python. At least five statistical data oriented/supported problems should be done from each course. Practical Record shall be maintained by each student and the same shall be submitted for verification at the time of external examination.

The Board of Examiners (BoE) shall decide the pattern of question paper and the duration of the external examination. The external examination at the centre shall be conducted and evaluated on the same day jointly by two examiners – one external and one internal, appointed by the College on the recommendation of the Chairman, BoE. The question paper for the external examination at the centre will be set by the external examiner in consultation with the Chairman, BoE and the H/Ds of the centre. The questions are to be evenly distributed over the entire syllabus. Evaluation shall be done by assessing each candidate on the scientific and experimental skills, the efficiency of the algorithm/program implemented, the presentation and interpretation of the results. The valuation shall be done by the direct grading system and grades will be finalized on the same day.

MST4P01: PROJECT/DISSERTATION (Credits: 5, Hours per week : 10)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Discuss the applications of various statistical techniques learned in the entire course in the form of project work.
CO 2: Manage a real practical situation where a statistical analysis is sought.
CO 3: Develop professional approach towards writing and presenting an academic report.
CO 4: Get more insight about the opportunities in research/career.

In partial fulfilment of the M.Sc. programme, during the fourth semester each student has to undertake a project work in a selected area of interest under a supervisor in the department. The topic could be a theoretical work or data analysis type. At the end of the fourth semester the student shall prepare a **report/dissertation** which summarizes the project work and submit to the H/D of the parent department positively before the deadline suggested in the Academic calendar. The project/ dissertation is of **5 credits** for which the following evaluation will be followed:

The valuation shall be jointly done by the supervisor of the project in the department and an External Expert appointed by the College, based on a well-defined scheme of valuation framed by them. The following break up of weightage is suggested for its valuation.

1. Review of literature, formulation of the problem and defining clearly the objective: 10%
2. Methodology and description of the techniques used: 10%
3. Analysis, programming/simulation and discussion of results: 20%
4. Presentation of the report, organization, linguistic style, reference etc.: 20%
5. Viva-voce examinations based on project/dissertation: 40%.

MST4V01: COMPREHENSIVE VIVA-VOCE (Credits:3)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO1: Communicate the concepts of each course precisely.

CO2: Communicate the importance and applications of the subject Statistics in a broad sense.

CO3: Get more insights into the subject areas.

CO4: Face interviews without fear and communicate their ideas effectively.

There shall be a comprehensive Viva Voce examination based on all courses of the programme with 3 credits, internal and external being in the ratio 1:4. The Viva-Voce shall be conducted by a board of examiners consisting of at least one external expert and internal examiner.

....

SYLLABI OF ELECTIVE COURSES

E01: OPERATIONS RESEARCH-I (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe the basic concepts of linear programming problem(LPP).
CO 2: Discuss different methods of solving LPP.
CO 3: Illustrate the concept of duality.
CO 4: Solve transportation problem, assignment problem and parametric programming problem.
CO 5: Explain Integer programming problems and the methods of solving it.
CO 6: Describe basics of game theory and solve game problem as LPP.

Unit-I: Operations Research.-definition and scope, Linear programming, graphical method, simplex method, artificial basis techniques, two phase simplex method, Big-M method, duality concepts, duality theorems, dual simplex methods. 30 hours

Unit-II: Transportation and assignment problems, sensitivity analysis, parametric programming. Sequencing and Scheduling problems-2 machine n-Job and 3- machine n-Job Problems. 25 hours

Unit-III: Integer programming: Cutting plane methods, branch and bound technique, application of zero – one programming. 20 hours

Unit-IV: Game theory: two person zero sum games, minimax theorem, game problem as a linear programming problem. Co-operative and competition games. 15 hours

Text Book

1**. K.V.Mital and Mohan, C (1996).** Optimization Methods in Operations Research and Systems Analysis, 3rd Edition, New Age International (Pvt.) Ltd.

References

1. **Hadley, G. (1964).** Linear Programming, Oxford & IBH Publishing Co, New Delhi.
2. **Taha. H.A. (1982).** Operation Research, An Instruction, Macmillan.
3. **Hiller FS. And Lieberman, G.J. (1995).** Introduction to Operations Research, McGraw Hill
4. **Kanti Swamp, Gupta, P.K and Manmohan.(1999).** Operations Research, Sultan Chand & Sons.

E02: OPERATIONS RESEARCH-II (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Discuss the Non linear programming problems and methods to solve the problems.
CO 2: Understand and solve quadratic programming problem.
CO 3: Explain Dynamic and Geometric programming.
CO 4: Discuss inventory management, deterministic and probability models.
CO 5: Understand Replacement models.
CO 6: Understand simulation modeling and random number generation

Unit-I. Non-linear programming, Lagrangian function, saddle point, Kuhn-Tucker Theorem, Kuhn-Tucker conditions, Quadratic programming, Wolfe's algorithm for solving quadratic programming problem. 20 hours

Unit-II. Dynamic and Geometric programming: A minimum path problem, single additive constraint, additively separable return; single multiplicative constraint, additively separable return; single additive constraint, multiplicatively separable return, computational economy in DP. Concept and examples of Geometric programming. 25 hours

Unit-III. Inventory management; Deterministic models, the classical economic order quantity, nonzero lead time, the EOQ with shortages allowed, the production lot-size model. Probabilistic models.the newsboy problem, a lot size. reorder point model. 20 hours

Unit-IV. Replacement models; capital equipment that deteriorates with time, Items that fail completely, mortality theorem, staffing problems, block and age replacement policies. Simulation modeling: Monte Carlo simulation, sampling from probability distributions. Inverse method, convolution method, acceptance-rejection methods, generation of random numbers, Mechanics of discrete simulation.

25 hours

Text Books

1. **K.V.Mital and Mohan, C (1996).** Optimization Methods in Operations Research and Systems Analysis, 3rd Edition, New Age International (Pvt.) Ltd.
2. **M.Sasieni, A.Yaspan and L.Friendman(1959).** Operations Research; Methods and Problems, Wiley, New York.
3. **Hamdy A. Taha (1997).** Operations Research – An Introduction, Prentice-Hall Inc., New Jersey.
4. **Ravindran, Philips and Solberg (1987).** Operations Research- Principles and Practice, John Wiley & Sons, New York.

References

1. **Sharma, J.K. (2003).** Operations Research, Theory & Applications, Macmillan India Ltd.
2. **Manmohan, Kantiswaroop and Gupta(1999).** Operation Research, Sultan Chand & Sons New Delhi.

E03: QUEUEING THEORY (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Discuss the basic characteristic features of a queuing system and acquire skills in analysing queuing models.
CO 2: Explain the theory of Markovian queueing models, M/M/I, M/M/∞ and M/M/c queueing systems.
CO 3: Define and explain basic concepts in the theory ofAdvanced Markovian models, M[X] /M/1, M/M[Y] /1, M/Ek/1, Ek/M/1 and Erlangian models.
CO 4: Analyze a network of queues with Jackson network.
CO 5: Apply models with general arrival pattern and M/G/1 queueing model.
CO 6: Discuss and apply queuing models in real life problem.

Unit-I. Introduction to queueing theory, Characteristics of queueing processes, Measures of effectiveness, Markovian queueing models, steady state solutions of the M/M/I model, waiting time distributions, Little's formula, queues with unlimited service, finite source queues. 20 hours

Unit-II. Transient behavior of M/M/1 queues, transient behavior of M/M/∞. Busy period analysis for M/M/1 and M/M/c models. Advanced Markovian models. Bulk input $M^{[X]}$ /M/1 model, Bulk service $M/M^{[Y]}$ /1 model, Erlangian models, $M/E_k/1$ and $E_k/M/1$. A brief discussion of priority queues.
25 hours

Unit-III. Queueing networks-series queues, open Jackson networks, closed Jackson network, Cyclic queues, Extension of Jackson networks. Non Jackson networks. 20 hours

Unit-IV. Models with general arrival pattern, The M/G/1 queueing model, The Pollaczek-khintchine formula, Departure point steady state systems size probabilities, ergodic theory, Special cases $M/E_k/1$ and M/D/1, waiting times, busy period analysis, general input and exponential service models, arrival point steady state system size probabilities.
25 hours

Text Books/References

1. **Gross, D. and Harris, C.M.(1985).** Fundamentals of Queueing Theory, 2nd Edition, John Wiley and Sons, new York.
2. **Kleinrock L (1975).** Queueing Systems, Vol. I &Vol 2, Joohn Wiley and Sons, New York.
3. **Ross, S.M. (2007).** Introduction to Probability Models. 9th Edition, Academic Press, New York.
4. **Bose, S.K. (2002).** An Introduction to Queueing Systems, Kluwer Academic/Plenum Publishers, New York.

