DIBRUGARH UNIVERSITY

Syllabus for FYIPGP in
Data Science
offered by
Department of Statistics, Dibrugarh University

(For Semesters I, II, & III)

As approved in the meeting of the BoS held on 07/03/2026

To be effective from the Session 2026-27

Programme Preamble

The Bachelor's Programme in Data Science course offered by the Department of Statistics, Dibrugarh University, is designed to integrate statistical foundations with modern computational and analytical techniques in alignment with the UGC FYIPGP framework. Recognizing the growing importance of data-driven decision-making across sectors, the programme emphasizes statistical reasoning, mathematical modeling, programming, and applied analytics. The curriculum progresses systematically from basic statistical methods and data handling to advanced topics such as machine learning, multivariate analysis, econometrics, big data analytics, and domain-specific applications. With strong experiential components including laboratories, internships, projects, and research methodology, the programme aims to develop competent data professionals grounded in statistical rigor, ethical practice, and interdisciplinary relevance.

Introduction to the Programme

This Data Science programme is structured as a multi-exit FYIPGP pathway leading from Undergraduate Certificate to Bachelor's Degree (Honours / Honours with Research). The curriculum begins with foundational courses in statistics and data science, followed by progressive exposure to probability, inference, regression, programming (R and Python), visualization tools, machine learning, and specialized analytics.

Distinctively anchored in Statistics, the programme equips learners with theoretical depth and practical competence, enabling them to model uncertainty, analyze complex datasets, and generate actionable insights. Through internships, community engagement, mini-projects, and dissertation work, students gain hands-on experience with real-world data, preparing them for careers in analytics, research, industry, and higher education

Aim of the Programme

The aim of the programme is to produce statistically grounded data science professionals capable of collecting, analyzing, modeling, and interpreting data using modern computational tools, while adhering to ethical standards and contributing meaningfully to society and industry.

Graduate Attributes (GA)

On successful completion of the programme, graduates will demonstrate:

GA1: Disciplinary Knowledge
Strong foundation in statistics, mathematics, and data science concepts.

GA2: Analytical and Critical Thinking
Ability to analyze complex datasets and draw evidence-based conclusions.

GA3: Problem Solving Skills
Apply statistical and computational techniques to real-world problems.

GA4: Digital and Technical Competence
Proficiency in programming, statistical software, visualization tools, and analytics platforms.

GA5: Communication Skills

Ability to present analytical findings clearly using visual, numerical, and written formats.

GA6: Ethical and Social Responsibility

Responsible handling of data with awareness of privacy, bias, and societal impact.

GA7: Research Orientation and Lifelong Learning

Preparedness for higher studies, research, and continuous professional development.

GA8: Teamwork and Professionalism

Capability to work collaboratively in multidisciplinary environments.

Programme Outcomes (POs)

Upon completion of the FYIPGP Data Science Programme, learners will be able to:

PO1: Demonstrate comprehensive knowledge of statistics, mathematics, and data science principles.

PO2: Collect, clean, organize, analyze, and interpret data using appropriate statistical and computational tools.

PO3: Apply probability, inference, regression, and machine learning techniques for data-driven decision making.

PO4: Develop analytical models and visualizations to communicate insights effectively.

PO5: Use programming languages (R and Python) and software platforms (Excel, SQL, Tableau, and Hadoop) for data analysis.

PO6: Conduct independent or collaborative projects, including research-oriented work.

PO7: Practice ethical data handling and recognize societal implications of analytics.

PO8: Pursue careers or higher education in data science, statistics, analytics, and allied fields.

DIBRUGARH UNIVERSITY, RAJABHETA, DIBRUGARH – 786004

FYIPGP Course Structure as per UGC Credit Framework of December, 2022

Year	Semester	Course	Title of the Course	Total Credit
Year 1	1st Semester	C 1	Introduction to Data Science and Basic Statistical Methods	4
		Minor 1	Introduction to Data Science and Basic Statistical Methods	4
		GEC 1	Statistical Methods	3
		AEC 1	Modern Indian Language	4
		VAC 1	Understanding India (Or) Health and Wellness	2
		SEC 1	Fundamentals of Computer Science	3
		Total		**20**
	2nd Semester	C 2	Introduction to Python and Data Visualization	4
		Minor 2	Introduction to Python and Data Visualization	4
		GEC 2	Basics of Probability Distribution and Inference	3
		AEC 2	English Language and Communication Skills	4
		VAC 2	Environmental Science (Or)Yoga Education	2
		SEC 2	Data Science Using MS-Excel and SQL	3
		Total		**20**
The students on exit shall be awarded Undergraduate Certificate (in the Field of Study/Discipline) after securing the requisite 40 Credits in Semester 1 and 2 provided they secure 4 credits in work based vocational courses offered during summer term or internship / Apprenticeship in addition to 6 credits from skill-based courses earned during 1st and 2nd Semester.				
Year 2	3rd Semester	C 3	Mathematics for Data Science	4
		C 4	Introduction to Probability and Distributions	4
		Minor 3	Introduction to Probability and Distributions	4
		GEC 3	Applied Statistics	3
		VAC 3	Digital and Technological Solutions/ Digital Fluency	2
		SEC 3	Statistical Data Processing using R	3
		Total		**20**
Grand Total (Semester I, II, and III)				**60**

DATA SCIENCE PROGRAMME (NEP)
DETAILED SYLLABUS OF 1st SEMESTER

Title of the Course	**:**	**Introduction to Data Science and Basic Statistical Methods**
Course Code	**:**	**C1**
Nature of the Course	**:**	**Core**
Total Credits	**:**	**4**
Distribution of Marks	**:**	**End Sem: 60 TH, In Sem: 40 TH**

COURSE OBJECTIVES:

Knowledge:

- Introduce the fundamental concepts of data science, including its evolution, lifecycle, and ethical considerations.
- Learn the different types of data, including structured, semi-structured, and unstructured data.
- Acquire knowledge of big data concepts, characteristics, and the importance of data quality.
- Understand the role of statistics in data science, including population and sample, variables, attributes, and levels of measurement.
- Learn methods of data representation through tabular, graphical, and categorical data analysis.
- Understand the principles of data engineering and data preprocessing, including data cleaning, missing data treatment, outlier detection, transformation, standardization, normalization.
- Gain knowledge of quantitative descriptive analytics, including measures of central tendency, dispersion, skewness, and kurtosis.
- Understand exploratory data analysis techniques such as five-number summaries, stem-and-leaf plots, box plots, and robust methods.

Skills:

- Classify different types of data and identify their appropriate use in data science applications.
- Apply statistical thinking to summarize, organize, and interpret data.
- Construct and interpret tabular, graphical, and categorical representations of data.
- Identify data quality issues and apply suitable preprocessing techniques to improve data reliability.
- Handle missing values, detect outliers, and perform data transformation, standardization, and normalization.
- Compute and interpret measures of central tendency, dispersion, skewness, and kurtosis.
- Use exploratory data analysis techniques to detect patterns, variability, and unusual observations in data.
- Draw meaningful conclusions from descriptive statistical summaries for data-driven decision making.

Attitude:

- Develop a logical and analytical approach to understanding and solving data-oriented problems.
- Appreciate the importance of data quality, accuracy, and consistency in statistical analysis.
- Cultivate awareness of ethical responsibility, privacy, and fairness in handling data.
- Develop confidence in interpreting and communicating statistical information.
- Show interest in applying statistical and data science methods to real-world problems.
- Build a systematic and organized approach to data preparation and analysis.
- Demonstrate critical thinking and perseverance while working with complex datasets.
- Develop a foundation for advanced study in statistics, data science, machine learning, and related fields.

COURSE OUTCOMES:

After completion of the course learners will be able to

CO1: Describe the fundamental concepts of data science, including data lifecycle, types of data, and ethical issues.

ILO1: Define data science and explain its scope and applications.

ILO2: Describe the stages of the data lifecycle.

ILO3: Identify different types of data and their characteristics.

ILO4: Explain ethical issues related to data collection, analysis, and use.

CO2: Classify data based on type, source, and measurement scales and represent data using appropriate tabular and graphical methods.

ILO1: Classify data according to type, source, and measurement scales.

ILO2: Organize data using frequency distributions and tabular methods.

ILO3: Construct appropriate graphical representations such as bar charts, histograms, pie charts, and ogives.

ILO4: Select suitable tabular and graphical methods for different types of data.

CO3: Apply statistical thinking to summarize and interpret data using tabular and graphical techniques.

ILO1: Summarize data using tables and graphical displays.

ILO2: Interpret patterns, trends, and variability from graphical representations.

ILO3: Compare datasets using appropriate tabular and graphical techniques.

ILO4: Draw preliminary conclusions from summarized data.

CO4: Perform data preprocessing tasks such as handling missing values, detecting outliers, and transforming data.

ILO1: Identify missing values and describe methods for handling them.

ILO2: Detect outliers using suitable statistical and graphical methods.

ILO3: Apply data transformation techniques such as scaling, normalization, and coding.

ILO4: Prepare datasets for further statistical analysis.

CO5: Perform exploratory data analysis using descriptive statistical measures and graphical techniques to draw meaningful conclusions.

ILO1: Compute measures of central tendency and dispersion.

ILO2: Use graphical techniques such as box plots, scatter plots, and histograms for exploratory analysis.

ILO3: Examine relationships, patterns, and distributions in data.

ILO4: Interpret exploratory analysis results and draw meaningful conclusions.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	**Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge	CO1					
Conceptual Knowledge	CO1	CO2, CO3				
Procedural Knowledge			CO2, CO3	CO4		
Metacognitive Knowledge					CO5	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO / PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**
CO1	S	–	–	–	–	–	L	–
CO2	M	S	–	M	–	–	–	–
CO3	M	S	M	–	–	–	–	–
CO4	L	M	S	–	–	–	–	–
CO5	L	M	M	M	–	S	L	–

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (12 marks)	**Foundations and Ecosystem of Data Science:** Evolution of Data Science and fundamental steps; Meaning and scope; DIKW hierarchy (Data-Information-Knowledge-Wisdom); Data Science Lifecycle (CRISP-DM); Roles and responsibilities in Data Science; Ethics, privacy, and algorithmic bias. **Types of Data:** Structured, semi-structured, and unstructured data; Concept of Big Data and its characteristics (5 V's); Introduction to data analytics and mining; Principles of data quality and its key components.	12	–	–	12
2 (12 Marks)	**Statistical Thinking and Data Characterization:** Definition and scope of Statistics in Data Science; Statistical population vs. sample. **Types of Data:** Quantitative and qualitative data; Primary vs. secondary sources; Attributes and variables; Levels of measurement (Nominal, Ordinal, Interval, and Ratio). **Data Representation:** Principles of tabular and graphical presentation. **Categorical Data Analysis:** Contingency tables and measures of association.	12	–	–	12
3 (18 marks)	**Data Engineering and Pre-processing:** Overview of data cleaning; Identifying and mitigating data quality issues; Handling missing data: identification and imputation strategies. Data Wrangling: Detection and treatment of outliers; Data consistency and removal of duplicate records. Transformation: Data formatting, standardization, and normalization; Scaling techniques; Data integration and linkage from multiple sources; Introduction to data warehousing and structured storage.	18	–	–	18
4 (18 Marks)	**Quantitative Descriptive Analytics:** Measures of Central Tendency: Mathematical and positional measures with geometric interpretations. Measures of Dispersion: Range, quartile deviation, mean deviation, standard deviation, Inter quartile range, and coefficient of variation. Distributional Characteristics: Raw, central, and factorial moments; Analysis of Skewness and Kurtosis; Sheppard's corrections. Exploratory Data Analysis (EDA): Five-point summary, Stem-and-Leaf plots, Box-plots, and robust outlier detection.	18	–	–	18

where ***L: Lectures*** ***T: Tutorials*** ***P: Practicals***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] Bruce, P., & Bruce, A., "Practical Statistics for Data Scientists: 50 Essential Concepts," O'Reilly Media, Inc., 2017.

[2] Singh, P., Mishra, A. R., & Garg, P., "Data Analytics and Machine Learning: Navigating the Big Data Landscape," Springer Nature, 2024.

[3] Jones, J. S., & Goldring, "Explanatory and Descriptive Statistics," SAGE Publications Limited, 2021.

[4] Hartwig, F., & Dearing, B. E., "Exploratory Data Analysis," SAGE, 1979.

[5] Kotu, V., & Deshpande, B., "Data Science: Concepts and Practice," Morgan Kaufmann, 2018.

[6] Gupta, S. P., "Statistical Methods," 43rd ed., Sultan Chand & Sons, 2018.

[7] Bagavathi, R. S. N. P., "Statistics: Theory and Practice," S. Chand Publishing, 2019.

