









5.00 credits

30.0 h + 7.5 h

Q1

|                     |  |
|---------------------|--|
| Teacher(s)          | Segers Johan ;   |
| Language :          | French   |
| Place of the course | Louvain-la-Neuve   |
| Prerequisites       | Concepts and tools equivalent to those taught in teaching units<br>LSTAT2011 <del>Eléments de mathématique pour la statistique</del><br>LSTAT2014 <del>Eléments de probabilités et de statistique mathématique</del>   |
| 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"> <li>• Reminders of algebra and geometry useful for data analysis</li> <li>• Basic principles of factorial methods</li> <li>• Principal component analysis</li> <li>• Classification: moving averages and hierarchical classification</li> <li>• Linear discriminant analysis</li> <li>• Simple and multiple correspondence analysis</li> <li>• Principal component regression</li> <li>• Partial least squares regression</li> </ul>  |
| Learning outcomes   | <p><b>At the end of this learning unit, the student is able to :</b></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>  |
| Evaluation methods  | <p>Exam (12/20):</p> <ul style="list-style-type: none"> <li>• written, closed book, with the help of a formula list and a pocket calculator</li> <li>• exercises and questions involving (small) calculations, interpretation of computer output, and understanding of the main results and formulas</li> </ul> <p>Tests during the lectures:</p> <ul style="list-style-type: none"> <li>• Test 1: Data matrices and principal component analysis</li> <li>• Test 2: Clustering and linear discriminant analysis</li> </ul> <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"> <li>• individually or in pairs</li> <li>• data application, the data being sought by the students themselves</li> <li>• written report, to be submitted at a date or at dates specified during the semester</li> <li>• detailed instructions will be provided in the exercise sessions and on the MoodleUCL course page</li> </ul> |

|                  |   |
|------------------|---|
|                  | <ul style="list-style-type: none"><li>• Simple and multiple correspondence analysis</li><li>• Principal component regression</li><li>• Partial least squares regression</li></ul> <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.  |
|                  |   |

| Programmes containing this learning unit (UE)                  |         |         |              |   |
|--|---------|---------|--------------|---|
| Program title  | Acronym | Credits | Prerequisite | Learning outcomes   |
| Master [120] in Data Science :<br>Statistic                    | DATS2M  | 5       |              |  |
| Master [120] in Biomedical<br>Engineering                      | GBIO2M  | 5       |              |  |
| Master [120] in Statistics:<br>Biostatistics                   | BSTA2M  | 5       |              |  |
| Master [120] in Mathematics                                    | MATH2M  | 5       |              |  |
| Master [120] in Statistics:<br>General                         | STAT2M  | 5       |              |  |
| Master [120] in Chemistry and<br>Bioindustries                 | BIRC2M  | 5       |              |  |
| Approfondissement en<br>statistique et sciences des<br>données | APPSTAT | 5       |              |  |
| Master [120] in Mathematical<br>Engineering                    | BIRC2M  | 5       |              |  |

BIRC2M