

Nirma University
Institute of Technology, School of Technology
MTech Computer Science and Engineering (Data Science)
Semester – II

L	T	P	C
2	0	2	3

Course Code	3CS4201
Course Name	Exploratory Data Analysis

Course Learning Outcomes (CLOs):

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

1. comprehend the basic concepts of probability and statistics and their need in engineering
2. apply concepts and methods of probability and statistics in simulation and modeling of various computer science problems
3. perform probabilistic and statistical analysis of data related to computer science research and projects

Syllabus:

**Teaching
Hours**

Unit I

The Role of Statistics in Engineering: The Engineering Method and Statistical Thinking, Collecting Engineering Data, Mechanistic and Empirical Models, Probability and Probability Models

3

Unit II

Probability and Probability Distributions: Independence, Bayes Theorem, Discrete & Continuous Random Variables, Probability Mass and Density Functions, Cumulative Distribution Functions, Mean and Variance of a Random Variable, Discrete & Continuous Uniform Distribution, Binomial Distribution, Geometric and Negative Binomial Distributions, Hypergeometric Distributions, Poisson Distributions, Normal Distribution, Normal Approximation to the Binomial and Poisson Distributions, Exponential Distributions, Erlang and Gamma Distributions, Lognormal Distribution, Two or more Random Variables, Covariance and Correlation, Multinomial and Bivariate Normal Distributions

10

Unit III

Descriptive Statistics and Point Estimation of Parameters: Numerical Summaries of Data, Frequency Distributions and Histograms, Box and Probability Plots, Point Estimation, Sampling Distributions and the Central Limit Theorem, Methods of Point Estimation

5

Unit IV

Statistical Intervals for a Single Sample: Confidence Interval on the mean of a Normal Distribution, Confidence Interval on the Variance and Standard Deviation of a Normal Distribution, Large-Sample Confidence Interval for a Population Proportion, Guidelines for Constructing Confidence Intervals, Tolerance and Prediction Intervals

6

Unit V

6

Hypothesis Testing: Statistical Hypothesis, P-Values in Hypothesis Test, Tests on the Mean of a Normal Distribution, Tests on the Variance and Standard Deviation of a Normal Distribution, Tests on a Population Proportion, Testing for Goodness of Fit, Chi Square Test for Nominal Values.

Self Study:

The self-study contents will be declared at the commencement of semester. Around 10% of the questions will be asked from self-study contents.

Laboratory Work:

Laboratory work will be based on above syllabus with minimum 5 experiments to be incorporated.

Suggested Readings[^]:

1. Applied Statistics and Probability for Engineers, Douglas C. Montgomery and George C. Runger, Wiley
2. Introduction to Probability and Statistics: Principles and Applications for Engineering and the Computing Sciences, J. Susan Milton and Jesse Arnold, McGraw Hill Education
3. Statistics in Plain English, Timothy C. Urdan, Routledge
4. Introduction to Probability, Bertsekas, Dimitri and John Tsitsiklis, Athena Scientific
5. Fundamentals of Applied Probability Theory, Alvin Drake. McGraw-Hill
6. A First Course in Probability, Sheldon Ross, Prentice Hall
7. Introductory Statistics with Randomization and Simulation, David M Diez, Christopher D Barr, Mine C etinkaya-Rundel, Openintro

L=Lecture, T=Tutorial, P=Practical, C=Credit

[^]this is not an exhaustive list