Anekant Education Society's

Tuljaram Chaturchand College

of Arts, Science and Commerce, Baramati

(Autonomous)

M.Sc. Degree Program in Data Science

(Faculty of Science & Technology)

CBCS Syllabus

M.Sc.(Data Science) Part – I Semester – II

For Department of Statistics

Tuljaram Chaturchand Collegeof Arts, Science and Commerce, Baramati

Choice Based Credit System Syllabus (2023 Pattern)

(As Per NEP 2020)

To be implemented from Academic Year 2024-2025

Program Outcomes for M.Sc.

PO1. Comprehensive Knowledge and Understanding:

Postgraduates will possess a profound understanding of their field, encompassing foundational theories, methodologies, and key concepts within a multidisciplinary context.

PO2. Practical, Professional, and Procedural Knowledge:

Postgraduates will acquire practical skills and expertise necessary for professional tasks, including industry standards, regulations, and ethical considerations, with effective application in real-world scenarios.

PO3. Entrepreneurial Mindset, Innovation, and Business Understanding:

Postgraduates will cultivate an entrepreneurial mindset, identify opportunities, foster innovation, and understand business principles, market dynamics, and risk management strategies.

PO4. Specialized Skills, Critical Thinking, and Problem-Solving:

Postgraduates will demonstrate proficiency in technical skills, analytical abilities, effective communication, and leadership, adapting and innovating in response to changing circumstances.

PO5. Research, Analytical Reasoning, and Ethical Conduct:

Postgraduates will exhibit observational and inquiry skills, formulate research questions, utilize appropriate methodologies for data analysis, and adhere to research ethics while effectively reporting findings.

PO6. Communication, Collaboration, and Leadership:

Postgraduates will effectively communicate complex information, collaborate in diverse teams, demonstrate leadership qualities, and facilitate cooperative efforts toward common goals.

PO7. Digital Proficiency and Technological Skills:

Postgraduates will demonstrate proficiency in using ICT, accessing information sources, analyzing data using appropriate software, and adapting to technological advancements.

PO8. Multicultural Competence, Inclusive Spirit, and Empathy:

Postgraduates will engage effectively in multicultural settings, respect diverse perspectives, lead diverse teams, and demonstrate empathy and understanding of others' perspectives and emotions.

PO9. Value Inculcation, Environmental Awareness, and Ethical Practices:

Postgraduates will embrace ethical and moral values, practice responsible citizenship, recognize and address ethical issues, and promote sustainability and environmental conservation.

PO10. Autonomy, Responsibility, and Accountability:

Postgraduates will apply knowledge and skills independently, manage projects effectively, and demonstrate responsibility and accountability in work and learning contexts, contributing to societal well-being.

Anekant Education Society's

Tuljaram Chaturchand College,Baramati

(Autonomous)

Board of Studies (BOS) in Statistics

From 2022-23 to 2024-25

Sr.No.	Name	Designation
1.	Prof. Dr. Vikas C. Kakade	Chairman
2.	Prin. Dr. Avinash S. Jagtap	Member
3.	Dr. Neeta K. Dhane	Member
4.	Dr.Vaishali V. Patil	Member
5.	Mrs.Sarita D. Wadkar	Member(Ad hoc)
6.	Mr.Chandrashekhar P. Swami	Member
7.	Ms.Priti M. Mohite	Member(Ad hoc)
8.	Ms.Nilambari A. Jagtap	Member (Ad hoc)
9.	Miss.Kalyani C. Kale	Member (Ad hoc)
10.	Ms.Pooja S. Zanjurne	Member (Ad hoc)
11.	Dr.Akanksha S. Kashikar	Vice-Chancellor Nominee
12.	Prin. Dr.Rajendra G. Gurao	Expert from other University
13.	Mr.RohanKoshti	Expert from other University
14.	Mr.SaurabhKadam	Industry Expert
15.	Dr. Jaya L. Limbore	Meritorious Alumni
16.	Miss. Priya N. Rakate	Invitee Member
17.	Ms.Ankita G. Deshmukh	Invitee Member
18.	Ms.Shital B. Choudhar	Invitee Member

19.	Miss. Kiran Banda (M.Sc. II)	Student Representative
20.	Mr.RushikeshPandhare (M.Sc. II)	Student Representative
21.	Mr. Bharat Jambhulkar (TYBSc)	Student Representative
22.	Miss. Prapti Mane (TYBSc)	Student Representative

Credit Distribution Structure for M.Sc.(Data Science)Part-I

<table>
<tr><th rowspan="2">Level</th><th rowspan="2">Semester</th><th colspan="2">Major</th><th rowspan="2">Research Methodology (RM)</th><th rowspan="2">OJT/FP</th><th rowspan="2">RP</th><th rowspan="2">Cum. Cr.</th><th rowspan="2">Degree</th></tr>
<tr><th>Mandatory</th><th>Electives</th></tr>
<tr><td rowspan="8">6.0</td><td rowspan="4">Sem-I</td><td>DSC-501-MJM: Probability and Statistics for Data Science(Credit 04)</td><td rowspan="4">DSC-511-MJE (A): Data Base Management System

DSC -511-MJE (B): Stochastic Models and Applications (Credit 04)</td><td rowspan="4">DSC -521-RM: Research Methodology (Credit 04)</td><td rowspan="4">--</td><td rowspan="4">--</td><td rowspan="4">20</td><td rowspan="8">PG Diploma (after 3 Year Degree)</td></tr>
<tr><td>DSC-502-MJM: Data Analytics Using R (Credit 04)</td></tr>
<tr><td>DSC -503-MJM: Data Science Practical – I (Credit 02)</td></tr>
<tr><td>DSC-504-MJM: Data Science Practical – II (Credit 02)</td></tr>
<tr><td rowspan="4">Sem-II</td><td>DSC -551-MJM: Machine Learning and Artificial intelligence(Credit 04)</td><td rowspan="4">DSC -561-MJE (A): Bayesian Inference
DSC -561-MJE (B): Computational Statistics (Credit 04)</td><td rowspan="4">--</td><td rowspan="4">DSC-581-OJT/FP: On Job Training/ Field Project</td><td rowspan="4">--</td><td rowspan="4">20</td></tr>
<tr><td>DSC-552-MJM: Regression Analysis and Predictive Models (Credit 04)</td></tr>
<tr><td>DSC-553-MJM: Data Science Practical – III (Credit 02)</td></tr>
<tr><td>STA -554-MJM: Data Science Practical – IV (Credit 02)</td></tr>
</table>

Credit Distribution Structure for M.Sc.(Data Science) Part-II

Level	Semester	Major: Mandatory	Major: Electives	Research Methodology (RM)	OJT /FP	RP	Cum. Cr.	Degree
6.5	Sem-III	DSC-601-MJM: Exploratory Multivariate Data Analysis (Credit 04)	DSC-611-MJE(A): Business and Project Management (Credit 02)	--	--	DSC-621-RP: Research Project (Credit 04)	**20**	PG Diploma (after 3 Year Degree)
		DSC-602-MJM: Time Series Analysis and Forecasting (Credit 04)	DSC-611-MJE(B): Text Mining and Natural Language Processing (Credit 02)					
		DSC-503-MJM: Statistics Practical – V (Credit 02)	DSC-612-MJE(A): Practical Based on Business and Project Management (Credit 02)					
		DSC-504-MJM: Statistics Practical – VI (Credit 02)	DSC-612-MJE(B): Practical Based on Text Mining and NLP (Credit 02)					
	Sem-IV	DSC-651-MJM: LLMs and Generative AI (Credit 04)	DSC-661-MJE (A): Supply Chain and Logistics Analytics (Credit 02)	--	--	DSC-681-RP: Research Project (Credit 06)	**20**	
		DSC-652-MJM: Deep Learning (Credit 04)	DSC-661-MJE(B): Discrete Data Analysis (Credit 02)					
		DSC-653-MJM: Data Science Practical – VI (Credit 02)	DSC-662-MJE (A): Introduction to Hadoop (Credit 02)					
			DSC-662-MJE (A): Web Application Development (Credit 02)					

Course Structure for M.Sc. Part-I (Data Science)(2023 Pattern)

<table>
<tr><th>Sem</th><th>Course Type</th><th>Course Code</th><th>Course Title</th><th>Theory/ Practical</th><th>No. of Credits</th></tr>
<tr><td rowspan="7">I</td><td>Major (Mandatory)</td><td>DSC-501-MJM</td><td>Probability Distributions</td><td>Theory</td><td>04</td></tr>
<tr><td>Major (Mandatory)</td><td>DSC -502-MJM</td><td>Statistical Inference</td><td>Theory</td><td>04</td></tr>
<tr><td>Major (Mandatory)</td><td>DSC -503-MJM</td><td>Data Science Practical – I</td><td>Practical</td><td>02</td></tr>
<tr><td>Major (Mandatory)</td><td>DSC -504-MJM</td><td>Data Science Practical – II</td><td>Practical</td><td>02</td></tr>
<tr><td rowspan="2">Major (Elective)</td><td>DSC-511-MJE (A)</td><td>Data Base Management System</td><td>Theory</td><td rowspan="2">04</td></tr>
<tr><td>DSC -511-MJE (B)</td><td>Stochastic Models and Applications</td><td>Theory</td></tr>
<tr><td>Research Methodology (RM)</td><td>DSC -521-RM</td><td>Research Methodology</td><td>Theory</td><td>04</td></tr>
<tr><td></td><td colspan="4">Total Credits Semester I</td><td>20</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td rowspan="7">II</td><td>Major (Mandatory)</td><td>DSC -551-MJM</td><td>Machine Learning and Artificial Intelligence</td><td>Theory</td><td>04</td></tr>
<tr><td>Major (Mandatory)</td><td>DSC-552-MJM</td><td>Regression Analysis and Predictive Models</td><td>Theory</td><td>04</td></tr>
<tr><td>Major (Mandatory)</td><td>DSC-553-MJM</td><td>Data Science Practical – III</td><td>Practical</td><td>02</td></tr>
<tr><td>Major (Mandatory)</td><td>DSC-554-MJM</td><td>Data Science Practical – IV</td><td>Practical</td><td>02</td></tr>
<tr><td rowspan="2">Major (Elective)</td><td>DSC -561-MJE (A)</td><td>Bayesian Inference</td><td>Theory</td><td rowspan="2">04</td></tr>
<tr><td>DSC -561-MJE (B)</td><td>Computational Statistics</td><td>Theory</td></tr>
<tr><td>On Job Training (OJT)/Field Project (FP)</td><td>DSC -581-OJT/FP</td><td>On Job Training
Field Project</td><td>Training/Project</td><td>04</td></tr>
<tr><td></td><td colspan="4">Total Credits Semester-II</td><td>20</td></tr>
<tr><td colspan="5">Cumulative Credits Semester I and II</td><td>40</td></tr>
</table>

