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 – I

For Department of Statistics

Tuljaram Chaturchand College of Arts, Science and Commerce, Baramati

Choice Based Credit System Syllabus (2023 Pattern)

(As Per NEP 2020)

To be implemented from Academic Year 2023-2024

Title of the Programme: M.Sc. (Data Science) Part – I

Preamble

AES's Tuljaram Chaturchand College has made the decision to change the syllabus of across various faculties from June, 2023 by incorporating the guidelines and provisions outlined in the National Education Policy (NEP), 2020. The NEP envisions making education more holistic and effective and to lay emphasis on the integration of general (academic) education, vocational education and experiential learning. The NEP introduces holistic and multidisciplinary education that would help to develop intellectual, scientific, social, physical, emotional, ethical and moral capacities of the students. The NEP 2020 envisages flexible curricular structures and learning based outcome approach for the development of the students. By establishing a nationally accepted and internationally comparable credit structure and courses framework, the NEP 2020 aims to promote educational excellence, facilitate seamless academic mobility, and enhance the global competitiveness of Indian students. It fosters a system where educational achievements can be recognized and valued not only within the country but also in the international arena, expanding opportunities and opening doors for students to pursue their aspirations on a global scale.

In response to the rapid advancements in science and technology and the evolving approaches in various domains of Statistics and related subjects, the Board of Studies in Statistics at Tuljaram Chaturchand College, Baramati - Pune, has developed the curriculum for the first semester of M.Sc. Data Science, which goes beyond traditional academic boundaries. The syllabus is aligned with the NEP 2020 guidelines to ensure that students receive an education that prepares them for the challenges and opportunities of the 21st century. This syllabus has been designed under the framework of the Choice Based Credit System (CBCS), taking into consideration the guidelines set forth by the National Education Policy (NEP) 2020, LOCF (UGC), NCrF, NHEQF, Prof. R.D. Kulkarni's Report, Government of Maharashtra's General Resolution dated 20th April and 16th May 2023, and the Circular issued by SPPU, Pune on 31st May 2023.

The preamble of an MSc Data Science course typically provides an overview and introduction to the program, outlining its objectives, structure, and key features. It sets the context and expectations for students pursuing a Master's degree in Data Science.

The Master of Science (MSc) in Data Science program is designed to equip students with the knowledge, skills, and expertise necessary to excel in the rapidly evolving field of

data science. This interdisciplinary program combines principles and techniques from statistics, computer science, and domain-specific areas to enable students to extract actionable insights and make data-driven decisions.

The MSc Data Science program is structured to provide a balance between theoretical foundations, practical skills, and hands-on experience. The curriculum consists of a combination of core courses, elective courses, and a capstone project. The program also offers opportunities for specialization in areas such as machine learning, big data analytics, natural language processing, or business analytics.

Upon successful completion of the MSc Data Science program, graduates will be equipped with the knowledge and skills to take on roles such as data scientists, data analysts, machine learning engineers, or data consultants in diverse industries, including finance, healthcare, e-commerce, and technology.

Overall, revising the M.Sc. Data Science syllabus in accordance with the NEP 2020 ensures that students receive an education that is relevant, comprehensive, and prepares them to navigate the dynamic and interconnected world of today. It equips them with the knowledge, skills, and competencies needed to contribute meaningfully to society and pursue their academic and professional goals in a rapidly changing global landscape.

Programme Specific Outcomes (PSOs)

PSO1. Advanced Data Analysis: Apply advanced statistical and machine learning techniques to analyze complex datasets, identify patterns, and derive actionable insights.

PSO2. Data Visualization and Communication: Effectively visualize and communicate data insights through compelling visualizations, reports, and presentations.

PSO3. Statistical Computing and Programming: Utilize statistical software packages, such as R, Python, Pawer BI, SQL etc. to implement statistical analyses and simulations.

PSO4. Research and Problem-Solving: Identify research problems, formulate appropriate hypotheses, and design research studies.

