KASHI NARESH UNIVERSITY, BHADOHI

DEPARTMENT OF STATISTICS AND DATA SCIENCE

(National Education Policy-2020)

Syllabus

of

STATISTICS AND DATA SCIENCE

(Effective from Academic Session 2026-2027)

For

B.A./B.Sc.

Approved by

Board of Studies
Department of Statistics and Data Science
Faculty of Science
Kashi Naresh University
Bhadohi

Convenor
Department of Statistics and Data Science
Kashi Naresh University, Bhadohi

Dean
Faculty of Science
Kashi Naresh University, Bhadohi

(Dr. Rahul Kumar)
Member of Board of Studies

(Prof. Brijesh Pratap Singh)
Member of Board of Studies

(Dr. Parul Yadav)
Member of Board of Studies

Course Title: Statistics and Data Science

Semester - I

Class: UG	**Year: First**	**Semester: First**
Subject: Statistics and Data Science		
Course Code:	**Course Title:** Descriptive statistics and Theory of Probability.	
Credits: 6	Core Compulsory	
MaxMarks:25(Internal)+75(External)	**Min Passing Marks:** As per University CBCS Norms	
Total No. of Lectures-Tutorials-Practical (in hours per week): L-T-P:4-0-2		
Course Prerequisites: To study this course, a student must have the subject Mathematics in class 12^{th}.		

Course objective:
The course aims to:
CO1: Develop a strong foundation in statistics and mathematics.
CO2: Ability to use appropriate measures of central tendency according to situation.
CO3: Learn data collection, analysis, visualization, and interpretation.
CO4: Develop skill to identify the situation to apply appropriate measure of dispersion as per nature and need of data and draw meaningful conclusions regarding behavior of data.
CO5: Develop a fundamental understanding of the concepts and principles of probability.
CO6: Introduce students to sample spaces, events, and axiomatic approach to probability.
CO7: Introduce random variables and their probability distributions.
CO8: Develop an understanding of discrete and continuous probability distributions.
CO9: Familiarize students with important probability distributions such as Binomial, Poisson, Normal and exponential distributions.

PART: A DESCRIPTIVE STATISTICS	
UNIT	**Topics**
I	**Introduction to Statistics**: Meaning, scope and importance of Statistics, types of data-primary and secondary, qualitative and quantitative, methods of data collection-questionnaires and schedules, scales of measurement-Nominal, Ordinal, Interval and Ratio.
II	**Presentation of Data**: Classification and tabulation of data, Diagrammatic and graphical representation of grouped and ungrouped data, frequency distribution, cumulative frequency distributions and their graphical representations, Bar diagram: simple, multiple and component, Pie chart Histogram, Frequency polygon and Ogives, Stem and Leaf plot, Box plot.(Only Introduction)
III	**Measures of Central Tendency**: Arithmetic mean, median, mode, Geometric mean and Harmonic mean, Merit and Demerits and comparison of these measures, Quartiles, Deciles and percentiles, Measure of Dispersion: range, Mean deviation, standard deviation, quartile deviation and variance, their properties, merit and demerits, Coefficient of variation. Moments: Raw moments and Central moments

	and relationship between moments, Skewness: Meaning, positive and negative skewness, Karl Pearson's coefficient and Bowley's Coefficient, Kurtosis: Mesokurtic, Leptokurtic and Platykurtic, properties and significance.
IV	**Correlation and Regression:** Bivariate data, Principle of least squares, Meanings of curve fitting, Fitting of straight line, parabola, logarithmic, power curves and other simple forms by method of least squares, Bi-variate frequency table, correlation, types of relationships, Scatter diagram, Karl-Pearson's correlation coefficient and its property, Rank correlation and its coefficient, Multiple and Partial correlation, Regression analysis. Coefficient of Determination and its uses.

PART B: THEORY OF PROBABILITY	
UNIT	**Topics**
I	**Basic Concept of Probability:** Random experiment, Trial, Sample point and sample space, Events, Operations of Events, Concept of equally likely, Mutually exclusive and Mutually exhaustive events, Definition of Probability: Classical, Empirical and Axiomatic approaches, Some theorems on probability, Conditional Probability, Multiplication theorem of probability, Independent events, Extension of multiplication theorem, Axioms of addition and axioms of continuity, Total and Compound Probability theorems , Bayes theorem and its application.
II	**Random Variables and Distribution Functions:** random variables: Discrete and Continuous, Probability Mass Functions (pmf) and Probability Density Functions(pdf), Cumulative density Function (cdf), Joint distribution of two random variables, Marginal and Conditional distributions, Independence of random variables, Expectation of random variable and its properties, Expectation of sum and product of random variables, Variance and Covariance and their properties, Conditional Expectation, Conditional Variance and Related problems.
III	**Generating Functions and Law of Large Numbers:** Moments: moments of Bivariate probability distributions, Moment Generating Functions and its limitations and Properties, Cumulants and its properties, Characteristic Function and its properties, Continuity theorem for m.g.f.(without proof), Chebyshev's Inequality, Weak Law of large numbers for sequence of independently and identically distributed random variables and their applications(statement only), Central Limit Theorem, Borel-Cantelli Lemma, Probability Generating Function.
IV	**Probability distributions:** Discrete Probability Distributions and their properties, Bernoulli distribution, Binomial distribution, Poisson Distribution, Negative Binomial Distribution, Geometric Distribution, Hypergeometric distribution. Continuous Probability Distribution and their properties, Normal distribution, Uniform distribution, Gamma distribution, Beta distribution of first and second kind, exponential distribution, Weibull distribution, Cauchy distribution.