Title of the Course	**:**	**Introduction to Data Science and Basic Statistical Methods**
Course Code	**:**	**Minor 1**
Nature of the Course	**:**	**Minor**
Total Credits	**:**	**4**
Distribution of Marks	**:**	**End Sem:60 TH, In Sem: 40 TH**

COURSE OBJECTIVES:

Knowledge:

- Introduce the fundamental concepts of data science, including its evolution, lifecycle, and ethical considerations.
- Learn the different types of data, including structured, semi-structured, and unstructured data.
- Acquire knowledge of big data concepts, characteristics, and the importance of data quality.
- Understand the role of statistics in data science, including population and sample, variables, attributes, and levels of measurement.
- Learn methods of data representation through tabular, graphical, and categorical data analysis.
- Understand the principles of data engineering and data preprocessing, including data cleaning, missing data treatment, outlier detection, transformation, standardization, normalization.
- Gain knowledge of quantitative descriptive analytics, including measures of central tendency, dispersion, skewness, and kurtosis.
- Understand exploratory data analysis techniques such as five-number summaries, stem-and-leaf plots, box plots, and robust methods.

Skills:

- Classify different types of data and identify their appropriate use in data science applications.
- Apply statistical thinking to summarize, organize, and interpret data.
- Construct and interpret tabular, graphical, and categorical representations of data.
- Identify data quality issues and apply suitable preprocessing techniques to improve data reliability.
- Handle missing values, detect outliers, and perform data transformation, standardization, and normalization.
- Compute and interpret measures of central tendency, dispersion, skewness, and kurtosis.
- Use exploratory data analysis techniques to detect patterns, variability, and unusual observations in data.
- Draw meaningful conclusions from descriptive statistical summaries for data-driven decision making.

Attitude:

- Develop a logical and analytical approach to understanding and solving data-oriented problems.
- Appreciate the importance of data quality, accuracy, and consistency in statistical analysis.
- Cultivate awareness of ethical responsibility, privacy, and fairness in handling data.
- Develop confidence in interpreting and communicating statistical information.
- Show interest in applying statistical and data science methods to real-world problems.
- Build a systematic and organized approach to data preparation and analysis.
- Demonstrate critical thinking and perseverance while working with complex datasets.
- Develop a foundation for advanced study in statistics, data science, machine learning, and related fields.

COURSE OUTCOMES:

After completion of the course learners will be able to

CO1: Describe the fundamental concepts of data science, including data lifecycle, types of data, and ethical issues.

ILO1: Define data science and explain its scope and applications.

ILO2: Describe the stages of the data lifecycle.

ILO3: Identify different types of data and their characteristics.

ILO4: Explain ethical issues related to data collection, analysis, and use.

CO2: Classify data based on type, source, and measurement scales and represent data using appropriate tabular and graphical methods.

ILO1: Classify data according to type, source, and measurement scales.

ILO2: Organize data using frequency distributions and tabular methods.
ILO3: Construct appropriate graphical representations such as bar charts, histograms, pie charts, and ogives.
ILO4: Select suitable tabular and graphical methods for different types of data.

CO3: Apply statistical thinking to summarize and interpret data using tabular and graphical techniques.
ILO1: Summarize data using tables and graphical displays.
ILO2: Interpret patterns, trends, and variability from graphical representations.
ILO3: Compare datasets using appropriate tabular and graphical techniques.
ILO4: Draw preliminary conclusions from summarized data.

CO4: Perform data preprocessing tasks such as handling missing values, detecting outliers, and transforming data.
ILO1: Identify missing values and describe methods for handling them.
ILO2: Detect outliers using suitable statistical and graphical methods.
ILO3: Apply data transformation techniques such as scaling, normalization, and coding.
ILO4: Prepare datasets for further statistical analysis.

CO5: Perform exploratory data analysis using descriptive statistical measures and graphical techniques to draw meaningful conclusions.
ILO1: Compute measures of central tendency and dispersion.
ILO2: Use graphical techniques such as box plots, scatter plots, and histograms for exploratory analysis.
ILO3: Examine relationships, patterns, and distributions in data.
ILO4: Interpret exploratory analysis results and draw meaningful conclusions.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	**Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge	CO1					
Conceptual Knowledge	CO1	CO2, CO3				
Procedural Knowledge			CO2, CO3	CO4		
Metacognitive Knowledge					CO5	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO / PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**
CO1	S	–	–	–	–	–	L	–
CO2	M	S	–	M	–	–	–	–
CO3	M	S	M	–	–	–	–	–
CO4	L	M	S	–	–	–	–	–
CO5	L	M	M	M	–	S	L	–

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 **(12 marks)**	**Foundations and Ecosystem of Data Science:** Evolution of Data Science and fundamental steps; Meaning and scope; DIKW hierarchy (Data-Information-Knowledge-Wisdom); Data Science Lifecycle (CRISP-DM); Roles and responsibilities in Data Science; Ethics, privacy, and algorithmic bias. **Types of Data:** Structured, semi-structured, and unstructured data; Concept of Big Data and its characteristics (5 V's); Introduction to data analytics and mining; Principles of data quality and its key components.	12	–	–	12
2 **(12 Marks)**	**Statistical Thinking and Data Characterization:** Definition and scope of Statistics in Data Science; Statistical population vs. sample. **Types of Data:** Quantitative and qualitative data; Primary vs. secondary sources; Attributes and variables; Levels of measurement (Nominal, Ordinal, Interval, and Ratio). **Data Representation:** Principles of tabular and graphical presentation. **Categorical Data Analysis:** Contingency tables and measures of association.	12	–	–	12
3 **(18 marks)**	**Data Engineering and Pre-processing:** Overview of data cleaning; Identifying and mitigating data quality issues; Handling missing data: identification and imputation strategies. Data Wrangling: Detection and treatment of outliers; Data consistency and removal of duplicate records. Transformation: Data formatting, standardization, and normalization; Scaling techniques; Data integration and linkage from multiple sources; Introduction to data warehousing and structured storage.	18	–	–	18
4 **(18 Marks)**	**Quantitative Descriptive Analytics:** Measures of Central Tendency: Mathematical and positional measures with geometric interpretations. Measures of Dispersion: Range, quartile deviation, mean deviation, standard deviation, Inter quartile range, and coefficient of variation. Distributional Characteristics: Raw, central, and factorial moments; Analysis of Skewness and Kurtosis; Sheppard's corrections. Exploratory Data Analysis (EDA): Five-point summary, Stem-and-Leaf plots, Box-plots, and robust outlier detection.	18	–	–	18

Where, ***L: Lectures*** ***T: Tutorials*** ***P: Practicals***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] Bruce, P., & Bruce, A., "Practical Statistics for Data Scientists: 50 Essential Concepts," O'Reilly Media, Inc., 2017.

[2] Singh, P., Mishra, A. R., & Garg, P., "Data Analytics and Machine Learning: Navigating the Big Data Landscape," Springer Nature, 2024.

[3] Jones, J. S., & Goldring, "Explanatory and Descriptive Statistics," SAGE Publications Limited, 2021.

[4] Hartwig, F., & Dearing, B. E., "Exploratory Data Analysis," SAGE, 1979.

[5] Kotu, V., & Deshpande, B., "Data Science: Concepts and Practice," Morgan Kaufmann, 2018.

[6] Gupta, S. P., "Statistical Methods," 43rd ed., Sultan Chand & Sons, 2018.

[7] Bagavathi, R. S. N. P., "Statistics: Theory and Practice," S. Chand Publishing, 2019.

Title of the Course	**:**	**Statistical Methods**
Course Code	**:**	**GEC1**
Nature of the Course	**:**	**Generic Elective Course**
Total Credits	**:**	**3**
Distribution of Marks	**:**	**End Sem:60 TH, In Sem: 40 TH**

COURSE OBJECTIVES:

Knowledge:

- Understand the definition, scope, and role of statistics in data science.
- Gain knowledge of statistical population and sample, and distinguish between qualitative and quantitative data.
- Understand primary and secondary data, attributes, variables, and levels of measurement such as nominal, ordinal, interval, and ratio scales.
- Learn methods of data presentation through tabular and graphical forms, including histograms and ogives.
- Acquire knowledge of measures of central tendency and measures of dispersion, together with their mathematical and geometric interpretations.
- Understand moments, skewness, kurtosis, and Sheppard's corrections.
- Gain knowledge of bivariate data analysis, including correlation, rank correlation, and simple linear regression.
- Understand the principle of least squares and the fitting of polynomial and exponential curves.
- Learn the concepts of multivariate data, including partial and multiple correlation, coefficient of multiple correlation, and coefficient of determination.
- Understand categorical data analysis, including consistency of data, independence and association of attributes, contingency, and the concept of odds ratio.

Skills:

- Classify different types of data and select appropriate scales of measurement.
- Organize and present data using tabular and graphical methods.
- Compute and interpret measures of central tendency, dispersion, moments, skewness, and kurtosis.
- Apply Sheppard's corrections where appropriate.
- Analyze relationships between two variables using correlation, rank correlation, and simple linear regression.
- Fit polynomial and exponential curves to observed data using the principle of least squares.
- Compute and interpret partial and multiple correlation coefficients and the coefficient of determination.
- Analyze categorical data using measures of association, contingency tables, and tests of independence of attributes.
- Solve statistical problems using both calculator-based and computer-based approaches.

Attitude:

- Develop logical and analytical thinking in interpreting statistical information.
- Appreciate the importance of accuracy, consistency, and careful interpretation of data.
- Build confidence in handling numerical data and drawing meaningful conclusions.
- Cultivate a systematic approach to problem-solving through statistical methods.
- Show interest in applying statistical tools to practical problems in data science, social sciences, and medical sciences.
- Develop critical thinking while examining relationships, patterns, and variability in data.

COURSE OUTCOMES:

After completion of the course learners will be able to

CO1: Explain fundamental statistical concepts and their relevance to Data Science.
- **ILO1:** Define basic statistical terms such as population, sample, variable, and parameter.
- **ILO2:** Distinguish between descriptive and inferential statistics.
- **ILO3:** Explain the role of statistics in data collection, analysis, and interpretation.
- **ILO4:** Describe the relevance of statistical concepts in data science applications.

CO2: Organize and present data using appropriate statistical methods.
- **ILO1:** Classify data according to type and scale of measurement.

ILO2: Organize data using frequency distributions and tabular methods.
ILO3: Construct graphical representations such as bar charts, histograms, pie charts, and ogives.
ILO4: Select suitable methods for presenting different types of data.

CO3: Compute and interpret measures of central tendency and dispersion.
ILO1: Compute measures of central tendency such as mean, median, and mode.
ILO2: Compute measures of dispersion such as range, variance, standard deviation, and coefficient of variation.
ILO3: Compare datasets using measures of central tendency and dispersion.
ILO4: Interpret the significance of summary measures in describing data.

CO4: Apply basic probability and inferential concepts to simple statistical problems.
ILO1: Define fundamental concepts of probability, including events, sample space, and probability rules.
ILO2: Solve simple problems involving probability distributions and expected values.
ILO3: Explain the basic ideas of sampling, estimation, and hypothesis testing.
ILO4: Apply elementary probability and inferential methods to simple statistical problems.

CO5: Interpret statistical results critically for informed decision making.
ILO1: Interpret statistical summaries and graphical outputs.
ILO2: Identify patterns, trends, and variability in data.
ILO3: Evaluate statistical results in relation to practical problems.
ILO4: Draw informed conclusions for decision making based on statistical evidence.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	**Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge	CO1					
Conceptual Knowledge	CO1	CO2, CO3				
Procedural Knowledge			CO2, CO3	CO4		
Metacognitive Knowledge					CO5	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO / PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**
CO1	S	M	L	–	–	–	-	–
CO2	M	S	M	M	–	–	–	–
CO3	M	S	M	–	–	–	–	–
CO4	L	S	S	–	–	–	–	–
CO5	L	M	M	M	M	L	–	–

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (12 Marks)	**Introduction**: Definition and scope of Statistics, role of statistics in data science, concepts of statistical population and sample. Data: quantitative and qualitative, primary and secondary, attributes, variables, scales of measurement - nominal, ordinal, interval and ratio. **Presentation**: tabular and graphical, including histogram and ogives.	17	01	–	08
2 (14 Marks)	**Measures of Central Tendency:** Mathematical and positional measures with their geometry. **Measures of Dispersion:** range, quartile deviation, mean deviation, standard deviation, coefficient of variation, Moments, absolute moments, factorial moments, skewness and kurtosis, Sheppard's corrections with their geometry.	11	01	–	12
3 (14 Marks)	**Bivariate Data:** Definition, scatter diagram, simple correlation, interpretation of r and rank correlation. Simple linear regression, principle of least squares and its geometry, fitting of polynomials and exponential curves. **Multivariate Data:** Definition, partial and multiple correlation, interpretation of R, and coefficient of determination.	11	01	–	12
4 (10 Marks)	**Categorical Data:** Attributes, consistency of data, independence and association of attributes, measures of association and contingency. concept of Odds ratio with example from medical science	07	01	–	08
5 (10 Marks)	**Co-requisites: (both calculator and computer based)** 1. Graphical representation of data. 2. Problems based on measures of central tendency. 3. Problems based on measures of dispersion. 4. Problems based on combined mean and variance and coefficient of variation. 5. Problems based on moments, skewness and kurtosis 6. Fitting of polynomials, exponential curves. 7. Karl Pearson correlation coefficient. 8. Partial and multiple correlations. 9. Spearman rank correlation with and without ties. 10. Correlation coefficient for a bivariate frequency distribution. 11. Fitting of lines of regression, and angle between lines 12. Testing the independence of attributes.	05	–	–	05

Where, ***L: Lectures*** ***T: Tutorials*** ***P: Practicals***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] S. P. Gupta, *Statistical Methods*, All-India ed., New Delhi, India: Sultan Chand & Sons, 2021.