Course Structure for M.Sc. Part-II (Data Science)(2023 Pattern)

Sem	Course Type	Course Code	Course Title	Theory/ Practical	No. of Credits
III	Major (Mandatory)	DSC-601-MJM	Exploratory Multivariate Data Analysis	Theory	04
	Major (Mandatory)	DSC -602-MJM	Time Series Analysis and Forecasting	Theory	04
	Major (Mandatory)	DSC -603-MJM	Data Science Practical – V	Practical	02
	Major (Mandatory)	DSC -604-MJM	Data Science Practical – VI	Practical	02
	Major (Elective)	DSC-611-MJE(A)	Business and Project Management	Theory	02
		DSC -611-MJE(B)	Text Mining and Natural Language Processing	Theory	
		DSC-612-MJE (A)	Practical Based on Business and Project Management	Practical	02
		DSC -612-MJE (B)	Practical Based on Text Mining and NLP	Practical	
	Research Project (RP)	DSC -621-RP	Research Project	Project	04
	Total Credits Semester III				**20**
IV	Major (Mandatory)	DSC -651-MJM	LLMs and Generative AI	Theory	04
	Major (Mandatory)	DSC-652-MJM	Deep Learning	Theory	04
	Major (Mandatory)	DSC-653-MJM	Data Science Practical – VI	Practical	02
	Major (Elective)	DSC -661-MJE (A)	Supply Chain and Logistics Analytics	Theory	02
		DSC -661-MJE (B)	Discrete Data Analysis	Theory	
		DSC -662-MJE (A)	Introduction to Hadoop	Practical	02
		DSC -662-MJE (B)	Web Application Development	Practical	
	Research Project (RP)	DSC -581-RP	Supply Chain and Logistics Analytics	Project	06
	Total Credits Semester-IV				**20**
Cumulative Credits Semester III and IV					**40**

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – I
Semester	: II
Course Type	:Major Mandatory Theory
Course Name	: Machine Learning and Artificial intelligence
Course Code	: DSC-551-MJM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. To comprehend the differences between structured, unstructured, and semi-structured data.
2. To distinguish between different types of data (discrete, continuous) and levels of measurement (nominal, ordinal, ratio, interval).
3. To identify and address data inconsistencies such as missing values and duplicates.
4. To understand the steps involved in model building.
5. To grasp the foundational concepts of supervised, unsupervised, and reinforcement learning.
6. To understand and apply different model evaluation techniques and selection methods, including AIC, BIC, CIC, DIC, and confusion matrix.
7. To differentiate AI techniques from non-AI techniques and understand the representation of intelligent agents and environments.

Course Outcomes:

Bytheendof thecourse,studentswillbeableto:

CO1 Students will be able to classify and differentiate between structured, unstructured, and semi-structured data.

CO2 Students will be able to identify and categorize various types of data and levels of measurement.

CO3 Students will be proficient in detecting and correcting data inconsistencies.

CO4 Students will be able to implement and evaluate classification and regression algorithms.

CO5 Students will gain hands-on experience with algorithms such as Linear

Regression, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, Gradient Boosting, Ada Boosting, and Bagging.

CO6 Students will be able to apply clustering and association techniques in practical scenarios.

CO7 Students will have a foundational understanding of AI, including its definition, significance, and historical development.

Topics and Learning Points

Unit 1

Introduction to big data- structured, unstructured, semi structured, Types of data- discrete, continuous, Levels of measurement- nominal, ordinal, ratio, interval, Data inconsistencies- missing values, duplicate, steps in model building, supervised learning, unsupervised learning, reinforcement learning **(15L)**

Unit 2

Supervised machine learning- classification, Regression

Linear Regression, Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Naïve Bayes, Gradient Boosting, Ada Boosting, Bagging.**(15L)**

Unit 3

Unsupervised machine learning- clustering, association

K means clustering, Principal component analysis, Apriori algorithm, Hierarchical, Gaussian mixture model.

Model evaluation techniques, model selection methods AIC,BIC, CIC, DIC, confusion matrix, self-organizing method (SOM), EM Model.**(15L)**

Unit 4

Introduction to Artificial Intelligence**-**Overview of AI: Definition, significance, and historical context, AI Representation: Differences between AI and Non-AI techniques, Intelligent Agents and Environments: Types, structure, and concept of rationality.

Problem Solving and Search Strategies- State Space Search: Problem formulation and state space representation, Uninformed Search Strategies- Breadth First Search (BFS), Depth First Search (DFS), Iterative Deepening DFS , Knowledge Representation and Logic**(15L)**

References:

1. Jiawei Han, MichelineKamber, Jian Pei, Data Mining: Concepts and Techniques, 3rd Edition.
2. Margaret H. Dunham, S. Sridhar, Data Mining - Introductory and Advanced Topics, Pearson Education
3. R.O. Duda, P.E. Hart, D.G. Stork., Pattern Classification, Second edition. John Wiley and Sons,2000.
4. Christopher M. Bishop, Pattern Recognition and Machine Learning, Springer 2006
5. Ian H.Witten, Data Mining: Practical Machine Learning Tools and Techniques, Eibe Frank Elsevier /(Morgan Kauffman)
6. Bing Liu: Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, Springer (2006).
7. Ivan Bratko : "Prolog Programming For Artificial Intelligence" , 2nd Edition Addison Wesley, 1440.
8. Eugene, Charniak, Drew Mcdermott: "Introduction to Artificial Intelligence.", Addison Wesley
9. Patterson: ―Introduction to AI and Expert Systems‖, PHI
10. Nilsson: ―Principles of Artificial Intelligence‖, Morgan Kaufmann.
11. Carl Townsend, ―Introduction to turbo Prolog‖, Paperback, 1483 6. Jacek M. Zurada, Introduction to artificial neural systems, Jaico Publication.

ProgrammeOutcomes andCourseOutcomesMapping:

	Programme Outcomes(POs)									
Course Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10
CO1	3	2	1	2	2	1	2	1	1	2
CO2	3	2	1	2	2	1	2	1	1	2
CO3	3	2	1	2	2	1	2	1	1	2
CO4	3	3	2	3	3	2	3	1	2	3
CO5	3	3	2	3	3	2	3	1	2	3
CO6	3	3	2	3	3	2	3	1	2	3
CO7	3	2	2	3	3	1	2	1	2	3

PO1: Comprehensive Knowledge and Understanding

CO1 (3): Strongly related as understanding different data types is fundamental.

CO2 (3): Strongly related as identifying and categorizing data types and levels is crucial.

CO3 (3): Strongly related due to the importance of data quality in foundational knowledge.

CO4 (3): Strongly related as implementing machine learning algorithms is a core component.

CO5 (3): Strongly related as practical experience with various algorithms enhances foundational knowledge.

CO6 (3): Strongly related as practical application of clustering and association techniques is key.

CO7 (3): Strongly related due to the foundational understanding of AI concepts.

PO2: Practical, Professional, and Procedural Knowledge

CO1 (2): Moderately related as it involves practical classification of data.

CO2 (2): Moderately related as it involves categorizing data for practical use.

CO3 (2): Moderately related due to the practical skill of detecting and correcting inconsistencies.

CO4 (3): Strongly related as implementing algorithms is a professional skill.

CO5 (3): Strongly related as hands-on experience is crucial for professional tasks.

CO6 (3): Strongly related as practical application is essential for professional scenarios.

CO7 (2): Moderately related due to practical implications of understanding AI.

PO3: Entrepreneurial Mindset, Innovation, and Business Understanding

CO1 (1): Partially related as understanding data types can help identify business opportunities.

CO2 (1): Partially related for similar reasons as CO1.

CO3 (1): Partially related as data quality can impact business decisions.

CO4 (2): Moderately related as machine learning can drive innovation.

CO5 (2): Moderately related as advanced algorithms can foster innovation.

CO6 (2): Moderately related as clustering and association techniques can identify new opportunities.

CO7 (2): Moderately related due to the potential for AI-driven innovation.

PO4: Specialized Skills, Critical Thinking, and Problem-Solving

CO1 (2): Moderately related as classification requires critical thinking.

CO2 (2): Moderately related as categorization involves analytical skills.

CO3 (2): Moderately related as detecting inconsistencies requires problem-solving.

CO4 (3): Strongly related due to the technical skills required for implementing algorithms.

CO5 (3): Strongly related due to the need for specialized skills in various algorithms.

CO6 (3): Strongly related as applying techniques involves critical thinking.

CO7 (3): Strongly related due to the problem-solving nature of AI.

PO5: Research, Analytical Reasoning, and Ethical Conduct

CO1 (2): Moderately related as it provides a basis for analytical reasoning.

CO2 (2): Moderately related for similar reasons as CO1.

CO3 (2): Moderately related due to the analytical aspect of detecting inconsistencies.

CO4 (3): Strongly related as machine learning involves research and analytical reasoning.

