VIMALA COLLEGE (AUTONOMOUS)

THRISSUR

(Affiliated to University of Calicut)

POSTGRADUATE PROGRAMME IN
STATISTICS
UNDER CHOICE BASED CREDIT AND SEMESTER SYSTEM

SCHEME AND SYLLABUS
2020 ADMISSION ONWARDS

M. Sc. Statistics Programme under CBCSS

Programme Structure &Syllabi

(With effect from 2020 Admission 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	Distribution Theory	4	5	1:4
MST1C04	Core	Probability Theory	4	5	1:4
MST1L01	Core (Practical)	Statistical Computing – I	4	5	1:4
MST1A01	Audit	Ability Enhancement Course	4 Credits (Not included in SGPA &CGPA)		
	II SEMESTER (Total Credits: 20)				
MST2C05	Core	Design and Analysis of Experiments	4	5	1:4
MST2C06	Core	Estimation Theory	4	5	1:4
MST2C07	Core	Sampling Theory	4	5	1:4
MST2C08	Core	Testing of Statistical Hypotheses	4	5	1:4
MST2L02	Core (Practical)	Statistical Computing - II	4	5	1:4
MST2A02	Audit	Professional Competency Course	4 Credits (Not included in SGPA & CGPA)		
	III SEMESTER (Total Credits:20)				
MST3C09	Core	Applied Regression Analysis	4	5	1:4
MST3C10	Core	Stochastic Processes	4	5	1:4
MST3E01	Elective	Operations Research - I	4	5	1:4
MST3E10	Elective	Statistical Quality Control	4	5	1:4
MST3L03	Core (Practical)	Statistical Computing - III	4	5	1:4
	IV SEMESTER (Total Credits: 20)				
MST4C11	Core	Multivariate Analysis	4	5	1:4
MST4E02	Elective	Time Series Analysis	4	5	1:4
MST4P01	Core	Project/Dissertation	5	10	1:4
MST4V01	Core (Viva Voce)	Comprehensive Viva Voce	3	---	1:4
MST4L04	Core (Practical)	Statistical Computing - IV	4	5	1:4
		Total	**80**		

Total credits: 80 (Core -60, Elective-12, Project /Dissertation-5, Comprehensive Viva-voce -3)

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

Internal Evaluation

(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

External Evaluation

The semester-end examinations in theory courses shall be conducted with question papers set by external experts. The question paper pattern is as follows:

a) 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**

b) 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

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**

(C) Comprehensive Viva-voce

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.

M.Sc. (STATISTICS) CBCSS: ACADEMIC PROGRAM.
(With effect from 2021 ADMISSION ONWARDS)

Objectives of the Program

The present program is intended to provide a platform for talented students to undergo higher studies in the subject as well as to train them to suit for the needs of the society. Apart from teaching core Statistics subjects the students can choose electives depending upon their interests, under the choice-based credit system. The students are also trained to handle real life problems through the practical classes and project work. As a part of the course the students are also exposed to various statistical softwares such as SPSS, MATLAB and R.

Program Outcomes (PO's):

On successful completion of M.Sc. Statistics program, the students will be able to:

P.O.1: Use probability and statistics in solving real life problems;

P.O.2: Acquire the knowledge on modern statistical techniques relevant for today's scientific community;

P.O.3: Acquire skills and competencies in statistical computing methods and develop algorithms and computer programs for analyzing complex data sets;

P.O.4: Handle real life problems using suitable statistical tools as well as they will be able to work in any industry which deals with data;

P.O.5: Become professionally inclined statistics teachers/statistician/data scientist who have sound knowledge of the subject matter and specialized in knowledge discovery through statistical methods;

P.O.6: Understand basic theoretical and applied principles of statistics with adequate preparation to pursue a Doctoral (Ph.D.) degree or enter job force as an applied statistician;

P.O.7: Communicate key statistical concepts to non-statisticians;

P.O.8: Gain proficiency in using statistical software/utility for data analysis. Also, gain proficiency in R and Python programming;

P.O.9: Continue to acquire relevant knowledge and skills appropriate to professional activities and demonstrate highest standards of ethical issues in statistical sciences;

P.O.10: Create awareness to become an enlightened citizen with commitment to deliver one's responsibilities within the scope of bestowed rights and privileges;

P.O.11: Create awareness on recent developments in statistical theory and practice.

SEMESTER-I

MST1C01: ANALYTICAL TOOLS FORSTATISTICS – I(4Credits)

Course Outcomes:

On completion of the course, students should be able to:
CO 1: Develop skills in generalizing the concepts in univariate calculus to multivariate setup
CO 2: Acquire the basic concepts of complex plane
CO 3: Determine derivatives and integrals in the case of functions in the complex plane
CO 4: Determine Poles and residue of complex functions.
CO 5: Find Laplace transform of a given function.
CO 6: Express a given function as a Fourier Series.

Unit-I-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.

Unit-II-Analytic functions and Complex Integration- Analytical functions, Harmonic functions, Necessary condition for a function to be analytic, Sufficient condition for a function to be analytic, Polar form of Cauchy- Riemann equation, Construction of analytic function. Complex integral, Cauchy's theorem, Cauchy's Integral formula and its generalized form, Poisson integral formula, Morera's theorem. Cauchy's inequality, Liouville's theorem, Taylor's theorem, Laurent's theorem.

Unit-III-Singularities and Calculus of Residues- Zeros of a function, singular point, different types of singularities. Residue at a pole, residue at infinity, Cauchy's residue theorem, Jordan's lemma, Integration around a unit circle. Poles on the real axis, Integration involving many valued functions.

Unit-IV-Laplace transform and Fourier Transform- Laplace transform, Inverse Laplace transform. Applications to differential equations, Infinite Fourier transform, Fourier integral theorem. Different forms of Fourier integral formula, Fourier series.

