Anekant Education Society's

Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati

(AUTONOMOUS)

Department of Statistics

Two Years Post Graduate Program

M.Sc. Data Science

Course Structure and Syllabus

Semester- I

Paper Code	Course Title	No. of Credits
PSDS111	Linear Algebra in Matlab	04
PSDS112	Probability Distributions	04
PSDS113	Optimization Techniques	04
PSDS114	Statistical Inference	04
PSDS115	Database Management System	04
PSDS116	Introduction to MATLAB and R	04

Semester- II

Paper Code	Course Title	No. of Credits
PSDS121	Design and Analysis of Experiments	04
PSDS122	Regression Analysis and Predictive Models	04
PSDS123	Statistical Quality Control	04
PSDS124	Computational Statistics	04
PSDS125	Bayesian Inference	04
PSDS126	Python and SQL Programming	04

Semester- III

Paper Code	Course Title	No. of Credits
PSDS231	Stochastic Models and Applications	04
PSDS232	Exploratory Multivariate Data Analysis	04
PSDS233	Time series analysis and Forecasting	04
PSDS234	Artificial Intelligence	04
PSDS235	Text Mining and Natural Language Processing	04
PSDS236	Data Visualization using Tableau	04

Semester- IV

Paper Code	Course Title	No. of Credits
PSDS241	Machine Learning	04
PSDS242	Discrete Data Analysis	04
PSDS243	Supply Chain & Logistics Analytics	04
PSDS244	Deep Learning	04
PSDS245	Thesis	08

PO1	**Disciplinary Knowledge:** Demonstrate comprehensive knowledge of the discipline that forms a part of a postgraduate programme. Execute strong theoretical and practical understanding generated from the specific programme in the area of work.
PO2	**Critical Thinking and Problem solving:** Exhibit the skill of critical thinking and understand scientific texts and place scientific statements and themes in contexts and also evaluate them in terms of generic conventions. Identify the problem by observing the situation closely, take actions and apply lateral thinking and analytical skills to design the solutions.
PO3	**Social competence:** Exhibit thoughts and ideas effectively in writing and orally; communicate with others using appropriate media, build effective interactive and presenting skills to meet global competencies. Elicit views of others, present complex information in a clear and concise way and help reach conclusions in group settings.
PO4	**Research-related skills and Scientific temper :** Infer scientific literature, build a sense of enquiry and able to formulate, test, analyse, interpret and establish hypothesis and research questions; and to identify and consult relevant sources to find answers. Plan and write a research paper/project while emphasizing on academics and research ethics, scientific conduct and creating awareness about intellectual property rights and issues of plagiarism.
PO5	**Trans-disciplinary knowledge:** Create new conceptual, theoretical and methodological understanding that integrates and transcends beyond discipline-specific approaches to address a common problem.
PO6	**Personal and professional competence:** Perform independently and also collaboratively as a part of a team to meet defined objectives and carry out work across interdisciplinary fields. Execute interpersonal relationships, self-motivation and adaptability skills and commit to professional ethics.
PO7	**Effective Citizenship and Ethics:** Demonstrate empathetic social concern and equity centred national development, and ability to act with an informed awareness of moral and ethical issues and commit to professional ethics and responsibility.
PO8	**Environment and Sustainability:** Understand the impact of the scientific solutions in societal and environmental contexts and demonstrate the knowledge of and need for sustainable development.
PO9	**Self-directed and Life-long learning:** Acquire the ability to engage in independent and life-long learning in the broadest context of socio-technological changes.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- II)
Paper Code : PSDS121
Paper : I
Title of Paper : Design and Analysis of Experiment
Credit : 4 credits
No. of lectures : 60

Learning Objectives:

1. The main objective of this course is to learn and understand various designs of experiments.
2. Students should be able to design and carryout various experiments and analyse the data.
3. Students should be able to apply appropriate design in real life situation.

Course outcomes:

CO 1. Students will be able to understand basic principles and various terms of Design of Experiments.

CO 2. Students will be able to apply Factorial design, fractional factorial design, confounding in real life problems.

CO 3. Students should be able to analyse the data of various experimental design.

CO 4. Make statistical inferences about population parameters based on experimental results.

CO 5. Demonstrate proficiency in implementing experimental designs using statistical software.

CO 6. Apply statistical methods to assess the significance of interactions between factors

CO 7. Understand and adhere to ethical considerations related to experimental design and data analysis.