CO5 (3): Strongly related due to the research required for understanding algorithms.

CO6 (3): Strongly related as unsupervised learning involves research skills.

CO7 (3): Strongly related due to the research aspect of AI understanding.

PO6: Communication, Collaboration, and Leadership

CO1 (1): Partially related as it aids in communicating data classifications.

CO2 (1): Partially related as it involves communicating data categorizations.

CO3 (1): Partially related due to the need to communicate data quality issues.

CO4 (2): Moderately related as explaining algorithms requires communication skills.

CO5 (2): Moderately related as collaboration is needed for implementing complex algorithms.

CO6 (2): Moderately related due to the collaborative aspect of applying techniques.

CO7 (1): Partially related as basic AI knowledge supports communication.

PO7: Digital Proficiency and Technological Skills

CO1 (2): Moderately related as it involves understanding digital data.

CO2 (2): Moderately related as it involves digital data categorization.

CO3 (2): Moderately related due to the use of technology in data correction.

CO4 (3): Strongly related as implementing algorithms requires technological skills.

CO5 (3): Strongly related as various algorithms are technology-driven.

CO6 (3): Strongly related as clustering and association use advanced software.

CO7 (2): Moderately related due to AI's technological nature.

PO8: Multicultural Competence, Inclusive Spirit, and Empathy

CO1 (1): Partially related as data types can be multicultural.

CO2 (1): Partially related for similar reasons as CO1.

CO3 (1): Partially related due to inclusive data practices.

CO4 (1): Partially related as algorithms can be applied to diverse datasets.

CO5 (1): Partially related due to the potential for inclusive applications.

CO6 (1): Partially related as clustering can identify diverse patterns.

CO7 (1): Partially related due to AI's broad applications.

PO9: Value Inculcation, Environmental Awareness, and Ethical Practices

CO1 (1): Partially related as ethical data classification is important.

CO2 (1): Partially related due to ethical considerations in data categorization.

CO3 (1): Partially related as ethical data practices involve correcting inconsistencies.

CO4 (2): Moderately related as algorithm implementation must be ethical.

CO5 (2): Moderately related due to ethical implications in algorithm use.

CO6 (2): Moderately related as unsupervised learning must be ethically conducted.

CO7 (2): Moderately related due to AI's ethical considerations.

PO10: Autonomy, Responsibility, and Accountability

CO1 (2): Moderately related as autonomous data classification is important.

CO2 (2): Moderately related due to the responsibility in data categorization.

CO3 (2): Moderately related as data correction involves accountability.

CO4 (3): Strongly related as algorithm implementation requires autonomy.

CO5 (3): Strongly related due to the responsibility in using algorithms.

CO6 (3): Strongly related as applying techniques requires accountability.

CO7 (3): Strongly related due to AI's autonomous nature.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – I
Semester	: II
Course Type	:Major Mandatory Theory
Course Name	: Regression Analysis and Predictive Models
Course Code	: DSC-552-MJM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. To provide a comprehensive understanding of the fundamental concepts of regression analysis.
2. To explain the importance of regression in statistical modeling and its applications in various fields.
3. To interpret and communicate the results of simple linear regression models.
4. Enable students to use multiple linear regression for practical problem-solving.
5. Explain the importance of residual analysis in assessing model adequacy.
6. Discuss when and how to apply these models to capture more complex relationships in data.
7. To provide an overview of Generalized Linear Models (GLM) and their significance in statistical analysis.

Course Outcomes:

By 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.

Topics and Learning Points

Unit1 **(10 L)**

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.

Unit 2 **(15 L)**

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.

Unit3 **(18 L)**

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:

Unit4 **(17 L)**

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.

References:

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)

ProgrammeOutcomes andCourseOutcomesMapping:

	Programme Outcomes(POs)									
Course Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10
CO1	3	2	3	3	2	3	1	2	3	3
CO2	3	3	2	3	3	2	3	1	2	3
CO3	3	3	2	3	3	2	3	1	2	3
CO4	3	3	2	3	3	2	3	1	2	3
CO5	3	3	2	3	3	2	3	1	2	3
CO6	3	3	2	3	3	2	3	1	2	3

PO1: Comprehensive Knowledge and Understanding

CO1 (3): Strongly related as it requires deep comprehension of both linear and nonlinear regression models, foundational in data science.

CO2 (3): Strongly related due to the necessity of understanding various model selection and regression modeling approaches.

CO3 (3): Strongly related as it involves examining the connections between dependent and independent variables, a key concept.

CO4 (3): Strongly related as estimating parameters and fitting models are fundamental concepts.

CO5 (3): Strongly related as investigating diagnostics is crucial for thorough model analysis.

CO6 (3): Strongly related as validating models with hypothesis testing and confidence intervals is essential.

CO7 (3): Strongly related as understanding advanced regression techniques is fundamental for comprehensive knowledge.

PO2: Practical, Professional, and Procedural Knowledge

CO1 (3): Strongly related due to practical application of linear and nonlinear regression models.

CO2 (3): Strongly related as model selection and regression modeling approaches are practical skills.

CO3 (3): Strongly related as examining variable connections has practical implications.

CO4 (3): Strongly related as parameter estimation and model fitting are practical tasks.

CO5 (3): Strongly related as regression diagnostics are practical tools in professional tasks.

CO6 (3): Strongly related as model validation using hypothesis testing is a practical procedure.

CO7 (3): Strongly related as advanced regression techniques are essential in professional practice.

PO3: Entrepreneurial Mindset, Innovation, and Business Understanding

CO1 (2): Moderately related as regression models can identify business opportunities.

CO2 (2): Moderately related as model selection can drive innovation in business processes.

CO3 (2): Moderately related as understanding variable connections can inform business strategies.

CO4 (2): Moderately related as model fitting can help in business decision-making.

CO5 (2): Moderately related as diagnostics can lead to innovative solutions.

CO6 (2): Moderately related as validating models can enhance business strategies.

CO7 (2): Moderately related as advanced techniques like logistic regression can be applied in business contexts.

PO4: Specialized Skills, Critical Thinking, and Problem-Solving

CO1 (3): Strongly related as understanding regression models involves critical thinking.

CO2 (3): Strongly related as model selection requires problem-solving skills.

CO3 (3): Strongly related as examining variable connections involves critical analysis.

CO4 (3): Strongly related as parameter estimation and model fitting require specialized skills.

CO5 (3): Strongly related as diagnostics require critical problem-solving.

CO6 (3): Strongly related as model validation involves critical thinking.

CO7 (3): Strongly related as advanced regression techniques require specialized knowledge.

PO5: Research, Analytical Reasoning, and Ethical Conduct

CO1 (3): Strongly related as regression models are fundamental in research.

CO2 (3): Strongly related as understanding model selection involves analytical reasoning.

CO3 (3): Strongly related as examining variable connections is key in research.

CO4 (3): Strongly related as estimating parameters involves analytical reasoning.

CO5 (3): Strongly related as investigating diagnostics is crucial for research accuracy.

CO6 (3): Strongly related as validating models is essential for ethical research.

CO7 (3): Strongly related as advanced techniques are crucial in research methodologies.

PO6: Communication, Collaboration, and Leadership

CO1 (2): Moderately related as understanding regression models aids in communicating complex concepts.

CO2 (2): Moderately related as model selection understanding enhances collaborative discussions.

CO3 (2): Moderately related as examining variable connections can improve team understanding.

CO4 (2): Moderately related as parameter estimation and model fitting results need effective communication.

CO5 (2): Moderately related as investigating diagnostics involves collaborative analysis.

CO6 (2): Moderately related as validating models requires clear communication of results.

CO7 (2): Moderately related as advanced regression techniques can be discussed in collaborative environments.

PO7: Digital Proficiency and Technological Skills

CO1 (3): Strongly related as regression models require the use of statistical software.

CO2 (3): Strongly related as model selection involves proficiency with technology.

CO3 (3): Strongly related as examining variable connections requires software tools.

CO4 (3): Strongly related as parameter estimation and model fitting require technological skills.

CO5 (3): Strongly related as investigating diagnostics involves using software.

CO6 (3): Strongly related as model validation requires technological proficiency.

CO7 (3): Strongly related as advanced regression techniques involve digital skills.

PO8: Multicultural Competence, Inclusive Spirit, and Empathy

CO1 (1): Partially related as understanding regression models can consider diverse datasets.

CO2 (1): Partially related as model selection can be applied to diverse contexts.

CO3 (1): Partially related as examining variable connections can include diverse perspectives.

CO4 (1): Partially related as model fitting can involve multicultural data.

CO5 (1): Partially related as diagnostics can be applied to diverse datasets.

CO6 (1): Partially related as model validation can consider multicultural contexts.

CO7 (1): Partially related as advanced regression techniques can be applied to diverse datasets.

PO9: Value Inculcation, Environmental Awareness, and Ethical Practices

CO1 **(2):** Moderately related as understanding regression models includes ethical considerations.

CO2 (2): Moderately related as model selection should follow ethical practices.

CO3 (2): Moderately related as examining variable connections involves ethical analysis.

CO4 (2): Moderately related as model fitting should follow ethical guidelines.

CO5 (2): Moderately related as diagnostics require ethical practices.

CO6 (2): Moderately related as validating models involves ethical considerations.

CO7 (2): Moderately related as advanced regression techniques should follow ethical guidelines.

PO10: Autonomy, Responsibility, and Accountability

CO1 (3): Strongly related as understanding regression models requires independent analysis.

CO2 (3): Strongly related as model selection involves responsibility in choosing the correct approach.

CO3 (3): Strongly related as examining variable connections requires accountable analysis.

CO4 (3): Strongly related as parameter estimation and model fitting involve autonomous work.

CO5 (3): Strongly related as diagnostics require responsibility and accountability.

CO6 (3): Strongly related as validating models involves autonomous and responsible analysis.

