Progressive Education Society's

Modern College of Arts, Science andCommerce Ganeshkhind, Pune 16

(An Autonomous College Affiliated to Savitribai Phule Pune University)

Two Year Degree Program in Statistics

(Faculty of Science & Technology)

M.Sc. (Statistics) Part-II

Choice Based Credit System Syllabus

To be implemented from Academic Year 2023-24

Structure and code of papers

T: Theory P: Practical O: Open C: Compulsory E: Elective

Course Code	T/P	O/C /E	Title	Credits	ESE Duration	Marks Assigned
Semester III						
23-ST-31	T	C	Applied Stochastic Processes	4	3 Hours	100
23-ST-32	T	C	Design and Analysis of Experiments	4	3 Hours	100
23-ST-33	T	C	Machine Learning	4	3 Hours	100
23-ST-34 (A) 23-ST 34 (B)	T T	E E	Bayesian Inference OR Statistical Quality Control	4	3 Hours	100
23-ST-35	P	C	Practical III	4	3 Hours	100
Total				24		500
Semester IV						
23-ST-41	T	C	Asymptotic Inference	4	3 Hours	100
23-ST-42 (A) 23-ST-42 (B)	T T	E E	Econometrics and Time series OR Operation Research	4	3 Hours	100
23-ST-43 (A) 23-ST-43 (B)	T T	E E	Survival Analysis OR Categorical Data Analysis	4	3 Hours	100
23-ST-44 (A) 23-ST-44 (B)	T T	E E	Computer Intensive Statistical Methods OR Analysis of clinical Trials	4	3 Hours	100
23-ST-45	P	C	Practical IV + Project	4	3 Hours	100
			Total	24		500

23-ST-31: Applied Stochastic Processes (from 2023-24) (4 credits)

Pre-Requisites: Linear Algebra, Differential equations.

Course Objectives: To study the sequences of dependent random variables, predictions, limiting behavior of such sequences.

Note: Emphasis be given to Applications of Stochastic Processes

Course Outcomes:

After completion of the course students will able to:

CO1) Understand the standard concept and apply the techniques and constructions of discrete andcontinuous time Markov chains to solve problems involving n-step transition probabilities, hitting probabilities, and stationary distributions.

CO2) Understand how to choose best stochastic process for specific situation.

CO3) Distinguish between transient and recurrent states in given finite and infinite Markov chains.

CO4) Apply the stochastic analysis to realistic problems.

CO5) Understand renewal theory and branching processes with applications.

Unit 1:

Stochastic processes, Markov property, Markov chains (MC), utility of stochastic processes and Markov Chain, finite MC, transition probabilities, initial distribution, illustrations such as random walk, Ehrenfest chain, gambler's ruin chain, queuing chain, birth death chain, branching chain, Chapman Kolmogorov equation, n-step transition probabilities, transition probability matrix (t. p. m.) hitting times, probability of ever return, transient and recurrent states, decomposition of state space, closed set of states, irreducible set of states, irreducible MC, absorption probabilities, martingales, classification of states of birth and death chains, , non-null and positive recurrent states, period of state, branching chain, queuing chain, random walk, gambler's ruin chain with absorbing , reflecting and elastic barrier, etc. probability of ruin cases (i) adversary is infinitely rich (ii) stakes are doubled or halved , expected gain, expected duration of the game. (15L)

Unit 2:

Elementary properties of stationary distributions, illustrations such as birth and death chains, Ehrenfest chain, particles in box, average number of visits to recurrent state, probability of absorption in persistent class starting from transient state, existence of uniqueness of stationary distributions, reducible chains, illustrations such as queuing chain finite chains, convergence to the stationary distribution. Steady state distribution, Concept of Ergodicity, utility and explanation with real life situation, Ergodic Markov chain, Ergodic theorem. (Without Proof)

Branching Chain: BGW (Bienayme-Galton-Watson) branching process, offspring distribution,mean and variance, generating function for

probability of ultimate extinction, nth generation size and related recurrence relations. (15L)

Unit 3:

Intensity rates, it's relation with transition probabilities. Kolmogorov consistency condition, Markov property in continuous time stochastic processes. Kolmogorov forward and backward equations.

Poisson process: Postulates and properties of Poisson process, probability distribution of N (t) the number of occurrences of the event in (0, t], Poisson process and probability distribution of inter-arrival time, Generalizations of Poisson process: pure birth process: Yule Furry process. Non- homogeneous Poisson processes.

Birth and death process: (i) Pure-Birth process, Yule Furry Process (ii) Pure death process, particular cases: Birth immigration process. (i) immigration-emigration process, (ii) linear growth process, (iii) linear growth with immigration, (iv) immigration death process. (15L)

Unit 4:

Renewal process: renewal process in continuous time, renewal function and renewal density,renewal equation, stopping time: Wald's equation, elementary renewal theorem and its applications: (i) Age and block replacement policies, (ii) Replacement on failure and block replacement, renewal theorems (Blackwell's and Smith's

Continuous time Markov chains: Markov processes with continuous state space: Introduction to Brownian motion and its properties, Transition probabilities Brownian motion process as limiting case of random walk. Wiener process and its properties. (15L)

Books Recommended:

1. Ross, S. (2000) Introduction to probability models, 7th edition (Academic Press)
2. Medhi J. (1982) Stochastic processes (Wiley Eastern)
3. Hoel P. G., Port, S.C. ,Stone, C.J. (1972) : Introduction to stochastic processes
4. Bhat B. R. (2000) stochastic models: Analysis and applications (New Age International)
5. Adke S.R., Manjunath, S.M. (1984) An introduction to finite Markov processes (Wiley Eastern)
6. Ross, S. (1996) Stochastic processes (John Wiley)
7. Taylor, H N and Karlin, S. (1984): An introduction to stochastic modeling (AcademicPress)
8. Vidyadhar G. Kulkarni: Modelling and Analysis of Stochastic systems CRS PressPublications.
9. Tijms S: Stochastic modeling and its Applications, Wiley Publishers.

ST- 32: Design and Analysis of Experiments (from 2023-24) (4 Credits)

Pre-Requisites: Probability distributions, sampling, testing of hypotheses, control charts and inspection sampling plans.

