Rayat Shikshan Sanstha's

Sadguru Gadage Maharaj College, Karad.

(An Autonomous College)

Undergraduate Programme

B. Sc. I Data Science

Syllabi of the course

(To be implemented w.e.f. from July 2024)

1) General Objectives of the Course: -

1. To create graduates with sound knowledge of Data Science, who can contribute towards recent advances in technology.
2. To provide advanced and in-depth knowledge of data science and specialization in one or two subjects of the new era of technology.
3. To prepare Graduates who will achieve peer-recognition, as an individual or in a team, through demonstration of good analytical, design, programming, and implementation skills.
4. To enable students, pursue a professional career in Data Science in related industry, business and research.
5. To impact industry knowledge and practical skills of current trends in IT field to the students.
6. To develop the ability among students to formulate, analyze and solve real life problems faced in Computer Science industry. To produce computer science professionals who can be directly employed or start his/her own work as
    - Data Scientist.
    - Business Analyst.
    - Data Analytics Manager.
    - Data Architect.
    - Data Administrator.
    - Business Intelligence Manager.
    - Entrepreneur in Computer Science industry.
7. To Develop designing, analyzing and critical thinking skill among students.

2) Duration:

The course shall be a full-time course.
The course shall be for **three** years, consisting of **six semesters**.

3) Fee Structure:

• **Course Fee:** as prescribed by the Institute.

4) Eligibility for Admission:

• As per Rule (2) for graduates of this Institute.

5) Medium of instruction:

-English

6) OUTCOMES:

After completing this courses student shall be expert in following things:

1. Avail yourself of Current trends in IT Industries and new Technologies.
2. Apply knowledge of programming platforms in Data Science and AI in real life.
3. Students should avail detailed knowledge of Data Science, Artificial Intelligence, Machine Learning, and Big Data etc.
4. Demonstrate their ability of advanced programming to design and develop innovative applications.
5. Access, evaluate, understand, and compare digital information from various sources and apply it for scientific knowledge acquisition as well as scientific data analysis and presentation.
6. Critically evaluate, analyze, and comprehend a scientific problem. Think creatively, experiment and mic research into innovation and creatively design scientific solutions to problems.
7. Exemplify generate a solution independently, check and validate it and modify if necessary.

8. Translate project plans, use management skills, and lead a team for planning and execution of a task.
9. Can start his own business or start up.

7) SCOPE:

After Successful completion of three years Degree in Data Science, we observed that the students have the ample opportunities in diversified areas such as:

1. Data Scientist
2. Data Analytics
3. Machine learning Engineer
4. Business Intelligence Analyst
5. Data Engineer.
6. Research Scientist.
7. Data Consultant

B.Sc. Data Science Part-I Semester- I
Computer Science Paper-I
Course Code-BDST24-101
Course Title: Fundamental of Data Science

Total Hours: 30 hrs (38 lectures of48 min)
Credits:02
Total Marks:40+10=50

UNIT I: Introduction to Data Science (7)

Defining data science and big data, Recognizing the different types of data, Gaining insight into the data science process, Data Science Process: Overview, Different steps, Machine Learning Definition and Relation with Data Science.

UNIT II Data management And Analysis (9)

Data collection and management: Introduction, Sources of data, Data collection and APIs, Exploring and fixing data, Data storage and management, using multiple data sources.

Data analysis: Introduction, Terminology and concepts, Introduction to statistics, Central tendencies and distributions, Variance, Distribution properties and arithmetic, Samples/CLT, Basic machine learning algorithms, Linear regression, SVM, Naive Bayes

UNIT III Data visualization (6)

Data visualization: Introduction, Types of data visualization, Data for visualization: Data types, Data encodings, Retinal variables, mapping variables to encodings, Visual encodings.

UNIT IV Applications of Data Science (8)

Technologies for visualization, Bokeh (Python), recent trends in various data collection and analysis techniques, various visualization techniques, application development methods of used in data science.

Computer Science Paper-II
Course Code-BDST24-102
Course Title: Programming in C

Total Hours: 30 hrs (38 lectures of48 min)
Credits:02
Total Marks:40+10=50

UNIT I Introduction to 'C' language

Problem Solving definition, Step involving in problem solving, Algorithm, Characteristics, Flowcharts, Definition, Symbol, features. History of 'C' language, Structure of 'C' programs, 'C' Tokens, Character set and keywords, Constant and its type, Variable and its type Data types, Operators and its types, Precedence rules, Input/output using standard functions.

UNIT II Branching and looping

Conditional branching, if, if else, else if ladder, switch, Nested statements. Looping – for, while do-while statements. Unconditional control statements- go to, break and continue.

UNIT III Arrays in C

Array definition and declaration, initialization of arrays, types of arrays, String handling functions, Arrays and functions

UNIT IV Functions in C

Definition, types & parts of functions, Local and global variable, Library functions and User defined functions, passing arguments to a function, return statement, recursion, Scope and lifetime of variables, Storage classes-Auto, Extern, Register, Static.