CO7 (3): Strongly related as advanced regression techniques require accountability in their application.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – I
Semester	: II
Course Type	:Major Mandatory Practical
Course Name	: Data Science Practical – III
Course Code	: DSC-553-MJM
No. of Credits	: 2
No. of Teaching Hours	: 60

CourseObjectives:

1. To develop skills in data understanding, cleaning, and visualization using advanced tools like Tableau and Power BI.
2. To master unsupervised learning techniques such as clustering, Principal Component Analysis (PCA), Factor Analysis, and Self-Organizing Maps (SOM).
3. To gain proficiency in various supervised learning techniques including Linear Discriminant Analysis, Logistic Regression, Bayes Classifier, k-Nearest Neighbors, Classification and Regression Tree (CART), Artificial Neural Networks (ANN), and Support Vector Machines (SVM).
4. To learn and apply techniques for enhancing model performance and improving classification accuracy.
5. To understand and apply association rule mining techniques for market basket analysis.
6. To gain practical experience in advanced machine learning techniques including Computer Vision using Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).
7. To gain proficiency in various supervised learning techniques including Artificial Neural Networks (ANN), and Support Vector Machines (SVM).

CourseOutcomes:

Bytheend of the course, students will be able to:

CO1. Students will be able to perform data understanding, cleaning, and visualization using Tableau.

CO2. Students will be able to perform data understanding, cleaning, and visualization using Power BI.

CO3. Students will be able to apply clustering techniques for data analysis.

CO4. Students will be proficient in Principal Component Analysis for dimensionality reduction

CO5. Students will be able to implement Self-Organizing Maps for data pattern recognition.

CO6. Students will be able to implement k-Nearest Neighbors for classification and regression.

CO7. Students will be able to utilize techniques to enhance model performance and improve classification accuracy.

TopicsandLearningPoints

Sr. No.	Title of Experiments	No. of Practical
1.	Data Understanding and Cleaning and Data Visualization Using Tableau	02
2.	Data Understanding and Cleaning and Data Visualization Using Power BI	02
3.	Unsupervised Techniques • Clustering • Principal Component Analysis • Factor Analysis • Self-Organizing Maps (SOM)	02
4.	Supervised Techniques • Linear Discriminant Analysis • Logistic Regression • Bayes Classifier • k-Nearest Neighbors • Classification and Regression tree (CART) • Artificial Neural Network (ANN) • Support Vector Machines (SVM)	04
5.	Model Enhancement Techniques to Improve Classification Accuracy	02
6.	Market Basket Analysis Using Association Rule Mining Techniques	01
7.	Computer vision (CNN)	01
8.	Recurrent Neural network	01

ProgrammeOutcomes andCourseOutcomesMapping:

	Programme Outcomes(POs)									
Course Outcomes	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO 10
CO1	3	2	3	2	2	3	1	2	2	
CO2	3	3	2	3	2	2	3	1	2	2
CO3	3	3	2	3	2	2	3	1	2	2
CO4	3	3	2	3	2	2	3	1	2	2
CO5	3	3	2	3	2	2	3	1	2	2
CO6	3	3	2	3	2	2	3	1	2	2
CO7	3	3	2	3	2	2	3	1	2	2

Justification:

PO1: Comprehensive Knowledge and Understanding

- **CO1** (3): Strongly related as it provides a fundamental understanding of data visualization, a key concept in data science.
- **CO2** (3): Strongly related for similar reasons as CO1.
- **CO3** (3): Strongly related as clustering techniques are foundational methodologies in data analysis.
- **CO4** (3): Strongly related due to the importance of PCA in understanding data structures.
- **CO5** (3): Strongly related as Self-Organizing Maps are key in data pattern recognition.
- **CO6** (3): Strongly related as k-Nearest Neighbors is a fundamental algorithm for classification and regression.
- **CO7** (3): Strongly related as model enhancement techniques are crucial for a profound understanding of improving classification accuracy.

PO2: Practical, Professional, and Procedural Knowledge

- **CO1** (3): Strongly related as it involves practical skills in data visualization using Tableau.
- **CO2** (3): Strongly related as it involves practical skills in data visualization using Power BI.
- **CO3** (3): Strongly related as clustering is a practical technique in data analysis.

- **CO4** (3): Strongly related as PCA is a practical technique for dimensionality reduction.
- **CO5** (3): Strongly related as implementing Self-Organizing Maps involves practical application.
- **CO6** (3): Strongly related due to the practical application of k-Nearest Neighbors in classification and regression.
- **CO7** (3): Strongly related as model enhancement techniques are essential for professional tasks.

PO3: Entrepreneurial Mindset, Innovation, and Business Understanding

- **CO1** (2): Moderately related as data visualization can help identify business opportunities.
- **CO2** (2): Moderately related for similar reasons as CO1.
- **CO3** (2): Moderately related as clustering can help in market segmentation.
- **CO4** (2): Moderately related as PCA can aid in identifying key business factors.
- **CO5** (2): Moderately related as Self-Organizing Maps can lead to innovative pattern recognition.
- **CO6** (2): Moderately related as k-Nearest Neighbors can be used in business predictions.
- **CO7** (2): Moderately related as enhancing model performance can drive business innovation.

PO4: Specialized Skills, Critical Thinking, and Problem-Solving

- **CO1** (3): Strongly related as data visualization requires critical thinking.
- **CO2** (3): Strongly related as data visualization requires critical thinking.
- **CO3** (3): Strongly related as clustering involves analytical skills and problem-solving.
- **CO4** (3): Strongly related as PCA requires critical thinking and problem-solving.
- **CO5** (3): Strongly related as Self-Organizing Maps involve specialized skills.
- **CO6** (3): Strongly related as k-Nearest Neighbors involves technical skills and critical thinking.
- **CO7** (3): Strongly related as enhancing model performance requires specialized skills and problem-solving.

PO5: Research, Analytical Reasoning, and Ethical Conduct

- **CO1** (2): Moderately related as data visualization involves research and analytical reasoning.
- **CO2** (2): Moderately related as data visualization involves research and analytical reasoning.
- **CO3** (2): Moderately related as clustering requires analytical reasoning.
- **CO4** (2): Moderately related as PCA involves research and analytical reasoning.
- **CO5** (2): Moderately related as Self-Organizing Maps require analytical reasoning.
- **CO6** (2): Moderately related as k-Nearest Neighbors involves research and analytical reasoning.
- **CO7** (2): Moderately related as model enhancement involves research and ethical conduct.

PO6: Communication, Collaboration, and Leadership

- **CO1** (2): Moderately related as data visualization enhances communication of complex data.
- **CO2** (2): Moderately related as data visualization enhances communication of complex data.
- **CO3** (2): Moderately related as clustering results need to be communicated effectively.
- **CO4** (2): Moderately related as PCA results need to be communicated effectively.
- **CO5** (2): Moderately related as results from Self-Organizing Maps need to be communicated effectively.
- **CO6** (2): Moderately related as k-Nearest Neighbors results need to be communicated effectively.
- **CO7** (2): Moderately related as model enhancement results need to be communicated effectively.

PO7: Digital Proficiency and Technological Skills

- **CO1** (3): Strongly related as proficiency in Tableau requires digital skills.
- **CO2** (3): Strongly related as proficiency in Power BI requires digital skills.
- **CO3** (3): Strongly related as clustering involves the use of advanced software.
- **CO4** (3): Strongly related as PCA requires proficiency in technological tools.
- **CO5** (3): Strongly related as Self-Organizing Maps involve advanced digital skills.

- **CO6** (3): Strongly related as k-Nearest Neighbors requires the use of technological tools.
- **CO7** (3): Strongly related as model enhancement involves advanced technological skills.

PO8: Multicultural Competence, Inclusive Spirit, and Empathy

- **CO1** (1): Partially related as data visualization can represent diverse perspectives.
- **CO2** (1): Partially related as data visualization can represent diverse perspectives.
- **CO3** (1): Partially related as clustering can reveal patterns in multicultural data.
- **CO4** (1): Partially related as PCA can be applied to diverse datasets.
- **CO5** (1): Partially related as Self-Organizing Maps can be applied to multicultural data.
- **CO6** (1): Partially related as k-Nearest Neighbors can be applied to diverse datasets.
- **CO7** (1): Partially related as model enhancement can involve diverse data sources.

PO9: Value Inculcation, Environmental Awareness, and Ethical Practices

- **CO1** (2): Moderately related as data visualization should follow ethical practices.
- **CO2** (2): Moderately related as data visualization should follow ethical practices.
- **CO3** (2): Moderately related as clustering should follow ethical practices.
- **CO4** (2): Moderately related as PCA should follow ethical practices.
- **CO5** (2): Moderately related as Self-Organizing Maps should follow ethical practices.
- **CO6** (2): Moderately related as k-Nearest Neighbors should follow ethical practices.
- **CO7** (2): Moderately related as model enhancement should follow ethical practices.

PO10: Autonomy, Responsibility, and Accountability

- **CO1** (2): Moderately related as data visualization requires autonomy.
- **CO2** (2): Moderately related as data visualization requires autonomy.
- **CO3** (2): Moderately related as clustering requires responsibility.
- **CO4** (2): Moderately related as PCA requires accountability.
- **CO5** (2): Moderately related as Self-Organizing Maps require autonomy.
- **CO6** (2): Moderately related as k-Nearest Neighbors requires responsibility.
- **CO7** (2): Moderately related as model enhancement requires accountability.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. Part – I
Semester	: II
Course Type	:Major Mandatory Practical
Course Name	: Data Science Practical – IV
Course Code	: DSC-554-MJM
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. Studentswilllearntousenumericalcomputingtoolsandprogramminglanguages,suchasMATLAB,Python, or R or Minitab, toimplementandsolveproblems.
2. Studentswill learn how to apply linear regression models in practice, identify situation where linear regression is appropriate, build and fit linear regression models with software also interpret estimates and diagnostic statistics, produce exploratory graphs.
3. Studentswill learn to predict an ordinal dependent variable given one or more independent variables.
4. Studentswill learn to diagnose the presence of Multicollinearity in a model.
5. Studentswill learn Multicollinearity refers to a state wherein there exists inter-association or inter-relation between two or more independent variables.
6. Student should be able to understand when it is relevant to choose logistic regression.
7. Students should able to predict the value of the dependent variable for individuals for whom some information concerning the explanatory variables is available, or in order to estimate the effect of some explanatory variable on the dependent variable.