E04: LIFETIME DATA ANALYSIS (Credits: 4, Hours per week: 5)

Course outcomes: *After successful completion of the course, students will be able to:*

CO 1: Discuss life time distributions and important parametric models.
CO 2: Explain censoring and estimation of parameters using censored data.
CO 3: Estimate the survival probabilities and cumulative hazard function using product limit and Nelson-Aalen estimate respectively.
CO 4: Describe inference under exponential model and discuss the comparison of distributions.
CO 5: Explain important hazard models and apply Rank test, Log-rank test and Generalized Wilcoxon test.
CO 6: Describe the concept of regression models and Cox PH model & AFT model.
CO 7: Discuss multivariate lifetime models.

Unit-I: Lifetime distributions-continuous and discrete models-important parametric models: Exponential, Weibull, Log-normal, Log-logistic, Gamma, Inverse Gaussian distributions, Log location scale models and mixture models. Censoring and statistical methods. 20 hours

Unit-II: The product-limit estimator and its properties. The Nelson-Aalen estimator, interval estimation of survival probabilities, descriptive and diagnostic plots, estimation of hazard function, methods for truncated and interval censored data, Life tables. 20 hours

Unit-III: Inference under exponential model – large sample theory, type-2 censored test plans, comparison of two distributions; inference procedures for Gamma distribution; models with threshold parameters, inference for log-location scale distribution: likelihood based methods: Exact methods under type-2 censoring; application to Weibull and extreme value distributions, comparison of distributions.

25 hours

Unit-IV: Log-location scale (Accelerated Failure time) regression models, Proportional hazard models, Methods for continuous multiplicative hazard models, Semi-parametric maximum likelihood-estimation of continuous observations, Rank test for comparing Distributions, Log-rank test, Generalized Wilcoxon test. A brief discussion on multivariate lifetime models and data.

25 hours

Text Books

1. **Lawless, J.F.(2003).** Statistical Methods for Lifetime (Second Edition), John Wiley & Sons Inc., New Jersey.
2. **Kalbfiesche, J.D. and Prentice, R.L. (1980).** The Statistical Analysis of Failure Time Data, John Wiley & Sons Inc. New Jersey.

References

1. **Miller, R.G.(1981).** Survival Analysis, John Wiley & Sons Inc.
2. **Bain, L.G.(1978).** Statistical Analysis of Reliability and Life testing Models, Marcel Decker.
3. **Nelson, W. (1982).** Applied Life Data Analysis.
4. **Cox, D.R and Oakes, D.(1984).** Analysis of Survival Data. Chapman and Hall.
5. **Lee, Elisa, T. (1992).** Statistical Methods for Survival Data Analysis, John Wiley & Sons.

E05: ADVANCED DISTRIBUTION THEORY (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO1: Generate generalized distribution model's by combining any two discrete distributions by the way of stopped sum.
CO2: Familiarize bivariate discrete distributions / family of distributions, their statistical properties and generation procedure,
CO3: Apply bivariate discrete distributions for modeling bivariate discrete data sets.
CO4: Describe bivariate and multivariate continuous distributions, derive its marginals, conditionals and sub cases and thereby able to apply these distributions for modeling bivariate/ multivariate continuous data sets.
CO5: Develop various record value distributions and analyze their statistical properties and through which, realize the importance of sole existence of extreme observations

Unit-I: Stopped sum distributions: Poisson stopped sum, Neyman type A, Poission-binomial, Poisson-negative binomial, Legrangian Poisson distributions, Distributions of order Poisson, negative binomial, Logarithmic series, Binomial. 25 hours

Unit-II: Bivariate discrete distributions: bivariate power series distributions, bivariate Poisson, negative binomial and logarithmic series distributions, properties of these distributions, bivariate hypergeometric distribution and its properties.

25 hours

Unit-III: Bivariate continuous models, bivariate Pearson system, Farlie Morgenstern distribution; distributions with specified conditionals, bivariate Pareto of I, II, III and IV kind, multivariate Liouville distributions.

20 hours

Unit-IV: Record values - definition, properties, distribution of nth record, record values from exponential, Weibull and logistic; Moments relationships, characterizations. 20 hours

Reference Books

1. **Johnson, N.L., Kotz, S. and Kemp, A.W. (1992).** Univariate discrete distributions, second edition, Wiley.

2. **Kocherlakota, S. and Kocharlakota, K. (1992).** Bivariate Discrete Distributions, Marcel-Dekker.
3. **Johnson, N.L., Kotz, S. and Balakrishnan, N. (1997).** Discrete multivariate distributions, second edition, Wiley.
4. **Kotz, S. ,Balakrishnan, N. and Johnson, N.L. (2000).** Continuous multivariate distributions, Volume I, John Wiley and Sons.
5. **Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1998).** Records, John Wiley and Sons.

E06: STATISTICAL DECISION THEORY (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Discuss the need and role of statistical decision theory to solve real decision problems.
CO 2: Explain various decision principles and associated loss functions.
CO 3: Differentiate between frequentist and Bayesian decision theory.
CO 4: Describe the prior and posterior distributions.
CO 5: Handle prior selection problem.
CO 6: Describe the development of optimal strategies for actions in competitive situations involving two or more intelligent antagonists.
CO 7: Explain the general techniques for solving games using game theory.

Unit-I: Statistical decision Problem – Decision rule and loss-randomized decision rule. Decision Principle – sufficient statistic and convexity. Utility and loss-loss functions- standard loss functions-vector valued loss functions. 25 hours

Unit-II: Prior information-subjective determination of prior density-Non-informative priors-maximum entropy priors, the marginal distribution to determine the prior-the ML-II approach to prior selection. Conjugate priors. 25 hours

Unit-III: The posterior distribution-Bayesian inference-Bayesian decision theory-empirical Bayes analysis – Hierarchical Bayes analysis-Bayesian robustness Admissibility of Bayes rules.
20 hours

Unit-IV: Game theory – basic concepts – general techniques for solving games Games with finite state of nature-the supporting and separating hyper plane theorems. The minimax theorem. Statistical games.
20 hours

Text Books

1. **Berger, O,J.(1985).** Statistical decision Theory and Bayesian Analysis, Second Edition Springer-Verlag.

References

1. **Ferguson, T.S. (1967).** Mathematical Statistics; A Decision-Theoretic Approach, Academic Press, New-York.
2. **Lehman, E.L.(1983).** Theory of Point Estimation. John-Wiley, New-York.
3. **Giovanni Parmigiani, Luroles, Y.T. Inouve and Hedibert F. Lopes (2009).** Decision Theory- Principles and Approaches, John Wiley.

E07: RELIABILITY MODELLING (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe the basic concepts of system reliability
CO 2: Discuss the lifetime of a system based on ageing properties
CO 3: Explain shock models and stress-strength models using reliability theory.
CO 4: Describe Maintenance and replacement policies
CO 5: Explain reliability growth models.

Unit-I: Reliability concepts and measures; components and systems; coherent systems; reliability of coherent systems; cuts and paths; modular decomposition; bounds on system reliability; structural and reliability importance of components. 20 hours

Unit-II: Life distributions; reliability function; hazard rate; common life distributions- exponential, Weibull, Gamma etc. Estimation of parameters and tests in these models.Notions of ageing; IFR, IFRA, NBU, DMRL, and NBUE Classes and their duals; closures or these classes under formation of coherent systems, convolutions and mixtures. 25 hours

Unit-III: Univariate shock models and life distributions arising out of them; bivariate shock models; common bivariate exponential distributions and their properties. Reliability estimation based on failure times in variously censored life tests and in tests with replacement of failed items; stress-strength reliability and its estimation. 25 hours

Unit-IV: 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; tests for HPP vs. NHPP with repairable systems. Basic ideas of accelerated life testing. 20 hours

Text Books / References

1. **Barlow R.E. and Proschan F.(1985).** Statistical Theory of Reliability and Life Testing; Holt, Rinehart and Winston.
2. **Bain L.J. and Engelhardt (1991).** Statistical Analysis of Reliability and Life Testing Models; Marcel Dekker.
3. **Aven, T. and Jensen,U. (1999).** Stochastic Models in Reliability, Springer-Verlag, New York, Inc.
4. **Lawless, J.F. (2003).** Statistical Models and Methods for Lifetime (Second Edition), John Wiley & Sons Inc., New Jersey.
5. **Nelson, W (1982).** Applied Life Data analysis; John Wiley.
6. **Zacks, S. (1992).** Introduction to Reliability Analysis: Probability Models and Statistics Methods. New York: Springer-Verlag.

.

E08: ACTUARIAL STATISTICS (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Apply the elements of interest.

CO 2: Discuss regular pattern of cash flows and related topics.

CO 3: Illustrate and apply individual and collective risk models for a short period.

CO 4: Discuss survival distributions and derive survival functions.

CO 5: Explain and apply life insurance models.

CO 6: Discuss and apply annuity models.