Text Books

1. **Khuri, A.T. (1993).** Advanced Calculus with Applications in Statistics. John Wiley &Sons, New York. (Chapter7).

2. **Pandey, H.D., Goyal, J.K. &Gupta K.P. (2003).** Complex Variables and Integral Transforms. Pragathi Prakashan, Meerut.

3. **Churchill, R. V. (1975).** Complex Variables and Applications. McGraw Hill, New York.

References

1. **Apostol, T.M. (1974).** Mathematical Analysis- Second Edition. Narosa Publications, New Delhi.

2. **Malik, S.C. & Arora, S. (2006).** Mathematical Analysis- Second Edition. New Age Internati International, New Delhi.

MST1C02: ANALYTICAL TOOLS FOR STATISTICS – II (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Illustrate vector space, subspaces, independence of vectors, basis and dimension, direct sum, complement and orthogonality with examples.

CO2: Examine linear independence and to construct orthogonal and orthonormal vectors.

CO3: Find rank and nullity, for analysis of matrices.

CO4: Determine eigen values and eigen vectors of a given matrix.

CO5: Establish the relation between algebraic and geometric multiplicity.

CO5: Execute the decomposition of a matrix.

CO6: Derive solution of homogeneous equations and their applications in real life situations and use of g inverse.

CO7: Classify quadratic forms.

Unit-I- Basics of Linear Algebra-Definition of vector space, sub spaces, linear dependence and independence, basis and dimensions, direct sum and compliment of a subspace, quotient space, Inner product and orthogonality

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 matrix, inverse of a matrix. Rank of product of matrix, rank factorization of a matrix, rank of a sum and projections, Inverse of a partitioned matrix, Rank of real and complex matrix.

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

Unit- IV- Linear Equations, Generalized Inverses and Quadratic Forms-Homogenous system, general system, Rank Nullity Theorem (statement only), generalized inverse, properties of g-inverse, Moore-Penrose inverse, properties, computation of g-inverse, definition of quadratic forms, classification of quadratic forms, rank and signature, positive definite and non-negative definite matrices, extreme of quadratic forms, simultaneous diagonalisation of matrices.

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: DISTRIBUTION THEORY (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe different types of discrete probability distributions

CO2: Explain the properties and applications of continuous distributions

CO3: Derive probability distributions of the different functions of discrete and continuous random variables

CO4: Describe different Sampling distributions and their interrelations

CO5: 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.

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

Unit-III- Functions of random variables- Joint and marginal distributions, Conditional distributions and independence, Bivariate transformations, Covariance and Correlations, Bivariate normal distributions, Hierarchical models and Mixture distributions, Multivariate distributions, Inequalities and Identities. Order statistics.

Unit-IV- 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.

Text Books

Rohatgi, V.K. (1976). Introduction to Probability Theory and Mathematical Statistics. John Wiley & Sons, New York.

Mood,A.F., Graybill, F.A. and Bose, D.C.(1973). Introduction to the Theory of Statistics. McGraw Hill, New York.

Mukhopadhyay, P. (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 & Sons, New York.
2. **Johnson, N.L., Kotz, S. and Kemp, A.W. (1992).** Univariate Discrete Distributions. John Wiley & Sons, New York.
3. **Kendall, M. and Stuart, A. (1977).** The Advanced Theory of Statistics- Vol. I: Distribution Theory- Fourth Edition. Charles Griffin & Company Ltd., London.

MST1C04: PROBABILITY THEORY (4Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Use algebra of sets in statistics

CO2: Describe basic concepts of Random variable from measure point of view

CO3: Explain the concept of distribution function, Characteristic function and their relationships and importance

CO4: Distinguish different types of convergence.

CO5: Acquire knowledge in some of the very important theorems like WLLN, CLT and their applications.

Unit-I- Sets and classes of events – Sequences of sets and their limits – Fields, Sigma fields, Borel field. Random variables, Sigma fields induced by random variables, Vector random variables, limits of sequence of random variables, Probability space, General Probability space, Induced probability space, Concepts of other measures.

Unit-II-Distribution functions of random variables- Decomposition of distribution functions, Distribution function of vector random variables, Correspondence theorem, Expectation and moments, Properties of expectations, Moments and inequalities, Characteristic functions, Properties, Inversion theorem, Characteristic functions and moments, Bochner's theorem (No proof required), Independence of classes of events; Independence of random variables; Kolmogorov0-1law; Borel 0-1 law.

Unit-III-Convergence of random variables- Convergence in probability, Convergence almost surely, Convergence in distribution, Convergence in r^{th} mean – their inter-relations- examples and counter-examples. Convergence of distribution functions; Weak convergence, Helly-Bray Lemma and Helly –Bray theorem, Levy continuity theorem.

Unit- IV- Law of Large Numbers – Kolmogorov inequality, Kolmogorov three series theorem; Weak law of large numbers (both IID and Non-IID cases). Strong Law of large numbers (Law of iterated logarithm not included), 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

Text books

1. **Bhat, B.R. (1999).** Modern Probability Theory. Wiley Eastern, New Delhi.
2. **Laha, R.G. & Rohatgi, V.K. (1979).** Probability Theory. John Wiley & Sons, New York.
3. **Mukhopadhyay, P. (2018).** Mathematical Statistics. Book and Allied Publishers Ltd., Calcutta.

References

1. **Billingsley, P. (1995).** Probability and Measure, John Wiley & Sons, New York
2. **Galambos, J. (1988).** Advanced Probability Theory. Marcel Dekker, New York.

MST1L01: STATISTICAL COMPUTING - I (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Develop scientific and experimental skills.

CO2: Apply the principles of Analytical Tools for Statistics- II and Distribution Theory using real data sets.

