

**SRI VENKATESWARA UNIVERSITY**  
**B.Sc. (DATA SCIENCE) – W.E.F. 2023-24**  
**SEMESTER – II**  
**(MAJOR)**

Sl.No.	Course	Name of the Subject	Total Marks	Internal Exam	Sem. End Exam	Teaching Hours	Credits
1.	First Language	English	100	25	75	4	3
2.	Second Language	(Telugu / Hindi / Sanskrit / Tamil / Urdu)	100	25	75	4	3
3.		Skill Enhancement Course-1	50	---	50	2	2
4.		Skill Enhancement Course-2	50	---	50	2	2
5.	Major - Course 3	Introduction to Data science and R Programming (T)	100	25	75	3	3
6.	Major - Course 3	Introduction to Data science and R Programming (P)	50	---	50	2	1
7.	Major - Course 4	Basic statistics for Data Science (T)	100	25	75	3	3
8.	Major - Course 4	Basic statistics for Data Science (P)	50	---	50	2	1
9.		Minor	100	25	75	3	3
10.		Courses with Practical s	50	---	50	2	1

**SRI VENKATESWARA UNIVERSITY**  
**B.Sc. (DATA SCIENCE) – W.E.F. 2023-24**

**SEMESTER – II**

**COURSE STRUCTURE**

**(MAJOR)**

<b>Year</b>	<b>Semester</b>	<b>Paper</b>	<b>Title of the Course</b>	<b>No. of Hrs./ Week</b>	<b>No. of Credits</b>
<b>1</b>	<b>II</b>	<b>3</b>	<b>Introduction to Data science and R Programming (T)</b>	<b>3</b>	<b>3</b>
			<b>Introduction to Data science and R Programming (P)</b>	<b>2</b>	<b>1</b>
		<b>4</b>	<b>Basic statistics for Data Science (T)</b>	<b>3</b>	<b>3</b>
			<b>Basic statistics for Data Science (P)</b>	<b>2</b>	<b>1</b>

**SRI VENKATESWARA UNIVERSITY**  
**B.Sc. (DATA SCIENCE) – W.E.F. 2023-24**

**SEMESTER – II**

**COURSE 3: INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING**

**Theory**

**Credits: 3**

**3 hrs/week**

**Aim and objectives of Course :**

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

**Learning outcomes of Course:**

- Recognize the various disciplines that contribute to a successful data science effort.
- Understand the processes of data science: identifying the problem to be solved, data collection, preparation, modeling, evaluation and visualization.
- Be aware of the challenges that arise in Data Sciences.
- Be able to identify the application of the type of algorithm based on the type of the problem.
- Be comfortable using commercial and open source tools such as the R/Python language and its associated libraries for data analytics and Visualization.

**UNIT I:**

Defining Data Science and Big data, Benefits and Uses, facets of Data, Data Science Process.

History And Overview of R, Getting Started with R, R Nuts and Bolts

**UNIT II:**

The Data Science Process: Overview of the Data Science Process-Setting the research goal,

Retrieving Data, Data Preparation, Exploration, Modeling, data Presentation and Automation. Getting Data in and out of R, Using reader package, Interfaces to the outside world.

**UNIT III:**

Machine Learning: Understanding why data scientists use machine learning-What is machine

learning and why we should care about, Applications of machine learning in data science, Where it is used in data science, The modeling process, Types of Machine Learning- Supervised and Unsupervised.

**UNIT IV:**

Handling large Data on a Single Computer: The problems we face when handling large data,

General Techniques for handling large volumes of data, Generating programming tips for dealing with large datasets.

**UNIT V:**

Sub setting R objects, Vectorised Operations, Managing Data Frames with the dplyr, Control

structures, functions, Scoping rules of R, Coding Standards in R, Loop Functions, Debugging, Simulation. Case studies on preliminary data analysis.



## TEXT BOOKS:

1. DavyCielen, Arno.D.B.Maysman, Mohamed Ali, "Introducing Data Science" ManningPublications, 2016.
2. Roger D. Peng, "R Programming for DataScience" Lean Publishing, 2015.

## REFERENCE BOOKS:

1. Nina Zumel, John Mount, "Practical Data Science with R", Manning Publications, 2014.
2. Tony Ojeda, Sean Patrick Murphy, Benjamin Bengfort, AbhijitDasgupta, "PracticalData Science Cookbook", Packt Publishing Ltd., 2014. WebReferences for case studies:
  - 1.<https://www.kaggle.com/datasets>
  - 2.<https://github.com/>

## RECOMMENDED TEXT BOOKS:

1. Mark Gardener, "Beginning R - The Statistical Programming Language", John Wiley & Sons, Inc., 2012.
2. W. N. Venables, D. M. Smith and the R Core Team, "An Introduction to R", 2013.  
Recommended Reference books:
3. The art of R Programming: A tour of Statistical Software design. Norman Matloff. Kindle Edition
4. The book of R : The first course in Programming and Statistics by Tilman M. Davies.