Course Code-BDSP24-103
Course Title: Practical I

Credits:02
Total Marks=50

Lab on Fundamental of Data Science and Programming in C

Fundamental of Data Science

1. Data Collection and Acquisition:

1. Identify relevant data sources and collect the required data.

2. Clean and pre-process the data to ensure its quality and integrity.

3. Convert the data into a suitable format for analysis, such as CSV, Excel, or a

database format.

2.Bar Charts for visualization:

1.Create a bar chart to compare categorical variables or display counts or frequencies.

2.Customize the chart with labels, colors, and legends to enhance readability.

3.Histograms for visualization:

1.Construct a histogram to display the distribution of a single variable.

2.Adjust the bin sizes to reveal different levels of granularity in the data.

4.Case study on Applications of Data Science

Programming in C

1. Write a program to accept 5 subject marks and calculate total marks, percentage, and grade of student.

2. Write a program to perform arithmetic operations.

3. Write a program to input n numbers and find the Odd and Even numbers.

4. Write a program to find the age of a person (Input birth date and today date).

5. Write a program to find the sum of first in natural numbers.

6. Write a program to accept the range and generate Fibonacci Series.

7. Write a program to calculate sum of numbers using simple function.

8. Write a program to calculate factorial of number using Recursion.

9. Write a program to enter array elements and perform arithmetic operations.

10. Write a program to sort the numbers in ascending and descending order using array.

Statistics

Preamble:

This syllabus is framed to give sound knowledge with understanding of Statistics to undergraduate students at first year of three years of B.Sc. (DS) degree course.

Students learn Statistics as a separate subject from B.Sc. I. (DS) .The goal of the syllabus is to make the study of Statistics popular, interesting and encouraging to the students for the higher studies.

General Objectives of the Program:

- Provide students with learning experiences that develop broad knowledge and understanding of key concepts of Statistics
- Develop abilities in students to design and develop innovative solutions for benefits of society, leadership, teamwork and lifelong learning..
- To develop scientific attitude among the student so as to make students curious and open minded.
- Provides Students with skills that enable them to get employment in industries or pursue higher studies.

General outcomes of Program:

1. The students will graduate with proficiency in the subject of their choice.

2. The students will be eligible to continue higher studies in the subject.

3. The students will be eligible to appear the examination for jobs in government organizations.

4. The students will be eligible to apply for jobs with a minimum requirement of B.Sc.(DS)

Program Specific Outcomes

The studies will acquire;

1. Knowledge of descriptive statistics and inferential statistics , sampling techniques.

2.Knowledge about the univariate, bivariate data analysis.

3. Knowledge about the correlation and regression analysis.

4. Knowledge of probability discrete and continuous probability distribution and various measures of these distributions.

5. Knowledge of different methods of estimation about inference of parameter of standard discrete and continuous probability distribution.

6. Knowledge of applied statistics such as ‘index number’, ‘time series’, ‘industrial statistics’.

B.Sc. Part I (DS)

1. **Title** : Statistics
2. **Year of implementation**: The syllabus will be implemented from June 2024 onwards
3. **Duration** : The course shall be a fulltime
4. **Pattern:** Semester examination
5. **Medium of Instruction** : English
6. **Structure of Course** :

B.Sc.-I (DS) : Semester –I

Subject Code	Title of theory Paper	Credit	Lectures /Practical per week	TH/PR	SEE		CCE		Total Theory/ Practical Marks
					Max	Min	Max	Min	
BDST24-104	Descriptive Statistics - I	02	2	TH	40	16	10	4	50
BDST24-105	Probability for Data Science	02	2	TH	40	16	10	4	50
BDSP24-106	Practical -I	02	4	PR	50	20	---	---	50

B.Sc.-I (DS) : Semester –II

Subject Code	Title of theory Paper	Credit	Lectures /Practical per week	TH/PR	SEE		CCE		Total Theory/ Practical Marks
					Max	Min	Max	Min	
BDST24-204	Descriptive Statistics - II	02	2	TH	40	16	10	4	50
BDST24-205	Data Science with R	02	2	TH	40	16	10	4	50
BDSP24-206	Practical -II	02	4	PR	50	20	---	---	50

Titles of Papers of B.Sc. course:

B.Sc.I(DS) Semester-I :

Theory:

BDST24-104:Descriptive Statistics - I
BDST24-105:Probability for Data Science

Practical (Semester) :

BDSP24-106 : Practical -I

B.Sc.I(DS) Semester-II :

Theory:

BDST24-204 :Descriptive Statistics - II
BDST24-205:Data Science with R

Practical (Semester) :

BDSP24-206 : Practical -II

B. Sc. I (DS) (Semester-I)

BDST24- 104: Descriptive Statistics - I

Theory-30 Hours Credits:02

Objectives:

The main objectives of this course are:

1. Methods in descriptive statistics
2. The use of concepts in descriptive statistics as applied to real data
3. Methods for finding correlation between variables
4. Fitting an equation for prediction and apply the same for real data.