CO3: Know the formulas to be applied for the analysis.
CO4: Write the R codes for the analysis of the given data.

CO5: To install and load the packages required to run the R codes.

CO6: Enter the data given for analysis

CO7: 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 TWO courses of the first semester.

1. MST1C02: Analytical Tools for Statistics – II
2. MST1C03: Distribution theory

Practical is to be done using R or Python Programming. At least five statistical data oriented/supported problems should be done from each course. Each student shall maintain practical Record and the same shall be submitted for verification at the time of external examination. Students are expected to acquire working knowledge of the statistical packages like EXCEL.

The Board of Examiners (BoE) shall decide the pattern of question paper and the duration of the external examination. The external examination at each centre shall be conducted and evaluated on the same day jointly by two examiners – one external and one internal, appointed at the centre ofthe examination by the University 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.

MST2C05: DESIGN AND ANALYSIS OF EXPERIMENTS (Credits: 4)

Course Outcomes:
On completion of the course, students should be able to:

CO1: Explain the Principles of planning of an experiment.

CO2: Discuss and compare different complete block designs with and without ancillary variables.
CO3: Analyze experiments with and without missing values.

CO4: Apply incomplete block designs and balanced incomplete block designs.
CO5: Explain factorial experiments, total confounding and partial confounding.

CO6: Describe Response surface design and method of steepest accent.

Unit-I- Randomization, Replication and local control, One way and two way classifications with equal and unequal number of observations per cell with and without interaction, Fixed effects and Random effects model. Model adequacy checking, CRD, RBD and Latin Square designs, Analysis of co-variance for completely randomized and randomized block designs. Analysis of experiments with missing observations.

Unit-II- 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, Lattice designs.

Unit-III- 2^n Factorial experiments. Analysis of 2^n factorial experiments. Total confounding of 2^n designs in 2^n blocks. Partial confounding in2^n blocks. Fractional factorial designs, Resolution of a design, 3^n factorial designs. Concepts of Split plot design and strip plot design

Unit-IV- Response surface designs, Orthogonality, Rotatability blocking and analysis - Method of Steepest accent, Models properties and Analysis.

Text Books

1. **Montgomery, D. C. (2001).** Design and Analysis of Experiments. John Wiley & Sons, New York.

2. **Das, M. N. and Giri, N. C. (1979).** Design and Analysis of Experiments, Second Edition. New Age International (Pvt.) Ltd, New Delhi.

3. **Hinkleman, K. and Kempthrone, O.** (1994). Design and Analysis of Experiments –
4. Vol. I. John Wiley & Sons, New York.

References

1. **Joshi, D.D. (1987).** Linear Estimation and Design of Experiments. Wiley Eastern, New Delhi.

2. **Chakrabarti, M.C. (1964).** Design of Experiments. ISI, Calcutta

MST2C06: ESTIMATION THEORY (Credits: 4)

Course Outcomes:

On completion of the course, students should be able to:
CO1: Describe the properties of estimators: unbiasedness, consistency and sufficiency.

CO2: Explain exponential family and Pitman family of distributions, with illustrations.

CO3: Describe the method of finding sufficient statistics, minimum variance unbiased estimators, consistent estimators and consistent and asymptotically normal estimators.

CO4: Relate sufficient statistic and ancillary statistic using Basu's thorem.

CO5: 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 Fiduicial 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.

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.

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.

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

Text books

1. **Kale, B.K. (2005).** A First Course in Parametric Inference- Second Edition.Narosa Publishing House, New Delhi.
2. **Casella, G.and Berger, R.L. (2002).** Statistical Inference- Second Edition. Duxbury, Australia.

References

1. **Lehmann, E.L. (1983).** Theory of Point Estimation. John Wiley & Sons, New York.
2. **Rohatgi, V.K. (1976).** An Introduction to Probability Theory and Mathematical Statistics. John Wiley & Sons, New York.

MST2C07: SAMPLING THEORY (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Distinguish between Probability and Non-Probability Sampling

CO2: Apply the sampling methods: simple random sampling, systematic sampling, stratified sampling and cluster sampling.

CO3: Estimate the population parameters for variables and attributes under the above procedures.

CO4: Estimate the population parameters concerning the study variables under auxiliary information (Ratio and regression methods)

CO5: Discuss probability proportional to size (PPS) sampling strategies.

CO6: Explain the concepts of ordered and unordered estimators and its properties.

CO7: Discuss the multi stage and multiphase sampling.

CO8: 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.

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.

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.

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.

Text books / References

1. **Cochran, W.G (1992):** Sampling Techniques, Wiley Eastern, New York.
2. **Singh, D. and F.S. Chowdhury, F.S.** (1986): Theory and Analysis of Sample Survey Design, New Age International, New Delhi.
3. **P.V.Sukhatme et.al.** (1984): Sampling Theory of Surveys with Applications. IOWA State University Press, USA.
4. **Des Raj** (1976): Sampling Theory. McGraw Hill
5. **Mukhopadhyay. P.** (1999). Theory and Methods of Survey Sampling. Prentice-Hall India, New- Delhi.

MST2C08: TESTING OF STATISTICAL HYPOTHESES (4 Credits)

Course Outcomes:

On completion of the course, students should 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.

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.

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.

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.

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. **Srivastava, M. and Srivastava, N.** (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 & Sons, New York.

2. **Lehman, E.L. (1983).** Theory of Point Estimation. John-Wiley & 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, J. P. (2005).** Testing Statistical Hypotheses- Third Edition, Springer, New York.

MST2L02: STATISTICAL COMPUTING - II (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Develop scientific and experimental skills of the students.

CO2: Apply the principles of Design of experiments, Estimation Theory, Sampling Theory, and Testing of Statistical Hypotheses using real data sets.

