UCLouvain

linge1222

2023

Multivariate Statistical Analysis

4.00 credits	30.0 h + 15.0 h	Q2

Teacher(s)	Segers Johan ;
Language :	French
Place of the course	Louvain-la-Neuve
Prerequisites	*The prerequisite(s) for this Teaching Unit (Unité d'enseignement – UE) for the programmes/courses that offer this Teaching Unit are specified at the end of this sheet.*
Main themes	Part 1: Basic descriptive methods and basic notations. In this part, students are taught how matrix notation facilitates treatment of multidimensional data and basic properties of random vectors. They will also learn that the basic (uni-and bivariate) descriptive tools have both their uses and limitations. Part 2: Techniques of multivariate data analysis. In this part, students learn about basic dimension reduction techniques for continuous and qualitative variables (principal components, correspondence analysis). Basic classification techniques are also presented. A wide range of examples is given to illustrate these methods and show when they should be used. Part 3: Multivariate analysis models. In this part, students see how to model inter-variable relations: linear models (including variance and variance-covariance analysis) which make it possible to use explanatory variables to explain response variable variation. Models adapted to categorical response variable are also introduced, log-linear models for contingency tables, the logit model and discrimination analysis models. Here too, a wide range of examples is given to illustrate these methods and show when they should be used.
Learning outcomes	**At the end of this learning unit, the student is able to :** 1 This course develops the elements introduced in the basic Probability and Statistics courses within a multivariate framework, the aim being to equip students with the instruments they need to analyse multidimensional data sets. By the end of the course, students should be able to use the most widely-used instruments to analyse real data. A key aim of the course will therefore be to give students a clear understanding of the methods and how to apply them, and how to use relevant analytical software.
Evaluation methods	• Project: near the end of the course, the students need to solve problems using real data sets and the statistical software environment R. This part is open-book, to be done at home by groups of 3 to 5 students. • Exam: written, closed book, with the help of a formula list and a pocket calculator. The exam part comprises both theory questions as well as exercises related to interpreting and reconstructing the output of the R software.
Teaching methods	• Lectures: the teacher introduces the concepts through an application and then presents the abstract form • Exercise sessions in computer rooms: the teacher gives students real-data problems to solve using the statistical software environment R.
Content	• Introduction to multivariate data analyis • Linear algebra and Euclidean geometry • Descriptive statistics for data matrices • Principal component analysis • Cluster analysis: k-means clustering and hierarchical cluster algorithms • Linear discriminant analysis • Distribution theory • Multiple linear regression • Logistic regression
Inline resources	The list of formulas, the slides used in the lectures and the computer labs, R software documentation and links to external web resources (videos, on-line courses, documents) are available on the Moodle course page.
Bibliography	• Härdle, W. and L. Simar (2007): Applied Multivariate Statistical Analysis, 2nd Edition, Springer-Verlag, Berlin. • James, G., Witten, D., Hastie, T. and R. Tibshirani (2013): An Introduction to Statistical Learning, Springer, New York. • Saporta, G. (2011): Probabilités, analyse des données et statistique, 3e édition révisée, Editions TECHNIP, Paris.
Faculty or entity in charge	ESPO

Programmes containing this learning unit (UE)

Program title	Acronym	Credits	Prerequisite	Learning outcomes
Master [120] in Data Science : Statistic	DATS2M	4		
Bachelor in Mathematics	MATH1BA	4	LMAT1271	
Approfondissement en statistique et sciences des données	APPSTAT	4		
Bachelor : Business Engineering	INGE1BA	4	LINGE1113	
Minor in Statistics, Actuarial Sciences and Data Sciences	MINSTAT	4		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	4		