Syllabus Contents

Unit-1 15 hrs.

Data condensation and Graphical methods–
Raw data, attributes and variables, discrete and continuous variables.
Presentation of data using frequency distribution and cumulative frequency distribution. (Construction of frequency is not expected)
Graphical Presentation of frequency distribution – histogram, stem and leaf chart, less than and more than type o give curves.
Numerical problems related to real life situations.

Descriptive Statistics - Measures of Central tendency: Mean, Mode and Median. Examples where each one of these is most appropriate.
Partition values: Quartiles, Box-Plot.

Measures of Dispersion: Range, Coefficient of range, Quartile deviation, Coefficient of quartile deviation, Variance, Standard Deviation, Coefficient of Variation. (for raw data, ungrouped frequency distribution and exclusive type grouped frequency distribution)

Moments
Raw and Central moments: definition, computations for ungrouped and grouped data (only up to first four moments).Relation between raw and central moments up to fourth order. Numerical problems related to real life situations.

Unit-2 15 hrs.

Skewness, Kurtosis and its measures.
Concept of symmetric frequency distribution, skewness, positive and negative skewness. Measures of skewness- Pearson's measure, Bowley's measure, β_1, γ_1.

Correlation and Linear Regression (for bivariate raw data)
Bivariate data, Scatter diagram, Correlation, Positive Correlation, Negative Correlation, Zero Correlation. Karl Pearson's coefficient of correlation (r), limits of r ($-1 \leq r \leq 1$), interpretation of r, Coefficient of determination (R^2), Auto-correlation coefficient up to lag 2.Meaning of regression. Fitting of line $y = a+bx$ using least square method. Concept of residual plot and mean residual sum of squares. Numerical Problems.

Reference Books:

1. Fundamental of mathematical statistics: Gupta &Kapoor
2. Statistical Methods: S.P.Gupta, Sultan Chand and Sons Publishers
3. Business Statistics: S.L. Agarwal , Kalyani Publishers
4.Introduction to Statistics Methods : C.B. Gupta and Vijay Gupta, Vikas Publication
5.Business Statistics :G.V Kumbhojkar, PhadkePrakashan

BDST24- 105: Probability for Data Science

Theory-30 Hours Credits: 02

Objectives:

The main objectives of this course are:

1. Data Science is a fast-growing interdisciplinary field, focusing on the analysis of data to extract knowledge and insight.
2. This course will introduce students to the collection. Preparation, analysis, modeling and visualization of data, covering both conceptual and practical issues.
3. Examples and case studies from diverse fields will be presented and hands-on use of statistical and data manipulation software will be included.

Syllabus Contents

Unit-1 **15 hrs.**

Basic Theory of Probability
Counting Principles, Permutations and Combinations, Deterministic and non-deterministic models.
Random Experiment, Sample Spaces (finite and countably infinite)
Events: types of events, Operations on events.
Probability-classical definition, probability models, axioms of probability, probability of an event.

Theorems of probability(with proof)
i) $0 \leq P(A) \leq 1$
ii) $P(A)+P(A') = 1$
iii) $P(A) \leq P(B)$ when $A \subset B$
iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Numerical problems related to real life situations.

Basic Notion of Probability Theory
Concepts and definitions of conditional probability, multiplication theorem $P(A \cap B) = P(A).P(B|A)$
Bayes' theorem(with proof)
Concept of Posterior probability, problems on posterior probability.
Definition of sensitivity of a procedure, specificity of a procedure. Application of Bayes' theorem to design a procedure for false positive and false negative. Concept and definition of independence of two events. Numerical problems related to real life situations.

Unit-2 **15 hrs.**

Discrete Random variable and Standard Discrete Distributions- Definition of a random variable

and discrete random variable.
Definition of probability mass function, distribution function and its properties. Definition of expectation and variance, theorem on expectation. Numerical problems related to real life situations.

Discrete Uniform Distribution: definition, mean, variance. Bernoulli Distribution: definition, mean, variance, additive property, real life applications.
Binomial Distribution: definition, mean, variance, additive property, real life applications.
Geometric Distribution(p.m.fp(x)=p q^x,x=0,1,2.):
definition, mean, variance, real life applications.
Poisson Distribution: definition, mean, variance, mode, additive property, limiting case of B (n, p) Illustration of real-life situations. Numerical problems related to real life situations.

Recommended Books:

1. Fundamental of mathematical statistics: Gupta & Kapoor
2. Probability & Statistics: T.K.V. Iyengar, S. Chand & Company Ltd.
3. Rohatgi V. K. and Sauh A. K. Md E. (2002) An Introduction to probability and statistics (John Wiley & Sons-Asia). 25
4. Meyer P. L. (1970): Introductory to probability and statistical Application. Addision wesly.

