

## Syllabus Minor B 214 –Multidimensional and Multivariate Statistical Methods

<b>Module : 214</b>	Multidimensional and Multivariate Statistical Methods
<b>Coordinator</b>	Mélanie Bertin, PhD Professor of Epidemiology/biostatistics, Department of METIS, EHESP <a href="mailto:melanie.bertin@ehesp.fr">melanie.bertin@ehesp.fr</a>
<b>Dates</b>	November 16 <sup>th</sup> to 20 <sup>th</sup> 2020
<b>ECTS</b>	3 ECTS
<b>Duration</b>	5 days of 6 hours = 30 hours
<b>Location</b>	EHESP 20 Avenue George Sand 93210 LA PLAINE ST DENIS
<b>Description</b>	<p>The goal is to provide knowledge and skills in performing and interpreting the results of multivariate analyses and of multidimensional exploratory methods including principal component analysis, correspondence analysis and cluster analyses. More specifically, multidimensional analyses will be view as a statistical methods used in public health to analyse high-dimensional data sets. This technic is particularly adequate to synthetizing variables highly correlated and eliminating collinearity problem in multiple regression (i.e. creation of a composite index). Complementarity between modelling approaches and multidimensional exploratory analysis will be presented.</p> <p>Each day is designed to alternate between lectures, exercices and computer lab on either Stata or R software depending on students' preferences.</p>
<b>Prerequisite</b>	Advanced core in biostatistics and in Epidemiology
<b>Course learning objectives</b>	<p><b>Learning objectives:</b> <i>at the end of the module, the students should be able to:</i></p> <ul style="list-style-type: none"> <li>- Perform advanced statistical methods : Run, interpret and compare multivariate models (revision advanced core biostatistics)</li> <li>- Identify the main difference between supervised learning (modelling approaches) and unsupervised learning (PCA/MCA/clustering) and their complementarity</li> <li>- Be familiar with the methodological concepts of the most common multidimensional methods: principal component analysis, cluster analysis and to understand the extension of PCA analysis to other variables types as factorial analysis or correspondence analysis.</li> <li>- To implement and to interpret the results including the statistical tables, such as contribution of variable on each components, correlation, eigen values and also correlation circle and dendogram, and also parameters estimates from statistical models</li> </ul>
<b>Structure</b> (details of sessions title/speaker/date/duration )	<p><b>Session 1:</b> Multivariate analysis and collinearity</p> <p><b>Session 2:</b> Multidimensional exploratory analyses with lectures, exercises and conferences, from Tuesday 17 Nov to Friday 20<sup>th</sup></p>
<b>Resources</b>	All readings and materials will be posted on REAL
<b>Course requirement</b>	Students will gain experience in using modern techniques to analyze high-dimensional public health data sets.
<b>Grading and assessment</b>	Group project (2/3 of the final mark) + Individual exam-2 hours (1/3 of the final mark) All team members will receive the same grade except if it is clear that a student has not participated effectively (attended and contributed to meetings; made timely, helpful contributions; been constructive, etc.). In that case, the student's grade will be lowered accordingly.
<b>Course policy</b>	<b>Attendance &amp; punctuality</b>

