Syllabus Minor B 214 –Multidimensional and Multivariate Statistical Methods

Module : 214	Multidimensional and Multivariate Statistical Methods
Coordinator	Séverine Deguen, PhD Professor of biostatistics, Department of EOHS, EHESP severine.deguen@ehesp.fr
Dates	November 5th to 9th 2018
ECTS	3 ECTS
Duration	5 days of 6 hours = 30 hours
Location	EHESP 20 Avenue George Sand 93210 LA PLAINE ST DENIS
Description	The goal is to provide knowledge and skills in performing and interpreting the results of a regression analysis and of multidimensional methods including principal component analysis and cluster analysis. More specifically, the principal component analysis will be view as a statistical method 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, for example, to constructing composite index. Different regression model are used to investigate existence of health risk factors; in this course, we will focus on the Poisson regression and on the negative binomial regression. Each day is designed to alternate between theory and computer lab on Stata software.
Prerequisite	Advanced core in biostatistics and in Epidemiology
Course learning objectives	*Learning objectives: at the end of the module, the students should be able to:* - Identify the main difference between supervised learning (regression) and unsupervised learning (PCA) and their complementarity - 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. - Perform multivariate Poisson regression and negative binomial regression - 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 the regression model
Structure (details of sessions title/spaeker/date/duration)	**Session 1**: Multivariate regression models with two lectures, one on the Poisson regression (Monday 5 Nov 9:00-11:00 pm) and another on the negative Binomial regression (Tuesday 6 Nov 9:00-11:00 pm), exercises (Monday 5 Nov 11:00-12:00 pm), conference (Tuesday 6 Nov 11:00-12:00 pm) and computer lab (Monday 5 Nov 1:00-4:00 am and Tuesday 6 Nov 1:00-4:00 am) **Session 2:** Multidimensional analysis with two lectures, one on PCA (Wednesday 7 Nov 9:00-12:00) and another on cluster analysis and Thursday 8 Nov 9:00-10:30 pm), exercises (Wednesday 7 Nov 1:00-2:30 am and Thursday 8 Nov 1:00-2:30 am), conferences (Thursday 8 Nov 10:30-12:00 pm and Friday 9 Nov 9:00-12:00) and computer lab (Wednesday 7 Nov 2:30-5:00 am, Thursday 8 Nov 2:30-5:00 am and Friday 9 Nov 1:00-4:00 Thursday November 23rd, 9:00 -12:00 and 1:00 – 4:30 pm
Resources	All readings and materials will be posted on REAL
Course requirement	Students will gain experience in using modern techniques to analyze high-dimensional public health data sets.
Grading and	Homework by group (1/3 of the final mark) + Individual exam-2 hours (2/3 of the final mark)

assessment	Note also that students will complete a questionnaire that assesses their own and their teammates' contributions to group work. 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.
Course policy	**Attendance & punctuality** **Regular and punctual class attendance is a prerequisite for receiving credit in a course.** Students are expected to attend each class. Attendance will be taken at each class. The obligations of attendance and punctuality cover every aspect of the course: - lectures, conferences, group projects, assessments, examinations, as described in EHESP Academic Regulations http://mph.ehesp.fr EHESP Academic Regulation Article. 3). 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. . 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. **Lateness**: Students who are more than 10 minutes late may be denied access to a class. Repeated late arrivals may be counted as absences (See http://mph.ehesp.fr EHESP Academic Regulation Article. 3 Attendance & Punctuality) **Maximum absences authorized & penalty otherwise** 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) **Exceptional circumstances** 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 http://mph.ehesp.fr EHESP Academic Regulation Article 4 Examinations). **Courtesy**: All cell phones/pages MUST be turned off during class time. 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.
Valuing diversity	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.
Course evaluation	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.

Session #1	Multivariate regression methods
Speakers	Lecturers: Séverine Deguen (EHESP), Cindy Padilla (EHESP)
Session Outline	This session introduces regression methods that are appropriate to analyse count (number of death for instance) or rate data (death rate). In public health, Poisson regression is mainly used to investigate associations between a count/rate health outcome and a number of predictor variables (including confounder, modifier and risk factor).More specifically, this course explores how Poisson regression models are formulated and how the results are interpreted using examples in public health. In addition, this course also details how Poisson regression can be extended when assumptions are not met (introducing negative binomial models). Then, the negative binomial regression will be introduced and explained.
Learning Objectives	At the end of the sessions, students will be able : - To conduct basic and advanced regression analyses - To interpret the results and select the crucial information - To write statistical analysis section, create tables and figures, and interpret statistical analysis results for public health papers.
Duration	14,5 hours
Training methods	Lectures alternate with in class discussions, conference and computer lab
Session #2	Multidimensional methods
Speakers	Lecturers: Séverine Deguen (EHESP), Wahida Kihal (CNRS) and Benoit Lalloué (IGR)
Session Outline	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. This course gives the main methodological aspects of the PCA and how the results can be considered in regression model using example in public health.
Learning Objectives	At the end of the session, the students should be able to: - To provide a general view of the different multidimensional approaches with a focus on PCA and cluster analysis - To perform a principal component analysis and to interpret the results - To conduct a cluster analysis, to draw a dendogram and interpret it - To select the appropriate number of classes from the cluster analysis results and to give the main characteristics of each class
Duration	13 hours
Training methods	Lectures alternate with in class discussions, conference and computer lab