Unit I: Elements of the Theory of Interest -Compound interest - Nominal rate - Discount and annuities -Accumulated value - Effective and nominal discount rates. Cash flows - An analogy with currencies - Discount functions - Calculating the discount function - Interest and discount rates - Constant interest - Values and actuarial equivalence – Regular pattern cash flows -Balances and reserves -Time shifting and the splitting identity - Change of discount function - Internal rates of return - Forward prices and term structure – Economics of Insurance – Utility – Insurance and Utility. 25 hours

Unit II: An Individual Risk Model for a Short Period: The distribution of individual payment – The aggregate payment (convolutions) – Premiums and solvency – Some general premium principles.
A Collective Risk Model for a Short Period: The distribution of aggregate claim (Single homogeneous and several homogeneous groups) – Premiums and solvency. 20 hours

Unit III: Survival Distributions - Survival functions and force of mortality - The time-until-death for a person of a given age - Curtate-future-lifetime- Survivorship groups- Life tables and interpolation- Analytical laws of mortality - A Multiple Decrement Model - Multiple Life Models.
20 hours

Unit IV: Life Insurance Models: The present value of a future payment- The present value of payments for a portfolio of many policies – Whole life insurance - Deferred whole life insurance - Term insurance – Endowments - Varying Benefits - Multiple Decrement and Multiple Life Models. Annuity Models: Continuous and discrete annuities - Level Annuities (certain and random annuities)- whole life annuities – Temporary annuities - Deferred annuities - Certain and life annuities - Varying Payments – annuities with m-thly payments - Multiple Decrement and Multiple Life Models – Premiums and reserves.

25 hours

Text Books/References

1. **Rotar, V.I. (2015).** Actuarial Models – The mathematics of Insurance – Second Edition. CRC Press, New York.
2. **Promislow, S.D. (2015).** Fundamentals of Actuarial Mathematics- Third Edition. John Wiley & Sons, New York.
3. **Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A.& Nesbitt**, **C.J. (1997).** Actuarial Mathematics, Society of Actuaries.

E09: STATISTICAL QUALITY CONTROL (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Distinguish the concepts quality, quality assurance and acceptance sampling.
CO 2: Explain and compare the methods of acceptance sampling for attributes.
CO 3: Explain acceptance sampling by variables and continuous sampling plans.
CO 4: Describe and apply the control chart for attributes.
CO 5: Explain and implement control chart for variables.
CO 6: Describe process capability analysis and CUSUM and EWMA control charts.

Unit-I: Quality and quality assurance, methods of quality assurance, Introduction to TQM. Acceptance sampling for attributes, Single sampling, Double sampling. Multiple sampling and Sequential sampling plans. Measuring the performance of these sampling plans.

20 hours

Unit-II: Acceptance sampling by variables, sampling plans for single specification limit with known and unknown and unknown variance, Sampling plans with double specification limits., comparison of sampling plans by variables and attributes, Continuous sampling plans I, II &III.

25 hours

Unit-III: Control charts, Basic ideas, Designing of control charts for the number of non-conformities. Mean charts. Median charts. Extreme value charts, R-charts, and S-charts ARI, Economic design of control charts.

25 hours

Unit-IV: Basic concepts of process monitoring and control; process capability and process optimization. Control charts with memory – CUSUM charts, EWMA mean charts, OC and ARI for control charts, Statistical process control, Modeling and quality programming. Orthogonal arrays and robust quality.

20 hours

Text Books

1. **Montgomory, R.C. (1985).** Introduction to Statistical Quality Control. 4th edition.Wiley, New-York.
2. **Mittage, H.J. and Rinne, H. (1993).**Statistical Methods for Quality Assurance. Chapman and Hall.Chapters13 and 14.
3. **Oakland, J.S. and Follorwel, R.F. (1990).** Statistical Process Control.East-West Press.Chapters 13 and 14.
4. **Schilling, E.G. (1982).** Acceptance Sampling in Quality Control. Marcel Dekker. 5. Duncan, A.J. (1986). Quality Control and Industrial Statistics.

References

1. **Gerant, E.L. and Leaven Worth, R.S. (1980).** Statistical Quality Control.Mc-Graw Hill
2. **Chin-Knei Chao (1987).** Quality Programming, John Wiley.
3. **Ott, E.R. (1975).** Process Quality Control; McGraw Hill. 4. Wetherill, G.B. and Brown, D.W ( ).: Statistical Process Control: Theory and Practice.

E10: ADVANCED PROBABILITY THEORY (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe the basic concepts of Probability, Mathematical Expectation and Lebesgue – Stieltjes Integrals
CO 2: Explain the concept of Weak and Complete convergence of random variables.
CO 3: Discuss the basic theorems based on Decomposition of normal distribution
CO 4: Describe the properties and applications of infinitely divisible distributions and stable distributions.
CO 5: *Explain the relevance's of Conditional expectations in Martingales and Study properties of martingales.*

Unit-I: Review of Elementary Probability theory, Basic properties of expectations, Sequences of integrals, Lebesgue-Stieltjes integrals, Weak convergence - Theorems. 25 hours

Unit-II: Complete convergence: Kolmogorov's three-series and two series theorems, Decomposition of normal distribution, Levy metric, Zolotarev and Lindeberg-Feller Theorems; Berry-Esseen Theorem.

25 hours

Unit-III: More on Infinitely divisible distributions, Convergence under UAN, Convergence to special distributions, Cauchy functional equation, Stable distributions. 20 hours

Unit-IV: Conditional expectations (general case), Random-Nikodyn theorem, Martingales, Doob's decomposition, L_P -spaces Martingales, Martingale limit theorems, Exchangeability, Definite's theorem.

20 hours

Text Books

1**. Galambos J (1988).** Advanced Probability Theory, Marcel Dekker, New York

References

1. **Ash R. B (2000).** Probability and Measure Theory, Second edition. Academic Press.
2**. Billingsley P (1985).** Probability and Measure, Second edition, John Wiley and Sons, NewYork.
3. **Laha R.G. and Rohatgi, V.K. (1979).** Probability Theory, John Wiley and Sons, NewYork.

E11: OFFICIAL STATISTICS (Credits: 4, Hours per week: 5)

Course Outcomes: After successful completion of the course, students will be able to:

CO1: Describe Indian and International Statistical systems, their role, functions and activities.
CO2: Discuss the scope and contents of population census of India;
CO3: Explain the population growth in developed and developing countries and evaluate the performance of family welfare programs.
CO4: Identify Statistics related to industries, foreign trade, balance of payment, cost of living inflation, educational and social statistics.
CO5: Illustrate economic development and national income estimation using product approach, income approach and expenditure approach.
CO6: Discuss the measures of inequality in income and measures of incidence and intensity.

Unit I: Introduction to Indian and International Statistical systems. Role, function and activities of Central and State Statistical organizations. Organization of large-scale sample surveys. Role of National Sample Survey Organization. General and special data dissemination systems. Scope and Contents of population census of India. 25 hours

Unit II: Population growth in developed and developing countries, Evaluation of performance of family welfare programmes, projections of labour force and man power. Statistics related to Industries, foreign trade, balance of payment, cost of living, inflation, educational and other social statistics. 25 hours

Unit III: Economic development: Growth in per capita income and distributive justice indices of development, human development index. National income estimation- Product approach, income approach and expenditure approach. 20 hours

Unit IV: Measuring inequality in incomes: Gini Coefficient, Theil's measure; Poverty measurements: Different issues, measures of incidence and intensity; Combined Measures: Indices due to Kakwani, Senetc. 20 hours

Text Books/References/Suggested Readings

1. Basic Statistics Relating to Indian Economy (CSO) 1990
2. Guide to Official Statistics (CSO) 1999
3. Statistical System in India (CSO) 1995
4. Principles and Accommodation of National Population Census, UNEDCO.
5. **Panse, V.G:** Estimation of Crop Yields (FAO)
6. Family Welfare Year Book. Annual Publication of D/O Family Welfare.
7. Monthly Statistics of Foreign Trade in India, DGCIS, Calcutta and other Govt. Publications.
8. **CSO (1989)a:** National Accounts Statistics- Sources and Methods.
9. **Keyfitz, N (1977):** Applied Mathematical Demography- Springer Verlag.
10. **Sen, A(1977):** Poverty and Inequality.
11. **UNESCO:** Principles for Vital Statistics Systems, Series M-12.
12. **CSO (1989)b:** Statistical System in India
13. **Chubey, P.K (1995):** Poverty Measurement, New Age International.

E12: BIOSTATISTICS (Credits: 4, Hours per week: 5)

Course Outcomes: After successful completion of the course, students will be able to:

CO 1: Discuss types of Biological data and Principles of Bio Statistical design of medical studies.

CO 2: Explain the concepts of survival time functions of important parametric models and comparetwo survival distributions using LR test and Cox's F-test.

CO 3: Explain censoring and estimation of parameters using censored data.

CO 4: Describe competing risk theory and estimate the probabilities of death by ML method.

CO 5: Discuss the Basic biological concepts in genetics and clinical trials.