BDSP24-106: Practical -I

Practical - 60 HoursCredits:02

Objectives:

The main objectives of this course are:
1. To represent statistical data.
2. To compute various measures of central tendency, dispersion, moments, Skewness and kurtosis.
3. To compute probability of various events.
4. To compute correlation coefficient for bivariate data and interpret it.
5. Predict value of dependent variable with the help of independent variable.

List of Practical's:

1. Diagrammatic & Graphical representation of the frequency distribution
2. Measures of Central Tendency
3. Measures of Dispersion
4. Moments, Skewness and Kurtosis
5. Correlation coefficient
6. Regression
7. Applications of Probability
8. Applications on Bayes' theorem.

9. Applications on Independence Probability
10. Applications of Binomial, Geometric and Poisson Distributions.

Learning Outcomes:

1) Students are able to draw diagram and graphs based on frequency distribution

2) Students are understand how to summarized data and find averages as well as spread of the data from central value (average).

3) Students get the knowledge about to compute moments and find out symmetry and skew symmetry of data

4) Students are become to find the probabilities of events and conditional probabilities.

Notes:

i) Students must complete all the practices to the satisfaction of the concerned teacher.

ii) Students must produce laboratory journal along with completion certificate signed by Head of the Department at the time of practical examination.

iii) Knowledge of MS-Excel / R Software should be tested on computer at the time of viva-voce.

Laboratory Requirement:

Laboratory should be well equipped with sufficient number of scientific calculators and computers along with necessary software's, UPS, and printers.

.

B.Sc. Data Science Part-I Semester- I
Mathematics Paper-I
Course Code-BDST24-107
Course Title: Linear Algebra
Total Hours: 30 hrs (38 lectures of48 min)
Credits: 02
Total Marks : 40+10=50

Course Outcomes:

After completion of this course students will be able to:

1. To understand the concept of matrices and solution of system of linear equations.
2. To understand the concept of vector space.
3. To understand the concept of Eigen values and Eigen vectors.
4. To understand the concept of Linear transformations.

Unit1:Matrices and System of Linear Equations (8)

Matrices, Types of matrices, Algebra of matrices, inverse of matrix, Matrix Transformation, Rank of matrix, System of Linear equations, Results on system of linear equations and (Statements Only),Solutions of System of Linear Equations: Gauss Elimination Method, Gauss Jordan method and examples,

Unit 2: Vector Space (7)

Vector Space: definition and examples, Basic properties of vector space,
Sub Space: definition and examples, Linear Dependence and Independence,
Linear combination and spanning, Basis and Dimension,
General inner product space: definition and examples

Unit3:Eigenvalues,Eigenvectors (7)

Eigen values and Eigen Vectors: Definition and examples.
Cayley-Hamiltonian theorem(Statement only) and examples,
Diagonalization: Definition and examples.

Unit4:Linear transformation (8)

Linear transformation: Definitions and examples
The Kernel and Range of a linear transformation, examples
The Matrix of a Linear Transformation and examples.

Reference Books:

1) A textbook of Linear Algebra, NiraliPrakashan.
2) A textbook of Linear Algebra, Vision Publication.
3) Elementary Linear Algebra with Applications, Howard Anton, Chris Rorres, John Wiley and sons., 7th Edition (1994).
4) Linear Algebra Linear Algebra with applications, Howard Anton, Chris Rorres

B.Sc. Data Science Part-I Semester- I
Mathematics Paper-II
Course Code-BDST24-108
Course Title: Numerical Methods
Total Contact Hours: 30 hrs (38 lectures of48 min)
Credits: 02
Total Marks: 40+10=50

Course Outcomes:

After completion of this course students will be able to:

1. To understand concept of non- linear equation and able to find its solution.
2. To understand the mathematical foundation of numerical interpolation.
3. Understand numerical differentiation and integration.
4. Grasping the basic of numerical method with application to differential equation.

Unit 1: Solution of Non –linear Equations (8)

Introduction, Bisection method : Algorithm ,graphical representation and examples
Regula – Falsi method: Algorithm ,graphical representation and examples
Newton Raphson method: Algorithm ,graphical representation and examples

Unit 2: Numerical Interpolation (7)

Introduction, Definitions of Difference operators∇, Δ, E, Basic results on Difference operators,
Newton - Gregory Forward &Backward Interpolation Formula and examples,
Lagrange's interpolation formula and examples

Unit 3: Numerical differentiation and Integration (7)

Introduction, Numerical differentiation using Forward interpolation formula and examples,
Numerical differentiation using Backward interpolation formula and examples,
Numerical Integration, General quadrature formula (without proof), Trapezoidal rule, Simpson's $1/3^{rd}$rule, Simpson's $3/8^{th}$ rule, Weddle's rule(without proof) & examples
Romberg method & examples