Course Objectives: To learn the basic principles in the design of simple experiments. To learn different tests for comparing pairs of treatment means, factorial experiments, fractional factorial experiments, confounding, BIBD, PBIBD.

Course Outcomes:

After completion of the course students will able to:

CO1) Understand the concept of BIBD, connectedness, balancedness and orthogonality of design.

CO2) Understand the difference between fixed and random effect models.

CO3) Compare the pairs of treatment means using different methods. Construct Fractional factorial experiments and apply confounding in real life problems.

CO4) To use appropriate design for solving real life examples.

CO5) To learn the applications of different designs in agricultural experiments

Unit 1:

One way classification with equal and unequal number of observations per cell, Lenege's test, Bartlet's test, Newman Keuls Test, Duncans Multiple Range Test (DMRT), Dunnet test, Non Parametric One way ANOVA (Kruskal Wallis Test), Friedman test (Non- parametric alternative to the one-way ANOVA with repeated measures), two way classification with equal number of observations per cell (with and without interaction), Missing plot techniques.

BIBD intra block analysis, incidence matrix, symmetric BIBD, resolvable BIBD, (Results related to all types of BIBD), PBIBD with 2 associate classes (PBIBD (2)) (18 L)

Unit 2:

Connectedness, balancedness and orthogonality of design, random effect models for one factor, estimation of variance components and confidence interval for intra class correlation coefficient, random effect model for the two factor, estimation of variance components. 2^k full factorial experiments: diagrammatic presentation of main effects, and first and second order interactions, model analysis using ANOVA, total confounding of 2^k design in 2^p blocks $p \geq 2$, partial confounding in 2^p blocks; p=2,3. (15L)

Unit 3:

Fractional factorial experiments, resolution of a design (III, IV & V), aberration of a design, Plackett- Burman designs. 3^2 designs: contrasts for linear and quadratic effects, statistical analysis of 3^2 design, 3^3 designs: contrasts for linear and quadratic effects, statistical analysis of 3^3 design, blocking of 3^2 in three blocks, blocking of 3^3 in 9 blocks, fractional factorial experiment in 3^p designs in $p = 2, 3$. (15L)

Unit 4:

Response surface methodology (RSM): linear and quadratic model, stationary point, canonical analysis, central composite designs(CCD), ridge systems, multiple responses, concept of rotatable designs, Box- Behnken design for 2 and 3 variables, blocking in Response surface design. Mixture experiments, Simplex lattice design and Simplex centroid design, Taguchi methods: concept of loss function, S/N ratio, orthogonal arrays, triangular tables, linear graphs, inner and outer arrays. (12L)

Books Recommended:

1. Dean, A. and Voss, D. (1999). Design and Analysis of Experiments, Springer.
2. George E. P. Box, Draper N.R. (1987). Empirical Model-Building and ResponseSurfaces, Wiley.
3. Hicks, C.R., Kenneth V. and Turner, Jr. (1999). Fundamental Concepts in theDesign of Experiments, Oxford University Press.
4. John P.W.M. (1971). Linear Models, Wiley.
5. Kshirsagar A.M. (1983). Linear Models, Marcel Dekker
6. Montgomery, D.C. (2001). Design and Analysis of Experiments, Wiley.
7. Ogawa J. (1974). Statistical Theory of the Analysis of Experimental Design,Marcel Dekker.

23-ST-33: Machine Learning (from 2023-24) (4 Credits)

Pre-requisites (Desirable): Knowledge of basic probability concepts and standard probability distributions, Basic concepts of non-linear programming. Knowledge of R-language or Python.

Course Objectives: Machine Learning (ML) deals with the design, analysis and validation of algorithms that enable computers (machines) to learn from data and automatically extract methods to perform intended tasks. Such algorithms that enable machines to learn methods from data are called learning algorithms. It is automatic algorithm development, which enables machines to develop their own algorithms. The input (data) to learning algorithms is called examples. The output from learning algorithms is called models or rules. These models (rules) are used in practice to extract hidden information in large data sets that arise in many application areas. Data Mining refers to this practical use of the ML algorithms. For an applied statistician (now a days known as a data scientist) ML algorithms provide additional set of tools for data analysis, particularly when data is large. The main objective of this course is to introduce some standard learning algorithms for tasks such as classification, regression, clustering, outlier detection and association finding. It may be noted that this is a course on Statistical Inference – model free Statistical Inference, which can be studied by any one from any discipline.

Course Outcomes:

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

CO1) Apply appropriate learning algorithm for analyzing data.

CO2) Use appropriate R-packages for data analysis.

CO3) Design learning algorithms for new tasks.

CO4) Self-learn many other ML techniques.

CO5) Be a better data scientist.

Unit 1: Introduction to Machine Learning

Need for and meaning of Machine Learning (ML). Various ML tasks. Framework of ML environment .Relationship with other fields such as Data Mining, Statistics, Data Science, Big Data Analytics. Data Cleaning and its importance. (10L)

Unit 2: Classification

Introduction to Classification, construction of Decision/classification tree, Decision Tree Learning – Impurity measures,, tree pruning modifications for regression trees. Ensemble learning-Bagging and boosting, random forests, Cross validation. Naïve Bayes classifier, optimality of Bayes rule. Generative and discriminative approaches to classification problem. Nearest neighbor classifier, SVM Learning. Regression And its various types. Neural Network Learning - basic concepts, Perceptron learning and its limitations. (25L)

Unit 3: Clustering

Introduction to clustering, Types of clustering Cluster learning- k-means algorithm, Agglomerative hierarchical clustering. Cluster quality. Association analysis, market basket analysis, Apriori algorithm. (25L)

****There is Lab work in Practical course for 12 hours in the entire term where in students get opportunity to test the algorithms using R/ Python/ SQL/Software Packages such as Weka.

