Liverpool John Moores University

Title:	MULTIVARIATE ANALYSIS AND DATA MINING
Status:	Definitive
Code:	**6106STATS** (124205)
Version Start Date:	01-08-2021
Owning School/Faculty:	Computer Science and Mathematics
Teaching School/Faculty:	Computer Science and Mathematics

Team	Leader
Sandra Ortega Martorell	Y
Ivo Siekmann	

Academic Level:	FHEQ6	**Credit Value:**	20	**Total Delivered Hours:**	57
Total Learning Hours:	200	**Private Study:**	143		

Delivery Options
Course typically offered: Semester 2

Component	Contact Hours
Lecture	49
Practical	6

Grading Basis: 40 %

Assessment Details

Category	Short Description	Description	Weighting (%)	Exam Duration
Exam	AS2	Examination.	60	2
Portfolio	AS1	Part A: Practical on Multivariate Analysis. Part B: Practical on Data Mining.	40	

Aims

To enable the student to explore the structure of multidimensional data sets.
To introduce the student to inferential procedures using multivariate data.
To provide the student with familiarity with linear and flexible methods for regression and classification.

Learning Outcomes

After completing the module the student should be able to:

1	Carry out an exploratory numerical and graphical analysis of a set of multivariate data.
2	Recognize situations in which a multivariate approach is required and carry out the appropriate inferential procedures.
3	Classify and cluster future multivariate observations into one of a number of known populations.
4	Apply linear methods and neural network algorithms.
5	Write a brief report.

Learning Outcomes of Assessments

The assessment item list is assessed via the learning outcomes listed:

Examination	1	3	4
Coursework portfolio	2	5	4

Outline Syllabus

Graphical display and numerical summary of multivariate data.
Investigation of the dependence among variables.
Discrimination and prediction. Error rate estimation.
Hypothesis construction and testing. Use of simultaneous confidence intervals.
Principal Components Analysis.
Use of R for data exploration, parameter estimation and significance testing.
Stages of the data analysis using data mining, including how to create and evaluate models
Neural networks, including backpropagation, gradient descent, the momentum method
Clustering with k-means and k-medoids methods
Blind Source Separation, including Independent Component Analysis and Non-negative Matrix Factorisation
Self-Organising Maps
Introduction to Big Data

Learning Activities

Lectures cover the complete syllabus
Practicals are carried out alongside lectures for the purpose of consolidating the material through application of key methods to synthetic and real-world data sets.

Notes

Module Overview

This final year module advances beyond univariate modelling with statistical methods, to the analysis of multivariate data. There are two generally accepted areas of statistics: exploratory and inferential.

How does this Module relate to the Programme overall?

The module begins with exploratory numerical and graphical methods, moving on to predictive inference with linear-in-the-parameter mainstream statistical methods. During the second semester, the module serves to introduce the use of non-linear methods. This begins with flexible extensions of generalised linear models, which are also feature in the machine learning literature under the name of neural networks. This reflects a link between statistics, computational learning and biological systems, whereby inference models are estimated by adjusting link-weights that control communication of information between simple neural cells. This leads to the relatively new field of data mining, which in essence comprises linear and non-linear exploration of data, through visualisation, clustering and inferential modelling.