[2] S. C. Gupta and V. K. Kapoor, *Fundamentals of Mathematical Statistics*, 12th ed., New Delhi, India: Sultan Chand & Sons, 2020.

[3] M. R. Spiegel and L. J. Stephens, *Schaum's Outline of Statistics*, 6th ed., New York, NY, USA: McGraw-Hill Education, 2018.

[4] D. Freedman, R. Pisani, and R. Purves, *Statistics*, 4th ed., New York, NY, USA: W. W. Norton & Company, 2007.

Title of the Course	**:**	**Fundamentals of Computer Science**
Course Code	**:**	**SEC 1**
Nature of the Course	**:**	**Skill Enhancement Course**
Total Credits	**:**	**3**
Distribution of Marks	**:**	**End Sem: 50 TH + 10 PR, In Sem: 30 TH + 10 PR**

COURSE OBJECTIVES:

Knowledge

- Understand the basic concepts, evolution, and classification of computer systems.
- Gain knowledge of the functional components of a computer, including CPU, control unit, ALU, memory hierarchy, input/output devices, and storage systems.
- Understand digital representation of data, number systems, and methods of base conversion and binary arithmetic.
- Learn the concepts and applications of professional word processing using Microsoft Word.
- Acquire knowledge of spreadsheet fundamentals, formulas, functions, data management, and visualization using Microsoft Excel.
- Understand the basic concepts of relational database systems and database objects using Microsoft Access.
- Gain knowledge of primary keys, relationships, queries, forms, and reports in database management.

Skills

- Identify the components of a computer system and explain their functions.
- Perform number system conversions and binary arithmetic operations.
- Create, edit, format, save, and manage professional documents using Microsoft Word.
- Use tables, charts, SmartArt, page layout features, and mail merge for document preparation.
- Organize, analyze, and present data using formulas, functions, logical expressions, and data validation in Microsoft Excel.
- Generate charts, apply sorting and filtering, and manage spreadsheet data effectively.
- Design and manage simple relational databases using Microsoft Access.
- Create tables, define relationships, perform queries, and generate forms and structured reports.

Attitude

- Develop a systematic and organized approach to handling digital information and office tasks.
- Appreciate the importance of accuracy, efficiency, and consistency in computer-based work.
- Build confidence in using computer applications for academic, professional, and administrative purposes.
- Cultivate problem-solving ability in organizing, processing, and managing data.
- Show responsibility in the secure handling, storage, and management of digital documents and data.
- Develop readiness to apply computer tools in real-world situations and advanced learning in data science, statistics, and information systems.

COURSE OUTCOMES:

After completion of the course learners will be able to

CO1: Understand the basic concepts, structure, and functioning of a computer system, including input/output devices and number systems.
- **ILO1:** Identify the basic components of a computer system and explain their functions.
- **ILO2:** Describe the role of input, output, storage, and processing devices.
- **ILO3:** Explain the basic operation and functioning of a computer system.
- **ILO4:** Represent and convert numbers among binary, decimal, octal, and hexadecimal number systems.

CO2: Demonstrate the ability to operate a Windows-based system and manage files and folders effectively.
- **ILO1:** Operate a Windows-based environment and use common desktop features.
- **ILO2:** Create, rename, copy, move, and delete files and folders.
- **ILO3:** Organize files and folders using appropriate directory structures.
- **ILO4:** Apply basic system settings and file management tools for efficient use of the operating system.

CO3: Create and format professional documents using MS Word, including text formatting, tables, images, and mail merge.
- **ILO1:** Create and edit text documents using MS Word.
- **ILO2:** Apply formatting features such as fonts, paragraphs, styles, and page layout.
- **ILO3:** Insert and format tables, images, and other document elements.

ILO4: Use mail merge to prepare personalized documents.

CO4: Utilize spreadsheet software to organize, analyze, and present data using formulas, functions, and charts.

ILO1: Enter, organize, and edit data in spreadsheet worksheets.

ILO2: Apply formulas and basic functions for data analysis.

ILO3: Format worksheets for clear presentation of data.

ILO4: Create charts and graphs to summarize and present data effectively.

CO5: Develop engaging presentations using MS PowerPoint and demonstrate basic internet usage skills, including email communication.

ILO1: Create and organize presentation slides using MS PowerPoint.

ILO2: Apply themes, layouts, animations, and multimedia elements to enhance presentations.

ILO3: Use basic internet tools for browsing, searching, and accessing online information.

ILO4: Compose, send, receive, and manage emails with attachments.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	**Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge	CO1					
Conceptual Knowledge		CO1				
Procedural Knowledge			CO2, CO3, CO4	CO4		
Metacognitive Knowledge					CO5	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO / PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**
CO1	S	M	L	–	–	–	–	L
CO2	M	M	M	S	–	–	–	–
CO3	M	M	M	S	M	–	–	–
CO4	M	S	S	S	–	–	–	–
CO5	L	M	M	S	S	–	M	M

UNITS	CONTENTS	L	T	P	Total Hours
1 (12 Marks)	**Introduction to Computer Systems:** Definition and evolution of computers; Classification (Digital, Analog, Hybrid); Core components: CPU, Control Unit, ALU. Hardware & Storage: Input/Output peripherals; Memory hierarchy (Primary, Secondary, Cache). Digital Representation: Introduction to Number Systems: Binary, Octal, Decimal, and Hexadecimal; Base conversions and binary arithmetic.	07	02	–	09
2 (16 Marks)	**Professional Word Processing (MS Word):** Document orchestration: creating, saving, and protecting files. Text & Paragraph Formatting: Styles, indentation, and alignment; Advanced editing: Find & Replace, Hyperlinks. Document Elements: Working with Tables, SmartArt, and Charts; Page Layout: Margins, Orientation, and Section breaks. Automation: Mail Merge fundamentals and printing workflows.	06	–	06	11
3 (16 Marks)	**Spreadsheet Analytics (MS Excel):** Grid architecture: Rows, Columns, and Cell addressing. Computation: Basic functions (SUM, AVERAGE, IF); Logic nesting. Data Integrity: Cell referencing (Relative, Absolute, and Mixed). Visualization & Management: Generating dynamic Charts and Graphs; Sorting, Filtering, and Data Validation techniques.	4 06	–	06	11
4 (16 Marks)	**Database Management Foundations (MS Access):** Introduction to Relational Databases; Database objects: Tables, Queries, Forms, and Reports. Data Architecture: Field types and properties; Primary Keys and Relationships. Querying: Basic SQL concepts within Access; Data entry through Forms and generating structured Reports.	06	–	06	11

Where, ***L: Lectures*** ***T: Tutorials*** ***P: Practicals***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] Rajaraman, V., "Fundamentals of Computers," PHI Learning Pvt. Ltd., 6th ed., 2014.
[2] Sinha, P. K., "Computer Fundamentals," BPB Publications, 6th ed., 2012.
[3] Srivastava, S. S., "MS-Office," Laxmi Publications, 2015.
[4] Thareja, R., "Fundamentals of Computers," Oxford University Press, 2014.
[5] Martin, E. G., "Discovering Microsoft Office 2021," Wiley, 2022.

DATA SCIENCE PROGRAMME (NEP)
DETAILED SYLLABUS OF 2nd SEMESTER

Title of the Course	**:**	**Introduction to Python and Data Visualization**
Course Code	**:**	**C2**
Nature of the Course	**:**	**Core**
Total Credits	**:**	**4**
Distribution of Marks	**:**	**End Sem: 60, In Sem: 40**

COURSE OBJECTIVES:

Knowledge:

- Understand the fundamental syntax of Python, including variables, data types, expressions, and operators for computational thinking.
- Gain knowledge of control flow mechanisms (if-else, loops) and modular code structures such as functions and return types.
- Comprehend the principles of Object-Oriented Programming (OOP), including classes, objects, inheritance, polymorphism, and scope.
- Acquire a theoretical understanding of numerical computing using NumPy and data structures like Series and Data Frames in Pandas.
- Understand the architecture of relational databases (SQLite) and the lifecycle of data retrieval through REST APIs and HTTP methods.

Skills:

- Develop modular and reusable Python code by applying proper function definitions and exception handling techniques.
- Perform efficient data manipulation and file handling (CSV, Excel, JSON) using specialized scientific computing libraries.
- Execute Exploratory Data Analysis (EDA) workflows, including data cleaning, handling missing values, and generating summary statistics.
- Create informative data visualizations (histograms, boxplots, heatmaps) to interpret and communicate complex data patterns.
- Design and implement basic web applications using the Flask framework to display processed data through web interfaces.
- Apply SQL queries for CRUD operations and integrate database management within Python scripts.
- Analyze computational efficiency using time and space complexity notations and perform debugging for error control.

Attitude:

- Develop a logical and systematic approach to solving real-world data problems through structured programming.
- Appreciate the necessity of accuracy, precision, and error propagation analysis in numerical and scientific computations.
- Cultivate critical thinking and perseverance when debugging complex logic or handling large-scale datasets.
- Recognize the importance of code efficiency and scalability in the design of data-driven algorithms.
- Build a professional foundation for advanced studies in computational data science and software engineering.

COURSE OUTCOMES:

After completion of the course learners will be able to

CO1: Students will be able to understand and apply Python programming foundations for modular code development.
- **ILO1:** Define and use variables, expressions, and operators in Python.
- **ILO2:** Implement control flow using if-else statements and loops.
- **ILO3:** Design modular code using functions with various argument and return types.
- **ILO4:** Perform simple I/O operations and understand the Python interpreter environment.

CO2: Students will be able to utilize numerical and scientific computing libraries for data manipulation.
- **ILO1:** Create and manipulate multi-dimensional arrays using NumPy.
- **ILO2:** Apply vectorization techniques to optimize computational performance.
- **ILO3:** Use Pandas for handling Series and Data Frames.

ILO4: Manage basic file handling for CSV, Excel, and JSON formats.

CO3: Students will be able to implement Object-Oriented Programming (OOP) principles.
- **ILO1:** Define classes and objects to represent state and behavior.
- **ILO2:** Apply principles of inheritance and polymorphism in software design.
- **ILO3:** Implement exception handling and debugging strategies.
- **ILO4:** Distinguish between shallow and deep copying and manage namespaces/scope.

CO4: Students will be able to perform Exploratory Data Analysis (EDA) and visualization.
- **ILO1:** Inspect data using Pandas functions like head(), info(), and describe().
- **ILO2:** Compute and interpret summary statistics including mean, median, and skewness.
- **ILO3:** Create data visualizations such as boxplots, heatmaps, and scatter plots using Matplotlib/Seaborn.
- **ILO4:** Identify and handle missing values, duplicates, and outliers in datasets.

CO5: Students will be able to integrate Python applications with databases and APIs.
- **ILO1:** Connect Python to SQLite databases and perform CRUD operations.
- **ILO2:** Manage basic transactions and manipulate database tables.
- **ILO3:** Fetch and parse JSON data from REST APIs using HTTP methods (GET, POST).
- **ILO4:** Process retrieved data using Pandas for further analysis.

CO6: Students will be able to design basic web interfaces and analyze computational efficiency.
- **ILO1:** Understand the basics of the Flask framework and routing.
- **ILO2:** Create simple web applications that display data from SQLite.
- **ILO3:** Define time and space complexity using asymptotic notations (Big O).
- **ILO4:** Identify types of errors and analyze error propagation in algorithms.

CO7: Students will be able to apply mathematical rigor to computational problem-solving.
- **ILO1:** Use statistical measures to evaluate feature relationships and multicollinearity.
- **ILO2:** Interpret the results of an EDA workflow in a real-world context.
- **ILO3:** Build a foundation for advanced computational fields through systematic algorithm design.