Descriptive Statistics and Theory of Probability Lab

- Problems based on graphical representation of data by Histogram, Frequency. Polygons, Frequency curves and Ogives, Stem and Leaf plot, Box plot.
- Problems based on calculation of Measures of Central Tendency.
- Problems based on calculation of Measures of Dispersion.

- Problems based on calculation of Measures of Skewness and Kurtosis.
- Problems based on computing correlation coefficient and regression.
- Computation of Conditional Probabilities based on Bayes Theorem.
- Fitting of Binomial and Poisson distribution.
- Develop an intuitive understanding of probability through experiments and simulation.
- Apply the concepts using Microsoft Excel.

Course Outcomes:

After completing this course, a student will have:

CO1: Knowledge of Statistics, its scope and importance in various fields.
CO2: Knowledge of methods of summarizing data sets, including common graphical tools (such as box-plots, histograms and stem-plots.
CO3: Ability to describe data with measures of central tendency and measure of dispersion.
CO4: Ability to understand measures of skewness and kurtosis and their utility and significance.
CO5: Ability to perform correlation and regression analysis.
CO6: Explain the basic concepts of probability, sample spaces, events, and probability axioms.
CO7: Define random variables and distinguish between discrete and continuous random variable.
CO8: Identify and apply appropriate discrete and continuous probability distributions to real world situations.
CO9: Solve problems involving Binomial, Poisson, Normal and Exponential distribution.

Suggested Readings:

1. **David, S. (1994): Elementary Probability, Cambridge University Press.**
2. **Gupta, S.C. and Kapoor, V.K. (2000). Fundamentals of Mathematical Statistics (10^{th} ed.), Sultan Chand and Sons**
3. **Goon, A.M., Gupta, M.K. and Dasgupta, B. (2013). Fundamentals of Statistics, Vol I, World Press, Kolkata**
4. **Mood, A.M. Graybill, F.A. and Bose, D.C. (2011). Introduction to the Theory of Statistics, 3^{rd} Edn., Tata McGraw-Hill Pub. Co. Ltd.**
5. **Meyer, P. (2017). Introductory Probability and Statistical Applications(2^{nd} ed.), New Delhi, Oxford&IBH Publishing Co. Pvt. Ltd.**
6. **Parzen, E.S. (1992). Modern Probability Theory and its Applications. Wiley Interscience.**
7. **Introduction of Statistical Methods, J. Medhi**

Semester - II

Class: UG	**Year: First**	**Semester: Second**
Subject: Statistics and Data Science		
Course Code:	**Course Title:** Foundations for Data Science and Statistical Computing	
Credits: 6	Core Compulsory	
MaxMarks:25(Internal)+75(External)	**Min Passing Marks:** As per University CBCS Norms	
Total No. of Lectures-Tutorials-Practical (in hours per week): L-T-P:4-0-2		
Course Prerequisites: To study this course, a student must have the subject Mathematics in class 12^{th}.		

Course Objectives:

The course aims to:

CO1: Develop mathematical foundations required for data science.
CO2: Introduce matrix algebra and calculus for data analysis.
CO3: Introduce students to statistical computing using modern software.
CO4: Teach data handling, visualization, and basic statistical data analysis.
CO5: Prepare students for advanced data science and machine learning.
CO6: Introduce students to the fundamental concepts and importance of statistical computing in modern data analysis.
CO7: Develop basic programming and computational skills required for statistical analysis.
CO8: Familiarize students with statistical computing environments such as Python and R.
CO9: Develop skills for data entry, data import, data export, data cleaning, and data management.
CO10: Introduce important data structures such as vectors, lists, arrays, matrices, and data frames.

PART: A FOUNDATIONS FOR DATA SCIENCE

Units	Topics
I	**Mathematical Foundations:** Set and Set Operations, Definition of set, Roster and Set builder forms, Union, Intersection, difference and compliment of a set, Cartesian Product, subsets and Power set, Axioms of sets, relations and functions, Composite and Inverse functions, Mathematical logic: statements and propositions, Truth tables, logical connectives, Quantifiers, Mathematical induction, Real and Complex numbers, Scientific notation, Binary representations, Floating point representation.
II	**Linear Algebra and Calculus for Data Science:** Matrices: Type of matrices, addition, multiplication, transpose, symmetric and skew-symmetric matrices, Determinant: Properties of determinants, minors and cofactors, Adjoint and inverse, System of linear equations: Gaussian elimination, Gauss-Jordan method, Matrix inverse method, Eigenvalues and Eigen vectors, Calculus for Data Science: Limit and Continuity: Concept of limit, Continuity of functions, basic Properties, differentiation, Partial differentiation, functions of several

