








| | | |
|--------------|----------------|----|
| 5.00 credits | 30.0 h + 7.5 h | Q1 |
|--------------|----------------|----|

| | |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Teacher(s) | Colling Benjamin ; |
| Language : | French |
| Place of the course | Louvain-la-Neuve |
| Prerequisites | <p>Concepts and tools equivalent to those taught in teaching units</p> <p>LSTAT2011 Eléments de mathématique pour la statistique</p> <p>LSTAT2014 Eléments de probabilités et de statistique mathématique</p> |
| 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</p> <p>1 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> |
| | |

| | |
|-----------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <ul style="list-style-type: none"> • Simple and multiple correspondence analysis • Principal component regression • Partial least squares regression <p>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).</p> |
| 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 Data Science : Statistic | DATS2M | 5 | |  |
| Master [120] in Biomedical Engineering | GBIO2M | 5 | |  |
| Master [120] in Statistics: Biostatistics | BSTA2M | 5 | |  |
| Master [120] in Mathematics | MATH2M | 5 | |  |
| Master [120] in Statistics: General | STAT2M | 5 | |  |
| Master [120] in Chemistry and Bioindustries | BIRC2M | 5 | |  |
| Approfondissement en statistique et sciences des données | APPSTAT | 5 | |  |
| Master [120] in Mathematical Engineering | MAP2M | 5 | |  |
| Master [120] in Economics: General | ECON2M | 5 | |  |
| Minor in Statistics, Actuarial Sciences and Data Sciences | MINSTAT | 5 | |  |
| Certificat d'université : Statistique et science des données (15/30 crédits) | STAT2FC | 5 | |  |