TOPICS/CONTENTS:

Unit 1

Analysis of one-way classification modal. Analysis of two-way classification model with equal number of observations per cell with and without interactions. Analysis of two-way classification model with unequal number of observations per cell without interactions. Introduction of the Design of Experiments (DOE), The Basic Principles of DOE, Steps for Planning, Conducting and Analysing an Experiment, Principles of scientific experimentation – Basic Designs: Completely Randomized Design (CRD), Randomized Block Design (RBD)

and Latin Square Design (LSD) – Analysis of RBD (with one observation per cell, more than one but equal number of observations per cell). **(15 L)**

Unit 2

Multiple Comparisons, Multiple Range Tests, Statistical analysis of Covariance. Analysis of non- normal data using: square root transformation for counts, $Sin^{-1}(.)$ transformation for proportions, Kruskal Wallis test. **(5 L)**

Unit 3

Full Factorial experiments and their analysis, concepts of main effects, interaction effect, their graphical representation, analysis of single replicate and more than one replicates of 2^kdesignand partial confounding of 2^k,fractional experiments in 2^k. Statistical analysis of single replicate and more than one replicates of 3^kdesign, confounding and fractional experiments in 3^k. **(15 L)**

Unit 4

Balanced Incomplete Block Design (BIBD) – Types of BIBD – Simple construction methods Concept of connectedness and balancing – Intra Block analysis of BIBD. Partially Balanced Incomplete Block Design with two associate classes – intra block analysis. Split plot and strip plot design and their analysis. **(15 L)**

Unit 5

Response surface methodology (RSM): linear and quadratic model, stationary point, central composite designs(CCD), ridge systems, multiple responses, concept of rotatable designs, Box-Behnken design, optimality of designs, simplex lattice designs, simplex centroid designs. **(10 L)**

References Books:

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 Response Surfaces, Wiley.
3. Kshirsagar A.M. (1983). Linear Models, Marcel Dekker.
4. Montgomery, D.C. (2001). Design and Analysis of Experiments, Wiley.
5. Phadke, M.S. (1989). Quality Engineering using Robust Design, Prentice Hall,Englewood Cliffs, New Jersey.
6. Wu, C.F. Jeff and Hamada M. (2000). Experiments: Planning, Analysis and Parameter Design Optimization, John Wiley and Sons.
7. Bapat, R. B. (2012). Linear algebra and linear models. Springer Science & Business.

Course Outcomes	Programme Outcomes (POs)								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	2	-	2	1	-	-	1
CO2	3	3	2	-	2	1	-	-	1
CO3	3	3	2	-	2	1	-	-	1
CO4	3	3	2	3	2	1	-	-	1
CO5	3	3	2	-	2	1	-	-	1
CO6	3	3	2	-	2	1	-	-	1
CO7	3	3	2	3	2	1	3	3	1

PO1: Disciplinary Knowledge

CO1: Understand basic principles and various terms of Design of Experiments.

Justification: Strongly Related (3) - The course outcome directly involves gaining comprehensive knowledge in the specific discipline of Design of Experiments.

CO2: Apply Factorial design, fractional factorial design, confounding in real-life problems.

Justification: Strongly Related (3) - Application of factorial design and related concepts is a practical demonstration of disciplinary knowledge in experimental design.

CO3: Analyse the data of various experimental design.

Justification: Strongly Related (3) - Data analysis is a key component of disciplinary knowledge, demonstrating the ability to apply theoretical understanding in a practical context.

CO4: Make statistical inferences about population parameters based on experimental results.

Justification: Strongly Related (3) - Making statistical inferences is a core aspect of disciplinary knowledge in experimental design and data analysis.

CO5: Demonstrate proficiency in implementing experimental designs using statistical software.

Justification: Strongly Related (3) - Proficiency in using statistical software is an essential skill within the discipline, indicating a strong understanding.

CO6: Apply statistical methods to assess the significance of interactions between factors.

Justification: Strongly Related (3) - Assessing the significance of interactions aligns with the core principles of experimental design and statistical analysis.

CO7: Understand and adhere to ethical considerations related to experimental design and data analysis.

Justification: Strongly Related (3) - Adherence to ethical considerations is an integral part of disciplinary knowledge, emphasizing responsible conduct in research.

PO2: Critical Thinking and Problem Solving

All Course Outcomes contribute to critical thinking and problem-solving skills, as they involve understanding, applying, and analyzing experimental designs and data.

PO3: Social Competence

All Course Outcomes are Moderately Related (2) - While the course outcomes focus on technical skills, effective communication is necessary for presenting findings and collaborating with others in research.

PO4: Research-related Skills and Scientific Temper

CO4: Make statistical inferences about population parameters based on experimental results.

Justification: Strongly Related (3) - Inference-making aligns with research-related skills and the scientific temper required for conducting experiments.

CO7: Understand and adhere to ethical considerations related to experimental design and data analysis.

Justification: Strongly Related (3) - Adherence to ethical considerations is crucial in research and demonstrates a scientific temper.

PO5: Trans-disciplinary Knowledge

All Course Outcomes are Moderately Related (2) - While the focus is on experimental design within a specific discipline, the skills developed can have applications in various fields.