Books Recommended:

1. Pang-Ning Tan, Michael Steinbach and Vipin Kumar (2013) Introduction to Data Mining. (Indian Edition) Pearson Education (Published by Dorling Kindersley (India)
2. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani (2013) An Introduction to Statistical Learning: With Applications in R. (Springer)
3. ShaiShalev-Schwartz and Shai Ben-David (2014) Understanding Machine Learning: From Theory to Algorithms (Cambridge University Press)
4. Han J., Kamber M., and Pei J (2012) Data Mining: Concepts and

Techniques. (Elsevier) (2)Alex Smola and S.V.N. Vishwanathan (2008) Introduction to Machine Learning. (Third Edition) (Cambridge University Press)

5. Ian H. Witten and Eibe Frank (2005) Data Mining: Practical Machine Learning Tools and Techniques. (Second Edition) (Elsevier)

23-ST- 34(A): Bayesian Inference (from 2023-24) (4 credits)

Pre-Requisites: Probability models, parametric and non-parametric inference.

Course Objectives: Estimation using pre-knowledge about the parameters.

Course outcomes: Use these methods and techniques in data mining.

Course Outcome:

After completion of the course students will able to:

CO1) Understand difference between classical and Bayesian appproach

CO2) Bayesian computation

CO3) Credible intervals

Unit 1:

Subjective and frequentist probability , Bayesian inference set up, prior and posterior distributions, loss functions, principles of minimum expected posterior loss, quadratic and other loss functions, advantages of being Bayesian , improper priors ,Common problems of Bayesian Inference, point estimation, maximum aposterior estimator (MAP) ,HPD confidence intervals, credible intervals, predictions of future observations, Bayesian testing. (18L)

Unit 2:

Bayesian analysis with subjective priors, classes priors, conjugate class of priors, Jeffrey's prior, probability matching prior, robustness and sensitivity. (12L)

Unit 3:

Bayesian model selection BIC, Bayes factors, limit of posterior distributions, consistency andasymptotic normality of posterior distributions (12L)

Unit4:

Bayesian computing, E-M Algorithm, MCMC, MH Algorithms, Gibb' sampling, convergence diagnostics. (Note: Minimum 10 hours of computational practice) (18L)

Books Recommended:

1. Bolstad W M (2007).Introduction to Baysian Statistics 2nd edition Wiley.
2. Christensen R. Johnson W. Branscum A. and Hanson T. E. (2011). Baysian Ideas and Data Analysis: An introduction for Scientists and Statisticians , Chapman and Hall
3. Congdon P (2006). Baysian Statistical Modeling, Wiley.

4. Ghosh J. K., Delampady M. and T. Samantha (2006): An Introduction to Baysian Analysis: Theory and Methods, Springer.
5. Jim A (2009) Baysian Computation with R 2nd edition .Springer.
6. Rao C. R. and Day D. (2006). Baysian Thinking Modelling and Computation, Hand book of Statistics Vol.25.Elsevier.

23-ST- 34 (B): Statistical Process control (SPC) (from 2023-24) (4 credits)

Pre-Requisites: Probability distributions, sampling, testing of hypotheses, control charts and inspection sampling plans.

Course Objectives: To understand the basics of TQM. To study different types of control charts and sampling plans. Comparison of control charts. Process Capability study.

Course Outcome:

After completion of the course students will able to:

CO1) To use appropriate control charts

CO2) To use different sampling plans

CO3) To draw inference about process capability

Unit 1:

a) TQM Total quality Management : meaning and dimensions of quality, Quality improvement, Quality Philosophy, Introduction to TQM, six sigma, DMAIC, and other extension of TQM, quality systems, The ISO 9000 and other Quality systems.
b) Control Chart: Revision of theory of control charts, Concepts of stable industrial processes, Systematic variation, random variation, variation within and between subgroups, estimation ofprocess parameters, Equivalence between control chart and testing of hypothesis problem. Choice of control limits Operating characteristic (O C curve) of control chart. Probability of false alarm, probability of catching shift in parameter. Concept of Run length, probability distribution of run length, average run length (ARL). Comparison of control chart using ARL,OC curve, criteria of detecting lack of controls (sensitizing rules),patterns on control charts with justification and its effect on Probability of false alarm. X -S chart with subgroup size (i) fixed (ii) variable, probability limits , S^2 chart Applications of control charts situations other than manufacturing. (15L)

Unit 2:

a) X-bar MR chart
b) CUSUM Chart: Chart statistic (Ci +, Ci −) and chart parameters (k, h), construction and working of tabular CUSUM chart for mean and variance, Statement of hypotheses. Estimation of shift in mean of process. Fast initial response or head start feature, Siegmund's approximation for ARL and determination of chart parameters. CUSUM chart for subgroup size n>1, comparison between Shewhart chart and CUSUM chart V mask procedure.
c) EWMA chart: Chart statistic its expectation and variance. Choice of chart parameters (λ, L). Construction and working of EWMA chart for mean and variance. EWMA chart for subgroup size n>1, Comparison of Shewhart

control charts with CUSUM charts. Simulation of ARL (δ).

d) Process capability: Different Process capability and performance indices Cp, Cpk. Cpm. Properties and relation between capability indices. Connection between proportion of defectives (DPPM) and Cp.Interval estimation of mean given Cpm ≥1.Estimation and confidence intervals of estimators of Cp
e) and Cpk Testing of hypothesis about Cp. (15L)

Unit 3: Other control charts

a) Synthetic control chart: Confirming run length (CRL) chart for attributes, Synthetic control chart , computations of chart parameters for given ARL(0),Steady state model, Computations of ARL (δ), ATS (δ). Comparison of with Shewhart control chart and CUSUM charts.
b) Non-parametric control chart: Concept, construction of non-parametric chart using signtest.Control charts for auto correlated observations: Need, constructions of control chart for residuals after fitting first order auto correlated model.
c) Hoteling T^2 Chart: Testing multivariate normality, Hotelling T^2 multivariate control chart for mean vector when (i) dispersion matrix is (i) known (ii) unknown ARL (0), ARL (δ). T^2 control chart when subgroup size n=1
d) Control chart for dispersion matrix when mean vector is (i) known (ii) unknown. (15L)