PSO5. Statistical Consulting and Collaboration: Collaborate with researchers, scientists, and professionals from various domains to provide statistical support and consultancy.

PSO6. Ethical and Legal Considerations: Understand and navigate ethical and legal challenges related to data privacy, security, and governance in the field of data science.

PSO7. Deep Learning: Apply deep learning algorithms and neural networks to solve complex data analysis problems, such as image recognition and natural language processing.

PSO8. Predictive Modeling: Build predictive models using machine learning algorithms to make accurate predictions and forecasts.

PSO8.Data Mining and Knowledge Discovery: Utilize data mining techniques to extract valuable knowledge and patterns from large and complex datasets.

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. Rohan Koshti	Expert from other University
14.	Mr. Saurabh Kadam	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. Rushikesh Pandhare (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

Level	Semester	Major		Research Methodology (RM)	OJT/FP	RP	Cum. Cr.	Degree
		Mandatory	Electives					
6.0	Sem-I	DSC-501-MJM: Probability Distributions (Credit 04) DSC-502-MJM: Statistical Inference (Credit 04) DSC -503-MJM: Data Science Practical – I (Credit 02) DSC-504-MJM: Data Science Practical – II (Credit 02)	DSC-511-MJE (A): Data Base Management System DSC -511-MJE (B): Stochastic Models and Applications (Credit 04)	DSC -521-RM: Research Methodology (Credit 04)	--	--	**20**	PG Diploma (after 3 Year Degree)
	Sem-II	DSC -551-MJM: Design and Analysis of Experiments (Credit 04) DSC-552-MJM: Regression Analysis and Predictive Models (Credit 04) DSC-553-MJM: Data Science Practical – III (Credit 02) STA -554-MJM: Data Science Practical – IV (Credit 02)	DSC -561-MJE (A): Bayesian Inference DSC -561-MJE (B): Computational Statistics (Credit 04)	--	DSC-581-OJT/FP: On Job Training/ Field Project	--	**20**	

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

Sem	Course Type	Course Code	Course Title	Theory/ Practical	No. of Credits
I	Major (Mandatory)	DSC-501-MJM	Probability Distributions	Theory	04
	Major (Mandatory)	DSC -502-MJM	Statistical Inference	Theory	04
	Major (Mandatory)	DSC -503-MJM	Data Science Practical – I	Practical	02
	Major (Mandatory)	DSC -504-MJM	Data Science Practical – II	Practical	02
	Major (Elective)	DSC-511-MJE (A)	Data Base Management System	Theory	04
		DSC -511-MJE (B)	Stochastic Models and Applications	Theory	
	Research Methodology (RM)	DSC -521-RM	Research Methodology	Theory	04
	Total Credits Semester I				**20**
II	Major (Mandatory)	DSC -551-MJM	Design and Analysis of Experiments	Theory	04
	Major (Mandatory)	DSC-552-MJM	Regression Analysis and Predictive Models	Theory	04
	Major (Mandatory)	DSC-553-MJM	Data Science Practical – III	Practical	02
	Major (Mandatory)	DSC-554-MJM	Data Science Practical – IV	Practical	02
	Major (Elective)	DSC -561-MJE (A)	Bayesian Inference	Theory	04
		DSC -561-MJE (B)	Computational Statistics	Theory	
	On Job Training (OJT)/Field Project (FP)	DSC -581-OJT/FP	On Job Training Field Project	Training/Project	04
	Total Credits Semester-II				**20**
	Cumulative Credits Semester I and II				**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	: I
Course Type	: Major Mandatory Theory
Course Name	: Probability Distributions
Course Code	: DSC-501-MJM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. Explore various types of probability distributions.
2. Learn how to calculate and interpret probabilities.
3. Study properties and characteristics of probability distributions.
4. Apply probability distributions to real-world scenarios.
5. Develop skills in data analysis and statistical inference: Probability distribution concepts are often essential in statistical inference and data analysis.
6. Students may learn how to use software like R, Python, or Excel to analyze and visualize data based on probability distributions.
7. How to apply mathematical concepts to real-world situations, analyze problems, and make informed decisions based on probability.

