CENTRAL UNIVERSITY OF TAMIL NADU, THIRUVARUR

School of Mathematics and Computer Sciences

Department of Statistics and Applied Mathematics

Ph.D. STATISTICS (COURSE WORK)

COURSE STRUCTURE (2021-2022)

(List of software to be used as applicable: R, Python and MATLAB)

* *To be decided by DRC as per student's requirement.*

** *The detail syllabus for this course is mentioned in UGC Letter No. D.O.No.F.1-1/2018 (Journal/CARE) dated December, 2019.*

List of Electives *(This list may be extended if needed)*

Sr. No.	Course code	Course title	Credits
1.	PSTAT01E	Advanced Statistical Inference	4
2.	PSTAT02E	Advanced Regression Analysis	4
3.	PSTAT03E	Nonparametric Inference	4
4.	PSTAT04E	Evolutionary Algorithms in Multi Objective Optimization	4
5.	PSTAT05E	Advanced Optimization Techniques	4
6.	PSTAT06E	Advanced Queueing Theory	4
7.	PSTAT07E	Neural Networks and Fuzzy Logic	4

Note: A student may be allowed to earn extra credits by taking courses mentioned in the list of electives of Ph.D. Statistics and Ph.D. Applied Mathematics Course work.

PSAM01 **RESEARCH METHODOLOGY** **4 Credits**

Unit I: Introduction to Research Methodology: Definition, Characteristics, Objectives, Research and Scientific method, Literature review, types of research, process of scientific research, Research design. Criteria of Good Research, Defining Research Problem and Problem Solving Technique.

(15L)

UNIT II: Solving Research Problems: concepts, constructs, definition, research questions, objectives and methodologies - Theoretical and Operational. Intellectual property rights (IPR) and Plagiarism; publications/patents/startups and its e-filing procedures, norms and policies. (15L)

Unit III: Art of writing: Research paper, Structure and component of research paper, research report, survey article, thesis - Presenting the research paper/thesis. Scientific Tools: Journal impact factor, Citation index, References and bibliography, Journal Communication, publishing a paper, writing a research proposal. (15L)

Unit IV: Technical Writing Tool: Introduction to LaTeX, PSTricks, Beamer, HTML and MathJaX**.** (15L)

References:

1. C. R. Kothari, Research Methodology (Methods and Techniques), New Age International Publishers, 2004**.**
2. R. Pratap, Getting Started with MATLAB (A Quick Introduction for Scientists and Engineers), Oxford University Press, Indian Edition, 2010**.**
3. N. J. Higham, Handbook of Writing for the Mathematical Sciences, 2nd Edition, SIAM, 1998**.**
4. D. E. Knuth, T. L. Larrabee, and P. M. Roberts, Mathematical Writing, Mathematical Association of America, Washington, D.C., 1989**.**
5. L. Lamport, LaTeX: A Document Preparation System, 2nd Edition (Addison-Wesley Series on Tools and Techniques for Computer T), 1994**.**
6. J. W. Best and J. V. Kahn, Research in Education, Pearson Edition, 2006**.**

PSTAT01 **ADVANCED STATISTICAL METHODS** **4 Credits**

Unit I: Multivariate Analysis

Multivariate Normal distribution - Hotellings T^2 and Wilk's Lambda criterion for significance tests of group differences(one way MANOVA)- Maximum Likelihood, Missing data- Outliers – Principal components – Factor Analysis – Rotation Techniques - Discriminant Analysis. (15L)

Unit II: Nonparametric Regression

Need of Nonparamteric regression, Notion of linear smoother-Regressogram-. Nearest neighbors algorithm, local averaging. Method "Super smoother".Cross-validation approach and the bass parameter-Local regression-Method "Loess" ("Lowess") - Nadaraya-Watson kernel estimator as a solution of the optimization problem and local polynomial estimate. Gasser-Muller estimate-Generalized cross-validation- Akaike criterion-Kullback- Leibler discrepancy. (15L)