Unit-I: Biostatistics-Example on statistical problems in Biomedical Research-Types of Biological data-Principles of Biostatistical design of medical studies- Functions of survival time, survival distributions and their applications viz. exponential, gamma, Weibull, Rayleigh, lognormal, distribution having bath-tub shape hazard function. Parametric methods for comparing two survival distributions (L.R test and Cox's F-test).

25 hours

Unit-II: Type I, Type II and progressive or random censoring with biological examples, Estimation of mean survival time and variance of the estimator for type I and type II censored data with numerical examples. Non-parametric methods for estimating survival function and variance of the estimator viz. Acturial and Kaplan –Meier methods. 25 hours

Unit-III: Categorical data analysis (logistic regression) - Competing risk theory, Indices for measurement of probability of death under competing risks and their inter-relations. Estimation of probabilities of death under competing risks by ML method. Stochastic epidemic models: Simple and general epidemic models. 20 hours

Unit-IV: Basic biological concepts in genetics, Mendel's law, Hardy- Weinberg equilibrium, random mating, natural selection, mutation, genetic drift, detection and estimation of linkage in heredity. Planning and design of clinical trials, Phase I, II, and III trials. Sample size determination in fixed sample designs. Planning of sequential, randomized clinical trials, designs for comparative trials; randomization techniques and associated distribution theory and permutation tests (basic ideas only); ethics behind randomized studies involving human subjects; randomized dose-response studies(concept only).

20 hours

Text Books / References

1. **Biswas, S. (1995).** Applied Stochastic Processes. A Biostatistical and Population OrientedApproach, Wiley Eastern Ltd.
2. **Cox, D.R.** and **Oakes, D. (1984).** Analysis of Survival Data, Chapman and Hall.
3. **Elandt, R.C.** and **Johnson (1975).** Probability Models and Statistical Methods in Genetics,John Wiley & Sons.
4. **Ewens, W. J.** and **Grant, G.R. (2001).** Statistical methods in Bioinformatics.: AnIntroduction, Springer.
5. **Friedman, L.M., Furburg, C.** and**DeMets, D.L. (1998).** Fundamentals of Clinical Trials,SpringerVerlag.
6. **Gross, A. J.** and **Clark V.A. (1975).** Survival Distribution; Reliability Applications inBiomedical Sciences, John Wiley & Sons.
7. **Lee, Elisa, T. (1992).** Statistical Methods for Survival Data Analysis, John Wiley & Sons.
8. **Li, C.C. (1976).** First Course of Population Genetics, Boxwood Press.
9. **Daniel, W.W.(2006).** Biostatistics: A Foundation for Analysis in the Health sciences, JohnWiley &sons.Inc.
10. **Fisher, L.D.** and **Belle, G.V. (1993).** Biostatistics: A Methodology for the Health Science,John Wiley & Sons Inc.
11. **Lawless, J.F.(2003).** Statistical Methods for Lifetime (Second Edition), John Wiley & Sons.
12. **Chow, Shein-Chung and Chang, Mark (2006).** Adaptive Design Methods inClinical Trials. Chapman & Hall/CRC Biostatistics Series.
13. **Chang, Mark (2007).** Adaptive Design Theory and Implementation Using SASand R. Chapman & Hall/CRC Biostatistics Series.
14. **Cox, D.R. and Snell, E.J. (1989).** Analysis of Binary Data, SecondEdition. Chapman & Hall / CRC Press.
15. **Hu, Feifang and Rosenberger, William (2006).** The Theory of Response-AdaptiveRandomization in Clinical Trials. John Wiley.

16. **Rosenberger, William and Lachin, John (2002).** Randomization in Clinical Trials: Theory and Practice. John Wiley.

E13: ECONOMETRIC MODELS (Credits: 4, Hours per week: 5)

Course Outcomes: *After* successful completion of the course, students will be able to:

CO 1: Explain the meaning and methodology of econometrics.
CO 2: Discuss the Leontief input output models and explain the optimization problems in Economics.
CO 3: Explain the optimization problems with equality constraints and discuss various production functions like Cobb-Douglas production function and CES production function.
CO 4: Discuss the Domar growth model, Solow growth model and Cobweb model.
CO 5: Explain the meaning of Multi collinearity, Heteroscedasticity, Autocorrelation and discuss various dynamic econometric models.
CO 6: Describe the Simultaneous equation models and Approaches to econometric forecasting..

Unit-I: Basic economic concepts: Demand, revenue, average revenue, marginal revenue, elasticity of demand, cost function, average cost, marginal cost. Equilibrium analysis: Partial market equilibrium-linear and nonlinear model, general market equilibrium, equilibrium in national income analysis. Leontief input output models. Optimization problems in economics, Optimization problems with more than one choice variable: multi product firm, price discrimination. 20 hours

Unit-II: Optimization problems with equality constraints: utility maximization and consumer demand, homogeneous functions, Cobb-Duglas production function, least cost combination of inputs, elasticity of substitution, CES production function. Dynamic analysis: Domar growth model, Solow growth model, Cobweb model. 20 hours

Unit-III: Meaning and methodology of econometrics, regression function, multiple regression model, assumptions, OLS and ML estimation, hypothesis testing, confidence interval and prediction. Multicollinearity, Heteroscedasticity, Autocorrelation: their nature, consequences, detection, remedial measures and estimation in the presence of them. Dynamic econometric models: Auto regressive and distributed lag- models, estimation of distributed lag- models, Koyck approach to distributed lag-models, adaptive expectation model, stock adjustment or partial adjustment model, estimation of auto regressive models, method of instrumental variables, detecting autocorrelation in auto regressive models: Durbin- h test, polynomial distributed lag model. 25 hours

Unit-IV: Simultaneous equation models: examples, inconsistency of OLS estimators, identification problem, rules for identification, method of indirect least squares, method of two stage least squares . Time series econometrics: Some basic concepts, stochastic processes, unit root stochastic processes, trend stationary and difference stationary stochastic processes, integrated stochastic processes, tests of stationarity, unit root test, transforming non-stationary time series, cointegration. Approaches to economic forecasting, AR, MA, ARMA and ARIMA modeling of time series data, the Box- Jenkins methodology. 25 hours

Text Books

1. **Alpha C Chiang (1984).** Fundamental Methods of Mathematical Economics(Third edition), McGraw –Hill, New York.
2. **Damodar N Gujarati (2007).** Basic Econometrics(Fourth Edition), McGraw-Hill, NewYork.

References

1. **Johnston, J (1984).** Econometric Methods(Third edition), McGraw–Hill, New York.
2. **Koutsoyiannis,A (1973).** Theory of Econometrics, Harper & Row, New York.
3. **Maddala,G.S. (2001).** Introduction to Econometrics(Third edition), John Wiley & Sons,New York.
4. **Taro Yamane (1968).** Mathematics for Economists an elementary survey(secondedition), Prentice-Hall, India.

E14: DEMOGRAPHIC TECHNIQUES (Credits: 4, Hours per week: 5)

Course Outcomes: After successful completion of the course, students will be able to:

CO 1: Describe the sources of demographic Statistics and explain basic demographic measures.
CO 2: Construct a life table.
CO 3: Explain the measures of fertility.
CO 4: Find the point estimates and population projections based on mortality, fertility and migration basis.
CO 5: Discuss the ageing of the population.
CO 6: Estimate the demographic measures from incomplete data.

Unit-I: Sources of demographic Statistics, Basic demographic measures: Ratios, Proportions and percentages, Population Pyramids, Sex ratio Crude rates, Specific rates, Labour force participation rates, Density of population, Probability of dying.

20 hours

Unit-II: Life tables: Construction of a life table, Graphs of lx, qx, dx, Functions Lx, Tx and Ex. Abridged life tables Mortality: Rates and Ratios, Infant mortality, Maternal mortality, Expected number of deaths, Direct and Indirect Standardization, Compound analysis, Morbidity.

25 hours

Unit-III: Fertility: Measures of Fertility, Reproductivity formulae, Rates of natural increase, Fertility Schedules, Differential fertility, Stable populations, Calculation of the age distribution of a stable population, Model Stable Populations. 25 hours

Unit-IV: Population estimates, Population Projections: Component method, Mortality basis for projections, Fertility basis for projections, Migration basis for projections, Ageing of the population, Estimation of demographic measures from incomplete data. 20 hours

Text Books

1. **Pollard, A.H. Yusuf, F. and Pollard, G.**N (1990). Demographic Techniques, Pergamon Press, Chapters 1-8, 12.

References

1. **Keyfitz, N. (1977)**. Applied Mathematical Demography A Wiley-Interscience Publication
2. **Keyfilz, N. (1968).** Introduction to the Mathematic of Population Ready, Mass: Addition-Wesley.
3. **Keyfilz, N. and Caswell, H. (2005).** Applied Mathematical Demography, Third edition, Springer.