Unit 4:

a) Attribute control charts: Revision of control charts for attributes, OC curve for P chart and C chart. Determination sample size for P chart by various criteria (i) probability of catching atleast 0.5 (ii) to get LCL >0 (iii) To have at least some defectives in sample with given confidence coefficient. (iv) minimizing ATS (δ) chart and O C Curve, U chart
b) Demerit control chart for number of defects.
c) Nelsons control chart for low defect counts.
d) General ideas of economic designing of control charts. Duncan’s model for the economic control chart
e) Acceptance Sampling Plan: Description of MIL STD and Dodge Roming sampling plans (i) Acceptance Sampling Plan for attributes with Curtailed inspection: Equivalence between sampling plans and testing of hypothesis problem.
f) Double and multiple and sequential sampling plans for attributes sampling plan, Operating characteristic functions. AOQL, ATI, ASN. Continuous Sampling Plans,
g) Chain sampling plan.
h) Acceptance sampling plan for variable: parameters of plan. (i) Critical distance method,(ii)critical proportion method. (15L)

Books Recommended:

1. Bourke P.D. (1991) Detecting shifts in fraction non – confirming using run length chartwith100% inspection. Journal of Quality Technology 23 (3) 225-230
2. Besterfield, D. H..Besterfield – Michana, c, Besterfield, G. H. Besterfield-Sace, M (2001)TotalQuality Management; Pearson Education (Singapore)

Pte. Ltd. India 2nd Edition.
3. Logotheris, N. (1992) Managing Total Quality; Prentic Hall of India.
4. Montgomery ,D.C. (1985) Introduction to Statistical Quality Control (Wiley)
5. Oakland J.S. (1989) Total Quality Management: Butterworth – Heinemann.
6. Raid W. Amin a Marion R. Reynolds Jr. b; BakirSaadc :Nonparametric quality control charts based on the sign statistic :Communications in Statistics - Theory and Methods Vol.34,2005. 7. Wu, Yeu and Spedding (2001) Asymptotic control charts for detecting fraction non confirming increases JQT 33 (1)104-111

23-ST-35 Practical –III (from 2023-24) (4 credits)

Sr. No.	Title of Experiment
	Practical based on Stochastic Process:
1	Realization of Markov chain when TPM is given and computation of transition probabilities andStationary distribution of Markov chain. Gambler's Ruin problem.
2	Realization of birth and death process.(with constant birth and death rates).
3	Realization of Poisson process by two ways (i) till n Poisson event occur (ii)for fixed length timeinterval(0, t]
4	Realization of Gaussian and Brownian process
	Practical based on Bayesian inference
5A	Bayesian Analysis : i. Plotting prior, posterior density function and likelihood function, ii. Generating random sample from posterior distribution **iii.** Constructing highest posterior density credible interval
6A	Testing of hypotheses by computing Bays factor.
	Practical based on Statistical Process control (SPC)
5B	CUSUM, EWMA charts
6B	Synthetic control chart and Hotelling T^2 chart for mean vector.
	Practical based on Design of Experiments
7	Analysis of one way and two way classification. Multiple comparison tests.
8	**Analysis of BIBD (Intra block analysis) and PBIBD.**
9	2^k Factorial Experiments, analysis of single replicate of 2^k factorial experiments
10	Total and partial confounding in 2^k factorial experiments.
11	Analysis of 3^2 factorial experiments
12	Random effect model with one factor, estimation of variance.
13	Fitting first and second order response surface model, central composite design(contour,surface plots, canonical analysis of stationery points).
14	Taguchi methods: S/N ratio, orthogonal arrays, triangular tables, linear graphs, innerandouter arrays.
	Practical based on Machine Learning (Test the algorithms using R/ Python/ SQL/Software Packages such as Weka.)
15	Implementation of nearest neighbor classifier16 Implementation of Naïve Bayes classifier
16	Implementation of Naïve Bayes' classifier
17	Implementation of k means clustering
18	Implementation of A priori algorithm
19, 20	**Project progress report (equivalent to 2 Practicals)**

23-ST-41: Asymptotic Inference (from 2023-24) (4 Credits)

Course Objectives: Asymptotic analysis has always been very useful for deriving distributions in statistics in cases wherethe exact distribution is unavailable. More importantly, asymptotic analysis can also provide insight into the inference process itself, suggesting what information is available and how this information may be extracted. The asymptotic theory proceeds by assuming that it is possible (in principle) to keep collecting additional data, so that the sample size grows infinitely. Bayesian inference has found application in a wide range of activities, including science, engineering, philosophy, medicine, sport, and law.

Course Outcomes:

After completion of the course students will able to:

CO1) Understand the concept of consistency and asymptotic normality.

CO2) Understand method of moments and percentiles, maximum likelihood to Find consistentestimator and Cramer Huzurbazar theorem.

CO3) Apply likelihood ratio tests, Wald, Score and Bartlett's test in real life situations.

CO4) Compare various tests through relative asymptotic efficiency.

Unit 1:

Consistency: real and vector parameters, Invariance under continuous transformation; Methods of obtaining consistent estimators: method of moments, method of percentiles, mean squared error criterion; Asymptotic relative efficiency, Comparison of consistent estimators, minimum sample size required by the estimator to attain certain level of accuracy, Asymptotic Normality; Consistent Asymptotic Normal (CAN)estimators: real and vector parameters; invariance of CAN property undernonvanishing differentiable transformation. Delta method, Methods of obtaining CAN estimators: method of moments and method of percentiles. **(20L)**

Unit-2

Maximum likelihood estimation, restricted parameter space, Inconsistent MLEs, MLEs in irregular cases. Asymptotic distribution of MLE in special class of distributions: Cramer regularity conditions, Cramer- Huzurbazar theorem, Extension to vector-valued parameters, Super-efficient estimators, BAN estimators, CAN and BAN estimation for multi-parameter exponential family and applications, Solution of likelihood equations, Method of scoring, Newton-Raphson and other iterative procedures. (20L)

Unit-3

Asymptotic theory of tests of hypotheses: Tests based on MLEs, Likelihood Ratio Test (LRT), asymptotic distribution of LRT statistic, Wald Test, Rao'ss coretest, Pearson Chi-square test for goodness of fit, Bartlett's test for

homogeneity of variances, locally most powerful tests.
Variance stabilizing transformations (VST): their existence, their applications in obtaining large sample tests and estimators Asymptotic Confidence Intervals: based on CAN estimators, based on VST. Asymptotic Confidence regions in multi-parameter families. **(20L)**