**Recommended Co-curricular activities:** (Co-curricular Activities should not promote copying from text book or from others' work and shall encourage self/independent and group learning)

### A. Measurable:

1. Assignments on:
2. Student seminars (Individual presentation of papers) on topics relating to:
3. Quiz Programmes on:
4. Individual Field Studies/projects:
5. Group discussion on:
6. Group/Team Projects on:

### B. General

1. Collection of news reports and maintaining a record of paper-cuttings relating to topics covered in syllabus
2. Group Discussions on:
3. Watching TV discussions and preparing summary points recording personal observations etc., under guidance from the Lecturers
4. Any similar activities with imaginative thinking.
5. Recommended Continuous Assessment methods:

**SRI VENKATESWARA UNIVERSITY**  
**B.Sc. (DATA SCIENCE) – W.E.F. 2023-24**  
**SEMESTER – II**  
**COURSE 3: PROBLEM SOLVING USING C**

**Practical s**

**Credits: 1**

**2 hrs/week**

---

**Lab/Practical/Experiments/Tutorials syllabus:**

1. Installing R and R studio, with proper notes on version management, cosmetic settings and different libraries.
2. Basic operations in R with arithmetic and statistics.
3. Getting data into R, Basic data manipulation, Loading Data into R
4. Basic plotting
5. Loops and functions
6. Create Vectors, Lists, Arrays, Matrices, Data frames and operations on them.
7. Demonstrate the visualization and graphics using visualization packages like ggplot2.
8. Implement Loop functions with lapply(), sapply(), tapply(), apply(), mapply().
9. Explore data using Single Variables: Unimodal, Bimodal, Histograms, Density Plots, Bar charts
10. Explore data using two Variables: Line plots, Scatter Plots, smoothing curves, Bar charts
11. Explore and implement commands using dplyr package
12. Download a dataset and work on basic data manipulation followed by inferential statistics.

**SRI VENKATESWARA UNIVERSITY**  
**B.Sc. (DATA SCIENCE) – W.E.F. 2023-24**

**SEMESTER – II**

**COURSE 3: INTRODUCTION TO DATA SCIENCE AND R PROGRAMMING**

**MODEL QUESTION PAPER**

Time: **3 Hours**

Max. Marks: **75**

**PART-A**

Answer any **FIVE** of the following. Each Question Carries 5 marks.

**(5X5=25)**

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.

**PART-B**

Answer any **FIVE** of the following. Each Question Carries 10 marks.

**(5X10=50)**

- 9.
- 10.
- 11.
- 12.
- 13.
- 14.
- 15.
- 16.
- 17.
- 18.

**SRI VENKATESWARA UNIVERSITY**  
**B.Sc. (DATA SCIENCE) – W.E.F. 2023-24**  
**SEMESTER – II**  
**COURSE 4: BASIC STATISTICS FOR DATA SCIENCE**

**Theory**

**Credits: 3**

**3 hrs/week**

**Course Learning Outcomes: Students will acquire:**

- knowledge of Statistics and its implementation through practical understanding for various domains related to data science.
- knowledge of various types of data, their organization and evaluation of summary measures such as measures of central tendency and dispersion etc.
- Knowledge of other types of data reflecting quality characteristics including concepts of independence and association between two attributes, insights into preliminary exploration of different types of data.
- Knowledge of correlation, regression analysis, regression diagnostics, partial and multiple correlations.

**UNIT I:**

**Introduction to Statistics:** Importance of Statistics. Scope of Statistics in different fields. Concepts of primary and secondary data. Diagrammatic and graphical representation of data: Histogram, frequency polygon, Pie. Measures of Central Tendency: Mean, Median, Mode, Geometric Mean and Harmonic Mean. Median and Mode through graph.

**UNIT II:**

**Measures of Dispersion:** Range, Quartile Deviation, Mean Deviation and Standard Deviation, Variance. Central and Non-Central moments and their interrelationship. Skewness and kurtosis.

**UNIT III:**

**Curve fitting:** Bi-variate data, Principle of least squares, fitting of degree polynomial. Fitting of straight line, Fitting of Second degree polynomial or parabola, Fitting of power curve and exponential curves.