Unit 4: Numerical solution of first order ordinary differential equations (8)

Introduction, Euler's method and examples, Modified Euler's method and examples, Runge – Kutta 2^{nd} order method and examples, Runge – Kutta 4^{th} order method and examples

Reference Books:

1. A textbook of Numerical Methods, Nirali Prakashan.
2. Introductory Methods of Numerical Analysis, S.S. Sastry, 3rd edition, Prentice Hall of India, 1999.
3. Finite differences and Numerical Analysis, H.C. Saxena, S. Chand and Company.
4. Numerical Analysis, Balguruswamy.
5. Calculus of Finite Differences and Numerical Analysis, P. P. Gupta,

Mathematics Practical- I

1. Gaussian Elimination Method.
2. Gauss Jordan Method.
3. Eigen values and Eigen vectors of a matrix
4. Examples on verification of Cayley Hamilton theorem.
5. Diagonalization of matrix
6. Newton - Gregory Forward & Backward Interpolation
7. Lagrange's interpolation formula and examples
8. Numerical differentiation using Forward & Backward interpolation
9. Weddel's Rule
10. Romberg method
11. Euler method & Euler modified method
12. Runge – Kutta method

SEM II

Course Code-BDST24-201
Course Title: Problem Solving and Python Programming

Total Hours: 30 hrs (38 lectures of48 min)

Credits:02
Total Marks: 40+10=50

Unit-I Introduction to Computing and Problem Solving: (9)

Fundamentals of Computing – Computing Devices – Identification of Computational Problems – Pseudo Code and Flowcharts – Instructions – Algorithms – Building Blocks of Algorithms.

Introduction to Python Programming: Python Interpreter and Interactive Mode– Variables and Identifiers – Arithmetic Operators – Values and Types – Statements, Reading Input, Print Output, Type Conversions, The type() Function and Is Operator, Dynamic and Strongly Typed Language.

Control Flow Statements: The if, The if…else, The if…elif…else Decision Control Statements, Nested if Statement, The while Loop, The for Loop, The continue and break Statements.

Unit-II (8)

Functions: Built-In Functions, Commonly Used Modules, Function Definition and Calling the Function, The return Statement and void Function, Scope and Lifetime of Variables, Default Parameters, Keyword Arguments, *args and **kwargs, Command Line Arguments. Strings: Creating and Storing Strings, Basic String Operations, Accessing Characters in String by Index Number, String Slicing and Joining, String Methods, Formatting Strings.

Unit-III (7)

Lists: list operations, list slices, list methods, list loop, mutability, aliasing, cloning lists, list parameters; Tuples: tuple assignment, tuple as return value; Dictionaries: operations and methods; advanced list processing - list comprehension; Illustrative programs: selection sort, insertion sort, merge sort, histogram.

Files and exception: text files, reading and writing files, format operator; command line arguments, errors and exceptions, handling exceptions, modules, packages; Illustrative programs: word count, copy file.

Unit-IV (6)

Object-Oriented Programming: Classes and Objects, Creating Classes in Python, Creating Objects in Python, The Constructor Method, Classes with Multiple Objects, Class Attributes versus Data Attributes, Encapsulation, Inheritance The Polymorphism.

Functional Programming: Lambda. Iterators, Generators, List Comprehensions.

Course Code-BDST24-202
Course Title: Database Management System

Total Hours: 30 hrs (38 lectures of48 min)
Credits:02
Total Marks:40+10=50

UNIT I: (08)

Database System Applications, Purpose of Database Systems, View of Data – Data Abstraction – Instances and Schemas – data Models – the ER Model – Relational Model – Other Models – Database Languages – DDL – DML – database Access for applications Programs – data base Usersand Administrator – Transaction Management – data base Architecture – Storage Manager – the Query Processor Data base design and ER diagrams – ER Model - Entities, Attributes and Entity sets – Relationships and Relationship sets – ER Design Issues – Concept Design – Conceptual Design for University Enterprise. Introduction to the Relational Model – Structure – Database Schema, Keys – Schema Diagrams

UNIT II: (07)

Relational Query Languages, Relational Operations. Relational Algebra – Selection and projection set operations – renaming – Joins – Division –Examples of Algebra overviews – Relational calculus – Tuple relational Calculus – Domain relational calculus. Overview of the SQL Query Language – Basic Structure of SQL Queries, Set Operations, Aggregate Functions – GROUPBY – HAVING, Nested Sub queries, Views, Triggers.