CO3: Know the formulas to be applied for the analysis.

CO4: Write the R codes for the analysis of the given data.

CO5: Enter the data given for the analysis.

CO7: 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 FOUR courses of the second semester.

The practical is based on the following FOUR courses of the second semester.

1. MST2C05: Design and Analysis of Experiments
2. MST2C06: Estimation Theory
3. MST2C07: Sampling Theory
4. MST2C08: Testing of Statistical Hypotheses

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
MST3C09: APPLIED REGRESSION ANALYSIS (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Illustrate the concept of linear regression model.

CO2: Estimate and test the significance of regression parameters and explain properties estimators.

CO3: Check the model adequacy of regression models using residual analysis.

CO4: Discuss polynomial, step-wise and non-parametric regression models.

CO5: Explain logistic and Poisson regression models for binary and count data and estimate their parameters.

CO6: 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, Generalized least squares; Hypothesis testing - likelihood ratio test, F-test; Confidence intervals.

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

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.

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.

Text Books

1. **Seber, A.F. and Lee, A.J. (2003).** Linear Regression Analysis. John Wiley & Sons, New York. Relevant Sections from Chapters 3, 4, 5, 6, 7, 9, 10.
2. **Montegomery, D.C., Peck, E.A. and Vining, G.G. (2001).** Introduction to Regression Analysis-Third Edition. John Wiley & Sons, New York.
3. **B. Abraham and Ledotter, J. (1983).** Statistical Methods for Forecasting. John Wiley & Sons, New York.

References

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

MST3C10: STOCHASTIC PROCESSES (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Recollect the basic concepts of random variables and conditional probabilities.

CO2: Explain Markov Chain with illustrations.

CO3: Classify the States of a Given Markov Chain.

CO4: Describe inter arrival time and waiting time distributions and their properties.

CO5: Explain generalized Poisson process and their properties.

CO6: Describe the concept and applications of renewal process.

CO7: Explain the basic characteristics of queues and the properties of Brownian motion.

Unit-I: Concept of Stochastic processes, examples, Specifications; Markov chains- Chapman Kolmogorov equations – classification of states – limiting probabilities; Gamblers ruin problem and Random Walk – Mean time spent in transient states – Branching processes (discrete time), Hidden Markov chains.

Unit-II: Exponential distribution – counting process – inter arrival time and waiting time distributions. Properties of Poisson processes – Conditional distribution of arrival times. Generalization of Poisson processes – non-homogenous Poisson process, compound Poisson process, conditional mixed Poisson process. Continuous time Markov Chains – Birth and death processes – transition probability functionlimiting probabilities.

Unit-III: Renewal processes-limit theorems and their applications. Renewal reward process. Regenerative processes, Semi-Markov process. The inspection paradox, Insurers ruin problem.

Unit-IV: Basic characteristics of queues – Markovian models – network of queues. The M/M/I, M/M/C, M/M/1/K, M/M/C/K models, Multi server queues. Brownian motion Process – hitting time – Maximum variable – variations on Brownian motion – Pricing stock options – Gaussian processes – stationary and weakly stationary processes.

Text Books

1. Ross, S.M. (2007). Introduction to Probability Models- Nineth Edition. Academic Press, New York.

2. Medhi, J. (1996). Stochastic Processes-Second Edition. 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.

MST3L03: STATISTICAL COMPUTING - III (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Develop scientific and experimental skills of the students.

CO2: Apply the principles of Design of experiments, Estimation Theory, Sampling Theory, and Testing of Statistical Hypotheses using real data sets.

CO3: Use the formulas to be applied for the analysis.
CO4: Write the R codes for the analysis of the given data.

CO5: Enter the data given for analysis.

CO6: Explain how to make conclusions and write the inference for the data analysis based on the output obtained.

Statistical Computing-III 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 THREE courses of the third semester.

1. MST3C09: Applied Regression Analysis
2. MST3E--: Elective -I
3. MST3E--: Elective -II

Practical is to be done by using R & 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 each centre shall be conducted and evaluated on the same day jointly by two examiners – one external and one internal, appointed at the centre of the examination by the University 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.

IV- SEMESTER
MST4C11: MULTIVARIATE ANALYSIS (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the development and uses of multivariate normal distribution.

CO2: Learn the various characterization properties of multivariate normal distributions

CO3: Get idea about sampling distributions of various multivariate statistics and know how the results are utilized in inference procedure.

CO4: Apply different aspects of testing of statistical hypothesis in multivariate set up.

CO5: Identify the most appropriate statistical techniques for a multivariate dataset.

CO6: Apply commonly used multivariate data analysis techniques, and interpret the results

Unit-I- Multivariate Normal Distribution – Definition and properties, conditional distribution, marginal distribution. Independence of a linear form and quadratic form, independence of two quadratic forms, distribution of quadratic form of a multivariate vector. Partial and multiple correlation coefficients, partial regression coefficients, Partial regression coefficient.

Unit-II-Estimation of mean vector and covariance vector – Maximum likelihood estimation of the mean vector and dispersion 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; canonical correlation. Wishart distribution – properties – generalized variance.

Unit-III-Testing Problems – Mahalanobis D^2 and Hotelling's T^2 Statistics, Likelihood ratio tests – Testing the equality of mean vector, equality of dispersion matrices, testing the independence of sub vectors, Sphericity test.

Unit-IV-The problem of classification – classification of one of two multivariate normal population when the parameters are known and unknown. Extension of this to several multivariate normal populations. Population principal components – Summarizing sample variation by principal components – Iterative procedure to calculate sample principal components; Factor analysis.

