Course ID 009457

Multivariate Statistics

MATH 68061

Credit rating 15

ECTS credits 7.5

Unit coordinator: Michael Tso

Semester 1

School of Mathematics

Postgraduate Taught

Level 6

FHEQ level ' Masters/Integrated Masters P4'

Marketing course unit overview

In practice most sets of data are multivariate in that they consist of observations on several different variables for each of a number of individuals or objects. Indeed, such data sets arise in many areas of science, the social sciences and medicine and techniques for their analysis form an important area of statistics. This course unit introduces a number of techniques, some of which are generalisation of univariate methods, while others are completely new (e.g. principal component analysis). The main part of the course focuses on continuous multivariate data in common with the level 3 module of the same name. The enhancement concentrates for the main part on multivariate methods for discrete data.

Course unit overview

In practice most sets of data are multivariate in that they consist of observations on several different variables for each of a number of individuals or objects. Indeed, such data sets arise in many areas of science, the social sciences and medicine and techniques for their analysis form an important area of statistics. This course unit introduces a number of techniques, some of which are generalisation of univariate methods, while others are completely new (e.g. principal component analysis). The main part of the course focuses on continuous multivariate data in common with the level 3 module of the same name. The enhancement concentrates for the main part on multivariate methods for discrete data.

Aims

To familiarise students with the ideas and methodology of certain multivariate methods together with their application in data analysis using the R statistical computing package.

Learning outcomes

On successful completion of the course students will:

- be familiar with multivariate random vectors and their probability distributions;
- have acquired skills in dimensionality reduction techniques, inferential methods based on the multivariate Normal distribution as an underlying model and in methods of data classification;
- be aware of how the statistical package R can be used as a tool for multivariate data analysis and graphical presentation.
- have skills in analyzing discrete multivariate data.

Syllabus

- Introductory ideas and basic concepts - random vectors and their distribution, linear transformations (including the Mahalanobis transformation), sample statistics and their properties, overall measures of dispersion in p-space, distances in p-space, simple graphical techniques. [3]
- Cluster Analysis - aims, hierarchical algorithms, the dendrogram. [2]
- Principal component analysis - definition and derivation of population PC's, sample PC's, practical considerations, geometrical properties, examples. [4]
- The Multivariate Normal (MVN) distribution - definition, properties, conditional distributions, the Wishart and Hotelling T-squared distributions, sampling distributions of the sample mean vector and covariance matrix, maximum likelihood estimation of the mean vector and covariance matrix. [4]
- Hypothesis testing and confidence intervals (one sample procedures) - the generalized likelihood ratio test, tests on the mean vector, CI's for the components of the mean vector. [4]
- Hypothesis testing and confidence intervals (two independent sample procedures) - tests on the difference between two mean vectors, testing equality of covariance matrices, CI's for the differences in the components of the mean vectors. [3]
- Profile Analysis. [2]
- Discriminant Analysis. [Guided Coursework]
- Techniques for discrete multivariate data incl. discrete multivariate vectors, two-way contingency tables, sampling distributions, odds ratio, testing independence, correspondence analysis, higher order contingency tables, conditional independence, introduction to log-linear models . [11]

Assessment methods

Other	20%
Written exam	80%

Coursework: weighting 20%End of semester examination: three hours, weighting 80%

Feedback methods

Tutorials will provide an opportunity for students' work to be discussed and provide feedback on their understanding.

Requisites

MATH20701	Probability 2	Pre-Requisite	Compulsory
MATH20802	Statistical Methods	Pre-Requisite	Compulsory

Students are not permitted to take more than one of MATH38061, MATH48061or MATH68061 for credit, either in the same or different undergraduate year or in an undergraduate programme and then a postgraduate programme, as the contents of the courses overlap significantly.

Available as free choice? N

Recommended reading

- Chatfield, C. and Collins, A. J., An Introduction to Multivariate Analysis, Chapman & Hall

1983.
- Krzanowski, W. J., Principles of Multivariate Analysis: A User's Perspective, Oxford University Press 1990.
- Johnson, R. A. and Wichern, D. W., Applied Multivariate Statistical Analysis 3rd edition, Prentice Hall 1992.
- Bishop, Y M, Fienberg, S E, and Holland P W (2007) Discrete Multivariate Analysis. Springer

Scheduled activity hours

Lectures	33
Tutorials	11

Independent study hours 106 hours

Version Nbr 005.0.0