

Module Synopsis

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.

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.

Introduction to Statistical Modelling

This module covers the theory and applications of statistical data modelling techniques. The module aims to equip students with good knowledge of the underlying theory, assumptions and applications of the techniques in statistical data modelling. Students will be exposed to the least squares theory required for modelling work. They will learn how to deal with various types of data using simple linear models, models for heteroscedastic data, model diagnostics, adequacy, comparison and building techniques as well as essentials of statistical simulation. Students will also be exposed to ideas of experimental design and system optimization in modelling work.

Generalised Modelling and Forecasting

This module aims to equip students with a greater breadth of skills in Predictive Analytics. It aims to build predictive analytics skills in the modelling of:

- i) Data with non-Gaussian distributions
- ii) Data of Gaussian but heteroscedastic structure
- iii) Categorical data and
- iv) Time-series data

Topics covered in the module include generalised linear models with a focus on Poisson and Gamma data, heteroscedastic regression with a focus on the generalised least squares approach, generalised estimating equations, analysis of categorical data with models for nominal and ordinal responses. The last part of the module focuses on analysis of time series data and discusses smoothing techniques, linear stationary and non-stationary models, model identification, estimation, diagnostics and forecasting.