Course Outcomes:

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

CO1. understand characteristics about discrete and continuous random variable and their probability distributions.

CO2. prepare students for modeling real data using distributions

CO3. develop understanding of distribution theory related for further advanced topics in statistical inference.

CO4. develop problem-solving techniques to solving real-world events.

CO5. apply selected probability distributions to solve problems.

CO6. present the analysis of derived statistics to all audiences.

CO7. develop problem-solving technique to calculate probability and conditional probability

Topics and Learning Points

Unit 1: (15 L)

Probability and Random variables

Introduction – Random Experiments, Empirical basis of probability, Algebra of events, laws of Probability; Conditional Probability, Independence, Bayes' law; Application of probability to business and economics. One-dimensional Random variable- Discrete and Continuous; Distribution functions and its properties; Bivariate Random Variables- Joint Probability functions, marginal distributions, conditional distribution functions, Notion of Independence of Random variables

Unit 2: (15L)

Functions of Random Variables

Functions of random variables: introduction, distribution function technique, transformation technique: one variable, transformation technique: several variables, theory and applications.

Mathematical Expectation

Expectation, Variance, and Co-variance of random variables; Conditional expectation and conditional variance; Markov, Holder, Jensen and Chebyshev's Inequality, Weak Law of Large numbers, Strong law of large numbers and Kolmogorov theorem, Central Limit Theorem.

Unit 3: (15L)

Generating Functions

Probability generating function (p.g.f.), moment generating function (m.g.f.), characteristic function (c.f.) Properties and Applications. Probability distributions of functions of random variables: one and two dimensions.

Sampling Distributions

Introduction, The sampling distribution of the Mean: Finite Populations, Sampling distribution of the proportion: Finite Populations, distribution of sample variance, the chi-square distribution, the t distribution, the F distribution, order statistics: properties, and applications, procedure of hypothesis testing.

Unit 4: (15L)

Discrete Distributions

Bernoulli, Binomial, Poisson, Geometric, Hypergeometric, Negative Binomial, Multinomial,

distributions and Discrete Uniform distribution - definition, properties and applications with numerical problems.

Continuous Distributions

Uniform, Normal distribution function, Exponential, Gamma, Beta distributions (First and Second kind), Weibull, Cauchy and Laplace distributions, lognormal, logistic, Pareto, Chi-sqaure and Rayleigh distribution functions - definition, properties and applications; concept of truncated distributions.

References:

1. Parimal Mukhopadhyay; An Introduction to the Theory of Probability, World scientific, 2012.
2. Irwin Miller, Marylees Miller, John E. Freund's; Mathematical Statistics, Pearson, 2017.
3. Fetsje Bijma, Marianne Jonker and Aadvander Vaart; Introduction to Mathematical Statistics, Amsterdam University Press, 2018.
4. Krishnamoorthy, K., Handbook of Statistical Distributions with Applications, Chapman & Hall/CRC, 2006.
5. Shanmugam, R., Chattamvelli, R. Statistics for scientists and engineers, John Wiley, 2015.

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	: I
Course Type	: Major Mandatory Theory
Course Name	: Statistical Inference
Course Code	: DSC-502-MJM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

Students successfully completing this course will be able to:

1. develop a solid understanding of foundational statistical concepts and principles.
2. study methods for estimating population parameters based on sample data.
3. gain proficiency in point estimation techniques, such as maximum likelihood estimation and method of moments.
4. acquire skills in formulating and conducting hypothesis tests. Learn how to apply different types of tests.
5. derive inference from different statistical data sets.
6. understand the role of statistical inference in machine learning algorithms.
7. apply statistical inference concepts and techniques to real-world data science problems.

Course Outcomes:

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

CO1. demonstrate a deep understanding of foundational statistical concepts, including probability theory, random variables, probability distributions.