PO6: Personal and Professional Competence

All Course Outcomes contribute to personal and professional competence, indicating the ability to perform independently and collaboratively, interpersonal skills, and commitment to professional ethics.

PO7: Effective Citizenship and Ethics

CO7: Understand and adhere to ethical considerations related to experimental design and data analysis.

Justification: Strongly Related (3) - Adherence to ethical considerations aligns with effective citizenship and ethical responsibilities.

PO8: Environment and Sustainability

CO8: Understand the impact of the scientific solutions in societal and environmental contexts and demonstrate the knowledge of and need for sustainable development.

Justification: Strongly Related (3) - Understanding the impact of experimental designs aligns with considering societal and environmental contexts.

PO9: Self-directed and Life-long Learning

All Course Outcomes contribute to self-directed and lifelong learning, indicating the ability to engage in independent learning in the context of socio-technological changes.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- II)
Paper Code : PSDS122
Paper : II
Title of Paper : Regression Analysis and Predictive Modelling
Credit : 4 credits
No. of lectures : 60

Learning Objectives:

1. Gain knowledge of model construction and regression analysis.
2. Give the capability of developing relationships between variables
3. Investigate possible diagnostics in regression techniques

Course Outcome:

At the end of the course students will be able to:

CO 1. Deep comprehension of the linear and nonlinear regression models.

CO 2. Demonstrate understanding of model selection and regression modeling approaches.

CO 3. The connections between dependent and independent variables should be examined.

CO 4. Estimate the parameters and fit a model.

CO 5. Investigate possible diagnostics in regression modeling and analysis.

CO 6. Validate the model using hypothesis testing and confidence interval approach.

CO 7. Understanding advanced regression techniques, such as logistic regression for binary outcomes or Poisson regression for count data.

Unit 1

Simple Linear Regression Analysis:

Simple linear regression model, Ordinary Least Square method, generalized and weighted least squares, validating simple regression model using t, F test, developing confidence interval.

(10 L)

Unit 2

Multiple linear Regression Analysis

Concept of Multiple regression model, Ordinary Least Square method, generalized and weighted least squares, Assessing the fit of the regression line, inferences from multiple regression analysis, problem of over fitting of a model, comparing two regression model, prediction with multiple regression equation.

(15 L)

Unit 3

Model Adequacy Checking and Transformation Techniques:

Residual analysis, PRESS statistics, detection and treatment of outliers, lack of fit of the regression model, test of lack of fit, Problem of autocorrelation and heteroscedasticity. Variance stabilizing transformations, transformations to linearize the model, Box-Cox methods, transformations on the repressors variables. Multicolinearity, sources of multicollinearity, effects of multicollinearity. Multicollinearity diagnostics: examination of correlation matrix, variance Inflation factors (VIF), Eigen system analysis of X'X. Methods of dealing with Multicollinearity: **(18 L)**

Unit 4

Polynomial regression, Non-linear regression: Non-linear least squares transformation to a linear model, their uses and limitations, examination of non-linearity, initial estimates, iterative procedure, Newton-Raphson method. Generalized linear model: Link function: normal, binomial, Poisson, exponential, gamma. Logistic regression: Logit transform, ML estimation, tests of hypothesis, Wald test, LR test, score test, test for overall regression.

(17 L)

Reference Book:

1. Draper, N. R. and Smith H. (1998) Applied regression analysis 3rd edition (John Wiley)
2. Hosmer, D. W. and Lemeshow, S. (1989) Applied logistic regression (John Wiley)
3. McCullagh, P. and Nelder, J. A.(1989) Generalized linear models (Chapman and Hall)
4. Montogomery D. C., Elizabeth a. Peck, G. Geoffrey.(2003) Introduction to linear regression analysis (Wiley Eastern)
5. Neter, J.; Wasserman, W. and Kutner, M.H.(1985) Applied linear statistical models
6. Ratkowsky, D. A.(1983) Nonlinear regression modeling (Marcel Dekker)

Course Outcomes	Programme Outcomes (POs)								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	-	-	-	-	-	-	2
CO2	3	3	-	-	-	-	-	-	2
CO3	3	3	2	-	-	-	-	-	2
CO4	3	3	-	2	-	-	-	-	2
CO5	3	3	-	-	2	-	-	-	2
CO6	3	3	-	-	-	2	-	1	2
CO7	3	3	-	-	-	-	2	-	2

PO1: Disciplinary Knowledge

Justification: All Course Outcomes are Strongly Related (3) - Deep comprehension of linear and nonlinear regression models aligns with comprehensive knowledge of the discipline.

PO2: Critical Thinking and Problem Solving

Justification: All Course Outcomes are Strongly Related (3) - Understanding model selection, regression modeling approaches, and examining connections between variables involve critical thinking and problem-solving skills.

PO3: Social Competence

CO3: The connections between dependent and independent variables should be examined.