**Correlation:** Meaning, Types of Correlation, Measures of Correlation: Scatter diagram, Karl Pearson's Coefficient of Correlation, Rank Correlation Coefficient (with and without ties), Bi-variate frequency distribution, correlation coefficient for bi-variate data and simple problems. Concept of multiple and partial correlation coefficients (three variables only) and properties

**UNIT IV:**

**Regression :** Concept of Regression, Linear Regression: Regression lines, Regression coefficients and its properties, Regressions lines for bi-variate data and simple problems. Correlation vs regression, sigmoid curve, derivation from linear regression to logistic regression.

**UNIT-V**

**Attributes :** Notations, Class, Order of class frequencies, Ultimate class frequencies, Consistency of data, Conditions for consistency of data for 2 and 3 attributes only, Independence of attributes, Association of attributes and its measures, Relationship between association and colligation of attributes, Contingency table: Square contingency, Mean square contingency, Coefficient of mean square contingency,



#### TEXT BOOKS:

1. V.K.Kapoor and S.C.Gupta: Fundamentals of Mathematical Statistics, Sultan Chand & Sons, NewDelhi.
2. BA/BSc I year statistics - descriptive statistics, probability distribution - Telugu Academy - Dr. M. Jaganmohan Rao, Dr N. Srinivasa Rao, Dr P. Tirupathi Rao, Smt.D.Vijayalakshmi.
3. K.V.S. Sarma: Statistics Made Simple: Do it yourself on PC. PHI

#### REFERENCE BOOKS:

1. Willam Feller: Introduction to Probabilitytheory and its applications. Volume –I,Wiley
2. Goon AM, Gupta MK, Das Gupta B : Fundamentals of Statistics , Vol-I, the World Press Pvt.Ltd.,Kolakota.
3. Hoel P.G: Introduction to mathematical statistics, Asia Publishing House.
4. M. JaganMohan Rao and Papa Rao: A Text book of StatisticsPaper-I.
5. Sanjay Arora and Bansi Lal: New Mathematical Statistics: Satya Prakashan ,NewDelhi

**Recommended Co-curricular activities:** (Co-curricular Activities should not promote copying from text book or from others' work and shall encourage self/independent and group learning)

#### A. Measurable:

1. Assignments on:
2. Student seminars (Individual presentation of papers) on topics relating to:
3. Quiz Programmes on:
4. Individual Field Studies/projects:
5. Group discussion on:
6. Group/Team Projects on:

#### B. General

1. Collection of news reports and maintaining a record of paper-cuttings relating to topics covered in syllabus
2. Group Discussions on:
3. Watching TV discussions and preparing summary points recording personal observations etc., under guidance from the Lecturers
4. Any similar activities with imaginative thinking.
5. Recommended Continuous Assessment methods:

**SRI VENKATESWARA UNIVERSITY**  
**B.Sc. (DATA SCIENCE) – W.E.F. 2023-24**  
**SEMESTER – II**  
**COURSE 4: BASIC STATISTICS FOR DATA SCIENCE**

**Practical s**

**Credits: 1**

**2 hrs/week**

---

List of the experiments:

1. Graphical presentation of data (Histogram, frequency polygon).
2. Diagrammatic presentation of data (Bar and Pie).
3. Computation of measures of central tendency (Mean, Median and Mode)
4. Computation of measures of dispersion (Q.D, M.D and S.D)
5. Computation of non-central, central moments, 1 and 2 for ungrouped data.
6. Computation of Karl Pearson's coefficients of Skewness and Bowley's coefficient of Skewness.
7. Fitting of straight line by the method of least squares
8. Fitting of parabola by the method of least squares
9. Fitting of power curve of the type by the method of least squares.
10. Fitting of the exponential curve of the type and by the method of least squares.
11. Computation of correlation coefficient and regression lines for ungrouped data.
12. Computation of correlation coefficient, forming regression lines for grouped data

**SRI VENKATESWARA UNIVERSITY**  
**B.Sc. (DATA SCIENCE) – W.E.F. 2023-24**  
**SEMESTER – II**  
**COURSE 4: BASIC STATISTICS FOR DATA SCIENCE**  
MODEL QUESTION PAPER  
Time: 3 Hours Max. Marks: 75

**PART-A**

Answer any **FIVE** of the following. Each Question Carries 5 marks.

**(5X5=25)**

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.
- 8.

**PART-B**

Answer any **FIVE** of the following. Each Question Carries 10 marks.

**(5X10=50)**

- 9.
- 10.
- 11.
- 12.
- 13.
- 14.
- 15.
- 16.
- 17.
- 18.