Text Books

1. **Anderson, T.W. (1984).** Multivariate Analysis. John Wiley & Sons, New York.
2. **Johnson, R.A. and Wichern, D.W. (2001).** Applied Multivariate Statistical Analysis- Third Edition. Prentice Hall of India, New Delhi.
3. **Rao, C.R. (2002).** Linear Statistical Inference and Its Applications- Second Edition. John Wiley & Sons, New York.

References

1. **Giri, N.C. (1996).** Multivariate Statistical Analysis. Marcel Dekker. Inc., New York.
2. **Kshirasagar, A.M. (1972).** Multivariate Analysis. Marcel Dekker. New York
3. **Rencher, A.C. (1998).** Multivariate Statistical Analysis. John Wiley & Sons, New York.
4. **Morrison, D.F. (1976).** Multivariate Statistical Methods. McGraw Hill, New York.

MST4P01: PROJECT/DISSERTATION (5 Credits)

Course Outcomes:

- On completion of the course, students should be able to:

CO1: Discuss the applications of various statistical techniques learned in the entire course in the form of project work.

CO2: Manage a real practical situation where a statistical analysis is sought.

CO3: Develop professional approach towards writing and presenting an academic report.

CO4: Get more insight about the opportunities in research/career.

In partial fulfillment 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 University, 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 (3 Credits)

Course Outcomes:

- On completion of the course, students should 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 examiners.

MST4L04: STATISTICAL COMPUTING - IV (4Credits)

<u>Course Outcomes:</u>

On completion of the course, students should be able to:

CO1: Develop scientific and experimental skills of the students and to correlate the theoretical principles with application-based studies.

CO2: Learn to apply the multivariate techniques using R or Python.

CO3: Validate results by simulation of artificial data sets using R or Python.

CO4: Learn to import and analyze multivariate data from other source of data files like spreadsheet or web page.

CO5: Prepare the complex raw data into manageable format to analyze.

CO6: Get basic knowledge about the avenues of further improvement of R packages and frontiers of ever-growing research on statistical computing.

Statistical Computing-IV 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 TWO courses of the fourth semester.

1. MST4C11: Multivariate Analysis
2. MST4E--: Elective -III

Practical is to be done by using R & 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 each centre shall be conducted and evaluated on the same day jointly by two examiners – one external and one internal, appointed at the centre of the examination by the University 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.

SYLLABI OF ELECTIVE COURSES

E01: OPERATIONS RESEARCH – I (Credits: 4)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the basic concepts of linear programming problem (LPP).

CO2: Discuss different methods of solving LPP.

CO3: Illustrate the concept of duality.

CO4: Solve transportation problem, assignment problem and parametric programming problem.

CO5: Explain Integer programming problems and the methods of solving it.

CO6: Describe basics of game theory and solve game problem as LPP.

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

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

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

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

Text Book

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

References

1. **Hadley, G. (1964).** Linear Programming. Oxford & IBH Publishing Co., New Delhi.
2. **Taha. H.A. (2007).** Operations Research -An Introduction-Eighth Edn. Pearson Printice Hall, New Jersey
3. **Hiller, F.S. And Lieberman, G.J. (1995).** Introduction to Operations Research. McGraw Hill, New York.
4. **Kanti Swarup, Gupta, P.K and Man Mohan**. (1999). Operations Research. Sultan Chand & Sons, New Delhi.

E02: TIME SERIES ANALYSIS (Credits: 4)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the basics of time series data, its auto-covariance, auto-correlation and stationarity.

CO2: Illustrate test for trend and seasonality.

CO3: Explain the smoothing methods for determining trend of the data.

CO4: Describe the properties of linear time series models.

CO5: Fit linear models for time series data sets.

CO6: Describe the maximum likelihood, Yule-Walker and least square estimation methods.

C07: Learn to validate a model using residual analysis.

CO8: Define ARCH and GARCH models and derive their properties.

CO9: 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.

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.

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 SPSS.

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

Text Books

1. **Box, G.E.Pand Jenkins, G.M. (1970).** Time Series Analysis, Forecasting and Control. Holden Day, San fransisco.
2. **Brockwell, P.J. and Davis, R.A. (1987).** Time Series: Theory and Methods. Springer Verlag. New York.
3. **Abraham, B and Ledolter, J.C. (1983).** Statistical Methods for Forecasting. John Wiley & Sons, New York.

References

1. **Anderson, T.W. (1971).** Statistical Analysis of Time Series. John Wiley & Sons, New York.
2. **Fuller, W.A. (1978).** Introduction to Statistical Time Series. John Wiley & Sons, New York..
3. **Kendall, M.G. (1978).** Time Series. Charles Griffin, London.
4. **Tanaka, K. (1996).** Time Series Analysis. Wiley Series, New York.

E03: OPERATIONS RESEARCH - II (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss the Nonlinear programming problems and methods to solve the problems.

CO2: Understand and solve quadratic programming problem.

CO3: Explain Dynamic and Geometric programming.

CO4: Discuss inventory management, deterministic and probability models.

CO5: Understand Replacement models.

CO6: 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

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.

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.

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.

Text Books

1. **Mital, K.V. and Mohan, C. (1996).** Optimization Methods in Operations Research and Systems Analysis- Third Edition. New Age International (Pvt.) Ltd., New Delhi.
2. **Sasieni, M., Yaspan, A. and Friendman, L. (1959).** Operations Research- Methods and Problems. John Wiley & Sons, New York.
3. **Taha. H.A. (2007).** Operations Research -An Introduction-Eighth Edn. Pearson Printice Hall, New Jersey.
4. **Ravindran, A., Philips, D.T. and Solberg, J.J. (2006).** Operations Research- Principles and Practice- Second Edition. John Wiley & Sons, New York.

References

1. **Sharma, J.K. (2003).** Operations Research-Theory & Applications, Macmillan India Ltd., New Delhi.