Justification: Moderately Related (2) - While the primary focus is on technical aspects, effective communication of findings may involve presenting complex information to others.

PO4: Research-related Skills and Scientific Temper

CO4: Estimate the parameters and fit a model.

Justification: Moderately Related (2) - Estimating parameters involves research-related skills, but the direct connection to scientific temper is not as strong.

PO5: Trans-disciplinary Knowledge

CO5: Investigate possible diagnostics in regression modeling and analysis.

Justification: Moderately Related (2) - Investigating diagnostics requires an understanding that can be applied beyond the discipline-specific context.

PO6: Personal and Professional Competence

CO6: Validate the model using hypothesis testing and confidence interval approach.

Justification: Moderately Related (2) - Validation using hypothesis testing involves both technical competence and a commitment to professional ethics.

PO7: Effective Citizenship and Ethics

CO7: Understanding advanced regression techniques, such as logistic regression for binary outcomes.

Justification: Moderately Related (2) - Understanding advanced techniques involves ethical considerations and aligns with effective citizenship.

PO8: Environment and Sustainability

CO7: Understand the impact of the scientific solutions in societal and environmental contexts and demonstrate the knowledge of and need for sustainable development.

Justification: Partially Related (1) - While regression models have applications, the direct connection to societal and environmental contexts is limited.

PO9 Self-directed and Life-long learning

Justifications: Moderately Related (2) Each Course Outcome contributes to a moderate level of relatedness to the Program Outcome of "Self-directed and Life-long Learning." The emphasis on foundational concepts, continuous improvement, and exploration in statistical modeling aligns with the principles of self-directed and life-long learning in the context of data science and statistics.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- II)
Paper Code : PSDS123
Paper : III
Title of Paper : Statistical Quality Control
Credit : 4 credits
No. of lectures : 60

Learning Objectives:

1. Students should be able Understand different control chart.
2. Students should be able use the methods of statistical process control.
3. To link and analyze the various sampling schemes to find the plan for quality inspection.

Course Outcomes:

Students will be able to

CO 1. Describe the DMAIC processes.

CO 2. Perform analysis of process capability and measurement system capability.

CO 3. Demonstrate the ability to design, use, and interpret multivariate control chart, synthetic control chart, non-parametric control chart.

CO 4. Learn about the construction and interpretation of control charts (also known as Shewhart charts) for monitoring process stability over time.

CO 5. Explore different types of control charts, including X-bar charts for the central tendency and R (or S) charts for variability.

CO 6. Learn about the Six Sigma methodology for process improvement.

CO 7. Calculate and interpret statistical tolerance limits to ensure product quality.

TOPICS/CONTENTS:

Unit 1

Six sigma

Concept of six sigma, methods of six sigma, DMAIC methodology, DFSS methodology, six-sigma control chart, case studies.

Control Chart:

Revision of control charts for attributes, S^2chart, X -S chart with subgroup size (i) fixed, (ii) variable. Equivalence between control chart and testing of hypothesis problem. Operating characteristic (OC curve) of control chart. Average run length (ARL).Probability of false alarm, probability of catching shift in parameter. Comparison of control chart using ARL and

OC curve. Patterns on control charts with justification and its effect on probability of false alarm. An application of control charts situations other than manufacturing.

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 at least 0.5 (ii) to get LCL >0 (iii) To have at least some defectives in sample with given confidence coefficient. (iv) Minimizing ATS () chart and OC Curve, U chart, Demerit control chart for number of defects. Nelsons control chart for low defect counts. General ideas of economic designing of control charts. Duncan's model for the economic control chart. **(15L)**

Unit 2

Process Capability analysis:

Meaning, Estimation technique for capability of a process –Capability Indices: Process capability ratios Cp, Cpk, Cpm, Cmk, Cpc – Process capability analysis using a control chart – Process capability analysis using design of experiments

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, Sigmund'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.

EWMA chart:

Chart statistic its expectation and variance. Choice of chart parameters (L). Construction and working of EWMA chart for mean and variance. EWMA chart for subgroup size > 1, Comparison of Shewhart control charts with CUSUM charts. Simulation of ARL. **(15L)**

Unit 3

Acceptance sampling –

Terminologies – Attribute sampling plan by attributes – Single sampling plan and Double sampling plan – OC, ASN, AOQ, AOQL and ATI curves –MILSTD -105E Tables

Double specification limits –

M-method, Double sampling by variables - MILSTD -414 Tables – Continuous Sampling plan – CSP-1, CSP-2, CSP-3, Wald and Wolfowitz SP-A.

