

MODULE DESCRIPTOR

MODULE TITLE	Exploratory Data Analysis		
MODULE CODE	CO4760 (L7)	CREDIT VALUE	20 Credits (10 ECTS)
SCHOOL	SCHOOL OF SCIENCE		

MODULE AIMS

The aim of this module is to provide all the necessary tools for analysing datasets and visualising their properties.

The main objectives of the module are to:

- Provide essential exploratory techniques to describe data
- Introduce computational methods for solving statistical problems
- Introduce the R programming language, its packages, statistical functions, plotting systems
- Demonstrate the principles for constructing visual representations of the data
- Evaluate the information discovered from data analysis

MODULE CONTENT

Data objects and Attribute Types: Nominal attributes, Binary Attributes, Ordinal Attributes, Numeric Attributes, Discrete and Continuous Attributes

Data Structures: Arrays, Vectors, Lists.

Programming: Variables, Conditional Statements, Loops, Functions.

Basic Statistical Descriptors: Mean, Median Mode, Midrange, Range, Quartiles, and Interquartile Range

Mathematical Calculations in R: Numbers, Vectors, Matrices, Random Numbers.

Data Visualization: Geometric Projection Visualisation Techniques, High-level plots, Low-level plots and the layout command par, Complex Data

Statistical Inference: Descriptive Statistics, Statistical Inference for one and two samples, test of goodness of fit, Contingency Tables.

Regression: Linear Regression, Logistic Regression: Logistic Model, Probit Model, Non-Parametric Regression: Local Polynomial Regression, Smoothing Splines, Additive Nonparametric Regression.

Analysis of Variance: One-Way ANOVA, Multiple-Factor ANOVA.

Timeseries Analysis

INTENDED LEARNING OUTCOMES

On successful completion of this module a student will be able to:

1. Summarize techniques for descriptive, exploratory and confirmatory data analysis
2. Evaluate modelling concepts based on the characteristics of the data
3. Develop programs in R for performing statistical computing and visualising data
4. Make judgements about case studies through the assessment of discovered criteria from data analysis

TEACHING METHODS

Lectures deliver factual material, introduce key concepts, direct reading and relate academic aspects to practical considerations.

Tutorial sessions allow students to apply the techniques and reinforce the material delivered in the lecture.

Practical sessions enable students to discuss material and complete online or paper-based exercises.

The module will be assessed by one written coursework. The coursework requires the student to analyse datasets using the statistical methods and tools studied in class to reveal important findings and summarise, organise and communicate the generated knowledge from this data analysis through a report.

ASSESSMENT METHODS

This module is assessed through a Written Coursework (100%).