Course Outcomes:

Bytheendof thecourse,studentswillbeableto:

CO1. Handle statistical software, packages such as R, Python, MATLAB, SPSS or Minitab toimplementandanalyzereal lifesituations.

CO2. Understand of model selection and regression modelling techniques should be demonstrated.

CO3. Understand **e**xcellent familiarity with both linear and nonlinear regression models.

CO4. Estimate the parameters and fit a model.

CO5. Investigate possible diagnostics in regression modeling and analysis concepts.

CO6. Predict categorical placement in or the probability of category membership on a dependent variable based on multiple independent variables.

CO7. Diagnose the presence of multicollinearity in a model.

Topics and Learning Points

Sr. No.	TitleofExperiments
1.	Simple regression and regression diagnostic
2.	Multiple regression and regression diagnostic
3.	Lack of fit of the regression model
4.	Multiple regression (selection of variable)
5.	Detection of multicollinearity
6.	Dealing with multicollinearity by using Principle Component Regression
7.	Dealing with multicollinearity by using Ridge Regression
8.	Polynomial regression
9.	Non-linear regression
10.	Poisson regression
11.	Logistic regression
12.	Ordinary logistic regression
13.	Multinomial logistic regression.
14.	Case Study (Two Practical’s).

ProgrammeOutcomes andCourseOutcomesMapping:

	Programme Outcomes(POs)									
Course Outcomes	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO 10**
CO1	CO1	3	3	2	3	3	3	3	1	2
CO2	CO2	3	3	2	3	3	2	3	1	2
CO3	CO3	3	3	2	3	3	2	3	1	2
CO4	CO4	3	3	2	3	3	2	3	1	2
CO5	CO5	3	3	2	3	3	2	3	1	2
CO6	CO6	3	3	2	3	3	2	3	1	2
CO7	CO7	3	3	2	3	3	2			

Justification:

PO1: Comprehensive Knowledge and Understanding

- **CO1** (3): Strongly related as handling statistical software and packages involves comprehensive knowledge of various tools and their applications in real-life situations.
- **CO2** (3): Strongly related due to the necessity of understanding model selection and regression modeling techniques, which are fundamental concepts.
- **CO3** (3): Strongly related as familiarity with linear and nonlinear regression models is essential for a comprehensive understanding of regression analysis.
- **CO4** (3): Strongly related as estimating parameters and fitting models are core aspects of regression analysis.
- **CO5** (3): Strongly related as investigating diagnostics in regression modeling requires a deep understanding of analysis concepts.
- **CO6** (3): Strongly related as predicting categorical placement and probabilities involves a comprehensive understanding of regression techniques.
- **CO7** (3): Strongly related as diagnosing multicollinearity is crucial for a thorough understanding of regression modeling.

PO2: Practical, Professional, and Procedural Knowledge

- **CO1** (3): Strongly related as practical skills in using statistical software for real-life data analysis are essential.
- **CO2** (3): Strongly related as understanding model selection and regression techniques is crucial for professional tasks.
- **CO3** (3): Strongly related as proficiency with both linear and nonlinear regression models is necessary for practical applications.

- **CO4** (3): Strongly related as parameter estimation and model fitting are practical skills.
- **CO5** (3): Strongly related as regression diagnostics are practical tools for model analysis.
- **CO6** (3): Strongly related as predicting categorical outcomes involves practical application of regression techniques.
- **CO7** (3): Strongly related as diagnosing and dealing with multicollinearity are important for practical regression analysis.

PO3: Entrepreneurial Mindset, Innovation, and Business Understanding

- **CO1** (2): Moderately related as using statistical software can aid in identifying business opportunities.
- **CO2** (2): Moderately related as model selection can drive innovation in business strategies.
- **CO3** (2): Moderately related as understanding regression models can be applied to business contexts.
- **CO4** (2): Moderately related as model fitting can be useful for business decision-making.
- **CO5** (2): Moderately related as diagnosing and addressing model issues can lead to innovative solutions.
- **CO6** (2): Moderately related as predicting outcomes based on regression can inform business strategies.
- **CO7** (2): Moderately related as diagnosing multicollinearity can impact business model accuracy.

PO4: Specialized Skills, Critical Thinking, and Problem-Solving

- **CO1** (3): Strongly related as using statistical software requires specialized skills and problem-solving.
- **CO2** (3): Strongly related as understanding model selection and techniques involves critical thinking.
- **CO3** (3): Strongly related as familiarity with regression models requires specialized knowledge.
- **CO4** (3): Strongly related as parameter estimation and model fitting involve problem-solving skills.

- **CO5** (3): Strongly related as investigating diagnostics involves critical analysis.
- **CO6** (3): Strongly related as predicting categorical outcomes requires problem-solving.
- **CO7** (3): Strongly related as diagnosing multicollinearity involves specialized skills and critical thinking.

PO5: Research, Analytical Reasoning, and Ethical Conduct

- **CO1** (3): Strongly related as using statistical software for real-life data analysis is essential in research.
- **CO2** (3): Strongly related as understanding model selection and techniques is crucial for analytical reasoning.
- **CO3** (3): Strongly related as familiarity with regression models supports research and analysis.
- **CO4** (3): Strongly related as parameter estimation and model fitting are fundamental in research.
- **CO5** (3): Strongly related as investigating diagnostics is crucial for accurate research.
- **CO6** (3): Strongly related as predicting outcomes based on regression supports research objectives.
- **CO7** (3): Strongly related as diagnosing multicollinearity is important for accurate research and analysis.

PO6: Communication, Collaboration, and Leadership

- **CO1** (2): Moderately related as using statistical software involves communicating complex data analysis results.
- **CO2** (2): Moderately related as understanding model selection can enhance collaborative discussions.
- **CO3** (2): Moderately related as familiarity with regression models can improve communication of analytical results.
- **CO4** (2): Moderately related as parameter estimation and model fitting results need effective communication.
- **CO5** (2): Moderately related as discussing diagnostics requires clear communication.
- **CO6** (2): Moderately related as predicting outcomes based on regression involves collaborative analysis.

- **CO7** (2): Moderately related as diagnosing multicollinearity requires effective communication of results.

PO7: Digital Proficiency and Technological Skills

- **CO1** (3): Strongly related as handling statistical software involves high digital proficiency.
- **CO2** (3): Strongly related as understanding model selection and techniques requires technological skills.
- **CO3** (3): Strongly related as familiarity with regression models involves technological expertise.
- **CO4** (3): Strongly related as parameter estimation and model fitting require digital skills.
- **CO5** (3): Strongly related as investigating diagnostics involves using technological tools.
- **CO6** (3): Strongly related as predicting categorical outcomes involves digital proficiency.
- **CO7** (3): Strongly related as diagnosing multicollinearity involves technological skills.

PO8: Multicultural Competence, Inclusive Spirit, and Empathy

- **CO1** (1): Partially related as using statistical software can include analyzing diverse datasets.
- **CO2** (1): Partially related as model selection can be applied in diverse contexts.
- **CO3** (1): Partially related as understanding regression models can involve multicultural data.
- **CO4** (1): Partially related as parameter estimation can include diverse datasets.
- **CO5** (1): Partially related as diagnostics can be applied to various contexts.
- **CO6** (1): Partially related as predicting outcomes based on regression can consider diverse contexts.
- **CO7** (1): Partially related as diagnosing multicollinearity can include diverse data perspectives.

PO9: Value Inculcation, Environmental Awareness, and Ethical Practices

- **CO1** (2): Moderately related as using statistical software involves ethical considerations.

- **CO2** (2): Moderately related as model selection should adhere to ethical practices.
- **CO3** (2): Moderately related as understanding regression models includes ethical considerations.
- **CO4** (2): Moderately related as parameter estimation and model fitting should follow ethical guidelines.
- **CO5** (2): Moderately related as investigating diagnostics involves ethical considerations.
- **CO6** (2): Moderately related as predicting outcomes requires ethical practices.
- **CO7** (2): Moderately related as diagnosing multicollinearity involves ethical considerations.

PO10: Autonomy, Responsibility, and Accountability

- **CO1** (3): Strongly related as using statistical software independently requires autonomy and responsibility.
- **CO2** (3): Strongly related as understanding model selection involves responsibility in choosing appropriate techniques.
- **CO3** (3): Strongly related as familiarity with regression models requires autonomous learning.
- **CO4** (3): Strongly related as parameter estimation and model fitting involve independent work.
- **CO5** (3): Strongly related as investigating diagnostics requires responsible analysis.
- **CO6** (3): Strongly related as predicting categorical outcomes involves accountability in analysis.
- **CO7** (3): Strongly related as diagnosing multicollinearity involves autonomous and responsible work.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. (Part – I)
Semester	: II
Course Type	: Major Elective Theory
Course Name	: Bayesian Inference
Course Code	: DSC-561-MJE(A)
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. To learn about Bayesian inference setup, including prior and posterior distributions.
2. To explore loss functions and the principles of minimum expected posterior loss.
3. To perform point estimation, construct highest posterior density (HPD) confidence intervals, and make predictions of future observations using Bayesian methods.
4. To understand different classes of priors and their role in Bayesian inference.
5. To learn about Bayesian computing techniques, including the Expectation-Maximization (E-M) algorithm.
6. To discuss convergence diagnostics in Bayesian computation and apply them to assess the convergence of MCMC chains.
7. To understand Bayesian model selection methods such as Bayesian Information Criterion (BIC) and Bayes factors.