Attribute Sampling plans

Producers risk, Consumers Risk, designing single sampling plan for stipulated Producers and consumers risk, OC curves under Normal, Tightened and reduces inspection, Single, Double and Multiple sampling plans in AQL systems **(15L)**

Unit 4

Other control charts

i. Synthetic control chart: Concept of run length, probability distribution of run length Confirming run length (CRL) chart for attributes, Synthetic control chart, computations of chart parameters for given ARL (0), Zero State Performance, Steady state performance, Computations of ARL (), ATS ()., Comparison of with Shewhart control chart and CUSUM charts.

ii. Non-parametric control chart: Concept, construction of non-parametric chart usingsign test.

ii. A distribution-free Shewhart Quality Control Chart Based on Singed-Rank

iii. Control charts for auto correlated observations: Need, constructions of control chart for residuals after fitting first order auto correlated model.

Hotelling T^2Chart:

Testing multivariate normality, Hotelling T^2multivariate control chart for mean vector when (i) dispersion matrix is (i) known (ii) unknown ARL (0), ARL (δ). Control chart for dispersion matrix when mean vector is (i) known (ii) unknown. T^2control chart when subgroup size n=1 **(15L)**

References Books:

1. Edward G. Schilling, Dean V. Neubauer, Acceptance Sampling in Quality Control, Second Edition, Taylor & Francis, 2009
2. Poornima M. Charantimath, Total Quality Management, 3/E, Pearson India Limited, 2017.
3. Eugene L. Grant Richard S. Leavenworth, Statistical Quality Control,7 edition, McGraw Hill Education, India, 2017.
4. Douglas C. Montgomery, Introduction to Statistical Quality Control, Seventh Edition, John Wiley and Sons, New York. 2013.
5. Wu, Yeu and Spedding (2001) A synthetic control chart for detecting fraction non confirming increases JQT Vol. 33 (1), 104-111

Course Outcomes	**Programme Outcomes (POs)**								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	-	-	-	3	3	-	3
CO2	3	3	-	-	-	3	3	-	3
CO3	3	3	2	-	-	3	3	-	3
CO4	3	3	-	2	-	3	3	-	3
CO5	3	3	-	-	2	3	3	-	3
CO6	3	3	-	-	-	3	3	-	3
CO7	3	3	-	-	-	3	3	1	3

PO1: Disciplinary Knowledge

All Course Outcomes are Strongly Related (3) - Describing the DMAIC processes and understanding process capability align with comprehensive knowledge in quality management.

PO2: Critical Thinking and Problem Solving

All Course Outcomes are Strongly Related (3) - Analyzing process and measurement system capability, interpreting control charts, and learning Six Sigma methodology involve critical thinking and problem-solving skills.

PO3: Social Competence

CO3: Demonstrate the ability to design, use, and interpret multivariate control chart, synthetic control chart, non-parametric control chart.

Justification: Moderately Related (2) - While the primary focus is on technical skills, effective communication of complex information may involve presenting findings to others.

PO4: Research-related Skills and Scientific Temper

CO4: Learn about the construction and interpretation of control charts (also known as Shewhart charts) for monitoring process stability over time.

Justification: Moderately Related (2) - Learning about control charts involves understanding scientific processes and principles of monitoring stability.

PO5: Trans-disciplinary Knowledge

CO5: Explore different types of control charts, including X-bar charts for the central tendency and R (or S) charts for variability.

Justification: Moderately Related (2) - Exploring different types of control charts contributes to knowledge that can be applied in various contexts beyond quality management.

PO6: Personal and Professional Competence

All Course Outcomes are Strongly Related (3) - Performing independently and collaboratively, executing interpersonal relationships, and committing to professional ethics are essential in quality management.

PO7: Effective Citizenship and Ethics

All Course Outcomes are Strongly Related (3) - Demonstrating empathetic social concern, equity-centered national development, and acting with awareness of moral and ethical issues align with ethical considerations in quality management.

PO8: Environment and Sustainability

CO7: Calculate.

Justification: Partially Related (1) - While the calculation aspect is more technical, understanding environmental impacts of scientific solutions may have limited direct connection.

PO9: Self-directed and Life-long Learning

All Course Outcomes are Strongly Related (3) - Acquiring the ability to engage in independent and life-long learning is inherent in the continuous improvement aspects of quality management and Six Sigma methodology.

Top of Form

SYLLABUS (CBCS) FOR M.Sc. Data Science

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- II)
Paper Code : PSDS124
Paper : IV
Title of Paper : Computational Statistics
Credit : 4 credits
No. of lectures : 60

Learning Objectives:

1. To familiarize students to learn the design of the algorithm used to apply statistical techniques like bootstrapping and simulation.
2. To identify problems that computational statistics can be used to solve include those involving optimization and resampling techniques.
3. To acquaint students with numerical integration and the simulation of random processes or variables.

Course Outcomes:

CO 1. Students can employ computational techniques to provide numerical solutions to statistical questions that are challenging or unsolvable analytically.

CO 2. Students can apply numerical techniques for transformations, for function approximation.

CO 3. Students will be able to understand and implement the Monte Carlo Studies in Statistics and random number generators.

