

Introduction to Statistics for Data Science

This module provides students with an introduction to elementary probability theory and statistical concepts and principles that lay the foundation to understand and learn the statistical procedures and methods in the subsequent modules. The topics covered include descriptive statistics, rules of probability, probability distributions of discrete and continuous random variables, sampling distributions, statistical estimation and hypothesis testing.

Learning Objectives:

- Descriptive Statistics
- Probability and Probability Distributions
- Sampling Techniques & Sampling Distribution for Sample Mean
- Estimation of Population Mean (Confidence Interval)
- Hypothesis Testing of Means
- One-Way Analysis of Variance (ANOVA)
- Chi Square Test
- Simple Linear Regression

Introduction to Programming for Data Science

This module provides students with the fundamental skills to code applications to retrieve, clean and visualize data using the Python programming language. Students learn key concepts such as what structured and unstructured data are, and how they can create and manipulate relational and NoSQL databases to explore data and to create visualizations that can help them gain useful insights from it.

Learning Objectives:

- To learn basic programming syntax such as variables/loops/if-else and data structures such as lists, dictionaries etc
- To learn how to code Python applications for retrieving, transforming and visualizing data
- To gain hands-on experience crafting data stories that support decision-making, using the output and visualizations generated by the Python programs

Data Mining Techniques

This module teaches students key concepts in data mining, including data exploration, data preparation, and model building. Students will learn how to prepare data from multiple sources, and develop classification models for applications such as direct marketing and customer retention. Modelling techniques covered include k-nearest neighbours, logistic regression, classification trees, and neural networks.

Learning Objectives:

- Data preparation
- Validation techniques
- Nearest neighbours algorithm
- Decision Trees
- Regression techniques
- Neural Networks

Applied Statistical Methods

This module equips students with the statistical procedures and methods commonly used in the analysis of information and data in industry. The coverage specifically discusses analysis techniques necessary for multivariate data. Topics include matrix algebra, analysis of variance, design of experiments, multivariate statistical analysis, principal component analysis, factor analysis, discriminant analysis and cluster analysis.

Learning Objectives:

- Matrix Algebra
- Analysis of Variance
- Principal Component Analysis
- Factor Analysis
- Discriminant Analysis
- Cluster Analysis