CO2. explain the principles and assumptions underlying statistical inference and the role it plays in data science.

CO3. understand the concepts of Type I and Type II errors, significance levels, and p-values.

CO4. formulate appropriate null and alternative hypotheses for hypothesis testing.

CO5. conduct hypothesis tests using appropriate statistical tests.

CO6. understand the concept of interval estimation and confidence intervals.

CO7. recognize the role of statistical inference in machine learning algorithms.

Topics and Learning Points

Unit 1: **(15 L)**

Introduction

Population, sample, parameter and statistic; characteristics of a good estimator, Unbiasedness, Sufficiency – Factorization Theorem – Minimal sufficiency, Efficiency – Most Efficient estimator, likelihood equivalence, applications of Lehmann-Scheffe's Theorem, Rao-Blackwell Theorem and, Consistency –Invariance property of Consistent estimator, sufficient condition for consistency uniformly minimum variance unbiased Estimator.

Unit 2: **(15 L)**

Point Estimation

Point Estimation- Estimator, Estimate, Methods of point estimation – Maximum likelihood Method (the asymptotic properties of ML estimators are not included), and large sample Properties of ML estimator (without proof) -applications, Method of moments, method of Least squares, method of minimum chi-square and modified minimum chi-square asymptotic, Maximum Likelihood Estimation and applications.

Unit 3 : **(15L)**

Testing of hypotheses

Types of errors, power of a test, most powerful tests, Neyman–Pearson Fundamental Lemma and its applications; Notion of Uniformly most powerful tests; Likelihood Ratio tests, Description and property of LR tests - Application to standard distributions.

Large sample tests

Large sample properties, Tests of significance (under normality assumption) Test for a Population mean, proportion, Test for equality of two means, proportions, Test for Variance, Test for correlation, Test for Regression.

Small sample tests

Student's t-test, test for a population mean, equality of two population means, paired t-test, F-test For equality of two population variances, Chi-square test for goodness of Fit and test for independence of attributes, χ^2 test for testing variance of a Normal Distribution Analysis of Variance.

Unit 4: **(15L)**

Interval estimation confidence limits and confidence coefficient;

Duality between acceptance region of a test and a confidence interval, Construction of confidence intervals for population proportion (Small and large samples) and between two population proportions (large samples), Confidence intervals for mean and variance of a normal population; Difference between the Mean and ratio of two normal populations.

Non-parametric tests Sign test, Signed rank test, Median test, Mann-Whitney test, Run test and one sample Kolmogorov – Smirnov test, Kruskal – Wallis H test (Description, properties and applications only).

All topics to be covered using R software. Manual calculations are not expected.
Reference Books:

References:

1. Casella G. and Beregar R.L. (2002) Statistical Inference, 2nd Edition (Duxbury Advanced Series)
2. Dudewitz E.J. & Mishra S.N.(1988) Modern Mathematical Statistics (John Wiley)
3. Kale B.K. (1999) A First course on Parametric Inference (Narosa)
4. Lehman E.L (1988) Theory of point estimation (John Wiley)
5. Lehman E.L(1986) Testing of Statistical hypotheses (John Wiley)
6. Rohatagi V.K. (1976) Introduction to theory of probability & mathematical statistics (John Wiley & sons)
7. Dasgupta A. (2008), Asymptotic Theory of Statistics and Probability, Spinger-Verlag, New York.

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	: I
Course Type	: Major Mandatory Practical
Course Name	: Data Science Practical – I
Course Code	: DSC-503-MJM
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. Students should be able to review the core topics in probability and statistics through the study and practice of data analysis and graphical interpretation using statistical software.
2. Students will learn to use numerical computing tools and programming languages, such as MATLAB, Python, or R, to implement and solve linear algebra problems
3. Students should be able to solve systems of linear equations using various methods.
4. Students will explore applications of linear algebra in data analysis and machine learning.
5. Students should be able to plots different probability distributions and draw a model sample from it.
6. Students will develop a deep understanding of common probability distributions. Students will gain experience in applying probability distributions to real-world data analysis problems.