CO 4. Construct interpolating polynomials using Lagrange interpolation and Newton's divided difference method.

CO 5. Calculating the estimator by using Jack-knife and Bootstrap, and comparing the average of these estimates to the original estimator, yielding a quantification of bias.

CO 6. Apply numerical methods to solve problems encountered in data science, machine learning, and statistical analysis.

CO 7. Use numerical optimization in the context of machine learning algorithms.

TOPICS/CONTENTS:

Unit 1

Theory of inverse transformation method (ITM) for random variate generation- definition of quantile function, its properties. Quantile function as a random variable and its distribution function. ITM based algorithms to generate random variates from standard discrete and continuous distributions. Generation of random variates using the relationships between distributions, composition and convolution methods. Algorithms for random variate generation from mixture distributions, Chi-square, t and F-distributions. Random variate generation from bivariate and conditional distributions. Theory of random number generation, testing a random number generator- run test, Kolmogorov-Smirnov test, sign test, rank test. Selection of a random number generator. **(15 L)**

Unit 2

Solutions to Non linear equations: Bisection method, Newton Raphson, Steepest descent, Quadrature interpolation, Jacobi and Gauss Seidel Methods. Simple Optimization method. Direct search, grid search, Hooke & Jeeves method Interpolatory search, Gradient search.

(15 L)

Unit 3

Numerical Differentiation: Forward and backward Difference, Error analysis: True solution, Approximate numerical solution, Causes of error. Numerical Integration: Trapezoidal rule, Simpson's Rule. Jack-knife and Bootstrap sampling. Bias and standard errors, Bootstrapping for estimation of sampling distribution. Confidence intervals, variance stabilizing transformation.

(15 L)

Unit 4

Methods to compute integrals- quadrature formula, double integration, Gaussian integration, Monte Carlo Methods: Monte Carlo integration and its application to compute expected values and probabilities, Verification of WLLN, CLT and other approximations through simulation.

(15 L)

References Books:

1. Atkinson K. E. (1989): An Introduction to Numerical Analysis. (Wiley)
2. Devroye L. (1986) : Non- Uniform Random Variate Generation. (Springer- Verlag New York)
3. Ephron B. and Tibshirani. R. J. (1994): An Introduction to the Bootstrap. (Chapman and Hall) 4. Morgan B. J. T.(1984) : Elements of Simulation. (Chapman and Hall)
4. Robert C. P. and Casella G. (1999): Monte Carlo Statistical Methods. (Springer Verlag New York, Inc.)
5. Ross. S. M. (2006): Simulation. (Academic Press Inc)
6. Rubinstein, R. Y. (1998) Modern Simulation and Modelling. (Wiley Series in Probability and Statistics)
7. William J., Kennedy, James E. Gentle. (1980): Statistical Computing. (Marcel Dekker)

Course Outcomes	**Programme Outcomes (POs)**								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	-	3	3	3	3	-	3
CO2	3	3	-	3	3	3	3	-	3
CO3	3	3	2	3	3	3	3	-	3
CO4	3	3	-	3	3	3	3	-	3
CO5	3	3	-	3	3	3	3	-	3
CO6	3	3	-	3	3	3	3	-	3
CO7	3	3	-	3	3	3	3	1	3

PO1: Disciplinary Knowledge

All Course Outcomes are Strongly Related (3) - Employing computational techniques in statistics and applying numerical methods require comprehensive knowledge in the field.

PO2: Critical Thinking and Problem Solving

All Course Outcomes are Strongly Related (3) - Applying numerical techniques, understanding Monte Carlo Studies, and using numerical methods in data science require critical thinking and problem-solving skills.

PO3: Social Competence

CO3: Students will be able to understand and implement the Monte Carlo Studies in Statistics and random number generators.

Justification: Moderately Related (2) - While the primary focus is on technical skills, effective communication may involve explaining Monte Carlo studies to others.

PO4: Research-related Skills and Scientific Temper

All Course Outcomes are Strongly Related (3) - Inferencing scientific literature, formulating hypotheses, and implementing numerical methods align with research-related skills and scientific temper.

PO5: Trans-disciplinary Knowledge

All Course Outcomes are Strongly Related (3) - Applying numerical methods in data science and machine learning contributes to trans-disciplinary knowledge.

PO6: Personal and Professional Competence

All Course Outcomes are Strongly Related (3) - Performing independently and collaboratively, executing interpersonal relationships, and committing to professional ethics are essential in computational statistics.

PO7: Effective Citizenship and Ethics

All Course Outcomes are Strongly Related (3) - Demonstrating empathy, social concern, and acting with an awareness of ethical issues align with ethical considerations in computational statistics.

PO8: Environment and Sustainability

CO7: Use numerical optimization in the context of machine learning algorithms.