Course Outcomes:

By the end of the course, students will be able to:

CO 1. 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 and Learning Points

Unit1 **(18L)**

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.

Unit2 **(12L)**

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

Unit3 **(12L)**

Bayesian model selection BIC, Bayes factors, limit of posterior distributions, consistency and asymptotic normality of posterior distributions.

Unit4 **(18L)**

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

References:

1. Abraham Silberschatz, Henry F. Korth, S. Sudarashan, Database System Concepts, McGraw-Hill International Edition, Sixth Edition
2. Elmasri, Navathe, Fundamentals of Database Systems, Pearson Education, Third Education
3. Ramakrishnan, Gehrke, Database Management Systems, McGraw Hill International Edition, Third Edition
4. Peter Rob, Carlos Coronel, Database System Concepts, Cengage Learning, India Edition
5. S.K. Singh, "Database Systems Concepts, Design and Applications", First Edition, Pearson Education, 2006

6. Redmond, E. & Wilson, Seven Databases in Seven Weeks: A Guide to Modern Databases and the No SQL Movement Edition:1st Edition.
7. Shamkant B. Navathe, RamezElmasri,(2010), Database Systems, ISBN:9780132144988, PEARSON HIGHER EDUCATION
8. Richard Stones, Neil Matthew, (2005), Beginning Databases with PostgreSQL: From Novice to Professional, ISBN:9781590594780, Apress
9. Korry, Douglas, (2005), Postgre SQL, ISBN:9780672327568, Sams Publishing.
10. Joshua D. Drake, John C. Worsley, Practical Postgre SQL, (2002), ISBN:9788173663925 O'Reilly Media, Inc., ISBN: 9781565928466.

ProgrammeOutcomes andCourseOutcomesMapping:

	Programme Outcomes(POs)									
Course Outcomes	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO 10**
CO1	3	2	1	2	2	1	1	1		
CO2	3	2	1	2	2	1	1	1		
CO3	2	1	1	2	2	1	1	1		
CO4	3	2	1	2	2	1	1	1		
CO5	3	2	1	2	3	1	1	1		
CO6	3	3	2	3	3	2				
CO7	3	2	1	2	2	1				

Justification:

PO 1 Comprehensive Knowledge and Understanding

CO1 Weightage: 3 (Strongly Related)

Justification: Bayesian prediction intervals are an essential component of Bayesian analysis, contributing to a profound understanding of statistical inference within a multidisciplinary context.

CO2 Weightage: 3 (Strongly Related)

Justification: Understanding the role of prior distributions in Bayesian analysis enhances comprehension of foundational theories and methodologies in statistics and data analysis.

CO3 Weightage: 2 (Moderately Related)

Justification: Demonstrating the robustness of Bayesian methods with small sample sizes reinforces the importance of practical, professional, and procedural knowledge in statistical analysis.

CO4 Weightage: 3 (Strongly Related)

Justification: Mastery of Bayes' theorem and its application in Bayesian inference is fundamental to developing a comprehensive understanding of statistical methodologies.

CO5 Weightage: 3 (Strongly Related)

Justification: Understanding the probabilistic nature of Bayesian analysis contributes to analytical reasoning and critical thinking skills, essential for research and data analysis.

CO6 Weightage: 3 (Strongly Related)

Justification: Application of Bayesian inference in real-world scenarios reinforces the interdisciplinary application of knowledge and skills, aligning with the multidisciplinary context of comprehensive knowledge and understanding.

CO7 Weightage: 3 (Strongly Related)

Justification: Utilizing posterior predictive checks demonstrates competency in research, analytical reasoning, and ethical conduct by ensuring the validity and reliability of Bayesian models in data analysis and reporting findings.

PO 2 Practical, Professional, and Procedural Knowledge

CO1 Weightage: 2 (Moderately Related)

Justification: Practical knowledge of constructing prediction intervals enhances skills necessary for professional tasks in statistical analysis and reporting.

CO2 Weightage: 2 (Moderately Related)

Justification: Understanding the role of prior distributions contributes to procedural knowledge in Bayesian analysis, aligning with industry standards and ethical considerations in statistical modeling.

CO3 Weightage: 1 (Partially Related)

Justification: While understanding the performance of Bayesian methods with small sample sizes is valuable for professional tasks, it may not directly correlate with practical or procedural knowledge.

CO4 Weightage: 2 (Moderately Related)

Justification: Application of Bayes' theorem in Bayesian inference contributes to procedural knowledge necessary for statistical analysis and modeling in real-world scenarios.

CO5 Weightage: 2 (Moderately Related)

Justification: Understanding probabilistic outputs in Bayesian analysis enhances practical knowledge of interpreting and communicating results effectively, aligning with professional standards in data analysis.

CO6 Weightage: 3 (Strongly Related)

Justification: Application of Bayesian inference in real-world contexts develops practical skills necessary for professional tasks across diverse domains, demonstrating proficiency in procedural knowledge and industry standards.

CO 7 Weightage: 2 (Moderately Related)

Justification: Mastery of posterior predictive checks contributes to professional knowledge by ensuring the validity and reliability of Bayesian models in practical applications, aligning with industry standards and ethical considerations.

PO 3 Entrepreneurial Mindset, Innovation, and Business Understanding

CO1 Weightage: 1 (Partially Related)

Justification: While Bayesian prediction intervals may not directly align with entrepreneurial mindset or business understanding, the ability to draw appropriate conclusions demonstrates analytical skills relevant to innovation and problem-solving.

CO 2 Weightage: 1 (Partially Related)

Justification: Understanding how prior distributions incorporate existing knowledge or beliefs may indirectly contribute to fostering an entrepreneurial mindset by encouraging critical thinking and creativity in decision-making processes.

CO 3 Weightage: 1 (Partially Related)

Justification: While the performance of Bayesian methods with small sample sizes may not directly relate to entrepreneurial mindset, the ability to adapt and make informed decisions under resource constraints aligns with entrepreneurial qualities.

CO4 Weightage: 1 (Partially Related)

Justification: Understanding Bayes' theorem and its application in Bayesian analysis may indirectly contribute to business understanding by fostering logical reasoning and decision-making skills.

CO 5 Weightage: 1 (Partially Related)

Justification: While probabilistic outputs may not directly relate to business understanding, the ability to quantify and communicate uncertainty in decision-making processes aligns with entrepreneurial mindset and risk management strategies.

CO6 Weightage: 2 (Moderately Related)

Justification: Application of Bayesian inference in diverse domains demonstrates innovation and adaptability, fostering an entrepreneurial mindset by addressing complex problems and identifying opportunities for improvement.

CO7 Weightage: 1 (Partially Related)

Justification: Mastery of posterior predictive checks contributes to analytical skills and critical thinking, indirectly supporting an entrepreneurial mindset by ensuring the validity and reliability of data-driven decisions.

PO 4 Specialized Skills, Critical Thinking, and Problem-Solving

CO1 Weightage: 2 (Moderately Related)

Justification: Bayesian prediction intervals require specialized skills in Bayesian analysis, which involve critical thinking and problem-solving to interpret and draw appropriate conclusions from the results.

CO2 Weightage: 2 (Moderately Related)

Justification: Incorporating prior knowledge or beliefs through the prior distribution requires critical thinking to assess the relevance and reliability of the prior information in the context of the problem being addressed.

CO3 Weightage: 2 (Moderately Related)

Justification: The ability of Bayesian analysis to perform well with small sample sizes, particularly with informative priors, highlights specialized skills in statistical modeling and critical thinking to effectively utilize available data for inference.

CO4 Weightage: 2 (Moderately Related)

Justification: Applying Bayes' theorem to combine prior knowledge with observed data to obtain the posterior distribution requires critical thinking and problem-solving skills to understand the underlying principles and make appropriate inferences.

CO5 Weightage: 2 (Moderately Related)

Justification: Understanding and interpreting probabilistic outputs in Bayesian analysis involve specialized skills in critical thinking to assess and communicate uncertainty, which is essential for effective problem-solving.

CO6 Weightage: 3 (Strongly Related)

Justification: Applying Bayesian inference to real-world problems requires specialized skills, critical thinking, and problem-solving abilities to adapt Bayesian methods to diverse domains and address complex challenges effectively.

CO7 Weightage: 2 (Moderately Related)

Justification: The ability to conduct posterior predictive checks involves critical thinking and problem-solving skills to evaluate the performance of Bayesian models and ensure their validity in capturing the underlying data patterns.

PO 5 Research, Analytical Reasoning, and Ethical Conduct

CO1 Weightage: 2 (Moderately Related)

Justification: Research and analytical reasoning skills are required to construct Bayesian prediction intervals, and ethical conduct is essential in accurately reporting conclusions drawn from the analysis.

CO2 Weightage: 2 (Moderately Related)

Justification: Research and analytical reasoning skills are necessary to evaluate existing knowledge or beliefs for incorporation into the prior distribution, ensuring that the Bayesian analysis reflects the relevant information available.

CO3 Weightage: 2 (Moderately Related)

Justification: Analytical reasoning skills are crucial in understanding the performance of Bayesian methods with small sample sizes, and ethical conduct is necessary to ensure the appropriate use of informative priors in such scenarios.

CO4 Weightage: 2 (Moderately Related)

Justification: Research and analytical reasoning skills are required to apply Bayes' theorem effectively in combining the prior and likelihood, while ethical conduct ensures the integrity of the posterior distribution obtained.

CO5 Weightage: 3 (Strongly Related)

Justification: Analytical reasoning skills are essential in understanding and interpreting probabilistic outputs in Bayesian analysis, which naturally express uncertainty in parameter estimates, contributing to ethical conduct in accurately conveying the level of uncertainty in research findings.

