On-line course materials

MATH38061 - Multivariate Statistics

Year: 3 - **Semester:** 1 - **Credit Rating:** 10

Requisites

Prerequisites
MATH20701 Probability 2

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.

Brief Description

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 generalisations of univariate methods, while others are completely new (e.g. principal component analysis). The course focuses on continuous multivariate data.

Learning Outcomes

On successful completion of the course students will:

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

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]

Teaching & Learning Process (Hours Allocated To)

Lectures	Tutorials/ Example Classes	Practical Work/ Laboratory	Private Study	Total
22	11	0	67	100

Assessment and Feedback

- Coursework: weighting 20%
- End of semester examination: two hours weighting 80%

Further 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.

Staff Involved

Mr Michael Tso - Lecturer

Data source is EPS system
Back To Top