2. **Man Mohan, Kanti Swarup and Gupta (1999).** Operation Research, Sultan Chand & Sons, New Delhi.

E04: QUEUEING THEORY (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss the basic characteristic features of a queuing system and acquire skills in analyzing queuing models.

CO2: Explain the theory of Markovian queueing models, M/M/I, M/M/∞ and M/M/c queueing systems.

CO3: Define and explain basic concepts in the theory of Advanced Markovian models, $M^{[X]}$ /M/1, $M/M^{[Y]}$ /1, $M/E_k/1$, $E_k/M/1$ and Erlangian models.

CO4: Analyze a network of queues with Jackson network.

CO5: Apply models with general arrival pattern and M/G/1 queueing model.

CO6: 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.

Unit-II- Transient behavior of M/M/1 queues, transient behavior of M/M/∞. Busy period analysis for M/M/1and 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/1and $E_k/M/1$. A brief discussion of priority queues.

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

Unit-IV- Models with general arrival pattern, The M/G/1 queueing model, The Pollaczek-Khintchine formula, Departure points 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.

Text Books/References

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

E05: LIFETIME DATA ANALYSIS (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss life time distributions and important parametric models.

CO2: Explain censoring and estimation of parameters using censored data.

CO3: Estimate the survival probabilities and cumulative hazard function using product limit and Nelson-Aalen estimate respectively.

CO4: Describe inference under exponential model and discuss the comparison of distributions.

CO5: Explain important hazard models and apply Rank test, Log-rank test and Generalized Wilcox on test

CO6: Describe the concept of regression models and Cox PH model & AFT model.

CO7: 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

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

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.

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

Text Books

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

References

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

E06: ADVANCED DISTRIBUTION THEORY (4 Credits)

Course Outcomes:

On completion of the course, students should 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 marginal's, 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, Poisson-binomial, Poisson-negative binomial, Legrangian Poisson distributions, Distributions of order Poisson, negative binomial, Logarithmic series, Binomial.

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.

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.

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

References

1. **Johnson, N.L., Kotz, S. and Kemp, A.W. (1992).** Univariate Discrete Distributions- Second Edition. John Wiley & Sons, New York.
2. **Kocherlakota, S. and Kocharlakota, K. (1992).** Bivariate Discrete Distributions. Marcel-Dekker, New York.
3. **Johnson, N.L., Kotz, S. and Balakrishnan, N. (1997).** Discrete Multivariate Distributions- Second Edition. John Wiley & Sons, New York.
4. **Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000).** Continuous Multivariate Distributions- Vol. I. John Wiley & Sons, New York.
5. **Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1998).** Records. John Wiley & Sons, New York.

E07: STATISTICAL DECISION THEORY (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss the need and role of statistical decision theory to solve real decision problems

CO2: Explain various decision principles and associated loss functions.

CO3: Differentiate between frequentist and Bayesian decision theory.

CO4: Describe the prior and posterior distributions.

CO5: Handle prior selection problem

CO6: Describe the development of optimal strategies for actions in competitive situations involving two or more intelligent antagonists.

CO7: 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.

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.

Unit-III-The posterior distribution-Bayesian inference-Bayesian decision theory-empirical Bayes analysis –

Hierarchical Bayes analysis-Bayesian robustness Admissibility of Bayes rules.

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.

Text Book

1. **Berger, O, J.** (1985). Statistical Decision Theory and Bayesian Analysis- Second Edition. Springer, New York.

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 & Sons, New York.
3. **Parmigiani, G., Inoue, L.Y.T. and Lopes, H.F.** (2009). Decision Theory- Principles and Approaches, John Wiley & Sons, New York.

E08: RELIABILITY MODELING (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:
CO1: Describe the basic concepts of system reliability

CO2: Discuss the lifetime of a system based on ageing properties

CO4: Explain shock models and stress-strength models using reliability theory.

CO5: Describe Maintenance and replacement policies

CO6: 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.

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.

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.

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.

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, New York.
3. **Aven, T. and Jensen, U. (1999).** Stochastic Models in Reliability, Springer, New York.
4. **Lawless, J.F. (2003).** Statistical Models and Methods for Lifetime -Second Edition. John Wiley & Sons., New York.
5. **Nelson, W (1982).** Applied Life Data analysis. John Wiley & Sons, New York.
6. **Zacks, S. (1992).** Introduction to Reliability Analysis: Probability Models and Statistics Methods. Springer, New York.

E09: ACTUARIAL STATISTICS (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Apply the elements of interest.

CO2: Discuss regular pattern of cash flows and related topics.

CO3: Illustrate and apply individual and collective risk models for a short period.

CO4: Discuss survival distributions and derive survival functions.

CO5: Explain and apply life insurance models.

CO6: 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.

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.

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

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 monthly payments - Multiple Decrement and Multiple Life Models – Premiums and reserves.

Text books:

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.

E10: STATISTICAL QUALITY CONTROL (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Distinguish the concepts quality, quality assurance and acceptance sampling.

CO2: Explain and compare the methods of acceptance sampling for attributes.

CO3: Explain acceptance sampling by variables and continuous sampling plans.

CO4: Describe and apply the control chart for attributes.

CO5: Explain and implement control chart for variables.

CO6: 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

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.

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.

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.

Text Books

1. **Montgomory, R.C. (1985).** Introduction to Statistical Quality Control- Fourth Edition. John Wiley & Sons, New York.

2. **Mittage, H.J. and Rinne, H. (1993).** Statistical Methods for Quality Assurance. Chapman & Hall, New York. Chapters 13 and 14.

3. **Oakland, J.S. and Follorwel, R.F. (1990).** Statistical Process Control. East-West Press. Chapters 13and 14.

4. **Schilling, E.G. (1982).** Acceptance Sampling in Quality Control. Marcel Dekker, New York.