Table: Learning Outcome Representation (CO)

Cognitive Knowledge Dimensions	Cognitive Process Dimension					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge	CO1					
Conceptual Knowledge		CO1, CO3				
Procedural Knowledge			CO1, CO3	CO2, CO5		
Metacognitive Knowledge					CO4, CO7	CO6

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO / PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	S	M	L	L	S	L	L	S
CO2	S	S	M	M	S	L	L	S
CO3	S	S	M	M	S	L	L	S
CO4	S	S	M	S	S	L	L	S
CO5	S	S	M	M	S	L	S	S
CO6	S	S	S	M	S	L	L	S
CO7	S	S	M	M	S	L	L	S

(*S= *Strong,* M= *Medium,* L= *Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (10 Marks)	**Python Foundations:** Introduction to the Python interpreter; Computational thinking and Flowcharts; Variables, Expressions, and Operators; Different data types in python; Control Flow (if-else, loops); Functions and modular code (definition, arguments, and return types); Simple I/O.	06		06	12
2 (10 Marks)	**Numerical & Scientific Computing:** Introduction to NumPy; Multi-dimensional arrays; Vectorization; Pandas basics: Series and DataFrames; Basic File Handling (Read/Write CSV/Excel/JSON etc.).	03		09	12
3 (12 Marks)	**Object-Oriented Programming (OOP):** Classes and Objects; State and Behavior; Principles of Inheritance and Polymorphism; Exception handling and Debugging; Namespaces and Scope; Shallow vs. Deep copying.	03		09	12
4 (14 Marks)	**Introduction to EDA** Data inspection using Pandas (head(), info(), describe()) **Summary statistics:** mean, median, standard deviation, skewness **Data visualization:** line, bar, histogram, boxplot, scatter, heatmap (Matplotlib/Seaborn) **Data cleaning:** handling missing values, duplicates, and outliers Feature relationships: correlation and basic multicollinearity Basic EDA workflow and interpretation of results	03		09	12
5 (14 Marks)	**Python Integration with Databases and APIs:** SQLite Integration: Introduction to SQLite; connecting Python using sqlite3; creation and manipulation of tables; execution of SQL queries (CRUD operations); basic transaction handling. **APIs and Data Retrieval:** Introduction to APIs and REST; HTTP methods (GET, POST); fetching data using requests; parsing JSON data; basic data processing using Pandas. **Flask (Introduction):** Basics of Flask framework; creating simple web applications; routing; integration with SQLite; displaying data through web interfaces.	03		09	12

where L: Lectures T: Tutorials P: Practicals

MODES OF IN-SEMESTER ASSESSMENT: (40 Marks)

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] McKinney, W. (2013). *Python for data analysis*. "O'Reilly Media, Inc."
[2] VanderPlas, J. (2022). *Python Data Science Handbook*. "O'Reilly Media, Inc."
[3] Tufte, E. R. (2001). *The visual display of quantitative information*.
[4] Murray, S. (2013). *Interactive Data visualization for the web*. "O'Reilly Media, Inc."

Title of the Course	**:**	**Introduction to Python and Data Visualization**
Course Code	**:**	**Minor 2**
Nature of the Course	**:**	**Minor**
Total Credits	**:**	**4**
Distribution of Marks	**:**	**End Sem: 60, In Sem: 40**

COURSE OBJECTIVES:

Knowledge:

- Understand the fundamental syntax of Python, including variables, data types, expressions, and operators for computational thinking.
- Gain knowledge of control flow mechanisms (if-else, loops) and modular code structures such as functions and return types.
- Comprehend the principles of Object-Oriented Programming (OOP), including classes, objects, inheritance, polymorphism, and scope.
- Acquire a theoretical understanding of numerical computing using NumPy and data structures like Series and DataFrames in Pandas.
- Understand the architecture of relational databases (SQLite) and the lifecycle of data retrieval through REST APIs and HTTP methods.

Skills:

- Develop modular and reusable Python code by applying proper function definitions and exception handling techniques.
- Perform efficient data manipulation and file handling (CSV, Excel, JSON) using specialized scientific computing libraries.
- Execute Exploratory Data Analysis (EDA) workflows, including data cleaning, handling missing values, and generating summary statistics.
- Create informative data visualizations (histograms, boxplots, heatmaps) to interpret and communicate complex data patterns.
- Design and implement basic web applications using the Flask framework to display processed data through web interfaces.
- Apply SQL queries for CRUD operations and integrate database management within Python scripts.
- Analyze computational efficiency using time and space complexity notations and perform debugging for error control.

Attitude:

- Develop a logical and systematic approach to solving real-world data problems through structured programming.
- Appreciate the necessity of accuracy, precision, and error propagation analysis in numerical and scientific computations.
- Cultivate critical thinking and perseverance when debugging complex logic or handling large-scale datasets.
- Recognize the importance of code efficiency and scalability in the design of data-driven algorithms.
- Build a professional foundation for advanced studies in computational data science and software engineering.

COURSE OUTCOMES:

After completion of the course learners will be able to

CO1: Students will be able to understand and apply Python programming foundations for modular code development.
- ILO1: Define and use variables, expressions, and operators in Python.
- ILO2: Implement control flow using if-else statements and loops.
- ILO3: Design modular code using functions with various argument and return types.
- ILO4: Perform simple I/O operations and understand the Python interpreter environment.

CO2: Students will be able to utilize numerical and scientific computing libraries for data manipulation.
- ILO1: Create and manipulate multi-dimensional arrays using NumPy.
- ILO2: Apply vectorization techniques to optimize computational performance.
- ILO3: Use Pandas for handling Series and DataFrames.
- ILO4: Manage basic file handling for CSV, Excel, and JSON formats.

CO3: Students will be able to implement Object-Oriented Programming (OOP) principles.
- ILO1: Define classes and objects to represent state and behavior.

ILO2: Apply principles of inheritance and polymorphism in software design.
ILO3: Implement exception handling and debugging strategies.
ILO4: Distinguish between shallow and deep copying and manage namespaces/scope.

CO4: Students will be able to perform Exploratory Data Analysis (EDA) and visualization.
ILO1: Inspect data using Pandas functions like head(), info(), and describe().
ILO2: Compute and interpret summary statistics including mean, median, and skewness.
ILO3: Create data visualizations such as boxplots, heatmaps, and scatter plots using Matplotlib/Seaborn.
ILO4: Identify and handle missing values, duplicates, and outliers in datasets.

CO5: Students will be able to integrate Python applications with databases and APIs.
ILO1: Connect Python to SQLite databases and perform CRUD operations.
ILO2: Manage basic transactions and manipulate database tables.
ILO3: Fetch and parse JSON data from REST APIs using HTTP methods (GET, POST).
ILO4: Process retrieved data using Pandas for further analysis.

CO6: Students will be able to design basic web interfaces and analyze computational efficiency.
ILO1: Understand the basics of the Flask framework and routing.
ILO2: Create simple web applications that display data from SQLite.
ILO3: Define time and space complexity using asymptotic notations (Big O).
ILO4: Identify types of errors and analyze error propagation in algorithms.

CO7: Students will be able to apply mathematical rigor to computational problem-solving.
ILO1: Use statistical measures to evaluate feature relationships and multicollinearity.
ILO2: Interpret the results of an EDA workflow in a real-world context.
ILO3: Build a foundation for advanced computational fields through systematic algorithm design.

Table: Learning Outcome Representation (CO)

Cognitive Knowledge Dimensions	**Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge	CO1					
Conceptual Knowledge		CO1, CO3				
Procedural Knowledge			CO1, CO3	CO2, CO5		
Metacognitive Knowledge					CO4, CO7	CO6

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO / PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**
CO1	S	M	L	L	S	L	L	S
CO2	S	S	M	M	S	L	L	S
CO3	S	S	M	M	S	L	L	S
CO4	S	S	M	S	S	L	L	S
CO5	S	S	M	M	S	L	S	S
CO6	S	S	S	M	S	L	L	S
CO7	S	S	M	M	S	L	L	S

(*S= *Strong,* M= *Medium,* L= *Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (10 Marks)	**Python Foundations:** Introduction to the Python interpreter; Computational thinking and Flowcharts; Variables, Expressions, and Operators; Different data types in python; Control Flow (if-else, loops); Functions and modular code (definition, arguments, and return types); Simple I/O.	06		06	12
2 (10 Marks)	**Numerical & Scientific Computing:** Introduction to NumPy; Multi-dimensional arrays; Vectorization; Pandas basics: Series and DataFrames; Basic File Handling (Read/Write CSV/Excel/JSON etc.).	03		09	12
3 (12 Marks)	**Object-Oriented Programming (OOP):** Classes and Objects; State and Behavior; Principles of Inheritance and Polymorphism; Exception handling and Debugging; Namespaces and Scope; Shallow vs. Deep copying.	03		09	12
4 (14 Marks)	**Introduction to EDA** Data inspection using Pandas (head(), info(), describe()) **Summary statistics:** mean, median, standard deviation, skewness **Data visualization:** line, bar, histogram, boxplot, scatter, heatmap (Matplotlib/Seaborn) **Data cleaning:** handling missing values, duplicates, and outliers Feature relationships: correlation and basic multicollinearity Basic EDA workflow and interpretation of results	03		09	12
5 (14 Marks)	**Python Integration with Databases and APIs:** SQLite Integration: Introduction to SQLite; connecting Python using sqlite3; creation and manipulation of tables; execution of SQL queries (CRUD operations); basic transaction handling. **APIs and Data Retrieval:** Introduction to APIs and REST; HTTP methods (GET, POST); fetching data using requests; parsing JSON data; basic data processing using Pandas. **Flask (Introduction):** Basics of Flask framework; creating simple web applications; routing; integration with SQLite; displaying data through web interfaces.	03		09	12

where L: Lectures T: Tutorials P: Practicals

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] McKinney, W. (2013). *Python for data analysis*. "O'Reilly Media, Inc."
[2] VanderPlas, J. (2022). *Python Data Science Handbook*. "O'Reilly Media, Inc."
[3] Tufte, E. R. (2001). *The visual display of quantitative information*.
Murray, S. (2013). *Interactive Data visualization for the web*. "O'Reilly Media, Inc."

Title of the Course	**:**	**Basics of Probability Distribution and Inference**
Course Code	**:**	**GEC 2**
Nature of the Course	**:**	**Generic Elective**
Total Credits	**:**	**3**
Distribution of Marks	**:**	**End Sem :60 TH, In Sem: 40 TH**

COURSE OBJECTIVES:

Knowledge:

- Understand the concepts of random samples, parameters, statistics, sampling distributions, and the Central Limit Theorem in statistical inference.
- Gain knowledge of standard error and its importance in estimation and hypothesis testing.
- Comprehend the principles of statistical hypothesis testing, including null and alternative hypotheses, level of significance, p-values, and types of errors.
- Understand the theoretical foundations and properties of exact sampling distributions such as χ^2, Student's t, and Snedecor's F distributions.
- Acquire knowledge of the relationships among χ^2, t, and F distributions and their applications in inferential statistics.

Skills

- Apply sampling distribution concepts and standard error calculations to practical statistical problems.
- Perform hypothesis testing using critical regions, acceptance regions, and one-tailed or two-tailed tests.
- Derive and compute probability density functions and statistical measures related to χ^2, t, and F distributions.
- Conduct statistical tests such as goodness-of-fit tests, tests of independence, variance tests, and tests concerning means.
- Analyze small and large sample data and select appropriate inferential procedures for statistical decision-making.

Attitude

- Develop a logical and analytical approach toward solving statistical inference problems.
- Cultivate accuracy and precision in statistical computations and interpretation of results.
- Appreciate the importance of evidence-based decision-making through hypothesis testing.
- Build confidence in applying theoretical distributions to real-world data analysis problems.
- Encourage critical thinking and systematic reasoning in statistical investigations and research.

COURSE OUTCOMES:

After completion of the course learners will be able to

CO1: Demonstrate understanding of sampling distributions, Central Limit Theorem, and standard error for statistical inference.

- **ILO1:** Define sampling distributions and explain their role in statistical inference.
- **ILO2:** Describe and apply the Central Limit Theorem for large sample analysis.
- **ILO3:** Compute and interpret standard errors for sample statistics.
- **ILO4:** Distinguish between population parameters and sample statistics.
- **ILO5:** Apply sampling distribution concepts to estimate population characteristics.

CO2: Formulate null and alternative hypotheses and apply hypothesis testing using p-values, critical regions, and error analysis.

- **ILO1:** Define null and alternative hypotheses for statistical problems.
- **ILO2:** Identify Type I and Type II errors in hypothesis testing.
- **ILO3:** Determine critical regions and decision rules for tests of significance.
- **ILO4:** Calculate and interpret p-values in statistical inference.
- **ILO5:** Draw conclusions from hypothesis tests in practical situations.

CO3: Derive and explain the properties of χ^2, t, and F distributions and describe their relationships.

- **ILO1:** Define χ^2, t, and F distributions and state their assumptions.

ILO2: Derive important properties of χ^2, t, and F distributions.

ILO3: Explain the relationships among normal, χ^2, t, and F distributions.

ILO4: Interpret the shapes and applications of χ^2, t, and F distributions.

ILO5: Identify appropriate situations for using χ^2, t, and F distributions.

CO4: Apply χ^2, t, and F tests to practical problems such as goodness-of-fit, independence, variance testing, and comparison of means.

ILO1: Apply χ^2 tests for goodness-of-fit and independence of attributes.

ILO2: Perform one-sample and two-sample t-tests for means.

ILO3: Apply paired t-tests for related samples.

ILO4: Conduct F-tests for comparison of population variances.

ILO5: Interpret the results of χ^2, t, and F tests in practical contexts.

CO5: Analyze statistical problems and select appropriate inferential techniques for small and large sample data.

ILO1: Distinguish between small-sample and large-sample statistical problems.

ILO2: Select suitable inferential methods for different types of data.

ILO3: Analyze statistical data using appropriate test procedures.

ILO4: Interpret statistical results and justify conclusions logically.