Justification: Partially Related (1) - While numerical optimization is technical, the direct connection to environmental sustainability may be limited.

PO9: Self-directed and Life-long Learning

All Course Outcomes are Strongly Related (3) - Acquiring the ability to engage in independent and life-long learning is inherent in the application of computational techniques and numerical methods.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- II)
Paper Code : PSDS125
Paper : V
Title of Paper : Bayesian Inference
Credit : 4 credits
No. of lectures : 60

Learning Objectives:

1. Introduction to Bayesian inference, Bayesian approach for data analysis in a variety of applications.
2. To obtain posterior distributions for the proportion and mean.
3. To Construct Bayesian prediction intervals and write appropriate conclusions.

Course Outcomes:

CO 1. Students can To Construct Bayesian prediction intervals and write appropriate conclusions.

CO 2. Allows the incorporation of existing knowledge or beliefs through the prior distribution.

CO 3. Performs well even with small sample sizes, especially when informative priors are available.

CO 4. Use Bayes' theorem to combine the prior and likelihood, yielding the posterior distribution.

CO 5. Provides probabilistic outputs, allowing for a natural expression of uncertainty in parameter estimates.

CO 6. Apply Bayesian inference to solve real-world problems in various domains, such as finance, healthcare, and social sciences.

CO 7. Understand and apply posterior predictive checks to assess the adequacy of Bayesian models in capturing the observed data patterns.

TOPICS/CONTENTS:

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, HPD confidence 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 and asymptotic normality of posterior distributions. **(12L)**

Unit 4

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

Reference Books:

1. Bayesian Data Analysis, by Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. CRC Press/Taylor & Francis, 2013, 3rd Edition. ISBN: 9781439840955
2. Bayesian Computation with R, by Jim Albert. Springer, 2009, 2nd Edition. ISBN: 0387922970
3. A First Course in Bayesian Statistical Methods, Peter D. Hoff, 2009, New York: Springer
4. Bayesian Data Analysis. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., & Rubin, D.B. (2013). CRC press.

Course Outcomes	**Programme Outcomes (POs)**								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	3	3	3	3	3	3	-	3
CO2	3	3	3	3	3	3	3	-	3
CO3	3	3	3	3	3	3	3	-	3
CO4	3	3	3	3	3	3	3	-	3
CO5	3	3	3	3	3	3	3	-	3
CO6	3	3	3	3	3	3	3	-	3
CO7	3	3	3	3	3	3	3	1	3

PO1: Disciplinary Knowledge

All Course Outcomes are Strongly Related (3) - Constructing Bayesian prediction intervals and applying Bayesian inference require comprehensive knowledge in statistics.

PO2: Critical Thinking and Problem Solving

All Course Outcomes are Strongly Related (3) - Applying Bayesian inference, incorporating knowledge through priors, and using posterior predictive checks involve critical thinking and problem-solving skills.

PO3: Social Competence

All Course Outcomes are Strongly Related (3) - Effectively communicating thoughts, presenting complex information, and helping reach conclusions align with social competence.

PO4: Research-related Skills and Scientific Temper

All Course Outcomes are Strongly Related (3) - Inferencing scientific literature, formulating hypotheses, and applying Bayesian inference align with research-related skills and scientific temper.

PO5: Trans-disciplinary Knowledge

All Course Outcomes are Strongly Related (3) - Applying Bayesian inference in various domains contributes to trans-disciplinary knowledge.

PO6: Personal and Professional Competence

All Course Outcomes are Strongly Related (3) - Performing independently and collaboratively, executing interpersonal relationships, and committing to professional ethics are essential in Bayesian inference.

PO7: Effective Citizenship and Ethics

All Course Outcomes are Strongly Related (3) - Demonstrating empathy, social concern, and acting with an awareness of ethical issues align with ethical considerations in Bayesian inference.

PO8: Environment and Sustainability

CO7: Understand and apply posterior predictive checks to assess the adequacy of Bayesian models in capturing the observed data patterns.

Justification: Partially Related (1) - While understanding and applying checks is technical, the direct connection to environmental sustainability may be limited.

PO9: Self-directed and Life-long Learning

All Course Outcomes are Strongly Related (3) - Acquiring the ability to engage in independent and life-long learning is inherent in the application of Bayesian inference and continuous improvement in statistical methodologies.

SYLLABUS (CBCS) FOR M.Sc. Data Science

(With effect from Academic Year 2022-2023)

Class : M. Sc. (Semester- II)
Paper Code : PSDS126
Paper : VI
Title of Paper : Python and SQL Programming
Credit : 4 credits
No. of lectures : 60

Learning Objectives:

1. To understand why Python is a useful scripting language for developers.
2. To learn how to design and program Python applications.
3. To learn how to write loops and decision statements in Python.
4. To learn how to use class inheritance in Python for reusability.
5. To learn how to use exception handling in Python applications for error handling.