Course Outcomes:

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

CO1. use statistical software, packages such as R, Python, MATLAB, SPSS or Minitab to implement and analyze real life situations.

CO2. acquire skills in solving systems of linear equations using various techniques.

CO3. construct the orthogonal matrix associated with a non-singular matrix through a Gram-Schmidth orthogonalization process, diagonalization of a symmetric matrix, the role of eigenvalues, eigenvectors, Cayley Hamilton theorem in theory of matrices etc.

CO4. develop critical thinking skills to analyze and solve problems by using linear algebra concepts.

CO5. understand various discrete and continuous probability distributions along with their real-life applications.

CO6. proficient in using simulation techniques to generate random samples from specific probability distributions.

CO7. apply appropriate probability distributions to model and analyze real-world data sets from various fields.

Topics and Learning Points

Sr. No.	Title of Experiments
1.	Introduction to Statistical Software (Minitab, R, Matlab, SPSS) -I
2.	Introduction to Statistical Software (Minitab, R, Matlab, SPSS)- II
3.	Matrices
4.	G-Inverse and MPG-Inverse
5.	Eigen value, Eigen vectors, Spectral decomposition, Power of matrix- I
6.	Eigen value, Eigen vectors, Spectral decomposition, Power of matrix- II
7.	Solution of system of linear equations using Gauss elimination and Gauss Jordon methods
8.	Solution of system of linear equations using Gauss Seidal and Gauss Jacobi methods
9.	Application of Calley- Hamilton Theorem
10.	Classification and reduction of quadratic forms
11.	Plotting of density function, distribution functions, univariate and bivariate probability distributions
12.	Computation of probability of events related to bivariate probability distribution, non-central χ^2, t, F-distributions
13.	Model sampling from Gamma, Chi-square, Weibull, Lognormal probability distribution
14.	Model sampling from discrete, continuous and mixture distribution
15.	Model sampling from bivariate probability distribution

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	: I
Course Type	: Major Mandatory Practical
Course Name	: Data Science Practical – II
Course Code	: DSC-504-MJM
No. of Credits	: 2
No. of Teaching Hours	: 60

Course Objectives:

1. Develop proficiency in Python programming language and its syntax.
2. Understand the core concepts of object-oriented programming (OOP) and apply them in Python.
3. Gain a solid understanding of Structured Query Language (SQL) and its role in database management.
4. Learn how to write SQL queries to retrieve, modify, and analyze data in relational databases.
5. Apply data wrangling techniques to handle missing values, outliers, and inconsistent data.
6. Understand how to connect Python with SQL databases for data extraction and manipulation.
7. Apply Python and SQL skills to develop end-to-end data science projects.

Course Outcomes:

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

CO1. develop a strong command of Python programming language and its syntax

CO2. apply Python to perform data manipulation, preprocessing, and cleaning tasks.

CO3. demonstrate expertise in writing SQL queries to interact with relational databases.

CO4. apply SQL to retrieve, manipulate, and aggregate data from databases effectively.

CO5. apply Python to preprocess data, train models, evaluate model performance, and perform model selection.

CO6. understand and apply ethical guidelines and principles in data collection, storage, and analysis using Python and SQL

CO7. demonstrate proficiency in using Python libraries and frameworks commonly used in data science think critically to evaluate existing research literature in the field of statistics.