E15: STOCHASTIC FINANCE (Credits: 4, Hours per week: 5)

Course Outcomes: After successful completion of the course, students will be able to:

CO 1.Expain basic concepts of financial markets and market lines.
CO 2. Learn the usage of Statistical models in modeling Financial data.
CO 3. Interpret and apply the black Scholes theorem and its properties.
CO 4. Describe the pricing of European and American options by Monte-Carlo and finite difference methods.
CO 5. Discuss on the modeling security market and price process models.
CO 6. Describe the special features of the financial time series, their models and its estimation.

Unit-I: Basic concepts of financial markets. Forward contracts, futures contracts, options-call and put options, European option and American options. Hedgers, speculators, arbitrageurs. Interest rates, compounding, present value analysis, risk free interest rates. Returns, gross returns and log returns. Portfolio theory – trading off expected return and risk, one risky asset and one risk free asset. Two risky assets, estimated expected return. Optimal mix of portfolio CAPM, capital market line, betas and security market line.

25 hours

Unit-II: Options, pricing via arbitrage, law of one price. Risk neutral valuation. Binomial model-single and multiperiod binomial model, martingale measure. Modelling returns: lognormal model, random walk model, geometric Brownian motion process. Ito lemma (without proof). Arbitrage theorem. The Black-Scholes formula. Properties of the Black-Scholes option cost, the delta hedging arbitrage strategy. Some derivatives, their interpretations and applications.

25 hours

Unit-III: Volatility and estimating the volatility parameter. Implied volatility. Pricing American options. Pricing of a European option using Monte-Carlo and pricing an American option using finite difference methods. Call options on dividend-paying securities. Pricing American put options, Modeling the prices by adding jumps to geometric Brownian motion. Valuing investments by expected utility. Modeling security market: Self-financing portfolio and no arbitrage, price process models, division rule, product rule

20 hours

Unit-IV: Financial Time Series – Special features of financial series, Linear time series models: AR(1), AR(p), ARMA(p,q) processes, the first and second order moments, estimation and forecasting methods. Models for Conditional heteroscedasticity: ARCH(1), ARCH(p), GARCH(p,q) models and their estimation. Comparison of ARMA and GARCH processes.

20 hours

Text Books/References

1. **Sheldon M. Ross (2003).** "An elementary introduction to Mathematical Finance",
2. **David Ruppert (2004).** "Statistics and Finance an Introduction" – Springer International Eddition.
3. **Masaaki Kijima (2003).** "Stochastic process with applications to finance", Chapman Hall.
4. **Ruey S. Tsay (2005).** "Analysis of Time Series III ed", John Wiley & Sons**Hull, J.C. (2008).** Options, Futures and other derivatives. Pearson Education India, New Delhi.
5. **Hull, J.C. (2008).** Options, Futures and other derivatives. Pearson Education India, New Delhi.
6. **Gourieroux, C. and Jasiak, J. (2005).** Financial Econometrics. New Age International (P) Ltd., New Delhi.
7. **Cuthbertson, K. and Nitzsche, D. (2001).** Financial Engineering - Derivatives and Risk Management. John Wiley & Sons, New York.
8. **Brockwell P.J. and Davis R.A. (2006).** Time Series: Theory and Methods, Springer – Verlag.
9. **Robert H Shumway and Davis S Stoffer(2016).** Time series analysis and its applications with R examples. Springer.

E16: LONGITUDINAL DATA ANALYSIS (Credits: 4, Hours per week: 5)

Course Outcomes: After successful completion of the course, students will be able to:

CO 1: Describe the basic concepts of Linear Model in longitudinal data analysis
CO 2: Analyze numerical methods to solve the problems in Linear Model
CO 3: Explain basic concepts of Generalized Linear Model
CO 4: Illustrate and study on missing data mechanism in longitudinal data analysis
CO5: Describe Multivariate and Time-dependent covariates in longitudinal data analysis

Unit-I: General Linear Model for Longitudinal Data.ML and REML estimation, EM algorithm: General linear mixed-effects model, Inference for ; the random effects, BLUPs, Empirical Bayes , Bayes, Shrinkage Model building and diagnostic, Relaxing parametric assumptions: generalized additive mixed model.

25 hours

Unit-II. Generalized Linear Model for Longitudinal Data: Marginal models, for binary, ordinal, and count data: Random effects models for binary ordinal and count data: Transition models: Likelihood-based models for categorical data; GEE; Models for mixed discrete and continuous responses. 25 hours

Unit-III. Dropouts and missing data: Classification missing data mechanism; Intermittent missing Values and dropouts; Weighted estimating equations; Modelling the dropout process (Selection and pattern mixture models). 20 hours

Unit-IV. Time-dependent covariates and special topics: Dangers of time-dependent covariates: Lagged covariates; Marginal Structural models; Joint models for longitudinal and survival data; Multivariate longitudinal data; Design of randomized and observational longitudinal studies. 20 hours

Text Books

1. **Diggle, P.J., Heagerty, P., Liang, K.Y and Zeger. S.L (2003).** Analysis of Longitudinal Data, 2nd Edn. Oxford University press, New York.
2. **Fitzmaurice, G.M., Laird, N.M and Ware, J.H.(2004).**Applied Longitudinal Analysis, John Wiley & Sons, New York.

References

1. **Crowder, M.J. and Hand, D.J. (1990).** Analysis of Repeated Measures. Chapman and Hall/CRC Press, London .
2. **Davidian,M. and Giltinan, D.M. (1995**). Nonlinear Models for Repeated Measurement Data. Chapman and Hall/CRC Press, London.
3. **Hand,D and Crowder, M. (1996).** Practical Longitudinal Data Analysis. Chapman and Hall/CRC Press, New York. Lindsey, J.K. (1993) Models for Repeated Measurements. Oxford University Press, New York.
4. **Little, R.J.A, and Rubin, O.B. (2002).** Statistical Analysis with Missing Data, 2nd edition, Wiley, New York.

5. **McCullagh,P. and Nelder.J.A (1989).** Generalized Linear Models. 2nd edition, Chapman and Hall/CRC Press, London.
6. **Weiss, R.E. (2005).** Modeling Longitudinal Data. Springer, New York.

E17: DATA MINING TECHNIQUES (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Apply classification techniques and concept of decision trees.
CO 2: Discuss clustering techniques in statistical and data mining viewpoints.
CO 3: Explain and apply unsupervised and unsupervised learning and data reduction techniques.
CO 4: Explain and apply artificial neural networks and extensions of regression models.
CO 5: Discuss data warehousing and online analytical data processing.
CO6: Explain and apply the techniques of association rules and prediction.

Unit-1: Review of classification methods from multivariate analysis; classification and decision trees. Clustering methods from both statistical and data mining viewpoints; vector quantization.

20 hours

Unit-II: Unsupervised learning from univariate and multivariate data; Dimension reduction and feature selection. Supervised learning from moderate to high dimensional input spaces.

25 hours

Unit-III: Artificial neural networks and extensions of regression models, regression trees. Introduction to databases, including simple relational databases.

20 hours

Unit-IV: Data warehouses and introduction to online analytical data processing. Association rules and prediction; data attributes, applications to electronic commerce.

25 hours

Text books / References

1. **Berson, A. and Smith, S.J. (1997).** Data Warehousing, Data Mining, and OLAP. (McGraw-Hill.)
2. **Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984).** Classification and Regression Trees. (Wadsworth and Brooks/Cole).
3. **Han, J. and Kamber.M. (2000).** Data Mining; Concepts and Techniques. (MorganKaufmann.)
4. **Mitchell, T.M. (1997).** Machine Learning. (McGraw-Hill.)
5. **Ripley, B.D. (1996).** Pattern Recognition and Neural Networks. (CambridgeUniversity Press).

E 18: STATISTICAL MACHINE LEARNING –I (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: To explain the concept of supervised learning and its connections to Statistics
CO 2: To use the linear methods of classification, linear regression, logistic regression, piecewise regression and lasso regression in machine learning.
CO 3: To understand kernel smothers and related concepts for machine learning.
CO 4: To learn various model assessment techniques and inferential procedures.

Unit-I: Introduction and overview of supervised learning: Variable Types and Terminology, Least Squares and Nearest Neighbours, Local Methods in High Dimension, Supervised Learning and Function Approximation, A Statistical Model for the Joint Distribution of input and output vectors, Function Approximation, Structured Regression Models, Classes of Restricted Estimators: Roughness Penalty and Bayesian Methods, Kernel Methods and Local Regression, Basis Functions and Dictionary Methods. Model Selection and the Bias–Variance Tradeoff. Linear Methods for Regression: Least squares, Subset selection, Shrinkage Methods, Methods using derived input directions, Multiple outcome shrinkage and selection, Lasso and related path algorithms.

25 hours

Unit-II: Linear methods for classification: Linear methods for classification using linear regression of an indicator matrix, linear discriminant analysis, logistic regression and separating hyper planes. Basis expansions and regularizations: Piecewise polynomials and splines, Filtering and feature extraction, smoothing splines, Automatic Selection of the Smoothing Parameters, Nonparametric Logistic Regression, Multidimensional Splines, Regularization and Reproducing Kernel Hilbert Spaces, Wavelet smoothing.