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, New York.

2. **Chin-Knei Chao (1987).** Quality Programming, John Wiley & Sons, New York.

3. **Ott, E.R. (1975).** Process Quality Control. Mc Graw Hill, New York.

4. **Wetherill, G.B. and Brown, D.W (1991).** Statistical Process Control: Theory and Practice, Chapman & Hall/CRC Press, New York.

E11: ADVANCED PROBABILITY THEORY (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the basic concepts of Probability, Mathematical Expectation and Lebesgue - Stieltjes Integrals

CO2: Explain the concept of Weak and Complete convergence of random variables.

CO3: Discuss the basic theorems based on Decomposition of normal distribution

CO4: Describe the properties and applications of Infinitely divisible distributions and Stable distributions.

CO5: 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.

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.

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

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

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, New York.
2. **Billingsley, P. (1985).** Probability and Measure- Second Edition. John Wiley & Sons, New York.
3. **Laha, R.G. and Rohatgi, V.K.** (1979). Probability Theory. John Wiley & Sons, New York.

E12: OFFICIAL STATISTICS (4 Credits)

Course Outcomes:

On completion of the course, students should 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.

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.

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.

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.

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.

E13: BIOSTATISTICS (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Discuss types of biological data and Principles of Bio Statistical design of medical studies.

CO2: Explain the concepts of survival time functions of important parametric models and compare two survival distributions using LR test and Cox's F-test.

CO3: Explain censoring and estimation of parameters using censored data.

CO4: Describe competing risk theory and estimate the probabilities of death by ML method.

CO6: 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).

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.

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.

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).

Text Books / References

1. **Biswas, S. (1995).** Applied Stochastic Processes. A Biostatistical and Population Oriented Approach. Wiley Eastern Ltd., New Delhi.

2. **Cox, D.R.** and **Oakes, D. (1984).** Analysis of Survival Data. Chapman & Hall, New York.
3. **Elandt, R.C.** and **Johnson (1975).** Probability Models and Statistical Methods in Genetics. John Wiley & Sons, New York.
4. **Ewens, W. J.** and Grant**, G.R. (2001).** Statistical methods in Bioinformatics: An Introduction. Springer, New York.
5. **Friedman, L.M., Furburg, C.** and **DeMets, D.L. (1998).** Fundamentals of Clinical Trials. Springer, New York.
6. **Gross, A. J.** and **Clark V.A. (1975).** Survival Distribution; Reliability Applications in Biomedical Sciences. John Wiley & Sons, New York.
7. **Lee, E. T. (1992).** Statistical Methods for Survival Data Analysis. John Wiley & Sons, New York.
8. **Li, C.C. (1976).** First Course of Population Genetics. Boxwood Press, California.
9. **Daniel, W.W. (2006).** Biostatistics: A Foundation for Analysis in the Health Sciences. John Wiley & Sons, New York.
10. **Fisher, L.D.** and **Belle, G.V. (1993).** Biostatistics: A Methodology for the Health Science. John Wiley & Sons, New York.
11. **Lawless, J.F. (2003).** Statistical Methods for Lifetime - Second Edition. John Wiley & Sons, New York.
12. **Chow, S.C. and Chang, M. (2006).** Adaptive Design Methods in Clinical Trials. Chapman & Hall/CRC Biostatistics Series, New York.
13. **Chang, M. (2007).** Adaptive Design Theory and Implementation Using SAS and R. Chapman & Hall/CRC Biostatistics Series, New York.
14. **Cox, D.R. and Snell, E.J. (1989).** Analysis of Binary Data- Second Edition. Chapman &Hall / CRC Press, New York.
15. **Hu, F. and Rosenberger, W.F. (2006).** The Theory of Response-Adaptive Randomization in Clinical Trials. John Wiley & Sons, New York.
16. **Rosenberger, W.F. and Lachin, J. (2002).** Randomization in Clinical Trials: Theory and Practice. John Wiley & Sons, New York.

E14: ECONOMETRIC MODELS (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Explain the meaning and methodology of econometrics.

CO2: Discuss the Leontief input output models and explain the optimization problems in Economics.

CO3: Explain the optimization problems with equality constraints and discuss various production functions like Cobb-Douglas production function and CES production function.

CO4: Discuss the Domar growth model, Solow growth model and Cobweb model.

CO5: Explain the meaning of Multi collinearity, Heteroscedasticity, Autocorrelation and discuss various dynamic econometric models.

CO6: 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.

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

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.

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

Text Books

1) Chiang, A.C. (1984). Fundamental Methods of Mathematical Economics –Third Edition. McGraw Hill, New York.

2). Gujarati, D.N. (2007). Basic Econometrics -Fourth Edition. McGraw-Hill, New York.

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**. Yamane, T. (1968).** Mathematics for Economists- An Elementary Survey- Second Edition. Prentice-Hall, India.

E15: DEMOGRAPHIC TECHNIQUES (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Describe the sources of demographic Statistics and explain basic demographic measures.

CO2: Construct a life table.

CO3: Explain the measures of fertility.

CO4: Find the point estimates and population projections based on mortality, fertility and migration basis.

CO5: Discuss the ageing of the population

CO6: 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.

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.

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.

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.

Text book

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

References

1. **Keyfitz, N. (1977)**. Applied Mathematical Demography. Wiley-Inter Science Publication.
2. **Keyfilz, N. (1968).** Introduction to the Mathematic of Population. Addition-Wesley, London.
3. **Keyfilz, N. and Caswell, H. (2005).** Applied Mathematical Demography- Third Edition. Springer, New York.

E16: STOCHASTIC FINANCE (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:
CO1. Expain basic concepts of financial markets and market lines.