Unit III: Multiple Hypotheses Testing

Introduction to multiple testing: testing the global null hypothesis, Family-wise error rate, false discovery rate, Methods for multiple testing to control FWER & FDR, FWER and FDR without independence. The empirical Bayes perspective on FDR, Local FDR Adaptive & structured testing: groups of hypotheses, ordered. Sign error, magnitude error, Confidence intervals, false coverage rate, conformal prediction. (15L)

Unit IV: Causal Inference

Fundamental Problem of Causal Inference, Estimands and causal quantities of interest, Causal identification versus estimation, Selection on observables, Instrumental Variables, Estimating effects under unobserved confounding using exogenous variation in treatment induced by an instrument · Assumptions behind the instrumental variable strategy – exogeneity, relevance, exclusion restriction· Estimation via the Wald Estimator and Two-Stage Least Squares. (15L)

References:

1. Alwin C. Rencher(2002): Methods of Multivariate Analysis, John Wiley & Sons, New York
2. Tabachnick,B.G., & Fidell,L.S (2012).Using Multivariate Statistica,6th Edition, Pearson
3. KunioTakezawa, (2005) Introduction to Non Parametric Regression, Wiley-Interscience; 1st edition.
4. Hsu, J. (1996). Multiple comparisons: theory and methods. CRC Press.
5. Rupert Jr, G. (2012). Simultaneous statistical inference. Springer Science & Business Media.
6. Imbens, Guido W. and Donald B. Rubin.(2010) Causal Inference for Statistics, Social, and Biomedical Sciences. Cambridge University Press.
7. Hernán, Miguel A. and James M. Robins. (2020) Causal Inference: What If. Chapman & Hall/CRC.

PSTAT01E **ADVANCED STATISTICAL INFERENCE** **4 Credits**

Unit I: Asymptotic Inference I

Weak and Strong law of large numbers, consistent estimator, methods for finding consistent estimator-invariance property, joint consistency, CAN estimator-invariance property, Fisher lower bound, BAN estimator. Cramer-Huzurbazar theorem.CAN and BAN for multi-parameter exponential family of distributions. (15L)

Unit II: Asymptotic Inference II

Variance stabilizing transformations; their existence; their applications in obtaining large sample tests and estimators. Asymptotic Distribution of likelihood ratio test statistics, Wald test, Rao's Score test, Pearson Chi-square test for goodness-of-fit, Asymptotic confidence intervals based on CAN and VST, Asymptotic Confidence regions in multi-parameter families. (15L)

Unit III: Bayesian Inference

Point estimation-Prior, types of prior, posterior, Bayes estimators under vaious loss functions-generalization to convex loss function. Evaluation of the estimate in terms of posterior risk-comparison with frequentist methods. Bayesian hypothesis testing.Interval estimation-credible interval,highest posterior density region. (15L)

Unit IV: Robust Inference

Types of Robustness, Basic Tools for Judging Robustness, Robust estimation of Location and Scale, Robust estimation in one and two sample problems. (15L)

References:

1. Kale, B. K., & Muralidharan, K. (2015). Parametric Inference: An Introduction. Alpha Science International Limited.
2. Rohatagi V.K. and Saleh A. K. Md. E.(2001) : Introduction to Probability Theory and Mathematical Statistics- John Wiley and sons Inc.
3. Ferguson, T.S. (1996): A Course in Large Sample Theory. Chapman and Hall.
4. Lehmann E L (1999): Elements of Large Sample Theory, Springer.
5. Berger, J.O(1985): Statistical Decision Theory and Bayesian Analysis(Second Edition),Springer Verlag, New Yordk
6. Bernardo J M and A.F.M.Smith (2000) : Bayesian Theory, John Wiley & Sons, New York
7. Wilcox, R. R. (2017) Introduction to Robust Estimation and Hypothesis Testing, 4th Ed Academic Press.
8. Huber, P. J., & Ronchetti, E. M. (2009). Robust Statistics. 2nd John Wiley & Sons. Hoboken, NJ.

PSTAT02E **ADVANCED REGRESSION ANALYSIS** **4 Credits**

Unit I: Linear and Generalized linear models

Multiple linear regression model: Least squares estimation, Maximum likelihood (ML) estimator, weighted least squares, testing of hypothesis and confidence estimation. Problem of multicollinearity: causes, consequences, detection, ridge estimator and its properties.