Topics and Learning Points

Sr. No.	Title of Experiments
1.	Basics of Python Language, When and why to use Python for Analytics. • Introduction & Installation of Python, Python Syntax, Strings, Lists and Dictionaries
2.	Loops and Regular Expressions
3.	Scientific Libraries in Python • Numpy, Scipy
4.	Introduction to Pandas • Selecting data from Pandas Data Frame, Slicing and dicing using Pandas
5.	Introduction to Pandas • GroupBY / Aggregate, Strings with Pandas, Cleaning up messy data with Pandas, Dropping Entries, Selecting Entries
6.	Data Manipulation using Pandas - I • Data Alignment, Sorting and Ranking, Summary Statistics, Missing values, Merging data
7.	Data Manipulation using Pandas - I • Concatenation, Combining Data Frames, Pivot, Duplicates, Binning
8.	Data visualization on using matplotlib and seaborn libraries • Scatter plot, Line plot, Bar plot, Histogram, Box plot, Pair plot
9.	Control structures using Toyota Corolla dataset • if-else family, for loop, for loop with if break, while loop
10.	Introduction to Database Management System This module introduces you to the database, the need for databases, and their examples. Further, you will learn about Database Management Systems and its history. Lastly, you will go through various Database Management System softwares
11.	Types of Database Management System • This chapter will cover various types of DBMS, including Hierarchical, Network, Relational, and Object-Oriented Databases. You will also be familiarized with several advantages of DBMS.
12.	Introduction to SQL • This chapter will brief you on the introduction to SQL and how to install it on your system. • In this chapter, you will learn how to implement various types of Commands in MySQL, such as DDL, DQL, DML, DCL, and TCL, with hands-on demos.
13.	Filter Record in MySQL • In this chapter, you will learn how to filter the records using the WHERE clause in MySQL, Operation in MySQL
14.	Pattern Matching in MySQL Here, you will learn to find patterns using the LIKE operator with the WHERE clause
15.	Null Values in MySQL In the final chapter, you will learn to insert a new record or update a record to an optional field without adding a value with the help of a Null value

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	: I
Course Type	: Major Elective Theory
Course Name	: Database Management System
Course Code	: DSC-511-MJE(A)
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. Students should gain a solid understanding of the basic concepts and principles of database management systems.
2. Students should learn how to design a relational database, including identifying entities, attributes, and relationships.
3. Students should become proficient in SQL, the standard language for interacting with relational databases.
4. Students should learn techniques for optimizing database queries to improve performance
5. Examine the logical, physical, and database modelling designs..
6. Students should be exposed to emerging trends and technologies in the field of database management systems.
7. Recognize how to create, modify, and query databases for data

Course Outcomes:

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

CO1. be familiar with the fundamentals of database concepts and database management systems.

CO2. understand the fundamental concepts and principles of database management systems, including data models, schemas, instances, and database architecture.

CO3. implement mechanisms for ensuring data integrity, such as primary keys, foreign keys, and constraints.utilize.

CO4. conceptual modelling techniques, like as the ER model and relational model, to model the data requirements for an application.

CO5. Write SQL commands to create tables, insert, update, delete and querying data.

CO6. create and manage database objects, such as tables, views, indexes, and constraints, using SQL.

CO7. normalize a database schema to eliminate redundancy and ensure data integrity.

Topics and Learning Points

Unit 1: (15L)

Introduction to file organization &DBMS, Database-system Applications, Purpose of Database Systems, Types of file Organization, File system Vs. DBMS, Data models, Levels of abstraction, Data in dependence, Structure of DBMS, Users of DBMS, Database Architecture, Speciality Databases.

Unit 2: (15L)

Structure of Relational Databases, Database Schema, Keys, Relational Operations, Conceptual Design (E-R model), Overview of DB design, ER data model (entities, attributes, entity sets, relations, relationship sets),Additional constraints (Key constraints, Mapping constraints), Conceptual design using ER modelling. Relational data model, Conversion of ER to Relational model, Integrity constraints, Relational algebra, Preliminaries.

Unit 3: (15L)

Introduction to SQL, Basic structure, Set operations, Aggregate functions, Null values , PL/PgSqL: Data types, Language structure, Operations with SQL, Nested Sub queries, Modifications to Database, DDL and DML commands with examples, SQL mechanisms for joining.

Unit 4 : (15L)

Intermediate and advanced SQL: Join Expressions- Join conditions, Outer joins, Join types and conditions, Views- View definition, using views in SQL queries, Materialized views, update a view 4.3 Create table extensions, Schemas, Catalogs and Environments, The relational Algebra, The tuple relational calculus.