	<p><b>Regular and punctual class attendance is a prerequisite for receiving credit in a course.</b> Students are expected to attend each class. Attendance will be taken at each class.</p> <p>The obligations of attendance and punctuality cover every aspect of the course: - lectures, conferences, group projects, assessments, examinations, as described in EHESP Academic Regulations <a href="http://mph.ehesp.fr">http://mph.ehesp.fr</a> EHESP Academic Regulation Article. 3).</p> <p>If students are not able to make it to class, they are required to send an email to the instructor and to the MPH program coordinating team explaining their absence prior to the scheduled class date. All supporting documents are provided to the end-of-year panel.</p> <p>.</p> <p>Students who miss class are responsible for content. Any student who misses a class has the responsibility for obtaining copies of notes, handouts and assignments. If additional assistance is still necessary, an appointment should be scheduled with the instructor. Class time is not to be used to go over material with students who have missed class.</p> <p><b>Lateness:</b> Students who are more than 10 minutes late may be denied access to a class. Repeated late arrivals may be counted as absences (See <a href="http://mph.ehesp.fr">http://mph.ehesp.fr</a> EHESP Academic Regulation Article. 3 Attendance &amp; Punctuality)</p> <p><b>Maximum absences authorized &amp; penalty otherwise</b></p> <p>Above 20% of absences will be designated a fail for a given class. The students will be entitled to be reassessed in any failed component(s). If they undertake a reassessment or they retake a module this means that they cannot normally obtain more than the minimum pass mark (i.e. 10 out of 20)</p> <p><b>Exceptional circumstances</b></p> <p>Absence from any examination or test, or late submission of assignments due to illness, psychological problems, or exceptional personal reasons must be justified; otherwise, students will be penalized, as above mentioned. Students must directly notify their professor or the MPH academic secretariat before the exam or before the assignment deadline. Before accepting the student's justification, the professor or the MPH academic secretariat has the right to request either a certificate from the attending physician or from a psychologist, or from any other relevant person (See <a href="http://mph.ehesp.fr">http://mph.ehesp.fr</a> EHESP Academic Regulation Article 4 Examinations).</p> <p><b>Courtesy:</b> <u>All cell phones/pages MUST be turned off during class time.</u></p> <p>Students are required to conduct themselves according to professional standards, eating during class time is not permitted during class time, such as course or group work.</p>
<p><b>Valuing diversity</b></p>	<p>Diversity enriches learning. It requires an atmosphere of inclusion and tolerance, which oftentimes challenges our own closely-held ideas, as well as our personal comfort zones. The results, however, create a sense of community and promote excellence in the learning environment. This class will follow principles of inclusion, respect, tolerance, and acceptance that support the values of diversity. Diversity includes consideration of: (1) life experiences, including type, variety, uniqueness, duration, personal values, political viewpoints, and intensity; and (2) factors related to "diversity of presence," including, among others, age, economic circumstances, ethnic identification, family educational attainment, disability, gender, geographic origin, maturity, race, religion, sexual orientation and social position.</p>
<p><b>Course evaluation</b></p>	<p>EHESP requests that you complete a course evaluation at the end of the school year. Your responses will be anonymous, with feedback provided in the aggregate. Open-ended comments will be shared with instructors, but not identified with individual students. Your participation in course evaluation is an expectation, since providing constructive feedback is a professional obligation. Feedback is critical, moreover, to improving the quality of our courses, as well as for instructor assessment.</p>

<p>Session #1</p>	<p><b>Multivariate regression methods</b></p>
<p><b>Speakers</b></p>	<p>Lecturers: Mélanie Bertin (EHESP), Séverine Deguen (EHESP)</p>

<b>Session Outline</b>	This session goes back to the concept of modelling approaches and the choice of the appropriate model and multivariate analysis.
<b>Learning Objectives</b>	At the end of the sessions, students will be able : <ul style="list-style-type: none"> <li>- To conduct basic and advanced multivariate models</li> <li>- To interpret the results and select the crucial information</li> <li>- To write statistical analysis section, create tables and figures, and interpret statistical analysis results for public health papers.</li> </ul>
<b>Duration</b>	6 hours
<b>Training methods</b>	Lectures alternate with exercises, conference and computer lab
Session #2	<b>Multidimensional methods</b>
<b>Speakers</b>	Lecturers: Mélanie Bertin (EHESP), Séverine Deguen (EHESP), Cindy Padilla (EHESP) Maxime Raffray (EHESP)
<b>Session Outline</b>	This session introduces the multidimensional methods that are particularly appropriate to deal with large dataset and when correlation exists between variables. Principal component analysis following by a cluster analysis is commonly use in public health to synthetize information by constructing composite index or to generate individuals' profiles. This course gives the main methodological aspects of the PCA, MCA, clustering and how the results can be associated to modelling approaches using example in public health.
<b>Learning Objectives</b>	At the end of the session, the students should be able to: <ul style="list-style-type: none"> <li>- To provide a general view of the different multidimensional approaches with a focus on PCA, MCA and cluster analysis</li> <li>- To perform a PCA and MCA and to interpret the results</li> <li>- To conduct a cluster analysis, to draw a dendogram and interpret it</li> <li>- To select the appropriate number of classes from the cluster analysis results and to give the main characteristics of each class</li> </ul>
<b>Duration</b>	24 hours
<b>Training methods</b>	Lectures alternate with in exercises, conference and computer lab