ILO5: Evaluate the applicability and limitations of inferential techniques.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	**Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge	CO1	CO1				
Conceptual Knowledge		CO1, CO2, CO3		CO3		
Procedural Knowledge			CO2, CO4	CO4		
Metacognitive Knowledge				CO5	CO5	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO / PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**
CO1	S	M	L	M	–	–	–	L
CO2	M	S	M	S	–	–	–	–
CO3	M	M	M	S	–	–	–	–
CO4	M	S	S	S	–	–	–	–
CO5	L	M	M	S	S	–	M	M

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (6 Marks)	**Sampling Distribution:** Idea of random sample, parameter and statistic, Concept of sampling distribution and central limit theorem. Standard error and its role in statistical inference.	4	02	-	06
2 (15 Marks)	**Testing of Hypotheses:** Null and alternative hypotheses, Level of significance and p-value, Type I and Type II errors, Power of a test, Critical region and acceptance region, One-tailed and two-tailed tests	10	05	-	15
3 (15 Marks)	**Exact sampling distribution (χ^2 distribution):** Definition and derivation of p.d.f. of χ^2 with n degrees of freedom (d.f.) using m.g.f., nature of p.d.f. curve for different degrees of freedom, mean, variance, m.g.f., cumulant generating function, mode, and limiting form of χ^2 distribution. Properties and applications (e.g., relation to gamma distribution, variance test, goodness-of-fit, test of independence)	13	03	-	16
4 (12 Marks)	***t*-distribution**: Student's and Fisher's *t*-distribution, Derivation of its p.d.f., nature of probability curve with different degrees of freedom, mean, variance, moments and limiting form of *t* distribution. Properties and applications in small sample inference (one-sample and two-sample problems)	10	03	-	13
5 (12 Marks)	**Snedecore's F-distribution**: Derivation of p.d.f., nature of p.d.f. curve with different degrees of freedom, mean, variance and mode. Relationship between *t*, *F*, and χ^2 distributions. Test of significance and confidence Intervals based F distributions.	08	02	-	10

where ***L: Lectures*** ***T: Tutorials*** ***P: Practicals***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] Casella, G., & Berger, R. (2024). *Statistical inference*. CRC Press.

[2] Devore, J. L. (2008). *Probability and statistics for Engineering and the Sciences*.

[3] Gupta, S. (2021). *Statistical methods: All India Edition*. Sultan Chand & Sons.

[4] Gupta, S., & Kapoor, V. (2020). *Fundamentals of Mathematical Statistics*. Sultan Chand & Sons.

Title of the Course	**:**	**Data Science Using MS-Excel and SQL**
Course Code	**:**	**SEC2**
Nature of the Course	**:**	**Skill Enhancement Course**
Total Credits	**:**	**3**
Distribution of Marks	**:**	**End Sem: 20 TH + 40 PR, In Sem: 10 TH + 30 PR**

COURSE OBJECTIVES:

Knowledge:

- Understand Spreadsheet Architecture, including workbook management, cell referencing, and protection/validation mechanisms in MS-Excel.
- Gain knowledge of functional logic using logical operators (IF, AND, OR) and advanced lookup functions (VLOOKUP, XLOOKUP).
- Comprehend Relational Database Foundations, including RDBMS concepts, entities, attributes, and schema design constraints.
- Learn the principles of database normalization (1NF, 2NF, 3NF) and the structure of Data Definition Language (DDL) and Data Manipulation Language (DML).
- Understand the Tableau interface and the core concepts of data storytelling and dashboarding.

Skills:

- Execute systematic data entry, formatting, and validation to ensure data integrity within spreadsheets.
- Perform statistical analysis using the Analysis ToolPak and create dynamic visualizations such as pivot tables and interactive dashboards.
- Construct complex SQL queries involving selection logic, aggregation, grouping, and relational JOIN operations.
- Implement ETL processes to import and export data between SQL databases and Excel for hybrid analysis.
- Develop interactive Tableau dashboards utilizing filters, parameters, and basic interactivity for professional data presentation.

Attitude:

- Develop an organized and meticulous mindset for managing database schemas and spreadsheet environments.
- Appreciate the role of data normalization and integrity in maintaining reliable information systems.
- Cultivate a creative yet analytical approach to data visualization and storytelling.
- Recognize the value of integrating multiple tools (SQL, Excel, Tableau) to solve end-to-end data ingestion and analysis problems.

COURSE OUTCOMES:

After completion of the course learners will be able to

CO1: Students will be able to manage and validate data within spreadsheet environments effectively.
- **ILO1:** Navigate the MS-Excel environment and manage workbooks/worksheets.
- **ILO2:** Apply cell referencing and custom data types for systematic data entry.
- **ILO3:** Implement protection and validation rules to secure worksheet data.

CO2: Students will be able to perform advanced data manipulation and statistical analysis in Excel.
- **ILO1:** Apply multi-level sorting, advanced filtering, and conditional formatting.
- **ILO2:** Use functional logic and nested IF statements to automate data processing.
- **ILO3:** Execute descriptive statistics and generate interactive dashboards using Pivot Tables.

CO3: Students will be able to design and modify relational database schemas.
- **ILO1:** Identify entities, attributes, and records within a relational model.
- **ILO2:** Design schemas using Primary Keys, Foreign Keys, and Unique constraints.
- **ILO3:** Apply normalization techniques (1NF to 3NF) to optimize database structure.
- **ILO4:** Use DDL commands (CREATE, DROP, ALTER) to manage database objects.

CO4: Students will be able to extract and manipulate data using SQL querying techniques.
- **ILO1:** Use SELECT, WHERE, and DISTINCT clauses for precise data retrieval.
- **ILO2:** Perform data aggregation and grouping using GROUP BY and HAVING clauses.
- **ILO3:** Execute various JOIN operations to combine data from multiple tables.
- **ILO4:** Modify records using INSERT, UPDATE, and DELETE commands.

CO5: Students will be able to visualize data and build interactive stories using Tableau.
- **ILO1:** Connect external data sources to the Tableau interface.
- **ILO2:** Create basic and advanced visualizations including maps and bar charts.

ILO3: Build interactive dashboards using filters and parameters for user engagement.

CO6: Students will be able to execute integrated ETL processes and hybrid data analysis.

ILO1: Import and export data between SQL databases and spreadsheets.

ILO2: Leverage SQL for data extraction and Excel for complex visualization.

ILO3: Evaluate the efficiency of data flow in an integrated analysis pipeline.

CO7: Students will be able to complete end-to-end data projects from ingestion to presentation.

ILO1: Integrate raw data ingestion, cleaning, and relational querying in a single project.

ILO2: Apply data storytelling techniques to present findings effectively.

ILO3: Demonstrate ethical data handling and accuracy in final project reporting.

Table: Learning Outcome Representation (CO)

Cognitive Knowledge Dimensions	**Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge	CO1					
Conceptual Knowledge		CO1, CO3				
Procedural Knowledge			CO1, CO2, CO3, CO4	CO2, CO4		C07
Metacognitive Knowledge					CO5, CO6	CO5

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO / PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**
CO1	S	M	L	M	S	L	L	L
CO2	M	S	M	S	S	L	L	M
CO3	M	M	M	S	S	L	M	M
CO4	M	S	S	S	S	L	M	M
CO5	L	M	M	S	S	S	M	S
CO6	S	S	S	S	S	S	S	S
CO7	S	S	S	S	S	S	S	S

(*S= *Strong,* M= *Medium,* L= *Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (6 Marks)	**Spreadsheet Architecture:** Introduction to the MS-Excel environment; Workbook and Worksheet management; Cell referencing and data types. Data Entry & Management: Systematic data entry; Cell formatting; Custom data types; Protection and validation of worksheets.	03	01	02	06
2 (12 Marks)	**Advanced Analytics in Excel:** Data Manipulation: Multi-level sorting, advanced filtering, and conditional formatting. Functional Logic: Text manipulation; Logical operators (IF, AND, OR, NESTED IF); Lookup and Reference (VLOOKUP, HLOOKUP, XLOOKUP). **Statistical Analysis:** Descriptive statistics using Analysis ToolPak; Pivot Tables and Slicers; Visualization: Dynamic Charts and Interactive Dashboards. **Introduction to Tableau:** Overview of Tableau interface, Connecting Excel data to Tableau, Creating basic visualizations (bar, line, maps); Concept of dashboards and storytelling	07	01	07	15
3 (14 Marks)	**Relational Database Foundations:** Concept of RDBMS; Relational model; Entities, Attributes, and Records. Schema Design: Primary Key, Foreign Key, and Unique constraints; Basics of Database Normalization (1NF, 2NF, 3NF). **Data Definition Language (DDL):** Introduction to SQL; Creating and modifying schemas using CREATE, DROP, and ALTER.	06	01	05	12
4 (14 Marks)	**Data Manipulation Language (DML) & Querying:** Selection Logic: SELECT, WHERE, and DISTINCT clauses; Logical and Comparison operators. Aggregation & Grouping: ORDER BY, GROUP BY, and HAVING; Aggregate functions (SUM, AVG, COUNT, MIN, MAX). Relational Operations: JOIN operations (INNER, LEFT, RIGHT, FULL); Subqueries; Modifying records (INSERT, UPDATE, DELETE); Exporting SQL result sets for downstream analysis.	07	01	06	14
5 (14 Marks)	**Integrated Data Analysis:** ETL Processes: Importing/Exporting data between SQL databases and Excel; Hybrid Analysis: Leveraging SQL for data extraction and Excel for complex visualization. **Data Visualization using Tableau:** Creating interactive dashboards Filters, parameters, and basic interactivity Publishing and sharing dashboards (intro) **Case Study:** End-to-end project involving raw data ingestion, cleaning, relational querying, and presentation insights.	04	01	08	13

where L: Lectures T: Tutorials P: Practicals

MODES OF IN-SEMESTER ASSESSMENT: (40 Marks)

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] Alexander, M., Kusleika, D., & Walkenbach, J. (2022). *Excel 365 Bible*. Wiley.

[2] Coronel, P. R. C. (2002). *Database Systems : Design, Implementation, and Management, 3rd Edition*.

[3] Nelson, S. L., & Nelson, E. C. (2014). *Excel Data analysis for Dummies*. John Wiley & Sons.

[4] Viescas, J. L., & Hernandez, M. J. (2014). *SQL queries for mere mortals: A Hands-on Guide to Data Manipulation in SQL*. Pearson Education.

DATA SCIENCE PROGRAMME (NEP)
DETAILED SYLLABUS OF 3rd SEMESTER

Title of the Course	**:**	**Mathematics for Data Science**
Course Code	**:**	**STSC3**
Nature of the Course	**:**	**Major**
Total Credits	**:**	**04**
Distribution of Marks	**:**	**End Sem: 60 TH, In Sem: 40 TH**

COURSE OBJECTIVES:
Knowledge:

- Understand the concept of functions, including one-to-one, onto, bijective functions, and identify domain, range, and co-domain.
- Gain knowledge of standard functions such as polynomial, exponential, logarithmic, and trigonometric functions.
- Comprehend the theory of real sequences and series, including convergence, divergence, monotonicity, and Cauchy's criteria.
- Understand infinite series and power series, including the concept of radius of convergence and approximations using Taylor and Maclaurin series.
- Acquire knowledge of matrix algebra, types of matrices, determinants, and their properties.
- Understand systems of linear equations, including rank, consistency, and solution methods such as Gaussian elimination and Gauss-Jordan methods.
- Learn about matrix inverses, including singular and non-singular matrices and the Moore-Penrose generalized inverse.
- Understand the structure of vector spaces, including basis, dimension, and linear dependence/independence.
- Gain knowledge of eigenvalues, eigenvectors, and the Cayley-Hamilton theorem.
- Understand orthogonality, inner product spaces, and Gram-Schmidt orthogonalization.
- Learn the theory of quadratic forms, classification, and geometric interpretation.
- Understand multivariate calculus concepts such as limits, continuity, differentiability, partial derivatives, gradient, and Hessian.
- Acquire knowledge of definite and improper integrals, including Beta and Gamma functions.
- Understand algorithm fundamentals, time and space complexity, asymptotic notations, and error analysis in computations.

Skills:

- Analyze and classify different types of functions and solve related problems.
- Test convergence/divergence of sequences and series using appropriate criteria.
- Expand functions using Taylor and Maclaurin series for approximation in computations.
- Perform matrix operations, compute determinants, and solve systems of linear equations using Gaussian and Gauss-Jordan methods.
- Apply generalized inverse techniques in solving least squares problems.
- Work with vector spaces, determine basis and dimension, and test linear dependence.
- Compute eigenvalues and eigenvectors and apply them in matrix decomposition problems.
- Apply Gram-Schmidt process to obtain orthonormal bases.
- Analyze and classify quadratic forms and perform transformations to principal axes.
- Solve problems in multivariate calculus, including partial derivatives, gradients, and optimization using Hessian.
- Evaluate definite and improper integrals, including Beta and Gamma functions.
- Design and analyze algorithms, evaluate their efficiency using asymptotic notations.
- Identify and analyze numerical errors and their propagation in computations.

Attitude:

- Develop a logical and analytical approach to solving mathematical and computational problems.
- Appreciate the importance of mathematical rigor and precision in scientific computations.
- Cultivate problem-solving confidence in handling abstract mathematical concepts.
- Show interest in applying mathematical techniques to real-world and computational problems.
- Develop a systematic and organized approach toward algorithm design and analysis.
- Recognize the importance of accuracy and error control in numerical computations.
- Demonstrate perseverance and critical thinking while working on complex mathematical problems.
- Build a foundation for advanced studies in mathematics, statistics, data science, and computational fields.