CO6 Weightage: 3 (Strongly Related)

Justification: Research and analytical reasoning skills are fundamental in applying Bayesian inference to real-world problems across different domains, ensuring that appropriate methodologies are utilized for data analysis while adhering to ethical conduct in research practices.

CO7 Weightage: 2 (Moderately Related)

Justification: Analytical reasoning skills are necessary to understand and apply posterior predictive checks effectively, ensuring that Bayesian models accurately capture the observed data patterns, which contributes to ethical conduct in research by validating the model's adequacy.

PO 6 Communication, Collaboration, and Leadership

CO1 Weightage: 1 (Partially Related)

Justification: While this objective focuses on analytical reasoning and reporting conclusions, effective communication skills are required to articulate findings clearly to various stakeholders.

CO2 Weightage: 1 (Partially Related)

Justification: Effective communication skills are necessary to understand and incorporate stakeholders' knowledge or beliefs into the Bayesian analysis, fostering collaboration and ensuring alignment with project objectives.

CO3 Weightage: 1 (Partially Related)

Justification: Collaboration skills are essential in discussing and determining the suitability of informative priors, especially in scenarios with limited data, to ensure the reliability of Bayesian analysis outcomes.

CO4 Weightage: 1 (Partially Related)

Justification: While this objective primarily involves analytical reasoning, effective communication skills are necessary to explain the Bayesian inference process to team members or stakeholders.

CO5 Weightage: 1 (Partially Related)

Justification: Communication skills are crucial in conveying the probabilistic nature of Bayesian outputs and the associated uncertainty to non-technical audiences, facilitating informed decision-making.

CO6 Weightage: 2 (Moderately Related)

Justification: Collaboration skills are required to work effectively with multidisciplinary teams in applying Bayesian inference to diverse real-world problems, ensuring effective communication and coordination to achieve common goals.

CO7 Weightage: 1 (Partially Related)

Justification: Effective communication skills are needed to explain the rationale behind posterior predictive checks and their implications for model adequacy to team members or stakeholders.

PO7: Digital Proficiency and Technological Skills

CO1: Weightage: 1 (Partially Related)

Justification: This objective involves utilizing statistical techniques and software tools, which may require digital proficiency and technological skills. However, it primarily focuses on statistical analysis rather than direct application of digital tools.

CO2: Weightage: 1 (Partially Related)

Justification: Understanding how to incorporate prior knowledge into Bayesian analysis may involve using software tools, but the emphasis is on statistical methodology rather than digital proficiency.

CO3: Weightage: 1 (Partially Related)

Justification: While digital tools may facilitate analysis with small sample sizes, this objective primarily concerns statistical methodology rather than digital proficiency.

CO4: Weightage: 1 (Partially Related)

Justification: This objective involves understanding Bayesian principles and mathematical concepts, which may involve using digital tools. However, the emphasis is on theoretical understanding rather than practical digital skills.

CO5: Weightage: 1 (Partially Related)

Justification: While digital tools may aid in generating probabilistic outputs, this objective primarily concerns statistical concepts and interpretation rather than direct digital proficiency.

PO8: Multicultural Competence, Inclusive Spirit, and Empathy

CO1: Weightage: 1 (Partially Related)

Justification: While this objective focuses on statistical analysis, it indirectly contributes to multicultural competence by fostering critical thinking and analytical skills, which are essential for understanding diverse perspectives.

CO2: Weightage: 1 (Partially Related)

Justification: Understanding how to incorporate prior knowledge into Bayesian analysis indirectly supports inclusive spirit by recognizing and respecting diverse viewpoints and beliefs in the decision-making process.

CO3: Weightage: 1 (Partially Related)

Justification: While statistical analysis techniques are the focus here, the ability to handle small sample sizes and incorporate informative priors demonstrates adaptability, a quality valuable in multicultural settings.

CO4: Weightage: 1 (Partially Related)

Justification: Understanding Bayesian principles involves considering prior beliefs and updating them based on new evidence, which aligns with the principles of inclusivity and empathy towards diverse perspectives.

CO5: Weightage: 1 (Partially Related)

Justification: The ability to express uncertainty in parameter estimates acknowledges the complexity of real-world scenarios and fosters empathy by recognizing and respecting uncertainty in decision-making.

CBCS Syllabus as per NEP 2020 for M.Sc. Part-I Data Science (2023 Pattern)

Name of the Programme	: M.Sc. Data Science
Program Code	: PSDSC
Class	: M.Sc. (Part – I)
Semester	: II
Course Type	: Major Elective Theory
Course Name	: Computational Statistics
Course Code	: DSC-561-MJE(B)
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. To understand the theory behind the inverse transformation method (ITM) for generating random variates.
2. To utilize iterative methods including Jacobi and Gauss-Seidel methods for solving systems of linear equations..
3. To explore variance stabilizing transformations and their application in statistical analysis.
4. To analyze the convergence properties and limitations of various numerical methods for nonlinear equations and optimization problems.
5. To perform numerical differentiation using forward and backward difference methods and analyze the associated errors.
6. To implement Monte Carlo methods, including Monte Carlo integration, to compute expected values and probabilities..
7. To analyze the strengths and limitations of Monte Carlo methods in various computational settings.

Course Outcomes:

By the end of the course, students will be able to:

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

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

CO 3. 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 and Learning Points

Unit1 **(15 L)**

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.

Unit2 **(15 L)**

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.

Unit3 **(15 L)**

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.

Unit4 **(15 L)**

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.

References:

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)

ProgrammeOutcomes andCourseOutcomesMapping:

	Programme Outcomes(POs)									
Course Outcomes	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO 10**
CO1	2	3	1	3	2	1	2	1	1	1
CO2	2	2	1	2	1	1	2	1	1	1
CO3	3	2	1	3	2	1	3	1	1	1
CO4	2	1	1	2	1	1	1	1	1	1
CO5	3	3	1	3	3	1	2	1	1	1
CO6	3	3	1	3	2	1	3	1	1	1
CO7	2	3	1	3	2	1	3	1	1	1

Justification:

PO1 Comprehensive Knowledge and Understanding:

CO1: Weightage: 2 (Moderately Related)

Justification: Computational techniques provide a practical approach to understanding complex statistical concepts, enhancing the depth of knowledge and understanding in the field.

CO2: Weightage: 2 (Moderately Related)

Justification: Understanding numerical techniques aids in grasping foundational theories and methodologies by providing practical applications and approximations.

CO3: Weightage: 3 (Strongly Related)

Justification: Monte Carlo studies and random number generation are fundamental aspects of statistical analysis, directly contributing to a profound understanding of statistical theories and methodologies.

CO4: Weightage: 2 (Moderately Related)

Justification: While polynomial interpolation techniques are specific mathematical methods, they enhance problem-solving skills and reinforce understanding of mathematical concepts underlying statistical methodologies.

CO5: Weightage: 3 (Strongly Related)

Justification: Understanding bias estimation techniques through Jack-knife and Bootstrap methods directly relates to observational and inquiry skills, essential for research and analytical reasoning.

CO6: Weightage: 3 (Strongly Related)

Justification: Applying numerical methods in data science and statistical analysis reinforces practical knowledge and enhances problem-solving abilities, aligning with the goal of comprehensive understanding.

CO7: Weightage: 2 (Moderately Related)

Justification: While numerical optimization focuses on specific techniques, its application in machine learning broadens the understanding of statistical methods in real-world scenarios, contributing to comprehensive knowledge and understanding.

PO2 Practical, Professional, and Procedural Knowledge:

CO1: Weightage: 3 (Strongly Related)

Justification: Computational techniques are practical skills necessary for addressing statistical questions in real-world scenarios, aligning with the acquisition of practical and procedural knowledge.

CO2: Weightage: 2 (Moderately Related)

Justification: Numerical techniques enhance practical knowledge by providing methods for approximating functions and transforming data, contributing to the acquisition of procedural knowledge in statistical analysis.

CO3: Weightage: 2 (Moderately Related)

Justification: Monte Carlo studies and random number generators are practical tools used in statistical analysis, contributing to the development of procedural knowledge and professional skills in the field.

CO4: Weightage: 1 (Partially Related)

Justification: While polynomial interpolation techniques are valuable mathematical skills, they are less directly applicable to practical statistical tasks, reducing their relevance to the acquisition of procedural knowledge.

CO5: Weightage: 3 (Strongly Related)

Justification: Estimation techniques like Jack-knife and Bootstrap are essential practical skills for assessing bias in statistical analysis, aligning closely with the acquisition of procedural knowledge and professional competency.

CO6: Weightage: 3 (Strongly Related)

Justification: Application of numerical methods in various domains, including data science and machine learning, demonstrates practical competence and procedural knowledge relevant to professional tasks in statistical analysis.

CO7: Weightage: 3 (Strongly Related)

Justification: Utilizing numerical optimization techniques in machine learning applications enhances practical knowledge and procedural skills, contributing to professional competence in statistical analysis.

PO3 Entrepreneurial Mindset, Innovation, and Business Understanding:

CO1: Weightage: 1 (Partially Related)

Justification: While computational techniques may foster innovation and problem-solving, they do not directly address entrepreneurial mindset or business understanding.

CO2: Weightage: 1 (Partially Related)

Justification: Numerical techniques for function approximation are essential for data analysis but do not directly contribute to entrepreneurial mindset or business understanding.

CO3: Weightage: 1 (Partially Related)

Justification: While Monte Carlo studies involve computational methods, they are not directly related to entrepreneurial mindset or business understanding.

CO4: Weightage: 1 (Partially Related)

Justification: Polynomial interpolation techniques are mathematical methods that may not directly contribute to entrepreneurial mindset or business understanding.

CO5: Weightage: 1 (Partially Related)

Justification: Estimation techniques like Jack-knife and Bootstrap are important for statistical analysis but do not directly relate to entrepreneurial mindset or business understanding.