20 hours

Unit- III: Kernel smoothing: One-Dimensional Kernel Smoothers, Selecting the Width of the Kernel, Structured Local Regression Models in Rp, Local Likelihood and Other Models, Kernel Density Estimation and Classification: Kernel Density Estimation, Kernel Density classification and the Naïve Bayes classifier. Radial Basis Functions and Kernels, Mixture Models for Density Estimation and Classification.

20 hours

Unit-IV: Model assessment, inference and averaging: Bias, Variance and Model Complexity, The Bias–Variance Decomposition, Optimism of the Training Error Rate, Estimates of In-Sample Prediction Error, The Effective Number of Parameters, The Bayesian Approach and BIC, Minimum Description Length, Vapnik–Chervonenkis Dimension, Cross-Validation, Bootstrap Methods, Conditional or Expected Test Error, introducing Model Inference and averaging: Local regression in IR, The Bootstrap and Maximum Likelihood Methods, Maximum Likelihood Inference, Bootstrap versus Maximum Likelihood, Bayesian Methods, Relationship Between the Bootstrap And Bayesian Inference, The EM Algorithm, MCMC for Sampling from the Posterior, Bagging, Model Averaging and Stacking, Stochastic Search: Bumping.

25 hours

Text Books/ References

1. **Hastie, T., Tshibirai, R. and Friedman, J. (2017) .** The Elements of Statistical Learning : Data Mining, Inference and Prediction, 2nd edition. Springer, New York.
2. **James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013) .** An Introduction to Statistical Learning with Applications in R. Springer, New York.
3. Introduction to Machine Learning The Wikipedia Guide.
4. **Vinoth, B, Rajarathian, A. and Manju Bargavi, S.K. (2016)** Nonlinear Regression and Artificial Neural Network Based Model for Forecasting Paddy (Oryza Sativa) Production in Tamil Nadu. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA)e-ISSN: 2394-0050, P-ISSN: 2394*-0042.*Volume 3, Issue 3. (May. - Jun. 2016), pp. 01-06.*

E19: STATISTICAL MACHINE LEARNING II (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: To explain the concept of unsupervised learning and its connections to Statistics

CO 2: To befamilarised with Random Forest techniques in unsupervised learning

CO 3: To beacquainted with the advanced unsupervised learning techniques like ensemble learning, undirected graphical method and to deal with high dimensional problems.

Unit-I: Unsupervised Learning: Introduction to unsupervised learning, Association Rules like the Market Basket Analysis, The a priori Algorithm, Unsupervised as Supervised Learning, Generalized Association Rules, Choice of Supervised Learning Method; Cluster Analysis: Proximity Matrices, Clustering algorithms, Combinatorial algorithms, K-means, Gaussian mixtures as Soft K-means clustering, Vector quantization, K-medoids, Hierarchical clustering. Self-Organizing Maps; Principal Components, Curves and Surfaces, Nonnegative Matrix Factorization: Archetypal Analysis. Independent Component Analysis(ICA) and Exploratory Projection Pursuit, Exploratory Projection Pursuit, Multidimensional Scaling, Nonlinear Dimension Reduction and Local Multidimensional Scaling, The Google Page Rank Algorithm.

25 hours

Unit- II: Random forests: Definition of Random Forests, Details of Random Forests: Out of bag samples, variable importance, proximity plots, Random Forests and Over fitting.

Analysis of Random Forests: Variance and the De- Correlation Effect, Bias, Adaptive Nearest Neighbours. 20 hours

Unit -III: Ensemble Learning and undirected graph models: Ensemble Learning, Boosting and Regularization Paths: Penalized Regression, The “Bet on Sparsity” Principle, Regularization Paths, Over-fitting and Margins. Learning Ensembles: Learning a Good Ensemble, Rule Ensembles. Undirected Graphical Models: Markov Graphs and Their Properties. Undirected Graphical Models for Continuous Variables: Estimation of the Parameters when the Graph Structure is Known, Estimation of the Graph Structure. Undirected Graphical Models for Discrete Variables: Estimation of the Parameters when the Graph Structure is Known, Hidden Nodes, Estimation of the Graph Structure, Restricted Boltzmann Machines. 20 hours

Unit -IV: High dimensional problem *p* (p is Much Bigger than N): Diagonal Linear Discriminant Analysis and Nearest Shrunken Centroids, Linear Classifiers with Quadratic Regularization: Regularized Discriminant Analysis, Logistic Regression with Quadratic Regularization, The Support Vector Classifier, Feature Selection, Computational Shortcuts When p . Linear Classifiers with $L1$ Regularization,The Fused Lasso for Functional Data. Classification When Features are Unavailable: Classification and Other Models Using Inner-Product Kernels and Pairwise Distances. High-Dimensional Regression: Supervised Principal Components, Connection to Latent-Variable Modelling, Relationship with Partial Least Squares, Pre-Conditioning for Feature Selection: Feature Assessment and the Multiple-Testing Problem: The False Discovery Rate(FDR), Asymmetric Cut points and the (significance analysis of microarrays) SAM Procedure, A Bayesian Interpretation of the FDR.

25 hours

Text Books/ References

1. **Hastie, T., Tshibirai, R. and Friedman, J. (2017)** The Elements of Statistical Learning : Data Mining, Inference and Prediction, 2nd edition. Springer, New York.
2. **James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013)** An Introduction to Statistical Learning with Applications in R. Springer, New York.
3. **James, G., Witten,D., Tibshirani, R. and Hastie, T.** Neural Networks and Deep Learning: A Textbook.
4. Introduction to Machine Learning The Wikipedia Guide.
5. **Vinoth, B, Rajarathian, A. and Manju Bargavi, S.K. (2016)** Nonlinear Regression and Artificial Neural Network Based Model for Forecasting Paddy (Oryza Sativa) Production in Tamil Nadu. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA)e-ISSN: 2394- 0050, P-ISSN: 2394*-0042.*Volume 3, Issue 3. (May. - Jun. 2016), pp. 01-06.*

E20: ADVANCED STATISTICAL MACHINE LEARNING TECHNIQUES

(Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: To familiarize with the use of generalized additive model, trees and related methods.
CO 2: To be well equipped with the Neural Networks and deep learning techniques.
CO 3: To be acquainted with the prototype methods like K-means clustering and nearest neighbours.

Unit-I: Additive models, Trees and related methods: Generalized Additive Models: Fitting Additive Models. Tree-Based Methods: Regression Trees, Classification Trees. The Patient Rule Induction Method(PRIM):Bump Hunting. Multivariate Adaptive Regression Splines (MARS), Hierarchical Mixtures of Experts, Missing Data, Boosting and Additive Trees: Boosting Methods, Boosting Fits an Additive Model, Forward Stagewise Additive Modelling, Exponential Loss and Ada Boost, Loss Functions and Robustness, "Off-the-Shelf" Procedures for Data Mining, Boosting Trees. Numerical Optimization via Gradient Boosting: Steepest Descent, Gradient Boosting, Implementations of Gradient Boosting. Right-Sized Trees for Boosting, Regularization: Shrinkage, Sub sampling. Interpretation: Relative Importance of Predictor Variables, Partial Dependence plots.

25 hours

Unit-II: Neural networks and deep learning: Projection Pursuit Regression, Neural Networks, Fitting Neural Networks. Some Issues in Training Neural Networks: Starting Values, Over fitting, Scaling of the Inputs, Number of Hidden Units and Layers, Multiple Minima. Bayesian Neural Nets and the NIPS 2003 Challenge: Bayes, Boosting and Bagging, Performance Comparisons. Artificial Neural networks, Non-linear regression and deep learning. 20 hours

Unit-III: Support vector machines & flexible discriminants: The Support Vector Classifier: Computing the Support Vector Classifier, Support Vector Machines and Kernels: Computing the SVM for Classification , The SVM as a Penalization Method, Function Estimation and Reproducing Kernels, SVMs and the Curse of Dimensionality. A Path Algorithm for the SVM Classifier, Support Vector Machines for Regression, Regression and Kernels, Discussion. Generalizing Linear Discriminant Analysis, Flexible Discriminant Analysis, Computing the FDA Estimates, Penalized Discriminant Analysis , Mixture Discriminant Analysis.

25 hours

Unit-IV: Prototype methods and nearest neighbours: Prototype Methods:- K-means Clustering, Learning Vector Quantization, Gaussian Mixtures; k-Nearest-Neighbour Classifiers , Examples, Invariant Metrics and Tangent Distance; Adaptive Nearest-Neighbour Methods, Global Dimension Reduction for Nearest- Neighbours, Computational Considerations.