Books Recommended:

Textbook:

1. Kale, B.K. and K. Muralidharan (2015), Parametric Inference: An Introduction, Alpha Science Intl Ltd.

Reference books:

2. Gupta Anirban Das (2008), Asymptotic Theory of Statistics and Probability, Springer, NewYork.
3. Manoj Kumar Srivastava, Abdul Hamid Khan and NamitaSrivastava (2014), Statistical Inference:Theory of Estimation, PHI Learning Pvt Ltd, Delhi.
4. Bolstad W. M. (2007) Introduction to Bayesian Statistics 2nd Ed. Wiley, New York.
5. Lee P.M. (2004) Bayesian Statistics: An Introduction, Hodder Arnold, New York.
6. Ferguson, T.S. (1996),A course on Large Sample Theory. Chapman and Hall, London.
7. Rao, C.R. (1973): Linear Statistical Inference and its Applications, Wiley, New York.
8. Lehmann, E.L. and Casella G. (1999), Theory of Point Estimation, Springer, New York.

23-ST- 42 (A): Econometrics and Time Series Analysis (from 2023-24) (4 credits)

Pre-Requisites: Linear Algebra, Differential equations.

Course Objectives: **To learn and develop scientific view to understand the time series data and its analysis. To learn stationary and non-stationary, and seasonal and non-seasonal time series models. Estimate model parameters and compare different models developed for the same dataset in terms of their estimation and prediction accuracy. To learn some basic concepts of econometrics.**

Course Outcomes:

After completion of the course students will able to:

CO1) Understand the concept of time series with its components and able to compute ACVF andACF.

CO2) Remove trend and seasonality using different methods to convert the time series **i**nto stationary.

CO3) Apply auto regressive, moving average, ARMA, ARIMA, SARIMA models, Box-Jenkinsapproach to forecast time-series data empirically.

CO4) Check and validate models with its residual analysis and diagnostic checking.

CO5) Apply econometrics concepts in real life data.

Unit 1:

Inference in Econometric Models: Simultaneous equation models – endogenous and exogenous models, Problems with OLS estimators, Identification problem and reduced models. Indirect Least Squares Method, 2 stage and 3 stage OLS estimation. Properties of the estimators. Indirect Inference in Econometric models. (15L)

Unit 2:

Time - series as a discrete parameter stochastic process. Exploratory time Series analysis, Auto covariance and autocorrelation functions and their properties. Methods of estimation and elimination of trend and seasonality: Graphical method, Moving average, exponential smoothing and least square method. Testing the estimated noise sequence: The sample ACF, the Portmanteau tests, the turning point test, the difference sign test and the rank test. Holt -Winters smoothing. Forecasting based on smoothing, adaptive smoothing. (15L)

Unit 3:

Stationary processes: General linear processes, moving average (MA), auto regressive (AR) and autoregressive moving average (ARMA) processes. Causal and non-causal process, Stationarity and inevitability conditions. Non-stationary and seasonal time series models: Auto regressive integrated moving average (ARIMA) models, Seasonal ARIMA (SARIMA) models, Transfer function models (Time series regression). (15L)

Unit 4:

Forecasting in time series models, Durbin-Levinson algorithm, innovation algorithm (without proof).Estimation of mean, auto covariance and autocorrelation functions, Yule-Walker estimation, Estimation of ARIMA models parameters, maximum likelihood method, large sample theory (without proofs). Choice of AR and MA periods, FPE, AIC, AIC, BIC, residual analysis and diagnostic checking. Unit-root non stationarity, unit-root tests (Dickey-Fuller). (15L)

Books Recommended:

1. Brockwell, P.J. and Davis, R. A. *Introduction to Time Series Analysis*, Springer
2. Chatfield, C. (2001). *Time Series Forecasting*, Chapmann& hall, London 3.Chatfield, C. (2007). *Time Series Analysis using R*, Chapmann& hall, London
4. Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd Ed. John Wiley,
5. Hamilton N. Y. (1994). *Time Series Analysis*.Princeton University press. Princeton
6. Kendall, Sir Maurice and Ord, J. K. (1990). Time Series (Third Edition), EdwardArnold.
7. Lutkepohl, H. and Kratzing, M. (Ed.) (2004). *Applied Time Series Econometrics*,Cambridge University Press, Cambridge
8. Shumway, R. H.andStoffer D. S. (2010) Time Series Analysis & Its Applications,Springer, New York.
9. Tsay, R. S. (2010). *Introduction to time series .Wiley.*

23-ST- 42 (B): Operations Research (from 2023-24) (4 credits)

Pre-Requisites: LPP and Transportation problem

Course Outcomes:

After completion of the course students will able to:

CO1) Understand basics and formulation of linear programming problems and appreciate their limitations; solve linear programming problems using graphical method.

CO2) Apply simplex method to solve real life problems.

CO3) Solve artificial variable technique, duality theory, revised simplex method, sensitivity analysis, and transportation and assignment problems.

CO4) Understand the concept of non-linear programming problem, PERT/ CPM, simulation, investment analysis with real life application

Unit 1:

Revision of LPP and Transportation problem : Basic theorems (with proofs **)** related to LPP. Two Phase method, Duality in Linear programming, The Dual simplex method, Revised simplex method, Sensitivity Analysis. Ellipsoid method , polynomial time algorithm, Karmakar's polynomial time algorithm. Convergence and complexity.

Revision of Transportation problem: Post optimality analysis in TP. Dual of the transportation problem, The trans-shipment problem, Assignment model, the travelling salesman problem. (15L)

Unit 2:

Integer linear programming problem: Gomory fractional All integer, Mixed IPP, Branch andBound Method, cutting planes.

Dynamic Programming problem: Formulation of Dynamic Programming problem, Optimal subdivision problem, Solution of LPP by Dynamic programming problem, Application of Dynamic programming problem.sensitivity analysis. Bellman's optimality principle , Stochastic dynamic programming.

