










5.00 credits

30.0 h + 7.5 h

Q1

Teacher(s)	Segers Johan ;
Language :	French
Place of the course	Louvain-la-Neuve
Main themes	<p>The course presents an overview of the main tools of exploratory multivariate data analysis via factorial methods. The data is projected onto a low-dimensional subspace while retaining maximum information. This reduction in dimension facilitates visualization and aids in the discovery of information and patterns in a data table.</p> <ul style="list-style-type: none"> • Reminders of algebra and geometry useful for data analysis • Basic principles of factorial methods • Principal component analysis • Classification: moving averages and hierarchical classification • Linear discriminant analysis • Simple and multiple correspondence analysis • Principal component regression • Partial least squares regression
Learning outcomes	<p>At the end of this learning unit, the student is able to :</p> <p>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.</p> <p>1</p>
Evaluation methods	<p>Exam (12/20):</p> <ul style="list-style-type: none"> • written, closed book, with the help of a formula list and a pocket calculator • exercises and questions involving (small) calculations, interpretation of computer output, and understanding of the main results and formulas <p>Tests during the lectures:</p> <ul style="list-style-type: none"> • Test 1: Data matrices and principal component analysis • Test 2: Clustering and linear discriminant analysis <p>Participation is optional. At the discretion of the student, each test can replace the part of the exam on the same topic.</p> <p>Project (8/20):</p> <ul style="list-style-type: none"> • individually or in pairs • data application, the data being sought by the students themselves • written report, to be submitted at a date or at dates specified during the semester • detailed instructions will be provided in the exercise sessions and on the MoodleUCL course page <p>Submitting a projet is a necessary requirement in order to participate at the exam and obtain an exam result. At a second exam inscription, a new project can be resubmitted.</p>
Teaching methods	<p>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.</p> <p>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.</p>
Content	<ul style="list-style-type: none"> • Data matrices • Principal component analysis • Classification: k-means clustering and hierarchical clustering • Linear discriminant analysis • Simple and multiple correspondence analysis • Principal component regression • Partial least squares regression

	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	<ul style="list-style-type: none"> • 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	<p>Prerequisites:</p> <ul style="list-style-type: none"> • 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	Learning outcomes
Master [120] in Statistics: General	STAT2M	5		
Master [120] in Mathematics	MATH2M	5		
Master [120] in Statistics: Biostatistics	BSTA2M	5		
Master [120] in Biomedical Engineering	GBIO2M	5		
Approfondissement en statistique et sciences des données	APPSTAT	5		
Minor in Statistics, Actuarial Sciences and Data Sciences	MINSTAT	5		
Master [120] in Economics: General	ECON2M	5		
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	5		
Master [120] in Mathematical Engineering	MAP2M	5		
Master [120] in Data Science : Statistic	DATS2M	5		