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, Ramez Elmasri,(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.

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	: I
Course Type	: Major Elective Theory
Course Name	: Stochastic Models and Applications
Course Code	: DSC-511-MJE(B)
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. students should acquire a fundamental understanding of stochastic processes, including the definition, types, and basic properties.
2. to understand discrete and continuous Markov chains models to compute the probability of events.
3. formulate and solve problems by computing the long-term probabilities of a Markov chain model.
4. write Python/R code to simulate Markov chains, and compute probabilities of events that may be difficult to derive by hand.
5. apply Poisson processes to model the occurrence of events in various applications.
6. students understand the practical relevance and utility of stochastic processes in modeling and analyzing complex systems.
7. students should learn about the properties of Markov chains, including the Markov property, transition probabilities, stationary distributions, and ergodicity.

Course Outcomes:

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

CO1. perform calculations involving transition probabilities, steady-state distributions, and inter-arrival times for Markov chains and Poisson processes.

CO2. explain the fundamental principles of probability theory and random variables as they pertain to stochastic processes.

CO3. present findings and results of stochastic process analyses in a clear and organized manner.

CO4. use simulation methods and computational tools to model and analyze stochastic

processes.

CO5. evaluate the appropriateness of different stochastic process models for specific applications and contexts.

CO6. to familiar with stochastic processes, including Poisson process, Wiener process and Renewal process, etc.

CO7. demonstrate the ability to independently research and explore advanced topics related to stochastic processes

Topics and Learning Points

Unit 1: **(15L)**

Notion of stochastic processes, Markov chain, one step transition probabilities, Chapman-Kolmogorov equations, evaluation of higher step transition probabilities, classification of states, periodicity of a Markov chain, concept of closed class, minimal closed class, stationary distribution. Some examples such as gamblers ruin problem and one-dimensional random walk. Concept of absorption probabilities, Use of these to compute probability of winning the game by a gambler having initial capital ‘a’

Unit 2: **(15L)**

Branching process, classification of states, identification of criticality parameter, extinction probability, relationship between criticality parameter and extinction probability of the process, Expression for mean and variance of the process. Extinction probability, Some epidemiological applications, Introduction to Markov chain in continuous time, concept of intensity rate, relationship between intensity matrix and transition probability matrix. Kolmogorov’s forward and backward equations

Unit 3: **(15L)**

Introduction to birth process, birth and death process, linear birth and death process, Growth model with immigration and related results, Expression for mean and variance of a birth process and, birth and death process, Applications of these processes.

Unit 4: **(15L)**

i) Poisson process, two definitions and their equivalence, Distribution of inter arrival times, conditional joint distribution of inter arrival times. Compound Poisson process,

Some applications. Introduction to renewal process, relationship with Poisson process, key and elementary renewal theorems associated with renewal processes.

References:

1. Bhat B.R. (2000). Stochastic Models: Analysis and Applications, New Age International.
2. Medhi, J. (2010) Stochastic Processes, New Age Science Ltd.
3. Pinsky M. A. and Karlin, S. (2010). An Introduction to Stochastic Modeling, 4thEdn. Academic Press.
4. Ross, S. (2014). Introduction to Probability Models, 11th Edn. Academic Press.
5. Feller, W. (1972). An Introduction to Probability Theory and its Applications, Vol. 1, Wiley Eastern.
6. Hoel, P.G. Port, S.C. & Stone, C.J. (1972). Introduction to Stochastic Processes, Houghton Mifflin
7. Karlin, S & Taylor, H.M. (1975). A First Course in Stochastic Processes (Second. Edition), Academic Press.
8. Serfozo, R. (2009). Basics of Applied Stochastic Processes, Springer.