Course Outcome:

When students complete Intro to Programming with Python, they will be able to:

CO 1. Build basic programs using fundamental programming constructs like variables, conditional logic, looping, and functions.

CO 2. Work with user input to create fun and interactive programs.

CO 3. Learn about SQL – Structured Query Language , Build database using Data Definition Language Statements Perform basic CRUD operations using Data Manipulation Language statements like Insert, Update and Delete Write and call Stored Procedures and Functions stored in database.

CO 4. Demonstrate a solid understanding of Python's basic syntax, data types, and control structures

CO 5. Utilize NumPy for numerical operations and Pandas for data manipulation and analysis.

CO 6. Connect Python to SQL databases, execute queries, and retrieve results.

CO 7. Comprehend the principles of normalization and apply them in designing relational databases.

Sr. No.	Title of Experiments
1.	Basics of Python Language When and why to use Python for Analytics • Introduction & Installation of Python • Python Syntax, Strings, Lists and Dictionaries • Loops • Regular Expressions
2.	Scientific Libraries in Python • Numpy, Scipy
3.	Introduction to Pandas • Selecting data from Pandas Data Frame, Slicing and dicing using Pandas • GroupBY / Aggregate, Strings with Pandas, Cleaning up messy data with Pandas, Dropping Entries, Selecting Entries
4.	Data Manipulation using Pandas • Data Alignment • Sorting and Ranking • Summary Statistics • Missing values • Merging data • Concatenation • Combining Data Frames • Pivot • Duplicates • Binning
5.	Data visualization on using matplotlib and seaborn libraries • Scatter plot, Line plot, Bar plot, Histogram, Box plot, Pair plot
6.	Control structures using Toyota Corolla dataset • if-else family, for loop, for loop with if break, while loop
7.	Introduction to Database Management System This module introduces you to the database, the need for databases, and their examples. Further, you will learn about Database Management Systems and its history. Lastly, you will go through various Database Management System softwares
8.	Types of Database Management System • This chapter will cover various types of DBMS, including Hierarchical, Network, Relational, and Object-Oriented Databases. You will also be familiarized with several advantages of DBMS.
9.	Introduction to SQL • This chapter will brief you on the introduction to SQL and how to install it on your system. • In this chapter, you will learn how to implement various types of Commands in MySQL, such as DDL, DQL, DML, DCL, and TCL, with hands-on demos.
10.	Filter Record in MySQL • In this chapter, you will learn how to filter the records using the WHERE clause in MySQL, Operation in MySQL
11.	Pattern Matching in MySQL

	Here, you will learn to find patterns using the LIKE operator with the WHERE clause
12.	Null Values in MySQL In the final chapter, you will learn to insert a new record or update a record to an optional field without adding a value with the help of a Null value

Course Outcomes	**Programme Outcomes (POs)**								
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9
CO1	3	-	-	3	-	3	3	-	3
CO2	3	2	1	3	-	3	3	-	3
CO3	3	-	-	3	-	3	3	-	3
CO4	3	-	-	3	-	3	3	1	3
CO5	3	-	-	3	-	3	3	-	3
CO6	3	-	-	3	2	3	3	-	3
CO7	3	-	-	3	-	3	3	-	3

PO1: Disciplinary Knowledge

All Course Outcomes are Strongly Related (3) - Building basic programs, understanding SQL, and working with Python demonstrate comprehensive knowledge in programming.

PO2: Critical Thinking and Problem Solving

CO2: Work with user input to create fun and interactive programs.

Justification: Moderately Related (2) - Creating interactive programs requires critical thinking about user needs and preferences.

PO3: Social Competence

CO2: Work with user input to create fun and interactive programs.

Justification: Partially Related (1) - While the focus is on technical skills, creating interactive programs involves considering user experience, aligning with social competence.

PO4: Research-related Skills and Scientific Temper

All Course Outcomes are Strongly Related (3) - Building a sense of inquiry, formulating hypotheses, and working with databases align with research-related skills and scientific temper.

PO5: Trans-disciplinary Knowledge

CO6: Connect Python to SQL databases, execute queries, and retrieve results.

Justification: Moderately Related (2) - Connecting Python to databases involves integration across different technologies, contributing to trans-disciplinary knowledge.

PO6: Personal and Professional Competence

All Course Outcomes are Strongly Related (3) - Performing independently, collaborating in teams, and committing to professional ethics are inherent in programming skills.

PO7: Effective Citizenship and Ethics

All Course Outcomes are Strongly Related (3) - Demonstrating empathy, social concern, and acting with an awareness of moral and ethical issues align with ethical considerations in programming.

PO8: Environment and Sustainability

PO9: Self-directed and Life-long Learning

All Course Outcomes are Strongly Related (3) - Acquiring the ability to engage in independent and life-long learning is inherent in programming skills and keeping up with evolving technologies.