Generalized linear model: Formal structure, ML estimation of parameters, testing of hypothesis and confidence estimation, deviance, residuals analysis, logistic regression, Poisson regression, Negative binomial regression. (15L)

Unit II: Variable selection

Problem of variable selection. All subset methods: R-square, adj. R square, Mallow's Cp, AIC, BIC. Stepwise procedures: Forward selection, Backward elimination. Regularized regression: Penalized least squares, penalized likelihood estimation, LASSO, LARS algorithm, elastic NET. (15L)

Unit III: Robust Regression

Influential observations, leverage, outliers, outliers and influential observations detection methods. Estimation in the presence ofvertical outliers: M-estimator, robust criterion functions (Huber, Hampel, Ramsey etc.) breakdown point, influence function, efficiency, Asymptotic distribution of M-estimator, rank estimation procedures when error follows non-normal distribution. (15L)

Unit IV: Zero-Inflated models

Zero-inflated Poisson regression-ML estimation, EM algorithm for estimation of regression parameters. Zero-inflated Negative binomial regression-ML estimation, EM algorithm for estimation of regression parameters. Variable selection in zero inflated models. Introduction to k-inflated models. (15L)

References:

1. Draper N. R. and Smith, H. (1998): Applied Regression Analysis. 3rd ed Wiley
2. Montgomery, D. C., Peck, E. A. and Vining, G. (2012): Introduction to Linear Regression Analysis, 5th Ed. Wiley.
3. Birkes, D. and Dodge, Y. (1993). Alternative methods of regression, John Wiley and Sons.
4. Huber, P. J. and Ronchetti, E. M. (2011) Robust Statistics, Wiley, 2nd Edition.
5. Efron, B., & Hastie, T. (2016). Computer age statistical inference (Vol. 5). Cambridge University Press.
6. Agresti A. (1990): Categorical Data Analysis. Wiley, New York.
7. Hilbe, J. (2011): Negative Binomial regression, Cambridge University, Press, 2nd Edition.
8. McCulloch, C. E.,& Searle, S. R. (2003). Generalized, linear, and mixed models, Wiley series in probability and statistics, New York.

PSTAT03E **NONPARAMETRIC INFERENCE** **4 Credits**

Unit I: One-sample procedures

Location functionals, geometry of inference in the one-sample location model, inference based on L1 norm, inference based on Wilcoxon signed-rank norm, one-sample U-statistics. (15L)

Unit II: Two-sample procedures

Geometry of two sample problems. Two-sample U-statistics. Mann-Whitney-Wilcoxon and related tests. (15L)

Unit III: Nonparametric functional estimation

Introduction to nonparametric functional estimation, nonparametric density estimation, nonparametric regression-kernel regression, nearest neighbor method, Rank regression. (15L)

Unit IV: Multisample and Mutivariate tests

R-fit of one-way design, rank-based test for $H_0 : \mu_1 = \cdot \cdot \cdot = \mu_k$, test of general contrast. Tests for multivariate location, scale and location-scale set up. Mardia's test, bivariate Wilcoxon sign-rank test, Peters and Randles test. Depth based ordering of vectors. Depth function, tests based on data depth, data depth plot. (15L)

References:

1. Randles and Wolfe (1994) Introduction to the Theory of Nonparametric Statistics– Krieger Publishing.
2. Hettmansperger, T. P. (1994) Nonparametric Inference Based on Ranks– Krieger Publishing.
3. Hajek, J. Sidak, Z. (1999). Theory of Rank Tests–Academic Press.
4. Serfling J, R.(1980). Approximation Theorems of Mathematical Statistics.
5. Gibbons, J., D., and Chakraborti, S. (2003). Nonparametric Statistical Inference, Marcel Decker, New York.
6. Liu, R., & McKean, J. W. (2015). Robust rank-based and nonparametric methods. Switzerland: Springer.
7. Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). Multivariate Analysis, Academic Press Inc. London) LTD, 15, 518.
8. Parelius, J. (1997). Multivariate analysis based on data depth. Ph.D. dissertation. Dept. Statistics, Rutgers Univ., New Jersey. Z. Z.
9. Härdle, W. (1990). Applied nonparametric regression. Cambridge university press.