CO6: Weightage: 1 (Partially Related)

Justification: While numerical methods are crucial for problem-solving in various domains, they do not directly enhance entrepreneurial mindset or business understanding.

CO7: Weightage: 1 (Partially Related)

Justification: Numerical optimization techniques are important for machine learning but do not directly contribute to entrepreneurial mindset or business understanding.

PO4 Specialized Skills, Critical Thinking, and Problem-Solving:

CO1: Weightage: 3 (Strongly Related)

Justification: Computational techniques are fundamental for problem-solving in statistical analysis, requiring specialized skills and critical thinking to implement numerical solutions effectively.

CO2: Weightage: 2 (Moderately Related)

Justification: While numerical techniques involve critical thinking and problem-solving skills, their direct application in statistical analysis may be more specialized, making them moderately related to specialized skills and problem-solving.

CO3: Weightage: 3 (Strongly Related)

Justification: Monte Carlo studies require specialized skills and critical thinking to implement effectively, contributing to problem-solving in statistical analysis and simulation.

CO4: Weightage: 2 (Moderately Related)

Justification: Polynomial interpolation techniques involve critical thinking and problem-solving skills but may not be directly related to specialized skills in statistical analysis, making them moderately related to this outcome.

CO5: Weightage: 3 (Strongly Related)

Justification: Estimation techniques such as Jack-knife and Bootstrap require specialized skills and critical thinking to assess bias and improve the accuracy of estimators, directly contributing to problem-solving in statistical analysis.

CO6: Weightage: 3 (Strongly Related)

Justification: Application of numerical methods in data science and statistical analysis demands specialized skills and critical thinking to solve complex problems effectively.

CO7: Weightage: 3 (Strongly Related)

Justification: Numerical optimization techniques are crucial for optimizing machine learning algorithms, requiring specialized skills and critical thinking to address optimization challenges effectively.

PO5 Research, Analytical Reasoning, and Ethical Conduct:

CO1: Weightage: 2 (Moderately Related)

Justification: Computational techniques are often utilized in research and analytical reasoning to solve statistical questions, albeit the direct focus is on numerical solutions rather than the broader aspects of research and analytical reasoning.

CO2: Weightage: 1 (Partially Related)

Justification: While numerical techniques are important for data analysis, they are not directly tied to the formulation of research questions or the analytical reasoning process.

CO3: Weightage: 2 (Moderately Related)

Justification: Monte Carlo studies involve utilizing computational methods to analyze statistical problems, contributing to the research and analytical reasoning process, although they are not the sole focus.

CO4: Weightage: 1 (Partially Related)

Justification: Polynomial interpolation techniques are mathematical methods that may not directly align with the research process or analytical reasoning in statistical analysis.

CO5: Weightage: 3 (Strongly Related)

Justification: Estimation techniques such as Jack-knife and Bootstrap are integral parts of research and analytical reasoning in statistical analysis, providing insights into bias and improving the accuracy of estimators.

CO6: Weightage: 2 (Moderately Related)

Justification: Numerical methods play a role in solving problems encountered in data science and statistical analysis, contributing indirectly to the research and analytical reasoning process.

CO7: Weightage: 2 (Moderately Related)

Justification: Numerical optimization techniques are relevant for optimizing machine learning algorithms, which can be part of the research and analytical reasoning process in statistical analysis, although they may not be the primary focus.

PO6 Communication, Collaboration, and Leadership:

CO1: Weightage: 1 (Partially Related)

Justification: While computational techniques may enhance communication and collaboration through the analysis of statistical data, they are not directly related to communication, collaboration, or leadership skills themselves.

CO2: Weightage: 1 (Partially Related)

Justification: Numerical techniques for transformations and function approximation are important for data analysis but do not directly contribute to communication, collaboration, or leadership skills.

CO3: Weightage: 1 (Partially Related)

Justification: While Monte Carlo studies involve computational methods, they are not directly tied to communication, collaboration, or leadership skills.

CO4: Weightage: 1 (Partially Related)

Justification: Polynomial interpolation techniques are mathematical methods that may not directly contribute to communication, collaboration, or leadership skills.

CO5: Weightage: 1 (Partially Related)

Justification: Estimation techniques like Jack-knife and Bootstrap are essential for statistical analysis but do not directly relate to communication, collaboration, or leadership skills.

CO6: Weightage: 1 (Partially Related)

Justification: While numerical methods are crucial for solving problems in data science and statistical analysis, they do not directly enhance communication, collaboration, or leadership skills.

CO7: Weightage: 1 (Partially Related)

Justification: Numerical optimization techniques are important for machine learning but do not directly contribute to communication, collaboration, or leadership skills.

PO7 Digital Proficiency and Technological Skills:

CO1: Weightage: 2 (Moderately Related)

Justification: Computational techniques involve the use of software and technology to analyze statistical data, demonstrating proficiency in utilizing ICT and software for data analysis.

CO2: Weightage: 2 (Moderately Related)

Justification: Numerical techniques often require the use of software tools for implementation, showcasing proficiency in utilizing technology for mathematical computations.

CO3: Weightage: 3 (Strongly Related)

Justification: Monte Carlo studies and random number generators are essential components of digital proficiency in statistical analysis, requiring knowledge and skills in using specialized software tools and algorithms.

CO4: Weightage: 1 (Partially Related)

Justification: While polynomial interpolation techniques may involve some computational implementation, they are not directly focused on digital proficiency or technological skills.

CO5: Weightage: 2 (Moderately Related)

Justification: Utilizing Jack-knife and Bootstrap methods often involves computational implementations, contributing to digital proficiency in statistical analysis.

CO6: Weightage: 3 (Strongly Related)

Justification: Applying numerical methods in data science and statistical analysis requires proficiency in utilizing software tools and technology for problem-solving.

CO7: Weightage: 3 (Strongly Related)

Justification: Numerical optimization techniques are crucial in machine learning, requiring proficiency in using software tools and algorithms for optimization tasks, thereby demonstrating digital proficiency and technological skills.

PO8 Multicultural Competence, Inclusive Spirit, and Empathy:

CO1: Weightage: 1 (Partially Related)

Justification: Computational techniques do not directly contribute to multicultural competence, inclusive spirit, or empathy.

CO2: Weightage: 1 (Partially Related)

Justification: While numerical techniques are important for data analysis, they do not directly address multicultural competence, inclusive spirit, or empathy.

CO3: Weightage: 1 (Partially Related)

Justification: Monte Carlo studies and random number generators are statistical tools that do not directly relate to multicultural competence, inclusive spirit, or empathy.

CO4: Weightage: 1 (Partially Related)

Justification: Polynomial interpolation techniques are mathematical methods that do not directly contribute to multicultural competence, inclusive spirit, or empathy.

CO5: Weightage: 1 (Partially Related)

Justification: Estimation techniques like Jack-knife and Bootstrap are important for statistical analysis but do not directly enhance multicultural competence, inclusive spirit, or empathy.

CO6: Weightage: 1 (Partially Related)

Justification: Numerical methods are crucial for problem-solving but do not directly address multicultural competence, inclusive spirit, or empathy.

CO7: Weightage: 1 (Partially Related)

Justification: Numerical optimization techniques are important for machine learning but do not directly contribute to multicultural competence, inclusive spirit, or empathy.

PO9 Value Inculcation, Environmental Awareness, and Ethical Practices:

CO1: Weightage: 1 (Partially Related)

Justification: While computational techniques are valuable in statistical analysis, they do not directly address value inculcation, environmental awareness, or ethical practices.

CO2: Weightage: 1 (Partially Related)

Justification: Numerical techniques, while important for various applications, do not directly contribute to value inculcation, environmental awareness, or ethical practices.

CO3: Weightage: 1 (Partially Related)

Justification: Monte Carlo studies and random number generators are statistical tools that do not directly relate to value inculcation, environmental awareness, or ethical practices.

CO4: Weightage: 1 (Partially Related)

Justification: Polynomial interpolation techniques are mathematical methods that do not directly contribute to value inculcation, environmental awareness, or ethical practices.

CO5: Weightage: 1 (Partially Related)

Justification: Estimation techniques like Jack-knife and Bootstrap are important for statistical analysis but do not directly enhance value inculcation, environmental awareness, or ethical practices.

CO6: Weightage: 1 (Partially Related)

Justification: While numerical methods are essential in problem-solving, they do not directly address value inculcation, environmental awareness, or ethical practices.

CO7: Weightage: 1 (Partially Related)

Justification: Numerical optimization techniques are crucial in machine learning but do not directly contribute to value inculcation, environmental awareness, or ethical practices.

PO10 Autonomy, Responsibility, and Accountability:

CO1: Weightage: 1 (Partially Related)

Justification: While computational techniques are important for problem-solving, they do not directly address autonomy, responsibility, or accountability.

CO2: Weightage: 1 (Partially Related)

Justification: Numerical techniques are valuable in various applications but do not directly contribute to autonomy, responsibility, or accountability.

CO3: Weightage: 1 (Partially Related)

Justification: Monte Carlo studies and random number generators are statistical tools that do not directly relate to autonomy, responsibility, or accountability.

CO4: Weightage: 1 (Partially Related)

Justification: Polynomial interpolation techniques are mathematical methods that do not directly enhance autonomy, responsibility, or accountability.

CO5: Weightage: 1 (Partially Related)

Justification: Estimation techniques like Jack-knife and Bootstrap are important for statistical analysis but do not directly address autonomy, responsibility, or accountability.

CO6: Weightage: 1 (Partially Related)

Justification: While numerical methods are crucial for problem-solving, they do not directly contribute to autonomy, responsibility, or accountability.

CO7: Weightage: 1 (Partially Related)

Justification: Numerical optimization techniques are essential in machine learning but do not directly enhance autonomy, responsibility, or accountability.