20 hours

Text Books/ References

1. **Hastie, T., Tshibirai, R. and Friedman, J. (2017)** The Elements of Statistical Learning : Data Mining, Inference and Prediction, 2nd edition. Springer, New York.
2. **James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013)** An Introduction to Statistical Learning with Applications in R. Springer, New York.
3. Introduction to Machine Learning The Wikipedia Guide.
4. **Vinoth, B, Rajarathian, A. and Manju Bargavi, S.K. (2016)** Nonlinear Regression and Artificial Neural Network Based Model for Forecasting Paddy (Oryza Sativa) Production in Tamil Nadu. *IOSR Journal of Mobile Computing & Application (IOSR-JMCA)e-ISSN: 2394- 0050, P-ISSN: 2394*-0042.*Volume 3, Issue 3. (May. - Jun. 2016), pp. 01-06.*

E21: NON-PARAMETRIC STATISTICAL METHODS TECHNIQUES

(Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Understand the concept of random and non parametric tests for randomness.
CO 2: Apply various nonparametric statistical methods for testing normality .
CO3: Use the nonparametric alternative of one way and two way analysis of variance.
CO 4: Get basic knowledge of nonparametric density estimation.

Unit-I: Measurement scales: Nominal Scale, Ordinal Scale, Interval Scale, Ra- tio Scale; Order Statistics, Quantiles, QQplot, Empirical distribution, Kaplan-Meier estimator; Measures of dependence: Cramer’s contingency coefficient, Pearson’s contingency coefficient, Pearson’s mean-square contingency coefficient, Phi coefficient.

25 hours

Unit-I: Lilliefors’s test for normal and exponential distributions, Cramer-Von Mises test for identical populations, Anderson Darling test, Shapiro-Wilk test, Jarque-Bera test, Test for randomness: Wald Wolfowitz run test. 20 hours

Unit-III: McNemar test, Cochran’s test, K samples procedures; Kruskal-Wallis test, Friedman’s test, Resampling techniques; Jackknife method for bias and variance of an estimator, Nonparametric Bootstrap method for estimating variance and distribution of an estimator. 25 hours

Unit-IV: Smoothing, Bias variance, trade off, Nonparametric density estimation; Cross validation, Histogram and Kernel density estimation, Kernel density rug plot in R.
(Practical problems Using R.) 20 hours

Text Books

1. **Gibbons, J. D. and Chakraborti, S. (2003)**. Nonparametric Statis- tical Inference, 4th edition(Sections 4.5,4.6,6.2 and 14.5). Marcel Dekker, New York.
2. **Wasserman, L. A.(2006)**. All of Nonparametric Statistics. Springer, New York(Sections 3.1, 3.2, 4.1,6.1,6.2 and 6.3).
3. **Hollander , M. , Wolfe, D. A. and Chicken, E. (2013)**. Non para- metric Statistical Methods. Wiley, New York(Section 11.6)
4. **Sprent, P. (1989)**. Applied Nonparametric Statistical Methods . Springer, Netherlands.(Section 5.5.2)
5. **Conover, W. J. (1999)**. Practical Nonparametric Statistics, 3rd Edi- tion. (Sections 2.1, 4.4, 4.6 and 6.2) Wiley.
6. **D'Agostino, R. B. and Stephans, M. A.(1986)**. Goodness-of-fit Techniques. Marcel Dekker. (Section 4.2.2)

References:

1. **Kloke, J. and McKean, J. W.(2015)**. Nonparametric Statistical Methods Using R . CRC Press, London.
2. **Siverman, B. W.(1986)**. Density Estimation for Statistics and Data Analysis . Springer, New York.
3. **Jarque, C.M and Bera, A. K(1987)**. A test for normality of ob- servations and regression residuals, *International Statistical Review* **55**, 163-172.
4. **Anderson, T. W., and Darling, D. A. (1954)**. A test of goodness of fit. *Journal of the American Statistical Association*, **49**, 765–769.

E22: STATISTICAL MODELING AND DATA MINING TECHNIQUES

(Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Acquire skills in advanced statistical modeling.
CO 2: Get acquainted with association analysis in data mining techniques.
CO 3: Learn the data pre-processing steps in data mining.
CO 4: Understand various measures of patterns, data warehousing and the concepts of online transaction and analytical processes in data mining.

Unit-I: Statistical Modeling, Steps in Statistical Modeling, Regression Analysis: transformations and weighting to correct model inadequacies, Analytical methods for selecting a transformation, The Box-Cox method, Transformation on the regressor variables, Ridge regression, Basic form of ridge regression, Robust regression; Least absolute deviation regression, Least median of squares regression. 30 hours

Unit-II: Data mining; Introduction, Data types for Data mining, Database and Data warehouse, Data mining functionalities - Concept/class description: characterization and discrimination, Association analysis, Classification and Prediction, Clustering analysis, Evolution and Deviation analysis, Data Pre-processing, Data cleaning, Data Integration and transformation, Data reduction, Discretization and concept hierarchy generation.

30 hours

Unit-III: Measures of pattern interestingness; Objective measures of pattern- sup- port and confidence, Classification of data mining systems; Classifications according to databases, Knowledge and techniques, Major issues in data mining. Unit IV Data warehouse, On-line transaction process(OLTP) and On-line analytical processing(OLAP), Distinguishing features between OLTP and OLAP.

30 hours

Text Books

1**. Montgomery, D.C., Peck, E.A. and Vining, G.G.(2007).** Introduction to Linear Regression
Analysis. Wiley, New York.

2. **Han, J., Kamber, M. and Pei, J.(2000).** Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers (Relevant sections of Chapters 1, 2 and 3).

References

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

2. **Berson, A. and Smith, S.J. (1997).** Data Warehousing, Data Mining, and OLAP. McGraw-
Hill, New York.

E23: APPLIED ALGORITHMS AND ANALYSIS OF MULTI TYPE AND BIG DATA TECHNIQUES (Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to*:

CO 1: Explain the concept of EM clustering algorithms.
CO 2: Understand the classification techniques and the concept of support vector machines.
CO 3: Get acquainted with basic concepts related to big data.
CO 4: Learn the multidimensional scaling techniques in unsupervised learning.

Unit-I: EM Algorithm: Two-Component Mixture Model, Gaussian Models, The EM Algorithm in General, EM as a Maximization–Maximization Procedure. 20 hours

Unit-II: Support Vector Machines: Maximal Margin Classifier, Support Vector Classifiers, Support Vector Machines, SVMs with More than Two Class- One- Versus-One Classification and One-Versus-All Classification.

25 hours

Unit-III: Big Data: Definition, Characteristics, Data Analytics, General Categories of Data Analytics, Structured, Unstructured and Semi Structured Data, Met data, Big Data Analytics Life Cycle.

20 hours

Unit-IV: Multi-Dimensional Scaling; Definition, Perceptual Map, Decision Frame- work for Perceptual Mapping, Non-metric versus Metric methods, Similarities Data, Preferences Data, Aggregate and Disaggregate Analysis, Decompositional and Compositional approaches, Interpreting the MDS results. (Practical problems using R).

25 hours

Text Books/References

1. **Hastie, T., Tibshirani, R. and Friedman, J. (2017)**. The Elements of Statistical Learning, Data Mining, Inference and Prediction, 2nd edition. Springer, New York (Section 8.5).
2. **James, G., Witten, D., Hastie, T. and Tibshirani, R.(2013)**. An Introduction to Statistical Learning with Applications in R. Springer, New York (Sections 9.1-9.4).
3. **Erl, T. and Khattak, W. (2016)**. Big Data Fundamentals Concepts, Drivers & Techniques. Prentice Hall. (Chapters 1 and 3) Page 57 of 62
4. **Hair, J. F., Black, W. C., Babin, B. J. and Anderson, R. E.(2009)**. Multivariate Data Analysis, 7th edition. Prentice Hall, New York (Chapter 1

E24: RISK MODELLING AND SURVIVAL ANALYSIS-I

(Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe the properties of the statistical distributions which are suitable for modelling individual and aggregate losses.

CO 2: Analyze and create solutions to problems that are of particular relevance to actuarial work using mathematical and statistical modelling techniques.

CO 3: Apply risk modeling theory to a variety of financial and insurance issues.

CO 4: Illustrate the use of Compound distributions in risk modelling.

Unit-I: Loss distributions: Simple loss distributions, The exponential distribution, The gamma distribution, The normal distribution, Other loss distributions-The lognormal distribution, The two-parameter Pareto distribution, The Burr distribution, The three-parameter Pareto distribution, The Weibull distribution, Estimation , The method of moments, Maximum likelihood estimation, The method of percentiles, Goodness-of-fit tests.