UNIT III: (07)

Normalization – Introduction, Non loss decomposition and functional dependencies, First, Second, and third normal forms – dependency preservation, Boyee/Codd normal form. Higher

Normal Forms -Introduction, Multi-valued dependencies and Fourth normal form, Join dependencies and Fifth normal form

UNIT IV: (08)

Transaction Concept- Transaction State- Implementation of Atomicity and Durability – Concurrent –Executions – Serializability- Recoverability – Implementation of Isolation – Testing forserializability- Lock –Based Protocols – Timestamp Based Protocols- Validation- Based Protocols –Multiple Granularity.Recovery and Atomicity – Log – Based Recovery – Recovery with Concurrent Transactions – BufferManagement – Failure with loss of nonvolatile storage- Advance Recovery systems- Remote Backup systems.File organization:– File organization – various kinds of indexes. Query Processing – Measures of query cost - Selection operation – Projection operation, - Join operation – set operation and aggregate operation – Relational Query Optimization – Transacting SQL queries – Estimating the cost –Equivalence Rules.

Course Code-BDST24-203
Course Title: Practical IV

Credits:02
Lab on DBMS and Problem Solving and Python Programming

DBMS LAB

1.Create table Student, Teacher, Book_dtls, Product andperform

all DDL and DML Commands.

2.Perform calculations on above created tables Condition

specification using Boolean and comparison operators (and,

or, not, =, <>,>, <,>=, <=)

3.Use Aggregate functions

4.Use String handling functions.

5.Create table and apply all constraints.

a. Create tables with relevant foreign key constraints.

b. Populate the tables with data.

6.Display all the details of all employees working in the

company and solve the following queries.

a. Display ssn, lname, fname, address of employees who working department no 7.

b. Retrieve the birthdate and address of the employee whose name is 'Franklin T. Wong.'

c. Retrieve the name and salary of every employee.

d. Retrieve all distinct salary values.

e. Retrieve all employee names whose address is in 'Bellaire.'

f. Retrieve all employees who were born during the 1950s.

g. Retrieve all employees in department 5 whose salary is between 50,000 and 60,000

Python Lab

1. Write a program to demonstrate basic data type in python

2. Create a list and perform the following methods

1) insert( ) 2) remove( ) 3) append( ) 4) len( ) 5) pop( ) 6) clear( )

3. Create a tuple and perform the following methods

1) Add items 2) len( ) 3) check for item in tuple 4)Access items

4. Create a dictionary and apply the following methods

1) Print the dictionary items 2) access items 3) use get( ) 4) change values 5) uselen( )

5. Write a program to create a menu with the following options

1. TO PERFORM ADDITITON 2. TO PERFORM SUBTRACTION
3. TO PERFORM MULTIPICATION 4. TO PERFORM DIVISION

Accepts users input and perform the operation accordingly. Use functions with arguments.

6. Write a python program to print a number is positive/negative using if-else.

7. Write a program for filter() to filter only even numbers from a given list.

8. Write a python program to print date, time for today and now

9. Write a python program to add some days to your present date and print the date added.

10. Write a program to count the numbers of characters in the string and store them in a dictionary data structure

11. Write a program to count frequency of characters in a given file.

B.Sc.I(DS) (Semester-II)

BDST24- 204: Descriptive Statistics - II

Theory-30 Hours Credits:02

Objectives: The main objectives of this course are:

1. To understand the concept of sampling techniques.
2. Know the concept and use of time series
3. Solve the examples on Index Numbers.
4. The meaning, purpose and use of Statistical Quality, Construction and working of control charts for variables and attributes.

Syllabus Contents

Unit-1 15 hrs.

Sampling Techniques

Introduction: Concept of distinguishable elementary units, sampling units, sampling frame, random sampling and non-random sampling. Advantages of sampling method over census method, objectives of a sample survey, Designing a questionnaire, Characteristics of a goodquestionnaire,Conceptofsamplingandnon-samplingerrors.Handling of non response cases. Simple random sampling: i)Simple random sampling from finite population of size N with replacement(SRSWR) and without replacement(SRSWOR), Stratified random sampling.

Time Series Analysis

Meaning and Need of Time Series, Use of Analysis of Time Series. Components of Time Series, Determination of Trend by i) Graphic method, ii) Method of Moving Averages, Numerical Examples.

Unit-2 15 hrs.

Index Number

Meaning and Need of Index Numbers, Use of Index Numbers, Problems in Construction of Index Numbers, Types of Index Numbers: Price, Quantity and Value Index Numbers, Methods of Constructing Index Numbers.

- Unweighted Index Numbers: By Simple Aggregate Method and Simple Average of Relative Method.

- Weighted Index Numbers: By Laspeyre's, Paasche's, Fisher's Price and Quantity Index Numbers.
  Value Index Numbers, Numerical Examples.

Statistical Quality Control (S.Q.C.)

Concept and Advantages of S.Q.C., Causes of Variation. Types of Control Process Control and Product Control, Construction of Control Chart.

Types of Control Chart: i) Variables ii) Attributes. Control charts for variables: Mean and Range Charts, Control charts for Attributes: Control chart for number of defectives (np-chart) for fixed sample size, Control chart for number of defects per unit (C-Chart).