PSTAT04E **4 Credits**

EVOLUTIONARY ALGORITHMS FOR MULTI OBJECTIVE OPTIMIZATION

Unit I

Introduction to Optimization; Single Objective Optimization (SOP); Deterministic Optimization Methods (Gradient Descent, LP and QP); Stochastic Optimization Methods (random search, Simulated Annealing, Evolutionary Algorithms); Difficulties in Single Objective Optimization; Difficulties with Classical Optimization algorithms; Need for Evolutionary Algorithms. (15L)

Unit II

Evolutionary Algorithm; EA operators (Selection, Recombination and Mutation operators);Single Objective Optimization (SOP) using EAs; Design & Parameterization for Single Objective Applications; Problem Formulation and representation issues for different real world engineering SOPs; some competent EAs. (15L)

Unit III

Constrained SOP; Discovery of innovative knowledge through Optimization; Difficulties in EAs; No Free Lunch Theorem; Enhancing efficiency of EAs through incorporation of domain specific information and hybridization with expressly designed algorithms. (15L)

Unit IV

Introduction to Multi-objective Optimization (MOP); Concept of Pareto optimality; Issues in Multi-objective Optimization; Multi-objective Evolutionary Approaches; Design & Parameterization for Multi-objective Applications- constrained Multi-objective Optimization; Dynamic Optimization; Robust Optimization; some real-world MOPs and their solution using MOEAs. (15L)

References:

1. A.E. Eiben and J.E. Smith (2003), Introduction to Evolutionary Computing, Springer, Heilderberg, Germany,.
2. Multi-objective Optimization Using Evolutionary Algorithms", Deb, Wiley,2005
3. T. Back, D.B. Fogel, and Z. Michalewicz (Eds), (2000) Evolutionary Computation: BasicAlgorithms and Operators, Vol. 1 and Vol. 2, Institute of Physics Publishing, Philadelphia, PA,
4. Y.C. Jin (Ed.), (2005) Knowledge Incorporation in Evolutionary Computation, Springer, New York.
5. R. Sarker and M. Mohammadian, and X. Yao (Eds),(2002). Evolutionary Optimization, Kluwer, Norwell, MA.

6. C.A. Coello,Coello, D. A. Van Veldhuitzen, and G. Lamont, (2002) Evolutionary Algorithms for Solving Multi-Objective Problems, Kluwer, New York.

PSTAT05E ADVANCED OPTIMIZATION TECHNIQUES 4 Credits

UNIT I: Quadratic Programming

Introduction to Quadratic programming – Wolfe's Modified Simplex Method – problems – Simulation – Models –Probability Distribution – Simulation Return. (15L)

UNIT II: Bounded Variable Problem

Introduction – Bounded Variable Simplex Algorithm – Problem to Solve LPP using Bounded Variable Technique Introduction to Integer Programming – problems - Gomory's All Integer Cutting Plane Method – problems - Gomory's Mixed Integer Method – Problems – Branch and Bound Method – problems – Optimum Integer Solution to LPP – Mixed Integer Programming Problems. (15L)

UNIT III: Post Optimal Analysis

Introduction – Discrete Variants – Structural Variations - problems – Limits of Variation for Feasibility – Effect of Discrete changes on pre optimality of the optimum basis feasible solution - Introduction to Parametric Linear Programming – solution – Critical values – Optimal Basic Feasible solution – Analysis. (15L)

UNIT IV: Dynamic Programming

Introduction – Recursive Algorithm – Development of Recursive Equation – solution to Recursive Equation – Bellman's Principle of Optimality – problems – Applications – Multistage problem. (15L)

References:

1. Man Mohan, Gupta and Kapoor, (1995) Operations Research, 4th edition-Springer Pub.
2. J. Lieberman and FederickS Hiller,-(2005) Operations Research, 8th edition, McGraw Hill
3. Taha (2007) Operations Research, III edition, University of Michigan
4. David Corne, Marco Dorigo, Fred Glover (1999) New Ideas in Optimization, McGraw Hill.
5. Melgon W Jeter (1996)Mathematical Programming:An Introduction to optimization, CRC Press