COURSE OUTCOMES:

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

CO1: Students will be able to understand, classify, and analyze different types of functions and their properties.
- **ILO1:** Define and distinguish between one-to-one, onto, and bijective functions.
- **ILO2:** Identify domain, range, and co-domain of functions.
- **ILO3:** Analyze and graph polynomial, exponential, logarithmic, and trigonometric functions.
- **ILO4:** Apply functional concepts to solve mathematical problems.

CO2: Students will be able to analyze real sequences and series and determine their convergence behavior.
- **ILO1:** Define real sequences and test for convergence and divergence.
- **ILO2:** Apply monotone sequence theorem and Cauchy's criteria.
- **ILO3:** Analyze infinite series and determine convergence.
- **ILO4:** Determine the radius of convergence of power series.
- **ILO5:** Use Taylor and Maclaurin series for function approximation in computations.

CO3: Students will be able to perform matrix operations and solve systems of linear equations efficiently.
- **ILO1:** Identify different types of matrices and their properties.
- **ILO2:** Perform matrix algebra and compute determinants.
- **ILO3:** Determine the rank of a matrix.
- **ILO4:** Analyze consistency and solvability of linear systems.
- **ILO5:** Solve systems using Gaussian elimination and Gauss-Jordan methods.
- **ILO6:** Distinguish between singular and non-singular matrices.
- **ILO7:** Apply Moore-Penrose inverse in solving least squares problems.

CO4: Students will be able to understand vector spaces and apply spectral techniques in linear algebra.
- **ILO1:** Define vector spaces, basis, and dimension.
- **ILO2:** Test for linear dependence and independence.
- **ILO3:** Compute eigenvalues and eigenvectors.
- **ILO4:** Apply the Cayley-Hamilton theorem.
- **ILO5:** Work with orthogonal and unitary matrices.
- **ILO6:** Apply inner product concepts and Gram-Schmidt orthogonalization.
- **ILO7:** Analyze and classify quadratic forms and interpret geometrically.
- **ILO8:** Understand and apply linear transformations.

CO5: Students will be able to apply concepts of multivariate calculus and evaluate integrals in various contexts.
- **ILO1:** Evaluate limits, continuity, and differentiability of functions.
- **ILO2:** Compute partial derivatives and apply the chain rule.
- **ILO3:** Determine and interpret gradient vectors and Hessian matrices.
- **ILO4:** Solve optimization problems using multivariate techniques.
- **ILO5:** Evaluate definite integrals and understand their properties.
- **ILO6:** Compute improper integrals, including Beta and Gamma functions.

CO6: Students will be able to design algorithms and analyze computational efficiency and numerical errors.
- **ILO1:** Define and explain algorithms and their characteristics.

ILO2: Analyze time and space complexity using asymptotic notations.
ILO3: Understand floating-point arithmetic and its limitations.
ILO4: Identify different types of errors in numerical computations.
ILO5: Analyze error propagation in algorithms.
ILO6: Apply concepts of efficiency and accuracy in computational problem-solving.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	Cognitive Process Dimension					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge				CO1, CO3	CO2, CO5	
Metacognitive Knowledge					CO4, CO6	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	S	M	L	L	L	L	L	L
CO2	S	S	M	M	L	L	L	L
CO3	S	S	M	M	L	L	L	L
CO4	S	S	M	S	L	L	L	L
CO5	S	S	M	M	L	L	L	L
CO6	S	S	S	M	M	L	L	L

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (12 Marks)	**Functions:** Definition and types of functions: one-to-one, onto, bijective, Domain, range, and co-domain, Polynomial, exponential, logarithmic, and trigonometric functions **Analysis of Sequences and Series:** Real Sequences: Definition, convergence, and divergence; Monotone sequences and Cauchy's criteria. Infinite Series; Power Series: Concept of radius of convergence; Introduction to Taylor and Maclaurin series for approximating functions in computation.	05	03	-	08
2 (12 Marks)	**Matrix Algebra:** Review of different matrix types and their properties; Algebra of matrices and determinants. **System of Linear Equations:** Rank of a matrix; Consistency and solvability of linear systems; Gaussian elimination and Gauss-Jordan methods. Inverses: Singular and non-singular matrices; Generalized inverse (Moore-Penrose) with applications in least squares.	06	03	-	09
3 (14 Marks)	**Vector Spaces and Spectral Theory:** Spaces: Definition, basis, dimension, and linear dependency. Decompositions: Characteristic roots and vectors (Eigenvalues and Eigenvectors); Cayley-Hamilton theorem. Orthogonality: Orthogonal and Unitary matrices; Inner product spaces; Gram-Schmidt orthogonalization. Quadratic forms: definition, classification (definiteness), and geometry of principal axes; Linear transformation.	07	03	-	10
4 (12 Marks)	**Multivariate Calculus and Integration:** Functions: Limits, continuity, and differentiability. Calculus of Several Variables: Partial derivatives; Gradient vector and Hessian matrix; Chain rule. Integration: Definite integrals and their properties. Improper Integrals: Beta and Gamma functions.	06	03		09
5 (10 Marks)	**Introduction to algorithm and error Analysis:** Definition and characteristics of algorithm, time and space complexity, asymptotic notations; Floating point arithmetic and propagation of errors in algorithms; Different types of error.	06	03	-	09
	Total	**30**	**15**	**-**	**45**

Where, ***L: Lectures*** ***T: Tutorials*** ***P: Practical***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] A Textbook of Algebra, Sudesh K Shah & Subhash Chand Garg

[2] A Textbook of Fundamental Calculus, Kumar Chaitanya , Kaur Bhavneet, Chawla Harinderjit Kaur

[3] Sridhar, S. (2015). *Design and analysis of algorithms*.

[4] Axler, S. (2014). *Linear algebra done right*. Springer.

[5] Haviv, M. (2023). *Linear algebra for data science*. World Scientific.

Title of the Course	**:**	**Introduction to Probability and Distributions**
Course Code	**:**	**STSC4**
Nature of the Course	**:**	**Major**
Total Credits	**:**	**04**
Distribution of Marks	**:**	**End Sem: 60 TH, In Sem: 40 TH**

COURSE OBJECTIVES:

Knowledge:

- Understand the fundamentals of set theory, including algebra of sets, types of sets, operations, Venn diagrams, and Cartesian products.
- Gain knowledge of relations and their properties, and basic concepts of algebraic structures and set functions.
- Understand combinatorics principles and their role in probability.
- Comprehend different approaches to probability theory: classical, statistical, and axiomatic.
- Learn addition and multiplication theorems of probability and their applications.
- Understand conditional probability, independence, Bayes' theorem, and the theorem of total probability.
- Gain knowledge of random variables (discrete and continuous), including PMF, PDF, and CDF.
- Understand joint, marginal, and conditional distributions in bivariate and multivariate cases.
- Learn the concept of mathematical expectation, variance, covariance, and conditional expectation.
- Understand generating functions: PGF, MGF, and characteristic functions.
- Acquire knowledge of important discrete distributions (Bernoulli, Binomial, Poisson, Geometric, Negative Binomial, Hypergeometric).
- Understand continuous distributions (Uniform, Exponential, Gamma, Beta, Normal) and their properties.
- Learn limit laws and types of convergence in probability theory.
- Understand key inequalities (Chebyshev, Markov, Cauchy-Schwarz, Jensen).
- Gain knowledge of limit theorems including LLN, CLT, and related results.
- Understand the theory of order statistics, including distributions and applications.

Skills:

- Perform set operations and represent them using Venn diagrams.
- Solve problems using counting techniques and combinatorics.
- Apply probability laws and theorems to compute probabilities in real-life situations.
- Use Bayes' theorem for decision-making and inference problems.
- Work with random variables, and compute PMF, PDF, and CDF.
- Derive and analyze joint, marginal, and conditional distributions.
- Compute and interpret expectation, variance, and covariance.
- Apply generating functions to find moments and probabilities.
- Identify and apply appropriate probability distributions to real-world data.
- Solve problems involving normal distribution and standardization.
- Analyze convergence of random variables using different modes of convergence.
- Apply inequalities to bound probabilities and expectations.
- Use limit theorems (LLN, CLT) for approximation and inference.
- Derive and analyze order statistics and their distributions.

Attitude:

- Develop a logical and analytical mindset for handling uncertainty and randomness.
- Appreciate the importance of probability theory in real-world decision-making.
- Cultivate critical thinking and problem-solving abilities in stochastic contexts.
- Show interest in applying statistical methods in diverse fields such as data science, economics, and

engineering.

- Develop precision and rigor in mathematical reasoning and interpretation.
- Build confidence in handling complex probabilistic models and data analysis.
- Recognize the importance of probabilistic thinking in research and innovation.
- Foster a habit of systematic analysis and interpretation of results.

COURSE OUTCOMES:

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

CO1: Students will be able to understand and apply concepts of set theory, relations, and combinatorics in probability.

ILO1: Define and perform operations on sets and represent them using Venn diagrams.

ILO2: Classify finite, infinite, and power sets.

ILO3: Understand ordered pairs and Cartesian products.

ILO4: Identify and analyze relations and their properties.

ILO5: Apply basic algebraic structures and set functions.

ILO6: Use fundamental counting principles in solving probability problems.

CO2: Students will be able to apply different approaches to probability and solve problems involving conditional probability.

ILO1: Explain classical, statistical, and axiomatic approaches to probability.

ILO2: Apply addition and multiplication theorems.

ILO3: Compute conditional probabilities.

ILO4: Analyze independent events.

ILO5: Apply Theorem of Total Probability.

ILO6: Use Bayes' Theorem in real-life decision-making problems.

CO3: Students will be able to analyze random variables and their probability distributions.

ILO1: Define discrete and continuous random variables.

ILO2: Compute and interpret PMF, PDF, and CDF.

ILO3: Perform transformations of random variables.

ILO4: Analyze joint, marginal, and conditional distributions.

ILO5: Determine independence of random variables.

ILO6: Extend concepts to multivariate distributions.

CO4: Students will be able to compute expectations and utilize generating functions in probability.

ILO1: Compute mathematical expectation and its properties.

ILO2: Calculate variance and covariance.

ILO3: Evaluate conditional expectation.

ILO4: Apply probability generating function (PGF).

ILO5: Use moment generating function (MGF) to find moments.

ILO6: Apply characteristic functions in probability analysis.

CO5: Students will be able to model real-world problems using standard probability distributions.

ILO1: Identify and apply discrete distributions (Bernoulli, Binomial, Poisson, etc.).

ILO2: Analyze properties and applications of discrete distributions.

ILO3: Apply continuous distributions (Uniform, Exponential, Gamma, Beta, Normal).

ILO4: Compute probabilities using normal distribution and standardization.

ILO5: Interpret real-life applications of probability distributions.

CO6: Students will be able to analyze convergence concepts and apply major probabilistic theorems.

ILO1: Define and distinguish between types of convergence (in probability, distribution, mean, almost sure).

ILO2: Analyze relationships among different modes of convergence.

ILO3: Apply Chebyshev's, Markov's, Cauchy-Schwarz, and Jensen's inequalities.

ILO4: Understand and apply Law of Large Numbers (LLN).

ILO5: Apply Central Limit Theorem (CLT) in approximation problems.

ILO6: Understand De-Moivre Laplace and Liapunov's Theorem.

CO7: Students will be able to analyze order statistics and their applications.

ILO1: Define and derive order statistics.

ILO2: Determine distribution of order statistics.

ILO3: Analyze joint distribution of order statistics.

ILO4: Compute and interpret range and related measures.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	**Cognitive Process Dimension**					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge				CO1, CO7	CO2, CO3, CO5	
Metacognitive Knowledge					CO4, CO6	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO/PO	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**
CO1	S	M	L	L	L	L	L	L
CO2	S	S	M	M	L	L	L	L
CO3	S	S	M	M	L	L	L	L
CO4	S	S	M	M	L	L	L	L
CO5	S	S	M	M	M	L	L	L
CO6	S	S	M	S	M	L	L	L
CO7	S	S	M	M	L	L	L	L

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (15 Marks)	**Set Theory and Probability Foundations:** Algebra of sets; Finite, infinite, and power sets; Operations on sets; Venn diagrams and their applications; Ordered pairs and Cartesian product of sets; Relations: Types and properties; Basic idea of field (algebra), set functions, and combinatorics. Fundamental principles of counting for probability. **Probability Theory:** Classical, statistical, and axiomatic approaches; Addition and multiplication theorems. Conditional Probability: Independent events; Theorem of total probability; Bayes' Theorem and its applications.	07	03	-	10
2 (20 Marks)	**Random Variables:** Univariate: Discrete and continuous random variables; cumulative distribution function, probability mass function, and probability density function; Univariate transformations. Bivariate: Joint, marginal, and conditional distributions; Independence of variables; Bivariate transformations; Multivariate extension. **Mathematical Expectation:** Theoretical properties; Variance and Covariance; Conditional expectation. **Generating Functions:** Probability generating function, Moment generating function, and Characteristic function.	11	05	-	16
3 (15 Marks)	**Discrete Probability Distributions:** Mathematical properties and real-world applications of: Bernoulli, Binomial, Poisson, Geometric, Negative Binomial, and Hypergeometric distributions. **Continuous Probability Distributions:** Mathematical properties and real-world applications of: Uniform, Exponential, Gamma, Beta (Type I & II), and Normal distributions; Areas under normal curve.	07	04	-	11
4 (10 Marks)	**Limit Laws:** Convergence in distribution, convergence in probability, convergence in mean, and almost sure convergence; Relations between different types of convergence. Inequalities: Chebyshev's, Markov's, Cauchy-Schwarz, and Jensen's inequalities. **Limit Theorems:** Weak and Strong Law of Large Numbers and their applications; De-Moivre Laplace Theorem; Central Limit Theorem for i.i.d. variates and its applications; Liapunov's Theorem. **Order Statistics:** Distribution of order statistic; Joint distribution of order statistics; Distribution of range.	05	03		08
	Total	**30**	**15**	**-**	**45**

Where, ***L: Lectures*** ***T: Tutorials*** ***P: Practical***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, 12th ed., New Delhi, India: Sultan Chand & Sons, 2020.