Recommended Books:

1. Cochran, W.G: Sampling Techniques, Wiley Eastern Ltd.,New Delhi
2. Chatfield C: "The Analysis of Time Series –An Introduction", Chapman& Hall.
3. Gupta S. C. & Kapoor V.K.: Applied Statistics. Sultan Chand & sons, New Delhi.
4. Kendall M.G. : "Time Series", Charles Griffin.
5. Dr. Kore B.G. and Dr. Dixit P.G.: "Statistical Methods-II", NiraliPrakashan, Pune.
6. Gupta S. C. &Kapoor V.K.: Fundamentals of Mathematical Statistics. Sultan Chand & sons, New Delhi.

BDST24-205: Data Science with R

Theory-30 Hours Credits:02

Objectives:

The students will acquire knowledge of

i) To understand programming fundamentals of R language
ii) To understand various data import methods in R
iii) To understand data manipulation in R
iv) To create visualizations and plots using R

Syllabus Contents

Unit-1 **15 hrs.**

Basics of R-Programming: Evolution of R, Features of R, Local Environment support, R Command prompt, R Script File, Comment, R Data types, R Variables, R Operators-function.

Understanding data: Introduction – Types of Data: Numeric – Categorical – Graphical – High Dimensional Data- Example Applications. Sources of Data: Time Series–Biological Data.

R Fundamentals: Introduction to R- Features of R - Environment - R Studio. R-Decision Making:- R-If statement, R-If….else statement, R- The if….else if…else statement-Switch Statement, R- Loop:- Repeat loop, While loop, for loop, Loop ,Control statement:- Break, Next. Basics of R-Assignment -Modes-Operators-special numbers- Logical values– Basic Functions - R help functions - R Data Structures - Control Structures.Vectors:Definition-Declaration-Generating-Indexing-Naming-Adding&Removing elements - Operations on Vectors.

Data Structures in R: Matrices - Creating Matrices - Adding or Removing rows/columns - Reshaping - Operations - Special functions on Matrices. Lists - Creating List– General List Operations – Special Functions - Recursive Lists. Data Frames - Creating Data Frames - Naming - Accessing -Adding-Removing-Applying Special functions to Data Frames –Merging Data Frames Factors and Tables.

Unit-2 15 hrs.

Statistics in R: central tendency – mean – median – mode – Dispersion – variance-standard deviation-coefficient of variation – moments – skewness – kurtosis – correlation – Regression – Probability-uniform distribution-binomial distribution-poisson distribution – mean - variance – model sampling – fitting of all these distributions

Application of R-programming: R- Pie charts: - Pie chart title and colour, 3-D Pie Chart. R- Bar Chart: - Bar Chart Labels, Title and colour, Group Bar chart and stacked bar chart. R-Box Plot: - Creating a box plot, Box plot with notch. R- Histogram: - Range of x and y values.

Reference Books

1. Statistics Using R – Sudha G. Purohit, Sharad Gore, Shailaja Deshmukh.

2. V.R. Pawagi and Saroj A. Ranade (2010) - Statistical methods using R software, Nirali. Publications.

3. Mark Gardener,“ Beginning R-The Statistical Programming Language”, John Wiley & Sons, Inc., 2012.

4. NinaZumel,JohnMount,“PracticalDataSciencewithR”,ManningPublications,2014.

BDSP24-206 : Practical - II

Practical - 60 HoursCredits:02

Objectives:

The main objectives of this course are:

i) To solve problems on time series, index number ,SQC using MS-Excel.

ii)The programming fundamentals of R language
iii) The understand various data import methods in R
iv) The understand and statistical methods using R.
v) To solve problems on Probability and probability distribution.

List of Practical's:

1. Time Series
2. Index Number
3. Construction of mean and range chart
4. Construction of np and C chart
5. Data Input
6. Diagrammatic and Graphical Representation
7. Measures of Central Tendencyusing R programming
8. Measures of Dispersionusing R programming
9. Correlation and Regression using R programming
10. fitting of uniform, binomial, and poisson distribution using R programming

Learning Outcomes:

1) The Concept and use of time series analysis.

2) The meaning, purpose and use of Statistical Quality, Construction andworking of control charts for variables and attributes

3) To understand programming fundamentals of R language

4) To understand various data import methods in R

5) The understand and implement basic Statistics and Various statistical methods

Notes:

i) Students must complete all the practices to the satisfaction of the concerned teacher.

ii) Students must produce laboratory journal along with completion certificate signed by Head of the Department at the time of practical examination.

iii) Knowledge of MS-Excel / R Software should be tested on computer at the time of viva-voce.

Laboratory Requirements:
Laboratory should be well equipped computers along with R software, UPS, and Printers.