Nonlinear Programming Problem: Introduction , Local and Global optimum, Concave and Convex function, Kuhn – Tucker Condition, quadratic programming problem, Wolef's Method, Beales Method. **(22L)**

Unit 3:

Queuing Theory: Introduction, Applications, Waiting time and idle time, classification ofqueuing models, Kendall's notation for queues various queue disciplines.

(i) Model: M/M/1:FCFS/ ∞/∞ Single channel Poisson arrival with exponential Service time, infinite population. Derivation of probability of queue length, system length, waiting time in queue and that in system. Little's formulae (relation)

Cases (a) M/M/1: FCFS/a/∞ **(b)** M/M/1: FCFS/ a/b (a, b are finite). **(c)** Arrival rate is function of queue length (n): $\lambda_n=\lambda/(n+1)$ **(d))** Service rate is proportional to queue length (n): $\mu_n=n\mu$

(ii) Model: M/M/c: FCFS/ ∞/∞ Multi channel Poisson arrival with exponential Service time, Derivation of probability of queue length, system length, waiting time in queue and that insystem

(iii) Model: M/M/∞: FCFS/ ∞/∞, expected system length and expected waiting time insystem.Equivqlence with Birth and Death process.

Simulation Technique: Introduction, definition, Advantages and disadvantages, Application of Simulation, Monte Carlo simulation, Generation of random number. Acceptance -Rejection method.Simulation from bivariate distributions

Network analysis: Review of CPM, PERT, Network flow, maximal flow, Transportation problemas network, Transshipment Problem as Network, linear programming formulation. **(23L)**

Books Recommended:

1.Taha. H.A.(1992) Operation Research 5^{th}ed Macmillan
2.Preamkumar Gupta and D.S.Hira
3.Bertsekas. D.(1999) : Nonlinear Programming 2^{nd}ed Athena Scientific
4.Hadley .G(1987) Linear Programming Addison –Wesley
5.J.K.Sharma : 5^{th}ed Theory and applications

23-ST-43 (A): SURVIVAL ANALYSIS (from 2023-24) (4 credits)

Course objectives: The objectives of this course are to study the different models from Survival Analysis, to understand different types of censoring, learn to estimate and interpret survival characteristics.To provide the construction of parametric and non-parametric estimators of survival distributions, and probability density functions based on incomplete data. The models with right- censored, truncated and interval censored data will be considered.

Course Outcomes:

After completion of the course students will able to:

CO1) Understand the concept of censoring, life distributions and ageing classes.

CO2) Gained the ability to recognize the difference between parametric and non-parametric survival models.

CO3) Estimate nonparametric survival function of the data.

CO4) To estimate survival function, cumulative hazard rate function using the so-called Kaplan-Meier estimator.

CO5) Use the test of exponentiality against nonparametric classes in real life problems.

Unit1: Revision

Concepts of time, order and random censoring (left and right), survival function density function, hazard function (rate), cumulative hazard rate, mean residual life function, Equilibrium distribution function. Exponential distribution & it’s no ageing properties: Lack of memory property, constant failure rate, Cauchy-function equation, constant mean residual life function, TTT transform, identity function as a TTT transform

Revision of Ageing classes - IFR, IFRA, NBU, NBUE, DMRL, HNBUE and their duals, and inter relationship among these classes. Bathtub Failure rate.

Life distributions - Exponential Gamma, Weibull, Lognormal, Pareto, linear Failure rate, Makeham family, Lehman families (proportional hazard rate family), spacing, normalized spacing and results of an exponential distribution based on normalized spacing.

Revision of Parametric inference for complete data:

a) Exponential distribution:

b) Weibull: Obtaining MLE of scale and shape parameter of Weibull distribution and sample information matrix.

c) Gamma: Obtaining MLE of scale and shape parameter of Gamma distribution and sample information matrix.

Graphical method for checking exponetiality of data. (20L)

Unit 2:

Parametric inference for censored data:

i) Type I censoring: Exponential distribution

ii) Type II censoring: Exponential, gamma, Lognormal

iii) Random censoring: Exponential, Lehman family, Weibull distribution,

Non-Parametricestimation of survival Function

i) For complete data: Non parametric estimator of distribution function and survival function, distribution of empirical survival function, confidence band for survival function (by Using Kolmogorov – Smirnov statistics)

ii) For censored data: Actuarial estimator of survival Function, Estimator of variance of actuarial estimator (Greenwoods formula), product limit estimator and its variance, redistribution to right algorithm. (20L)

Unit 3:

Test for Exponetiality: Estimable function of degree r, Kernel, symmetric Kernel,U- statistic, variance of U- Statistic, one sample U-Statistic theorem, Hollander and Proschan Test, Test for exponetiality against positive ageing based n sample spacing, Analytical test for exponentiality against NBUE, Deshpande's Test, Two sample U- statistic theorem, Wilcoxon and Mann-Whitney test, Gehan'stest,Mantel-Haenzeltest, Log rank test, concept of covariates Semi-parametric regression for failure rate-Cox's proportional hazards model with one and several covariates. Base line model, link function, likelihood function proportional Hazard Rate model

Nelson-Aalen estimators, introduction to frailty models. (20L)

Books Recommended:

1. Deshpande, J.V, Purohit, S.G.,(2005),Life Time Data :Statistical Models and Methods
2. Klein J. P. and Moeschberger M.L. (1997) Survival Analysis: Techniques for censored andtruncated data.Springer, New York.
3. Collett D (2003) Modelling Survival Data in Medical

research 2nd edition ,Chapman andHall/CRC

4. Cox, D.R. and Oakes, D. (1984) Analysis of Survival Data, Chapman and Hall, New York.
5. Elandt-Johnson, R.E., Johnson N.L. (1980) Survival models and Data Analysis, John WileyandSons
6. Gross A.J. and Clark, V. A. (1975) Survival Distributions: Reliability Applications in theBiomedical Sciences, John Wiley andSons.
7. Miller, R.G. (1981) Survival Analysis,John Wiley andSons.
8. Therneau T M and Grambsch P M (2000) Modeling Survival data extending the Cox model.Springer, New York.
9. Duchateau L Johnson P (2008) The Frailty model. Springer, New York
10. Hanagal D D (2011) Modeling Survival Data using frailty models CRC press

23-ST 43(B) Categorical Data Analysis (from 2023-24) (4 credits)

Course Outcomes:

CO1) Appreciation of difference between linear models and logistic and log-linear models.