PSTAT06E **ADVANCED QUEUEING THEORY** **4 Credits**

UNIT I : Introduction - Semi-Markov and Markov Renewal Processes - Regenerative Processes and Long Run Probabilities - Random Walks and the G/G/1 Queue - Applications to the M/M/1, M/G/1 and G/M/m Queues. (15L)

UNIT II : Continuous Time Markov Chains and Markovian Queues - Transforms and Transient Behavior of the M/M/1 Queue **-** Birth/Death Processes and the M/M/m Queue - M/M/1 Queue and its Variants - M/G/1 Queue and Busy Period Analysis in G/G/1 Queue. (15L)

UNIT III: Markovian queues – Birth and Death processes – Single and multiple server queueing models – Little‟s formula – Queues with finite waiting rooms – Queues with impatient customers: Balking and reneging - General Markovian Queues. (15L)

UNIT IV: Finite source models – M/G/1 queue – Pollaczek Khinchin formula – M/D/1 and M/EK/1 as special cases – Series queues – Open Jackson networks. Bulk Input ($M^{[X]}/M/1$) ,Bulk Service($M/M^{[Y]}/1$), Erlangian Models, Priority Queue Disciplines, Retrial Queues and applications. (15L)

References:

1. John F. Shortle, James M. Thompson, Donald Gross, Carl M. Harris (2018) Fundamentals of Queueing theory, 5th Edition, Wiley India Pvt. Ltd.
2. Moshe Haviv (2013) Queues: A Course in Queueing Theory (International Series in Operations Research & Management Science, Springer.
3. M. L. Chaudhry & J.G.C. Templeton (1983) A First Course in Bulk Queues, New York:John Wiley & Sons.
4. MorHarchol-Balter, (2013) Performance Modeling and Design of Computer Systems: Queueing Theory in Action, Cambridge University Press.

PSTAT07E **NEURAL NETWORKS AND FUZZY LOGIC** **4 Credits**

UNIT I: Evolution of neural networks; Artificial Neural Network: Basic model, Classification, Feed forward and Recurrent topologies, Activation functions; Learning algorithms: Supervised, Un-supervised and Reinforcement; Fundamentals of connectionist modeling: McCulloach – Pits model, Perceptron, Adaline, Madaline. (15L)

UNITII: Topology of Multi-layer perceptron, Back propagation learning algorithm, limitations of Multi-layer perceptron. Radial Basis Function networks: Topology, learning algorithm; Kohenen's self-organising network: Topology, learning algorithm; Bidirectional associative memory Topology, learning algorithm, Applications. (15L)

UNITIII: Recurrent neural networks: Basic concepts, Dynamics, Architecture and training algorithms, Applications; Hopfield network: Topology, learning algorithm, Applications; Industrial and commercial applications of Neural networks: Semiconductor manufacturing processes, Communication, Process monitoring and optimal control, Robotics, Decision fusion and pattern recognition. (15L)

UNIT IV: Classical and fuzzy sets: Introduction, Operations and Properties, Fuzzy Relations: Cardinality, Operations and Properties, Equivalence and tolerance relation, Value assignment: cosine amplitude and max-min method; Fuzzification: Membership value assignment- Inference, rank ordering, angular fuzzy sets. Defuzzification methods, Fuzzy measures, Fuzzy integrals, Fuzziness and fuzzy resolution; possibility theory and Fuzzy arithmetic; composition and inference; Considerations of fuzzy decision-making. (15L)

References:

1. T. J. Ross, (1995) Fuzzy logic with engineering applications , 1 ed. New York, NY: McGraw-Hill, .
2. S. Rajasekaran and G.A.Vijaylakshmi Pai.(2003) Neural Networks Fuzzy Logic, and Genetic Algorithms, Prentice Hall of India.
3. J. Yen and R. Langari (1998). Fuzzy Logic, Intelligence, Control and Information, Pearson Education
4. Zimmermann, (1996) Fuzzy set theory and its Applications, 3 ed. Norwell, MA: Kluwer.