UCLouvain

lstat2110

2018

Data Analysis

5 credits	22.5 h + 7.5 h	Q1

Teacher(s)	Segers Johan ;
Language :	French
Place of the course	Louvain-la-Neuve
Main themes	Contents: - Reminders of algebra and geometry useful for multivariate data analysis - Basic principles of factorial methods - Principal components analysis (PCA) - Canonical correlation - Factorial discriminant analysis (FDA) - Factorial correspondence analysis (FCA simple and multiple) - Cluster analysis - Data analysis in practice
Aims	1 General objectives. Presentation of the modern techniques for the analysis of huge multivariate data sets. Developing the basic tools for " data mining ". Specific objectives. At the end of this course, the students should be able to: - Manipulate and describe the information contained in huge data sets; - Understand why such or such method is appropriate; - Give a correct interpretation of the resulting pictures and of the output of the software; - Solve problems with real data sets. - - - - *The contribution of this Teaching Unit to the development and command of the skills and learning outcomes of the programme(s) can be accessed at the end of this sheet, in the section entitled "Programmes/courses offering this Teaching Unit".*
Evaluation methods	Tests during the lectures: • Test 1: Data matrices and principal component analysis • Test 2: Clustering and linear discriminant analysis Participation is optional. At the discretion of the student, each test can replace the part of the exam on the same topic. Exam (12/20): • written, closed book, with the help of a formula list and a pocket calculator • exercises and questions involving (small) calculcations, interpretation of computer output, and understanding of the main results and formulas Project (8/20): • individually or in pairs • data application, the data being sought by the students themselves • written report in R Markdown, to be submitted before the exam session • detailed instructions will be provided in the exercise sessions and on the MoodleUCL course page
Teaching methods	During the lectures, the teacher presents the various statistical methods, covering the questions and data-sets to which they apply, the underlying mathematical theory, and how to program them in R. Homework assignments are given, the solution of which is discussed in the lectures too. The tutorials take place in computer rooms and have as primary objective to allow the students to train themselves in applying the method on real data-sets in R.
Content	• Data matrices • Principal component analysis • Classification: k-means clustering and hierarchical clustering • Linear discriminant analysis • Simple and multiple correspondence analysis Implementation of the methods is done in the R language using the RStudio integrated development environment, and the R Markdown framework is used to combine text, mathematical formulas, R code and R output (tables, graphs).
Inline resources	All teaching material is made available through the MoodleUCL cours page: slides, exercises, software scripts. In addition, links to interesting external material are given too: on-line courses, videos, software documentation.
Bibliography	• Escofier, B. et Pagès, J. (2016): Analyses factorielles simples et multiples, 5e édition, Dunod, Paris. • Lebart, L., Piron, M. et Morineau, A. (2006): Statistique exploratoire multidimensionnelle, 4e édition, Dunod, Paris. • Saporta, G. (2011): Probabilités, analyse des données et statistique, 3e édition révisée, Editions TECHNIP, Paris.

Other infos	Prerequisities: • vector and matrix calculus • Euclidean geometry: points, spaces, orthogonality, distances, angles • basic notions in statistiques: sample mean, (co)variance, correlation, covariance matrix, conditional probabilities, normal distribution, chi-square distribution
Faculty or entity in charge	LSBA

Programmes containing this learning unit (UE)

Program title	Acronym	Credits	Prerequisite	Aims
Master [120] in Mathematics	MATH2M	5		
Master [120] in Statistic: Biostatistics	BSTA2M	5		
Master [120] in Economics: General	ECON2M	5		
Master [120] in Biomedical Engineering	GBIO2M	5		
Master [120] in Statistic: General	STAT2M	5		
Master [120] in Mathematical Engineering	MAP2M	5		
Master [120] in data Science: Statistic	DATS2M	5		
Minor in Statistics and data sciences	LSTAT100I	5		
	LSTAT100P	5		