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	: I
Course Type	: Research Methodology
Course Name	: Research Methodology
Course Code	: DSC-521-RM
No. of Credits	: 4
No. of Teaching Hours	: 60

Course Objectives:

1. To introduce the statistical aspects associated with the design and analysis of sample surveys, and to develop your understanding of the principles and methods used to design survey sampling schemes.
2. Understand the steps in developing a sampling plan.
3. Distinguish between probability and non-probability sampling.
4. Develop critical thinking on sampling methods and results.
5. Understand potential sources of error and limitations of different sampling techniques.
6. To introduce the fundamental concepts and principles of research, including the scientific method, research questions and research designs.
7. To develop skills in designing research studies, including formulating research questions, selecting appropriate research designs.
8. To develop critical thinking to evaluate research studies, methodologies, and findings.

Course Outcomes:

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

CO1. define principal concepts about sampling.

CO2. lists the stages of sampling process.

CO3. understand the distinctive features of different sampling techniques and their related estimation problems.

CO4. learn the practical applications of the various sampling techniques in real life situations.

CO5. develop an appreciation for research ethics and demonstrate an understanding of

ethical principles and guidelines in conducting research.

CO6. apply appropriate research design principles to formulate research questions.

CO7. develop research proposals that demonstrate clear research objectives, appropriate methodologies, and justifications for the significance of the research.

Topics and Learning Points

Unit – 1 **(15L)**

Introduction to Sampling , review of basic finite population sampling techniques SRSWR, SRSWOR, Stratified, Systematic, Probability Proportional to Size With Replacement (PPSWR) methods, cumulative total method and Lahiri's method for estimation problem, estimation of finite population mean and total, PPSWOR methods and related estimation of a finite population mean (Horvitz-Thompson and Des Raj estimators for a general sample size and Murthy's estimator for a sample of size 2), midzuno scheme of sampling.

Unit – 2 **(15L)**

Use of supplementary information for estimation, ratio and regression estimators using separate strata and combined strata, unbiased and almost unbiased ratio type estimators of population mean, post stratification, variance of estimator of population mean under it. Cluster sampling with clusters of equal sizes and unequal size, estimation of population mean and its standard error, two stage sampling with equal first stage units, expected value and the variance of sample mean, multistage-sampling, Multiphase sampling.

Unit – 3 **(10L)**

Meaning of research, objective of research, motivation in research, types of research, research approaches, significance of research, defining the research problem, selecting the problem, necessity of defining the problem, techniques involved in defining a problem, designing a questionnaire.

Unit – 4 **(20L)**

Layout of the research report, types of reports, construction of title and preparation of abstract for research paper / proposed project, writing of materials and methods, results discussions, conclusion etc., writing of research proposals, significance of report writing, different steps in writing report, oral presentation, mechanics of writing research report, precautions for writing research reports, research ethics. Use of tools or techniques for research: methods to search required information effectively, reference management software like Zotero/ Mendeley, software for paper formatting like LaTex/ MS office, software for detection of plagiarism.

References:

1. Des Raj & Chandhok P. (1998), Sample survey theory. (Narosa)
2. Murthy M.N. (1977) Sampling theory and methods. (Statistical Publishing Society)
3. Parimal Mukhopadhyay, Theory and methods of survey sampling, Prentice Hall of India private limited, 2nd Edition, 2008.
4. W.G.Cochran, (1977) Sampling techniques.(John Wiley and sons)
5. Sukhatme P. V. Sukhatme B. V. and C. Ashok Sampling theory of survey and applications.(Indian society for Agricultural statistics)
6. Research Methodology: Methods and Techniques, Kothari C.R., 1990. New Age International.
7. An introduction to Research Methodology; Garg B.L., Karadia, R., Agarwal, F. and Agarwal, U.K., 2002., RBSA Publishers.
8. Research Methodology; Sinha S.C. and Dhiman, A.K., 2002. Ess Publications. 2 volumes.
9. Research Methods: the concise knowledge base; Trochim W.M.K., 2005. Atomic Dog Publishing. 270p.
10. Research Methodology ; Panneerselvam R., PHI, Learning Pvt. Ltd., New Delhi – 2009.