



SRI VENKATESWARA UNIVERSITY

I Year Syllabus for 4-Year UG Honours in B.Sc. (Data Science) as Major in consonance with Curriculum framework w.e.f. AY 2025-26

Year	Semester	Course	Title of the Course	No. of Hrs /Week	No. of Credits
I	I	1	Computer Fundamentals and Office Automation	3	3
			Computer Fundamentals and Office Automation Lab	2	1
		2	Problem Solving Using C	3	3
			Problem Solving Using C Lab	2	1
	II	3	Python Programming and Data Structures	3	3
			Python programming and Data Structures lab	2	1
		4	Statistical Foundations for Data Science	3	3
			Statistical Foundations for Data Science lab	2	1

SEMESTER-II
COURSE 4: STATISTICAL FOUNDATIONS FOR DATA SCIENCE

Theory

Credits: 3

3 hrs/week

Course Objectives

1. To introduce the fundamental concepts of probability and statistics for quantifying and analyzing uncertainty in real-world problems.
2. To develop an understanding of random variables, expectations, and common probability distributions (discrete and continuous).
3. To build the ability to summarize and describe data using measures of central tendency, dispersion, correlation, and visualization techniques.
4. To equip students with statistical tools for modeling relationships using correlation and regression analysis.
5. To provide knowledge of estimation and hypothesis testing for making valid inferences from sample data about populations.

Course Outcomes

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

1. **Apply** the basic rules of probability, conditional probability, and Bayes' theorem to solve problems involving uncertainty.
2. **Compute and interpret** descriptive statistics (mean, median, mode, variance, standard deviation, correlation, covariance) and represent data effectively using histograms, bar charts, and scatter plots.
3. **Analyze** random variables and probability distributions (Binomial, Poisson, Normal, Exponential, etc.) to model real-life situations.
4. **Perform** correlation and regression analysis to identify and interpret relationships between variables.
5. **Conduct** statistical inference through confidence intervals and hypothesis testing (z-test, t-test, chi-square, F-test) for decision-making.

Unit 1: Fundamentals of Probability & Basic Statistics

Probability: Concept of Uncertainty, Axioms and rules of probability, Conditional probability and independence, Law of total probability and Bayes' theorem.

Measures of central tendency: Mean, Median, Mode

Measures of dispersion: range, interquartile range, variance, standard deviation Introduction to correlation and covariance

Data representation: histograms, bar charts, scatter plots

Unit 2: Random Variables, Expectation, and Variance

Random variables: definition, types (discrete & continuous), and properties, Probability mass function (PMF) and probability density function (PDF), Cumulative distribution function (CDF), Mathematical expectation (mean), variance, and standard deviation, Moments and moment-generating functions

Unit 3: Probability Distributions

Discrete distributions: Binomial, Poisson, Geometric, Negative Binomial distributions - definitions, properties, and examples

Continuous distributions: Uniform, Normal (Gaussian), Exponential, Gamma distributions - definitions, properties, and applications

Joint, marginal, and conditional distributions, Introduction to Central Limit Theorem

Unit 4: Correlation and Regression Bivariate data and scatter plots

Correlation: Pearson and Spearman coefficients, interpretation

Simple linear regression: model, estimation, properties, and analysis of variance Multiple linear regression basics (conceptual understanding), Residuals and goodness of fit

Unit 5: Statistical Inference, Estimation, and Hypothesis Testing

Population and sample, parameters and statistics, Sampling distributions, Point and interval estimation (confidence intervals), Tests of significance: z-test, t-test, chi-square test, and F-test, p-values and errors (Type I & II), Power of a statistical test

Textbooks:

1. Probability and Statistics for Engineers and Scientists, Ronald E. Walpole, Wiley.
2. Sheldon M. Ross, Introduction to Probability and Statistics for Engineers and Scientists
3. Douglas C. Montgomery & George C. Runger, Applied Statistics and Probability for Engineers

Reference Books:

1. D.C. Agarwal, Statistics for Data Science and AI
2. Larry J. Stephens, Excel Data Analysis: Your visual blueprint for analyzing data, statistics, and AI

Activities:

Outcome: Apply probability rules, conditional probability, and Bayes' theorem

Activity: Classroom **Quiz** (MCQs & short problems on probability, conditional probability, Bayes).