[2] J. L. Devore, Probability and Statistics for Engineering and the Sciences, 9th ed., Boston, MA, USA: Cengage Learning.

[3] S. M. Ross, A First Course in Probability, 10th ed., Boston, MA, USA: Pearson Education, 2019/2020.

[4] G. Casella and R. L. Berger, Statistical Inference, 2nd ed. (International), Belmont, CA, USA: Cengage Learning (reprint), 2024

Title of the Course	**:**	**Introduction to Probability and Distributions**
Course Code	**:**	**Minor3**
Nature of the Course	**:**	**Minor**
Total Credits	**:**	**04**
Distribution of Marks	**:**	**End Sem: 60 TH, In Sem: 40 TH**

COURSE OBJECTIVES:

Knowledge:

- Understand the fundamentals of set theory, including algebra of sets, types of sets, operations, Venn diagrams, and Cartesian products.
- Gain knowledge of relations and their properties, and basic concepts of algebraic structures and set functions.
- Understand combinatorics principles and their role in probability.
- Comprehend different approaches to probability theory: classical, statistical, and axiomatic.
- Learn addition and multiplication theorems of probability and their applications.
- Understand conditional probability, independence, Bayes' theorem, and the theorem of total probability.
- Gain knowledge of random variables (discrete and continuous), including PMF, PDF, and CDF.
- Understand joint, marginal, and conditional distributions in bivariate and multivariate cases.
- Learn the concept of mathematical expectation, variance, covariance, and conditional expectation.
- Understand generating functions: PGF, MGF, and characteristic functions.
- Acquire knowledge of important discrete distributions (Bernoulli, Binomial, Poisson, Geometric, Negative Binomial, Hypergeometric).
- Understand continuous distributions (Uniform, Exponential, Gamma, Beta, Normal) and their properties.
- Learn limit laws and types of convergence in probability theory.
- Understand key inequalities (Chebyshev, Markov, Cauchy-Schwarz, Jensen).
- Gain knowledge of limit theorems including LLN, CLT, and related results.
- Understand the theory of order statistics, including distributions and applications.

Skills:

- Perform set operations and represent them using Venn diagrams.
- Solve problems using counting techniques and combinatorics.
- Apply probability laws and theorems to compute probabilities in real-life situations.
- Use Bayes' theorem for decision-making and inference problems.
- Work with random variables, and compute PMF, PDF, and CDF.
- Derive and analyze joint, marginal, and conditional distributions.
- Compute and interpret expectation, variance, and covariance.
- Apply generating functions to find moments and probabilities.
- Identify and apply appropriate probability distributions to real-world data.
- Solve problems involving normal distribution and standardization.
- Analyze convergence of random variables using different modes of convergence.
- Apply inequalities to bound probabilities and expectations.
- Use limit theorems (LLN, CLT) for approximation and inference.
- Derive and analyze order statistics and their distributions.

Attitude:

- Develop a logical and analytical mindset for handling uncertainty and randomness.
- Appreciate the importance of probability theory in real-world decision-making.
- Cultivate critical thinking and problem-solving abilities in stochastic contexts.
- Show interest in applying statistical methods in diverse fields such as data science, economics, and engineering.

- Develop precision and rigor in mathematical reasoning and interpretation.
- Build confidence in handling complex probabilistic models and data analysis.
- Recognize the importance of probabilistic thinking in research and innovation.
- Foster a habit of systematic analysis and interpretation of results.

COURSE OUTCOMES:

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

CO1: Students will be able to understand and apply concepts of set theory, relations, and combinatorics in probability.
- **ILO1:** Define and perform operations on sets and represent them using Venn diagrams.
- **ILO2:** Classify finite, infinite, and power sets.
- **ILO3:** Understand ordered pairs and Cartesian products.
- **ILO4:** Identify and analyze relations and their properties.
- **ILO5:** Apply basic algebraic structures and set functions.
- **ILO6:** ILO6: Use fundamental counting principles in solving probability problems.

CO2: Students will be able to apply different approaches to probability and solve problems involving conditional probability.
- **ILO1:** Explain classical, statistical, and axiomatic approaches to probability.
- **ILO2:** Apply addition and multiplication theorems.
- **ILO3:** Compute conditional probabilities.
- **ILO4:** Analyze independent events.
- **ILO5:** Apply Theorem of Total Probability.
- **ILO6:** Use Bayes' Theorem in real-life decision-making problems.

CO3: Students will be able to analyze random variables and their probability distributions.
- **ILO1:** Define discrete and continuous random variables.
- **ILO2:** Compute and interpret PMF, PDF, and CDF.
- **ILO3:** Perform transformations of random variables.
- **ILO4:** Analyze joint, marginal, and conditional distributions.
- **ILO5:** Determine independence of random variables.
- **ILO6:** Extend concepts to multivariate distributions.

CO4: Students will be able to compute expectations and utilize generating functions in probability.
- **ILO1:** Compute mathematical expectation and its properties.
- **ILO2:** Calculate variance and covariance.
- **ILO3:** Evaluate conditional expectation.
- **ILO4:** Apply probability generating function (PGF).
- **ILO5:** Use moment generating function (MGF) to find moments.
- **ILO6:** Apply characteristic functions in probability analysis.

CO5: Students will be able to model real-world problems using standard probability distributions.
- **ILO1:** Identify and apply discrete distributions (Bernoulli, Binomial, Poisson, etc.).
- **ILO2:** Analyze properties and applications of discrete distributions.
- **ILO3:** Apply continuous distributions (Uniform, Exponential, Gamma, Beta, Normal).
- **ILO4:** Compute probabilities using normal distribution and standardization.
- **ILO5:** Interpret real-life applications of probability distributions.

CO6: Students will be able to analyze convergence concepts and apply major probabilistic theorems.
- **ILO1:** Define and distinguish between types of convergence (in probability, distribution, mean, almost sure).
- **ILO2:** Analyze relationships among different modes of convergence.
- **ILO3:** Apply Chebyshev's, Markov's, Cauchy-Schwarz, and Jensen's inequalities.
- **ILO4:** Understand and apply Law of Large Numbers (LLN).
- **ILO5:** Apply Central Limit Theorem (CLT) in approximation problems.
- **ILO6:** Understand De-Moivre Laplace and Liapunov's Theorem.

CO7: Students will be able to analyze order statistics and their applications.
- **ILO1:** Define and derive order statistics.
- **ILO2:** Determine distribution of order statistics.
- **ILO3:** Analyze joint distribution of order statistics.
- **ILO4:** Compute and interpret range and related measures.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	Cognitive Process Dimension					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge				CO1, CO7	CO2, CO3, CO5	
Metacognitive Knowledge					CO4, CO6	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	S	M	L	L	L	L	L	L
CO2	S	S	M	M	L	L	L	L
CO3	S	S	M	M	L	L	L	L
CO4	S	S	M	M	L	L	L	L
CO5	S	S	M	M	M	L	L	L
CO6	S	S	M	S	M	L	L	L
CO7	S	S	M	M	L	L	L	L

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (15 Marks)	**Set Theory and Probability Foundations:** Algebra of sets; Finite, infinite, and power sets; Operations on sets; Venn diagrams and their applications; Ordered pairs and Cartesian product of sets; Relations: Types and properties; Basic idea of field (algebra), set functions, and combinatorics. Combinatorics: Fundamental principles of counting for probability. **Probability Theory:** Classical, statistical, and axiomatic approaches; Addition and multiplication theorems. Conditional Probability: Independent events; Theorem of total probability; Bayes' Theorem and its applications.	07	03	-	10
2 (20 Marks)	**Random Variables:** Univariate: Discrete and continuous random variables; cumulative distribution function, probability mass function, and probability density function; Univariate transformations. Bivariate: Joint, marginal, and conditional distributions; Independence of variables; Bivariate transformations; Multivariate extension. **Mathematical Expectation:** Theoretical properties; Variance and Covariance; Conditional expectation. **Generating Functions:** Probability generating function, Moment generating function, and Characteristic function.	11	05	-	16
3 (15 Marks)	**Discrete Probability Distributions:** Mathematical properties and real-world applications of: Bernoulli, Binomial, Poisson, Geometric, Negative Binomial, and Hypergeometric distributions. **Continuous Probability Distributions:**Mathematical properties and real-world applications of: Uniform, Exponential, Gamma, Beta (Type I & II), and Normal distributions; Areas under normal curve.	07	04	-	11
4 (10 Marks)	**Limit Laws:** Convergence in distribution, convergence in probability, convergence in mean, andalmost sure convergence; Relations between different types of convergence. Inequalities: Chebyshev's, Markov's, Cauchy-Schwarz, and Jensen's inequalities. **Limit Theorems:** Weak and Strong Law of Large Numbers and their applications; De-Moivre Laplace Theorem; Central Limit Theorem for i.i.d. variates and its applications; Liapunov's Theorem. **Order Statistics:** Distribution of order statistic; Joint distribution of order statistics; Distribution of range.	05	03		08
	Total	**30**	**15**	**-**	**45**

Where, ***L: Lectures*** ***T: Tutorials*** ***P: Practical***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] S. C. Gupta and V. K. Kapoor, Fundamentals of Mathematical Statistics, 12th ed., New Delhi, India: Sultan Chand & Sons, 2020.

[2] J. L. Devore, Probability and Statistics for Engineering and the Sciences, 9th ed., Boston, MA, USA: Cengage Learning.

[3] S. M. Ross, A First Course in Probability, 10th ed., Boston, MA, USA: Pearson Education, 2019/2020.

[4] G. Casella and R. L. Berger, Statistical Inference, 2nd ed. (International), Belmont, CA, USA: Cengage Learning (reprint), 2024.

Title of the Course	**:**	**Applied Statistics**
Course Code	**:**	**GECSTS3**
Nature of the Course	**:**	**Generic Elective**
Total Credits	**:**	**03**
Distribution of Marks	**:**	**End Sem: 60 TH, In Sem: 40 TH**

COURSE OBJECTIVES:

Knowledge:

- Understand the construction and application of index numbers in economics.
- Learn the principles and methods of statistical quality control.
- Gain knowledge about demographic methods and their significance.

Skills:

- Develop the ability to compute and interpret various index numbers.
- Acquire skills in constructing and analyzing control charts for variables and attributes.
- Enhance capabilities in calculating and interpreting demographic measures.

Attitude:

- Appreciate the role of statistics in economics, industry, and society.
- Develop a methodical and analytical approach to solving statistical problems.
- Foster a critical mindset towards data interpretation and quality control processes.

COURSE OUTCOMES:

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

CO1: Develop a comprehensive understanding of index numbers.
- **ILO1:** Ability to define and describe various types of index numbers.
- **ILO2:** Competence in constructing price and quantity index numbers using Laspeyre's, Paasche's, Marshall-Edgeworth's, and Fisher's formulas.
- **ILO3:** Skill in interpreting consumer price index numbers and wholesale price index numbers.
- **ILO4:** Understanding the uses and limitations of index numbers.
- **ILO5:** Capability to perform base shifting, splicing, and deflating of index numbers.

CO2: Gain proficiency in statistical quality control techniques.
- **ILO1:** ILO1: Knowledge of the importance of statistical methods in industrial research and practice.
- **ILO2:** ILO2: Understanding the causes of variations in quality: chance and assignable.
- **ILO3:** ILO3: Ability to construct and interpret X-bar and R-charts for variables.
- **ILO4:** ILO4: Proficiency in creating and analyzing p-charts and c-charts for attributes.
- **ILO5:** ILO5: Skill in estimating process capability and understanding its implications.

CO3: Acquire in-depth knowledge of demographic methods.
- **ILO1:** Understanding the measurement of population, rates, and ratios of vital events.
- **ILO2:** Ability to compute and interpret measures of mortality, including CDR, SDR, IMR, and standardized death rates.
- **ILO3:** Competence in measuring fertility and reproduction rates, including CBR, GFR, and TFR.
- **ILO4:** Proficiency in constructing and using life tables and understanding their uses.
- **ILO5:** Skill in measuring population growth using GRR and NRR.