	variables, Chain rules, Optimization: Maxima and Minima, Constrained optimization, Lagrange Multipliers,
III	**Introduction to Computational Mathematics:** Python for Statistical Computing: NumPy arrays, vector operations, Matrix operations, Built-in mathematical functions, solving mathematical problems: solving equations, matrix inverse computation, Numerical calculations, Graphical Representation: Plotting basic functions, line graphs, Coordinate systems.
IV	**Applications in Data Science:** Variables and parameters, Linear models, Growth models, Matrix Application: Data representation, Image representation, Recommendation systems(conceptual), Introduction to Regression: Dependent and Independent variable, Line fitting concepts, Prediction idea.

PART B: STATISTICAL COMPUTING	
Unit	**Topics**
I	**Introduction to Statistical Computing and Programming:** Introduction to statistical Computing: Meaning and importance of Statistical computing, Evolution of Statistical software, applications in research, business, healthcare, finance, agriculture, and government, Introduction to Python/R Environment: Installing Python, Jupyter Notebook, R and R studio, Working with notebook/scripts, Basic syntax and commands, Variables and Data types, Input and Output operations, Operators and expressions: arithmetic, Relational, Logical and Assignment operators, Control Structures: Conditional statements(if, else, elif), Loops(for, while), Functions (basic concepts).
II	**Data Handling and Management:** Data Structures: Lists, Tuples, Dictionaries, Arrays, Matrices, Data Frames, Data Import and Export: CSV files, Excel files, Text Files, JSON(Introduction), Data Cleaning: Missing values, Duplicate observations, Data type conversion, Renaming Variables, Sorting and Filtering, Data Manipulation: Indexing, selecting variables, Grouping, Aggregation, Merging datasets.
III	**Descriptive Statistics and Data Visualization:** Descriptive statistics: frequency distributions, Measures of central tendency, Measure of Dispersion, Quartiles and Percentiles, Exploratory Data Analysis: Summary Statistics, Outlier detection, Data distribution, Data Visualization: Bar charts, Pie charts, Histograms, Frequency polygons, Line graphs, Scatter plots, Box plots, Visualization Libraries: Matplotlib, Seaborn (Python), ggplot2(R).
IV	**Statistical Analysis and Applications:** Correlation Analysis: Scatter diagram, Pearson Correlation coefficient, Spearman rank correlation (Introduction), Interpretation, Simple Linear Regression: Concept of regression, Regression equation, least squares method (basic idea), Prediction using regression, Interpretation of regression coefficients, Introduction to Statistical Packages: SciPy, Statsmodels, Basic use of statistical functions.

Mini Data Analysis Projects:

- Import a real dataset
- Clean the data
- Generate descriptive statistics
- Create visualizations
- Perform correlation and simple regression
- Write a brief report with conclusions.

Foundations for Data Science and Statistical Computing Lab

- Solve the problem using NumPy matrix operation.
- Solving equations computationally
- Plotting mathematical functions
- Problem based on fitting of curves by method of least squares e.g. fitting of straight-line, second-degree polynomial, power curve, exponential curve.
- Installing Python/R and required packages.
- Reading CSV and Excel files.
- Data cleaning and processing.
- Descriptive Statistics.
- Frequency tables. Scatter plots
- Correlation analysis
- Simple linear regression
- Data visualization dashboard
- Build computational skills using Python, R, or Excel
- Perform simulations and analyze random phenomenon.
- Mini project using a real dataset

Course Outcome:

After completing this course, a student will have:

CO1: Apply algebra and calculus in data science.

CO2: Solve matrix based computational problems.

CO3: Familiarity with computational techniques and Statistical software including programming language R and Python.

CO4: Create charts and dashboards for data visualization.

CO5: Understand the fundamentals of data science and ethical data practice.

CO6: Use basic programming concepts, operators, functions, and control structures for statistical computations.

CO7: Work with statistical computing environments such as Python, R, and Jupyter note book.

CO8: Create and manipulate basic data structures, including vectors, arrays, metrices, lists, and data frames.

CO9: Import, Export, Organize, and manage data from common sources such as CSV and Excel files.

Suggested Readings:

1. **Wes McKinney-Python for Data Analysis**
2. **Hadley Wickham & Garret Grolemund-R for Data Science**
3. **Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani- An Introduction to Statistical Learning**
4. **Joel Grus-data Science from Scratch**
5. **Jake VanderPlas-Python Data Science Handbook**
6. **Gilbert Strang-Introduction to Linear Algebra**
7. **Charles Severance-Python for everybody**
8. **Senior Secondary School Mathematics, RS Agrawal, Bharti Bhawan, 1995**
9. **Kenneth Rosen- Discrete Mathematics**
10. **Advanced Engineering Mathematics, Erwin Kreysig, Wiley, 2015**