Evaluation Method: Individual quiz marks (accuracy of solutions).

Outcome: Compute and interpret descriptive statistics & visualize data

Activity: Poster Presentation – students prepare posters with a dataset summary (mean, median, variance, histograms, scatter plot).

Evaluation Method: Rubric-based evaluation (clarity, correctness, creativity, interpretation).

Outcome: Analyze random variables and probability distributions

Activity: Seminar/Presentation – each student (or group) explains one distribution (Binomial, Poisson, Normal, Exponential, etc.) with real-life examples and graphs. **Evaluation Method:** Seminar grading (content accuracy, explanation, use of examples, presentation skills).

Outcome: Perform correlation and regression analysis

Activity: Case Study Assignment – students are given real-world data (e.g., sales vs advertising, study hours vs marks) to compute correlation/regression and interpret. **Evaluation Method:** Submission of case study report + viva (correctness of analysis and interpretation).

Outcome: Conduct statistical inference (estimation & hypothesis testing)

Activity: Group Discussion/Debate – teams test a given hypothesis (e.g., “Male and Female students score equally”) using sample data and statistical tests, then defend their conclusion. **Evaluation Method:** Marks based on correct application of test, interpretation of p-value, and clarity in argument.

SEMESTER-II
COURSE 4: STATISTICAL FOUNDATIONS FOR DATA SCIENCE

Practical

Credits: 1

2 hrs/week

Advanced Spreadsheets/Excel Lab/PSPP Open Source

1. Construct a contingency table from sales data and compute conditional probabilities. Verify independence of variables.
2. Apply Bayes' theorem on medical test data to compute the probability of disease given a positive result.
3. Calculate measures of central tendency (mean, median, mode) for student marks dataset.
4. Compute measures of dispersion (range, variance, standard deviation, IQR) for the same dataset and interpret variability.
5. Create a histogram of student marks and comment on the shape of the distribution.
6. Prepare bar charts of categorical data (e.g., Gender vs Section) and interpret group comparisons.
7. Generate scatter plots between Hours Studied and Exam Score, compute correlation and covariance, and interpret the relationship.
8. Random Variable Simulation: Simulate and visualize discrete/continuous random variables using Excel functions and Data Analysis Tool pack.
9. Expectation & Variance Calculation: Use Excel formulas to compute expected value, variance, and standard deviation from given datasets.
10. Modeling Discrete Probability Distributions: Generate and plot Binomial and Poisson distributions; analyze probabilities and mean/variance.
11. Modeling Continuous Distributions: Simulate Normal and Exponential distributions; use NORM.DIST, NORM.INV, EXPON.DIST functions.
12. Correlation Analysis: Calculate Pearson/Spearman correlation coefficients; visualize with scatter plots and trendlines.
13. Linear Regression in Excel: Fit a linear regression model, interpret coefficients, predict new values; use REGRESSION tool.
14. Statistical Inference & Estimation: Create confidence intervals using Excel formulas; visualize sampling distributions
15. Hypothesis Testing: Perform z-test, t-test, chi-square tests in Excel; interpret p-values and results.

SRI VENKATESWARA UNIVERSITY::TIRUPATI
B.Sc. Data Science Honours
I year II Semester
Course 4: STATISTICAL FOUNDATIONS FOR DATA SCIENCE

MODEL QUESTION PAPER

Time: **3 Hours**

Max. Marks: **75**

PART-A

Answer any **FIVE** of the following. Each Question Carries 3 marks.

(5X3=15)

1. Short answer question from unit-I
2. Short answer question from unit-I
3. Short answer question from unit-II
4. Short answer question from unit-II
5. Short answer question from unit-III
6. Short answer question from unit-III
7. Short answer question from unit-IV
8. Short answer question from unit-IV
9. Short answer question from unit-V
10. Short answer question from unit-V

PART-B

Answer any **FIVE** of the following. Each Question Carries 12 marks.

(5X12=60)

11. a) Essay question from Unit-I
OR
b) Essay question from Unit-I
12. a) Essay question from Unit-II
OR
b) Essay question from Unit-II
13. a) Essay question from Unit-III
OR
b) Essay question from Unit-III
14. a) Essay question from Unit-IV
OR
b) Essay question from Unit-IV
15. a) Essay question from Unit-V
OR
b) Essay question from Unit-V