CO4: Develop practical skills in statistical analysis and interpretation.
- **ILO1:** Ability to construct and interpret statistical control charts (X-bar & R-chart, s-chart, np-chart, p-chart, c-chart, u-chart).
- **ILO2:** Skill in constructing and interpreting single sample inspection plans, including OC, AQL, LTPD, ASN, ATI, AOQ, AOQL curves.
- **ILO3:** Competence in calculating process capability and comparing 3-sigma control limits with specification limits.
- **ILO4:** Proficiency in performing practical calculations related to demographic measures and index numbers.

CO5: Foster a critical and analytical approach towards statistical data and methods.
- **ILO1:** Appreciation for the role of statistics in economics, industry, and society.
- **ILO2:** Development of a methodical and analytical approach to solving statistical problems.
- **ILO3:** Ability to critically analyze data and interpret results in the context of quality control and demographic analysis.
- **ILO4:** Understanding the application of statistical methods in real-world scenarios.

ILO5: Skill in presenting statistical findings effectively through various modes of assessment (assignments, presentations, laboratory work).

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	Cognitive Process Dimension					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge						
Conceptual Knowledge			CO1, CO2			
Procedural Knowledge				CO3, CO4	CO5	
Metacognitive Knowledge						

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	S	S	M	S	M	S	M	M
CO2	S	S	M	S	M	M	M	M
CO3	S	S	M	S	M	S	M	M
CO4	S	S	M	S	M	S	M	M
CO5	S	S	M	S	M	M	M	M

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (12 Marks)	**Index Numbers**: Definition, Criteria for a good index number, different types of index numbers. Construction of index numbers of prices and quantities, consumer price index number. Uses and limitations of index numbers. Base shifting, Splicing and deflating of Index numbers, idea of inflation.	05	03	-	08
2 (12 Marks)	**Statistical Quality Control:** Importance of statistical methods in industrial research and practice. Causes of variations in quality: chance and assignable. General theory of control charts, process & product control, Control charts for variables: X-bar and R-charts. Control charts for attributes: p- andc-charts.	06	03	-	09
3 (14 Marks)	**Demographic Methods:** Introduction, measurement of population, rates and ratios of vital events. Measurement of mortality, fertility and reproduction. Basic concepts and construction of life tables.	07	03	-	10
4 (12 Marks)	**Educational Statistics:** Scales of test score (Z scores, standard scores, normalized scores etc.), Reliability of test scores (Definition, reliability index, parallel test etc.), Validity of test scores (estimation and types of validity). Intelligence tests and intelligence quotient (definition only).	06	03		09
5 (10 Marks)	**Co-requisite: (computer based)** 1. Construction of price and quantity index numbers by Laspeyre's formula, Paasche's formula, Marshall-Edgeworth's formula, Fisher's Formula. Comparison and interpretation. 2. Construction of wholesale price index number, fixed base index number and consumer price index number with interpretation. 3. Construction and interpretation of X bar &R-chart. 4. Construction and interpretation p-chart (fixed sample size) and c-chart. 5. Computation of measures of mortality. 6. Completion of lifetable. 7. Computation of measures of fertility and population growth. 8. Computation of test scores.	06	03	-	09
	Total	**30**	**15**	**-**	**45**

Where, ***L: Lectures*** ***T: Tutorials*** ***P: Practical***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] Gun, A. M., Gupta, M. K., and Dasgupta, B. (2008). *Fundamentals of Statistics, Vol. II (9th Edn.)*. World Press, Kolkata.

[2] Gupta, S. C. (1990). *Fundamentals of Statistics*. Himalaya Publishing House, Mumbai, India.

[3] Gupta, S. C., and Kapoor, V. K. (2008). *Fundamentals of Applied Statistics (4th Edn., Reprint)*. Sultan Chand & Sons.

[4] Keyfitz, N., and Beckman, J. A. *Demography through Problems*. S-Verlag, New York.

[5] Montgomery, D. C. (2009). *Introduction to Statistical Quality Control (6th Edn.)*. Wiley India Pvt. Ltd.

[6] Mukhopadhyay, P. (1999). *Applied Statistics*. New Central Book Agency, Calcutta.

[7] Pathak, K. B., and Ram, F. (2016). *Techniques of Demographic Analysis*. Himalaya Publishing House.

Title of the Course	**:**	**Statistical Data Processing using R**
Course Code	**:**	**SEC 3**
Nature of the Course	**:**	**Skill Enhancement Course**
Total Credits	**:**	**03**
Distribution of Marks	**:**	**End Sem: 60, In Sem: 40**

COURSE OBJECTIVES:

Knowledge:

- Understand the R ecosystem, including RStudio, scripting, and project workflow.
- Gain knowledge of R as a functional programming language and its syntax.
- Understand data types (numeric, character, logical) and data structures (vectors, matrices, lists, data frames).
- Learn the structure and functionality of data frames and tibbles.
- Understand methods of data ingestion and export (CSV, Excel, JSON).
- Gain knowledge of scoping rules and environment handling in R.
- Understand principles of data wrangling using tidyverse packages (dplyr, tidyr).
- Learn core data manipulation functions such as filter(), select(), mutate(), arrange(), summarize().
- Understand data cleaning techniques, including handling missing values, duplicates, and inconsistencies.
- Gain knowledge of data visualization principles and basic plotting techniques.
- Understand the Grammar of Graphics (ggplot2) and its components.
- Learn the concepts of statistical computing and simulation in R.
- Understand how to create dynamic reports using R Markdown integrating code, text, and visualizations.

Skills:

- Write and execute R scripts and manage projects efficiently in RStudio.
- Create and manipulate data structures such as vectors, matrices, lists, and data frames.
- Import, clean, and preprocess real-world datasets from multiple sources.
- Apply data wrangling techniques using tidyverse functions.
- Use piping (%>%) to build efficient and readable workflows.
- Reshape data using pivoting techniques (long ↔ wide formats).
- Detect and handle missing values, duplicates, and inconsistencies in datasets.
- Generate and customize data visualizations (bar plots, histograms, boxplots, scatter plots).
- Create advanced visualizations using ggplot2, including faceting and layering.
- Perform statistical computations, including descriptive statistics and correlation analysis.
- Conduct basic simulations and sampling from probability distributions.
- Develop reproducible reports using R Markdown in HTML/PDF formats.

Attitude:

- Develop a data-driven mindset for problem-solving and decision-making.
- Appreciate the importance of clean, well-structured, and reproducible code.
- Cultivate attention to detail in data cleaning and preprocessing.
- Show interest in data visualization and storytelling with data.
- Develop analytical thinking in interpreting graphical and statistical outputs.
- Build confidence in working with real-world datasets and tools.
- Demonstrate a systematic and organized approach to data analysis projects.
- Value reproducibility and transparency in reporting and research.
- Foster continuous learning in data science tools and techniques.

COURSE OUTCOMES:

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

CO1: Students will be able to understand and work efficiently within the R ecosystem and write basic R programs.

- **ILO1:** Explain the R ecosystem and use RStudio interface effectively.
- **ILO2:** Define and use data types (numeric, character, logical).
- **ILO3:** Create and manipulate data structures (vectors, matrices, lists, data frames).
- **ILO4:** Perform basic operations, indexing, and subsetting.
- **ILO5:** Develop and manage R scripts and projects.
- **ILO6:** Understand workspace and environment management.

CO2: Students will be able to import, manage, and manipulate structured datasets in R.

- **ILO1:** Describe the structure of data frames and tibbles.
- **ILO2:** Import data from CSV, Excel, and JSON formats.
- **ILO3:** Export processed datasets.
- **ILO4:** Apply scoping rules and manage environments.
- **ILO5:** Organize and prepare datasets for analysis.

CO3: Students will be able to clean, transform, and reshape data using modern data manipulation tools.

- **ILO1:** Apply dplyr functions: filter(), select(), mutate(), arrange(), summarize().
- **ILO2:** Use pipe operator (%>%) for efficient workflows.
- **ILO3:** Perform data reshaping (long ↔ wide).
- **ILO4:** Handle missing values and detect inconsistencies.
- **ILO5:** Identify and remove duplicate records.
- **ILO6:** Apply data validation and preprocessing techniques.

CO4: Students will be able to create and interpret visualizations for data analysis.

- **ILO1:** Explain principles of data visualization.
- **ILO2:** Create basic plots (bar plot, histogram, boxplot, scatter plot).
- **ILO3:** Customize plots using labels and themes.
- **ILO4:** Apply ggplot2 grammar of graphics.
- **ILO5:** Use aesthetic mapping and geometric objects.
- **ILO6:** Perform faceting and layering for multivariate analysis.
- **ILO7:** Interpret visual patterns and insights from data.

CO5: Students will be able to perform statistical analysis and simulations using R.

- **ILO1:** Generate random samples from probability distributions.
- **ILO2:** Compute descriptive statistics (mean, variance, etc.).
- **ILO3:** Calculate and interpret correlation matrices.
- **ILO4:** Perform basic simulations for statistical understanding.

CO6: Students will be able to create reproducible and professional reports integrating code and analysis.

- **ILO1:** Explain the concept of reproducible research.
- **ILO2:** Use R Markdown for documentation.
- **ILO3:** Integrate code, text, LaTeX equations, and visualizations.
- **ILO4:** Generate HTML/PDF reports.
- **ILO5:** Present analysis results in a clear and structured format.

Table: Learning Outcome Representation (CO): Bloom's Taxonomy Table

Cognitive Knowledge Dimensions	Cognitive Process Dimension					
	Remember	**Understand**	**Apply**	**Analyze**	**Evaluate**	**Create**
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge			CO1	CO2	CO5	
Metacognitive Knowledge					CO3, CO4, CO6	

Table: Course Outcome (CO) and Program Outcome (PO) mapping

CO/PO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8
CO1	M	M	L	L	S	L	L	L
CO2	M	M	M	M	S	L	L	L
CO3	M	S	S	M	S	L	L	M
CO4	M	S	S	M	S	L	L	M
CO5	S	S	M	S	S	L	L	L
CO6	M	M	M	M	S	L	L	M

(S= Strong, M= Medium, L= Low)

UNITS	CONTENTS	L	T	P	Total Hours
1 (10 Marks)	**R Ecosystem and Syntax:** Overview of R and RStudio; R as a functional programming language; Data types: numeric, character, logical; Data structures: vectors, matrices, lists, data frames; Basic operations, indexing, and subsetting; Workflow: Creating scripts; Project management; Workspace management.	05	01	02	08
2 (12 Marks)	**The Data Frame:** Anatomy of a data frame; Tibbles vs. Data Frames. **Data Ingestion:** Importing data from CSV, Excel, and JSON; Exporting processed datasets; Scoping rules and environment. **Data Wrangling (The Tidyverse):** Introduction to the dplyr and tidyr packages. Core Verbs: filter(), select(), mutate(), arrange(), and summarize(); Piping: Streamlining code using the pipe operator; Reshaping: Pivoting data (Long to Wide and vice versa); Handling missing values; Detection and treatment of duplicates and inconsistencies; Data validation and basic preprocessing techniques	06	01	04	11
3 (15 Marks)	**Data Visualization:** Principles of data visualization; Basic plots: bar plot, histogram, boxplot, scatter plot; Customization of plots (labels, themes). **Exploratory Graphics (ggplot2):** The Grammar of Graphics; Mapping aesthetics; Geometric objects; Visual Layering: Statistical transformations; Faceting for multivariate data; Customizing scales, themes, and labels for publication-quality output; Visual interpretation of data patterns.	07	01	05	13
4 (13 Marks)	**Statistical Computing:** Basic simulations (Sampling from distributions); Calculating descriptive statistics and correlation matrices.**Reporting:** Introduction to R Markdown; Creating dynamic HTML/PDF reports that combine code, LaTeX equations, and visualizations.	07	02	04	13
	Total	**25**	**05**	**15**	**45**

Where, ***L: Lectures*** ***T: Tutorials*** ***P: Practical***

MODES OF IN-SEMESTER ASSESSMENT: **(40 Marks)**

- Internal Examination **20 Marks**
- Others (Quiz, Seminar presentation, Assignment) **20 Marks**

SUGGESTED READINGS:

[1] H. Wickham, M. Çetinkaya-Rundel, and G. Grolemund, R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, 2nd ed., Sebastopol, CA, USA: O'Reilly Media, 2023.

[2] W. McKinney, Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd ed., Sebastopol, CA, USA: O'Reilly Media, 2022.

[3] J. VanderPlas, Python Data Science Handbook: Essential Tools for Working with Data, Sebastopol, CA, USA: O'Reilly Media.

[4] J. L. Devore, Probability and Statistics for Engineering and the Sciences, 9th ed., Boston, MA, USA: Cengage Learning.

[5] H. Wickham, D. Navarro, and T. L. Pedersen, ggplot2: Elegant Graphics for Data Analysis, 3rd ed., Cham, Switzerland: Springer International Publish

[6] Baumer, B. S., Kaplan, D. T., & Horton, N. J. (2017). *Modern Data Science with R*. CRC Press.

[7] Horton, N. J., & Kleinman, K. (2015). *Using R and RStudio for data management, statistical analysis, and graphics*. CRC Press.