Mathematics Paper-III

B.Sc. Data Science Part-I Semester- II

Course Title: Graph Theory
Total Hours:48 hrs (38 lectures of 48 min)

Credits:02

TotalMarks:40+10=50

Course Outcomes:

After completion of this course students will be able to:

1. To understand the concepts of Graph and able to perform operations on graphs with an examples.
2. To understand the concepts of connected graphs and solve the problems using algorithms.
3. To understand the concepts of Tree graphs with an examples
4. To understand the concepts of directed graphs with an examples

Unit 1: Graphs and operations on graphs (10)

Graph: Definition and elementary results, Types of graphs, Matrix representation of graphs: Adjacency matrix and incidence matrix, Isomorphism, Sub graphs and induced graphs, Complement of a graph, Self complementary graphs, Union, intersection of graphs, Ring sum of two graphs.

Unit 2: Connected graphs (08)

Definitions: walk, trail, tour, path and circuit, Definitions of connected, disconnected graphs,
Dijkstra's shortest path algorithm, Definition of Euler's and Hamiltan Graph and Example.

Unit 3: Tree Graphs (08)

Tree: Definition, Theorem: A tree with n vertices has n -1 edges.
Theorem: A Connected graph G with n vertices and n – 1 edges is a tree,
Theorem: A graph with n Vertices is a tree if and only if it is circuit free and has n – 1 edges,

Theorem : A graph G is a tree if and only if it is minimally connected, Center of a tree, Spanning tree: Definition and examples, Fundamental circuit and cut – set: Definition and examples, Binary trees and elementary results, Kruskal's algorithm.

Unit 4: Directed Graphs (10)

Definition, types of directed graphs, Directed (rooted) trees, arborescence and Polish notation,

Isomorphism of digraphs, Connectedness in digraphs, Euler digraph, Network and flows:

Definition and examples, Maximal flow algorithm.

Reference Books:

1. A Text book of Discrete mathematics by S.R.Patil and others, NIRALI Prakashan
2. Elements of Discrete Mathematics by C.L. Liu.
3. Discrete Mathematics by Olympia Nicodemi
4. Discrete Mathematical Structure for Computer Science by Alan Doer and K.Levasicur.
5. Discrete and Combinatorial Mathematics by R.m. Grassl
6. Discrete Mathematics by Kenneth Rosen, Tata McGraw Hill
7. Graph Theory with Applications to Computer Science and Engineering by
8. A First Step in Graph Theory by Raghunathan, Nimkar and Solapurrkar
9. Discrete mathematics by Bhopatkar, Nimbkar, Joglekar, VISION Publication.
10. Discrete mathematics by Naik and Patil, PHADAKE Prakashan

Mathematics Paper-IV

B.Sc. Data Science Part-I Semester- II

Course Title: Operation Research

TotalContactHours:48 hrs (60lectures of48 min)

Credits:02 TeachingScheme:Theory:02Lect./WeekTotalMarks:40+10=50

Course Outcomes:

After completion of this course students will be able to:

1. Describe the history of Operation research.
2. Student be able to formulate L.P.P.
3. Formulation of a Transportation problem.
4. Understand the concept of game.

Unit 1: Introduction to operation Research (5)

Basics of operation research, Different definitions of operation research, Characteristics, scope, limitations of operation research

Unit 2: Linear Programming Problem (12)

Basic definitions, Feasible solution, Basic solution, Optimal solution, Solution of L.P.P by Simplex method &examples, Solution of L.P.P by Big-M method &examples

Unit 3: Transportation and Assignment Problems (15)

Introduction to Transportation Problem, Initial solution: North-West corner rule, Matrix-Minima method, Vogel approximation method, Optimal solution by MODI method, Maximization in Transportation Problem, Unbalanced Transportation Problem,
Introduction to Assignment problem, Hungarian Method & examples, Maximization in assignment
problem, Assignment problem with restrictions, Unbalanced assignment problem.

Unit 4: Theory of games (13)

Introduction, Two person zero sum game, Pure and mixed strategies, statement of min - max and max - min theorem, Saddle point, Solution of 2x2 game by arithmetic and algebraic methods, examples, Dominance method: Principle of dominance and examples, Sub game method and examples.

Reference Books:

1. A text book of Operations Research, M. D. Bhagat and others, NiraliPrakashan.
2. Operations Research, S. D. Sharma.
3. Principles of Operations Research, H. M. Wagner, Prentice Hall of India.
4. Operations Research, Gupta and Hira.
5. Operations Research, J K Sharma (second edition.)

Mathematics Practical- II

1. Operations on graph
2. Dijkstra's Shortest path algorithm with examples
3. Fundamental circuit and fundamental cut set with example
4. Kruskal's algorithm with examples
5. Polish notations with examples
6. Maximal flow algorithm
7. Linear Programming Problem(LPP)
8. Transportation Problem: Initial Solution
9. Transportation Problem: MODI Method
10. Assignment Problem
11. Game Theory: Arithmetic and Algebraic methods
12. Game Theory: Dominance method and Sub game method
13.