CO2) Knowledge of models for categorical data analysis and ability to fit them and interpret the results.

CO3) Awareness of dependence relationships amongst categorical variables.

CO4) Ability to use any related software to fit models for categorical data

Unit 1:

Introduction to Categorical data analysis: categorical response data, Probability distributions for categorical data, statistical inference for discrete data. Contingency tables: Probability structure for contingency tables, comparing proportions with 2x2 tables, odds ratio, tests for independence, exact inference, extension to three way and larger tables (15L)

Unit 2:

Generalized linear models (GLM): GLM for binary data and count data, Statistical inference and model checking, fitting GLMs. Logistic Regression: interpretation, inference, logistic regression withcategorical predictors (15L)

Unit 3:

Multiple logistic regression, building and applying logistic regression model, multi category logit models. Log-linear models for two way and three way tables, inference for log linear models, log linear-logistic connection, independence graphs and collapsibility (15L)

Unit 4:

Models for matched pairs: comparing dependent proportions, logistic regression for matched pairs, comparing margins of square contingency tables. Random effects modeling of clustered categorical data, extension to multinomial responses, hierarchical models. (15L)

Books Recommended:

1. A. Agresti, Analysis of Categorical Data, Wiley, 1990.
2. A. Agresti, An Introduction to Categorical Data Analysis, Wiley, New York, 1996.

3. E.B. Andersen, The Statistical Analysis of Categorical Data, Springer-Verlag, 1990.
4. T.J. Santner and D. Duffy, The Statistical Analysis of Discrete Data, Springer-Verlag, 1989.

23-ST-44(A): Computer Intensive Statistical Methods (from 2023-24) (4 credits)

Course objective: To study the various techniques of computation.

Course outcomes:

After completion of the course students will able to:

CO1) to apply various methods like Bootstrap, Jackknife method.

CO2) To imderstamd MCMC methods for missing values

CO3) Smoothing techniques.

Unit 1:

Bootstrap methods, estimation of sampling distribution, various types of confidence intervals,variance stabilizing transformation, Jackknife and cross-validation, Permutation tests. Bagging and Boosting methods with applications. Cross validation **(15L)**

Unit 2:

Missing Values and Imputations Techniques: Missing values and types of missingness,imputations methods for missing values, single and multiple imputations. MCMC methods formissing values, EM Algorithm and Applications: EM algorithm for incomplete data, EM algorithmfor mixture models, EM algorithm for missing values, stochastic EM algorithm. **(15L)**

Unit 3:

Review of ITM and ARM, Importance Sampling, Metropolis-Hastings andGibbs Sampling algorithms. Particle Filtering, Rejection algorithms for Approximate Bayes Computation(ABC-Rejection). **(15L)**

Unit 4:

Smoothing techniques: Kernel estimators, nearest neighbor estimators,

orthogonal and local polynomial estimators, wavelet estimators, Splines, Choice of bandwidth and other smoothing parameters. Statistical methods for Big Data analysis **(15L)**

Books Recommended:

1. Burren, Stef van (2012). Flexible Imputation of Missing Data. Chapman and Hall.
2. Chihara, L. and Hesterberg, T. (2011) Mathematical Statistics with Resampling and R. Wiley.
3. Davison, A.C. and Hinkley, D.V. (1997) Bootstrap methods and their Applications. Chapman andHall.
4. Effron, B and Hastie, T (2016). Computer-Age Statistical Inference-Algorithms, Evidence andData Science, Cambridge University Press.
5. Gilks, W. R., Richardson, S., and Spiegelhalter, D. (eds.) (1995) Markov Chain Monte Carlo inPractice. Chapman and Hall.

6. Good, P. I. (2005) Resampling Methods: A Practical Guide to Data Analysis. BirkhauserBosel.
7. Jim, A. (2009). Bayesian Computation with R, 2nd Edn, Springer.
8. McLachlan, G.J. and Krishnan, T. (2008) The EM Algorithms and Extensions. Wiley

23-ST-44(B): Statistical Analysis of Clinical Trials (from 2023-24) (4 Credits)

Course objective: To learn ethics of clinical trials, phases of clinical trials, cross over designs

Course Outcomes:

After completion of the course students will able to:

CO1) Learn data collection systems for good clinical practice

CO2) Knowledge of Pharmokintics, pharmodynamics

Unit 1:

Introduction to clinical trials: need and ethics of clinical trials, bias and random error in clinical studies, conduct of clinical trials, overview of Phase I-IV trials, multicenter trials.

Data management: data definitions, case report forms, database design, data collection systems for good clinical practice. Bioavailability, pharmacokinetics and pharmacodynamics, two-compartment model. **(15L)**

Unit 2:

Design of clinical trials: parallel vs. cross-over designs, hybrid design, cross-sectional vs. longitudinal designs, response surface experiments and group allocation design, objectives and endpoints of clinical trials, design of Phase I trials, design of single-stage and multi-stage Phase II trials. Design and monitoring of Phase III trials with sequential stopping, design of bio- equivalence trials. Inference for2x2 crossover design: Classical methods of interval hypothesis testing for bioequivalence, Bayesian methods, nonparametric methods. **(15L)**

Unit 3:

Power and sample size determination, multiplicative (or log-transformed) model, ML method ofestimation, assessment of inter and intra subject variabilities, detection of outlying subjects.