CO2. Learn the usage of Statistical models in modeling financial data. CO3. Interpret and apply the black Scholes theorem and its properties.

CO4. Describe the pricing of European and American options by Monte-Carlo and finite difference methods.

CO5. Discuss on the modeling security market and price process models

CO6. 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.

Unit-II: Options, pricing via arbitrage, law of one price. Risk neutral valuation. Binomial model- single and multi period binomial model, martingale measure. Modeling 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.

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.

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.

References

1. **Ross, S.M. (2003).** An Elementary Introduction to Mathematical Finance- Second Edition. Cambridge University Press, UK.
2. **Ruppert, D. (2004).** Statistics and Finance- An Introduction. Springer, New York.
3. **Kijima, M. (2003).** Stochastic Process with Applications to Finance. Chapman Hall/CRC Press, New York.
4. **Tsay, R.S. (2005).** Analysis of Time Series – Third Edition. John Wiley & Sons, New York.
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 Manage- ment. John Wiley & Sons, New York.

E17: LONGITUDINAL DATA ANALYSIS (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:
CO1: Describe the basic concepts of Linear Model in longitudinal data analysis **CO2:** Analyze numerical methods to solve the problems in Linear Model **CO3:** Explain basic concepts of Generalized Linear Model

CO4: 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.

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.

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).

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.

Text book

1. **Diggle, P.J., Heagerty, P., Liang, K.Y and Zeger. S.L (2003).** Analysis of Longitudinal Data- Second Edition. Oxford University Press, London.

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, London.
4. **Lindsey, J.K. (1993)** Models for Repeated Measurements. Oxford University Press, London.
5. **Little, R.J.A, and Rubin, O.B. (2002).** Statistical Analysis with Missing Data- Second Edition. John Wiley & Sons, New York.
6. **McCullagh, P. and Nelder, J.A (1989).** Generalized Linear Models- Second Edition. Chapman and Hall/CRC Press, London.
7. **Weiss, R.E. (2005).** Modeling Longitudinal Data. Springer, New York.

E18: DATA MINING TECHNIQUES (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1: Apply classification techniques and concept of decision trees.

CO2: Discuss clustering techniques in statistical and data mining viewpoints.

CO3: Explain and apply unsupervised and unsupervised learning and data reduction techniques.

CO4: Explain and apply artificial neural networks and extensions of regression models.

CO5: Discuss data warehousing and online analytical data processing.

CO6: Explain and apply the techniques of association rules and prediction.

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

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

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

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

Text books / References

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

E 19: Statistical Machine Learning – I (4 Credits)

Course Outcome

On completion of the course, students should be able:

CO1: To explain the concept of supervised learning and its connections to Statistics

CO2: To use the linear methods of classification, linear regression, logistic regression, piecewise regression and lasso regression in machine learning.

CO3: To understand kernel smothers and related concepts for machine learning.

CO4: 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

Unit II: Linear methods for classification

Linear methods for classification using linear regression of an indicator matrix, linear discriminant analysis, logistic regression and separating hyperplanes. 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.

Unit III: Kernel smoothing

One-Dimensional Kernel Smoothers, Selecting the Width of the Kernel, Structured Local Regression Models in R^p, 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.

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.

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.*

E 20: Statistical machine Learning II (4 Credits)

Course Outcome

On completion of the course, students should be able:

CO1: To explain the concept of unsupervised learning and its connections to Statistics

CO2: To be familiarized with Random Forest techniques in unsupervised learning

CO3: To be acquainted 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 Apriori 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, Non-negative 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.

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 Neighbors.

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.

Unit IV: High dimensional problem $p \gg N$ (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 QuadraticRegularization, The Support Vector Classifier, Feature Selection, Computational Shortcuts When $p \gg N$. 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.

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.*

E 21: Advanced Statistical Machine Learning Techniques (4 Credits)

Course Outcomes:

On completion of the course, students should be able:

CO1: To familiarize with the use of generalized additive model, trees and related methods.

CO2: To be well equipped with the Neural Networks and deep learning techniques

CO3: 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.

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 deeplearning

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

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

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.*

Elective 22: Non-Parametric Statistical Methods (4 Credits)

Course Outcomes:
On completion of the course, students should be able to:

CO1 Understand the concept of random and non-parametric tests for random- ness.

CO2 Apply various nonparametric statistical methods for testing normality.

CO3 Use the nonparametric alternative of one way and two-way analysis of variance.

CO4 Get basic knowledge of nonparametric density estimation.

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

Unit-II 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.

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.

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

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-parametric 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 Edition. (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 observations 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.

E23: Statistical Modeling and Data Mining Techniques (4 credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1 Acquire skills in advanced statistical modeling.

CO2 Get acquainted with association analysis in data mining techniques.

CO3 Learn the data pre-processing steps in data mining.

CO4 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.

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.

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.

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, 3rd Edition, Wiley.
2. **Berson, A. and Smith, S.J. (1997)**. Data Warehousing, Data Mining, and OLAP. McGraw-Hill, New York.

E24: Applied Algorithms and Analysis of Multi type and Big Data (4 Credits)

Course Outcomes:

On completion of the course, students should be able to:

CO1 Explain the concept of EM clustering algorithms.

CO2 Understand the classification techniques and the concept of support vector machines.

CO3 Get acquainted with basic concepts related to big data.

CO4 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.

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.

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.

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.

Text Books:

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)
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 10).

MODEL QUESTION PAPER

I/II/III/IV SEMESTER M. Sc. DEGREE EXAMINATION (CBCSS) 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 CBCSS

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.

MST1A01: Ability Enhancement Course (AEC) 4 Credits

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.

MST2A02: Professional Competency Course (PCC) 4 Credits

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 patterns, as in the theory courses and the GPA and overall grade of the PCC should be determined.