20 Hours

Unit-II: Reinsurance: Proportional & Non-proportional reinsurance, Reinsurance arrangements, Excess of loss reinsurance, The reinsurer's conditional claims distribution, Normal and lognormal distributions, Inflation, Estimation, Policy excess. 25 Hours

Unit-III: Risk models 1: General features of a product, Insurable interest, Insurable risk, Models for short-term insurance contracts-The basic model, Discussion of the simplifications in the basic model, Notation and assumptions, The collective risk model, Distribution functions and convolutions, Moments of compound distributions, The compound Poisson distribution, Sums of independent compound Poisson random variables, The compound binomial distribution, The compound negative binomial distribution. 20 Hours

Unit-IV: Risk models 2: Aggregate claim distributions under proportional and individual excess of loss reinsurance, Aggregate excess of loss reinsurance, The individual risk model, Parameter variability / uncertainty, Variability in a heterogeneous portfolio, Variability in a homogeneous portfolio, Variability in claim numbers and claim amounts and parameter uncertainty. 25 Hours

Text Books/References

1. **Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J. F. (2007).** Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems. John Wiley & Sons.
2. **Edward W Frees(2010).** Regression Modeling with Actuarial and Financial Applications. (International Series on Actuarial Science).
3. **Blacker & Yang (1983).** Actuaries in Micro insurance: Managing Risk for the Underserved ACTEX
4. **Annette J Dobson (1983).** An introduction to statistical modelling. Springer.

5. **Hossack,Ian B; Pollard, John H; Zehnwirth, Benjamin (1999).** Introductory statistics with applications in general insurance- – 2nd ed. – Cambridge University Press.
6. **Klugman, Stuart A; Panjer, Harry H; Willmot, Gordon E; Venter, Gary G (1998).** Loss models: from data to decisions. .- John Wiley & Sons

E25: RISK MODELLING AND SURVIVAL ANALYSIS-II

(Credits: 4, Hours per week: 5)

Course Outcomes: *After successful completion of the course, students will be able to:*

CO 1: Describe the various ways in which lifetime data might be censored.

CO 2: Explain the concepts of ruin theory for a risk model

CO 3: Explain the main branches of machine learning and describe examples of the types of problems typically addressed by machine learning.

Unit-I: Survival models- A simple model of survival, Future lifetime, Probabilities of death and survival, The force of mortality, Survival probabilities, The probability density function of T_x, Life table functions, Initial and central rates of mortality, Expected future lifetime- Complete and curtate expectation of life, Simple parametric survival models, The Gompertz and Makeham laws of mortality. 20 hours

Unit-II: Estimating the lifetime distribution function, Censoring mechanisms, Right censoring, Left censoring, Interval censoring, Random censoring, Type I censoring, Type II censoring, Informative and non-informative censoring, The Kaplan-Meier (product-limit) model- Assumptions and notation, Extending the force of mortality to discrete distributions, Calculating the Kaplan-Meier estimate of the survival function, A graphical approach, Comparing lifetime distributions, The Nelson-Aalen model, Relationship between the Kaplan-Meier and Nelson-Aalen estimates, Parametric estimation of the survival function.
25 hours

Module –III :Ruin Theory: Introduction - Basis concepts and notation - The surplus process – Equations. The probability of ruin in discrete and continues time –The Poisson and compound Poisson Processes, Probability of ruin in the short term - premium security loadings. Reinsurance and ruin – Introduction – Proportional reinsurance – Excess of loss reinsurance – Examples. 20 hours

Unit III: Machine Learning - An overview of machine learning – Concepts in machine learning – The loss function – Model evaluation – Examples. Generalization error and model validation – Train validation test – Validation and over-fitting – Regularization – Branches of machine learning – Supervised Learning - Unsupervised Learning – Semi - Supervised Learning - Reinforcement Learning. Stages of analysis in machine learning – Collecting data – Types of data - Exploring and Preparing Data – Splitting the table into the training and validation sets. 25 hours

Text Books/References

1. **Blacker & Yang (1983).** Actuaries in Micro insurance: Managing Risk for the Underserved ACTEX
2. **Annette J Dobson (1983).** An introduction to statistical modelling. Springer.
3. **Hossack,Ian B; Pollard, John H; Zehnwirth, Benjamin (1999).** Introductory statistics with applications in general insurance- – 2nd ed. – Cambridge University Press.
4. **Klugman, Stuart A; Panjer, Harry H; Willmot, Gordon E; Venter, Gary G (1998).** Loss models: from data to decisions. .- John Wiley & Sons
5. **Denuit, M., Maréchal, X., Pitrebois, S., & Walhin, J. F. (2007).** Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems. John Wiley & Sons.
6. **Edward W Frees(2010).** Regression Modeling with Actuarial and Financial Applications. (International Series on Actuarial Science)
7. **Ethem Alpaydın (2014).** Introduction to Machine Learning. Massachusetts Institute of Technology.

MODEL QUESTION PAPER

I/II/III/IV SEMESTER M. Sc. DEGREE EXAMINATION (FCCBCSS) Month & Year

Branch: Statistics

Course Code & Course Name

Time: 3 Hours **Maximum Weightage: 30**

PART A

Answer any **four (**2 weightages each)

1. .
2. .
3. .
4. .
5. .
6. .
7. .

(2 x 4=8 weightages)

PART B

Answer any **four** (3 weightages each)

8. .
9. .
10. .
11. .
12. .
13. .
14. .

(3x 4=12 weightages)

PART C

Answer any **two** (5 weightages each)

15. .
16. .
17. .
18. .

(5x2=10 weightages)

M. Sc. Statistics Programme under FCCBCSS - 2022

Audit courses:

In addition to the core and elective courses of the programme there will be two Audit Courses (Ability Enhancement Course & Professional Competency Course) with 4 credits each. These have to be done one each in the first two semesters. These courses are mandatory for all programmes but their credits will not be counted for evaluating the overall SGPA & CGPA. The Department/College shall conduct examination for these courses and have to intimate /upload the results of the same to the University on the stipulated date during the Third Semester. Students have to obtain only minimum pass requirements in the Audit Courses. The details of Audit courses are given below.

SEMESTER I

MST1A01: Ability Enhancement Course (AEC) (Credits :4)

The objective of this course is to enhance the ability and skill of students in the core and elective areas of statistics, through hands on experience, internship, industrial visits, case study, community linkage, book/research paper review, scientific word processing etc.

The faculty members in the department collectively or a particular faculty member shall be in charge of this course for students of the semester, which shall be decided by the Department council. The following are the requirements in this course:-

1. Short term internships at research institutions/R&D centre/Industry.
2. Seminar presentation on a topic in statistics or related fields that is not normally covered in the in the syllabi of the programme.
3. Case study and analysis on any relevant issues in the nearby society
4. Publication of articles in statistical magazines/journals
5. Interaction with Statistical Organizations/ Industries/ Research Institutions.
6. Any community linking programme relevant to the area of study

7. Book/paper review and summary.
8. English communication skills and technical writing with LATEX.
9. Survey methodology and Data collection- sampling frames and coverage error, non-response.
10. Developing a questionnaire, collect survey data pertaining to a research problem (such as gender discrimination in private vs government sector, unemployment rates, removal of subsidy, impact on service class). Formats and presentation of reports.

After conducting the AEC, the evaluation/examination should be done either common for all students of the semester or individually depending upon the AEC conducted. Evaluation/examination on AEC must contain the following components: MCQ type written examination, Report on study/investigation, Presentation, Viva voce etc. as decided by the Department council. Evaluation/examination must be conducted by 30 weightage pattern, as in the theory courses and the GPA and overall grade of the AEC should be determined.

SEMESTER II

MST2A02: Professional Competency Course (PCC) (Credits:4)

The objective of this course is to get professional competency and exposure in the core areas of statistics. It particularly aims to improve the skill level of students, especially for using software useful in their respective field of study, both related to the core and elective subject area. Also it is a platform for the student community to undertake socially committed statistical investigations and thereby developing a method of learning process by doing through the involvement with society.

The faculty members in the department collectively or a particular faculty member shall be in charge of this course for students of the semester, which shall be decided by the Department council. The following are the requirements in this course:-

1. Working knowledge on different statistical software/utilities like SPSS (or GNU PSPP), R, Python. (Introduction of the software- Use of the software as a calculator, as a graphing (plotting) utility, for matrix operations and for problems on probability distributions)
2. Use of Internet and other technologies - Internet and www, applications, internet protocols.
3. E-commerce and financial statistics- Electronic fund transfer, payment portal, e-commerce security.
4. Mobile commerce, Bluetooth and Wi-Fi
5. Introduction to Data Science and Big-data issues.
6. Trend Analysis (elementary time series analysis) and Index numbers
7. Official Statistics: An outline of present official statistical systems in India, Methods of collection of official statistics, their reliability and limitations, Role of MoSPI, CSO, NSSO and NSC.
8. Monte Carlo methods: Brief look at some popular approaches- simulating a coin toss, a die roll and a card shuffle.
9. CDF inversion method- simulation of standard distributions

10. Monte Carlo Integration- Basic ideas of importance sampling.

After conducting the PCC, the evaluation/examination should be done either common for all students of the semester or individually depending upon the PCC conducted. Evaluation/examination on PCC must contain the following components: MCQ type written examination, Report on study/investigation, Presentation, Viva voce etc. as decided by the Department council. Evaluation/examination must be conducted by 30 weightage pattern, as in the theory courses and the GPA and overall grade of the PCC should be determined.

....