Optimal crossover designs: Balaam's design, two-sequence dual design, optimal four-period designs, assessment of bioequivalence for more than two drugs, Williams design. **(15L)**

Unit 4:

Designs based on clinical endpoints: Weighted least squares method, log-linear models, generalized estimating equations. Drug interaction study, Dose proportionality study and steady-state analysis, Interim analysis and group sequential tests, alpha spending functions, Analysis of categorical data. (12L)

Books Recommended:

1. Chow S.C. and Liu J.P. (2009). Design and Analysis of Bioavailability and bioequivalence. 3rd Ed. CRC Press.
2. Chow S.C. and Liu J.P. (2004). *Design and Analysis of Clinical Trials*. 2nd Ed. Marcel Dekker
3. Fleiss J. L. (1989). *The Design and Analysis of Clinical Experiments*. Wiley.
4. Friedman L. M. Furburg C. Demets D. L. (1998). *Fundamentals of Clinical Trials*, Springer.
5. Jennison .C. and Turnbull B. W. (1999). *Group Sequential Methods with Applications toClinical Trails,* CRC Press.
6. Marubeni .E. and Valsecchi M. G. (1994). *Analyzing Survival Data from Clinical Trials andObservational Studies,* Wiley.

23-ST 45: Practical + Project (from 2023-24) (4 credits)

Sr. No.	Title of Experiment
	Practical based on Asymptotic Inference:
1	Verification of consistency and asymptotic normality of the estimators
2	Comparing methods of estimation, MSE and sample size considerations
3	Power functions and comparison of tests & confidence intervals (LR, Wald, Rao),
	Practicals based on Econometrics and Time Series Analysis:
4A	a) Smoothing the series using various filters :Other filters, data transfer, Box- Coxtransformation,differencing, checking stationarity and normality after transformation. (b)ACF/PACF, Analysis of series and residuals, residual analysis.
5A	(a) Order selection in time series : Use of A CF/PACF and ATC,BIC, fitting of AR, MAmodels(conditional least squares or maximum likelihood) (b) Fitting of ARMA, ARIMA, SARIMA, models (conditional least squares or maximumlikelihood)
6A	(a) Forecasting using fitted linear models (recurresively) ,Holt-Winters forecasts, constructionofforecast intervals. (b) Fitting heterioscadastic models : Checking for heterioscadecity from residuals , ARCH,GARCH
	Practical based on Operations Research:
4B	Integer programming, Non linear programming
5B	Dynamic programming,
6B	CPM and PERT, Simulation, Simulation of M/M/1 queue.
	Practicals based on Survival analysis:
7A	Parametric analysis of complete data and censored data
8A	Computation of Actuarial estimator of survival function and PL – estimator and their variances.
9A	Tests for exponentiality, Nelson- Aalen estimators.
	Practical based on Categorical Data Analysis:
7B	FittingGLMs and multiple logisticregression
8B	Logistic regression for matched pairs
9B	Log leaner models for two way and three way tables
	Practicals based on Statistical Analysis of Clinical Trials:
10	Testing of hypothesis for various type of clinical trial, Power of the test and sample sizedetermination of clinical trials.
11	Estimate survival function, cumulative survival function using Kaplan and Meier estimatorand comparison of survival function using log rank test.
12	(a). Hypothesis testing and estimation of confidences interval for bioequivalence study,nonparametric methods. (b) Estimation of the Pharmacokinetic parameters clinical trial

	Project (equivalent to 8 Practicals)

Project Guidelines:

1. Project may be done by individual student or in a group of students not more than 3.
2. As far as possible, students should use **LaTeX** for dissertation document preparation and paper presentation.
3. All the students should give their details in writing to the Project Coordinator/HOD just before the commencement of the semester or during the first week of semester. This should contain Names of the group members. Title of the project and name of the guide.
4. The project coordinator/ guide shall declare the dates of two rounds of internal presentations at the beginning of semester itself.

The project guide should meet his/ her group(s) at least once in a week and keep record the meeting, an attendance and the weekly progress of the project. Submit monthly progress report toHOD/ Project co-ordinator.

5. Students should try to use real data sets for their project problems. To the maximum possible extent, text book dada sets should be avoided. If possible, students should conduct actual experiment and generate data (not by simulation) or contact some of the research organizations/ industry to get real data sets or real statistical problems they are trying to solve and be part of that.
6. I**nternal Evaluation:** Guide will work as an internal examiner. There will be two presentations rounds for the Continuous Internal Assessment (CIA). These presentations will be graded out of 15 marks each (by Guide). Students are expected to submit the presentation to Guide at least one day before the presentation. It is advisable that both the presentations shall be attended by all the faculty members and students from other groups.
7. In the first presentation, students are expected to describe their project problem, the data they are going to analyze and the objectives of their project. In addition to this, they should also mention their methodology (without much detail). **Students are to read at least TWO research papers which addresses similar kind of problems and they should include main contents of the papers in their first presentation.**
8. In the second presentation, students should discuss the results of their analysis, finding and new methodology they have introduced (if any). Students should make sure that they have something innovative in their project work.
9. Assessment will also be done on followings aspects. Timely submission of the draft of project report in the proper format which includes (title,

abstract, key words, methodology, conclusion, references, limitations and source of data etc.) is essential.

10. The completed project report in **two copies (one for the candidate and other for examiner)** should be submitted to the Project coordinator / HOD on or before the last day of teaching of the semester. HOD in consultation with guide should ensure that whether the project is **free from plagiarism** and the project is worth presenting finally for end semester examination.
11. There shall be end semester examination for 70 marks (with guidelines similar to practical examination). The individual student or group (as the case may be) will give presentation of the project as per time table for 25 minutes .The final draft of the project report copy should be given to external examiner before presentation. The project shall be evaluated by **one external examiner and (other than the guide; the examiner will be from the other college appointed by SPPU) one internal examiner jointly.**

Aspects of Assessment and marks assigned (The following are the guidelines, somemodification can be done as and when required by the examiners)

Sr. No.	Description	Marks
1	Dissertation in the proper pro forma which includes (Title, abstract, Key words, Methodology, conclusion, references, limitations and source of data etc.)	5 M
2	Appropriateness of tools used for analysis, testing the assumptions needed for analysis. methodology, program coding (if any) and numerical computations.	10 M
3	General understanding about the problem in the project and the two research papers studied	10 M
4	Presentation	5 M
5	Validity of conclusions	5 M
Total		35 M

****After the presentation within two days , students should incorporate all suggestions / corrections suggested by the external examiner as well as the guide and submit the final copy to the department (within two days after the final presentation), failing which he /she will fail the course.

****