lmafy1101

2018

Data exploration and introduction to statiscal inference

UCLouvain

5 credits	30.0 h + 30.0 h	Q2

Teacher(s)	El Ghouch Anouar ;
Language :	French
Place of the course	Louvain-la-Neuve
Main themes	This teaching unit provides an active introduction to the exploratory methods and fundamental principles of probabilistic and statistical modeling essential to the analysis of observational and experimental data. Real data will be used to present the numerical indexes and graphical descriptive analysis tools and to allow the student to discover and develop his operational skills in data processing using specialized software (for example, the free software R: https://www.r-project.org/). Simulations of data, also done on software, following basic probability laws will develop an intuitive understanding of the notion of chance in modeling. Once these notions of probability (and random variable) are assimilated, fundamental notions of statistical inference such as sampling, point and interval estimation will be presented empirically via simulations and applied to different real contexts. The emphasis will be deliberately not on the demonstration of mathematical results but on the meaning and interpretation of concepts. The teaching unit will also include a module on the analysis of the uncertainties of experimental measurements, their sources, their propagation and their modeling. These aspects are illustrated on concrete cases in fields such as physics or biology.
Aims	**a. Contribution of the teaching unit to the learning outcomes of the programme** With regard to the AA reference system of the Bachelor programme in mathematics, this teaching unit allows students to master: · as a priority, the following LOs: x.x, ...; · in a secondary way the following LO: x.x, ... With regard to the AA reference system of the Bachelor programme in physics, this teaching unit allows students to master: · as a priority, the following LOs: A1.2, A1.5, A2.3, A3.4; · in a secondary way the following LOs: A5.3, A6.5, A4.4. **b. Specific learning outcomes of the teaching unit** At the end of this teaching unit, the student will be able to: ' select descriptive statistics tools (numerical and graphical indexes) adapted to effectively summarize and answer questions about a dataset; ' analyze a potentially large data set via descriptive tools using specialized software; ' read and interpret the results of a descriptive data analysis by formalizing them in the context of the study; ' know how to explain and define the basic concepts of probability on events and related to univariate random variables; ' perform basic probability calculations in a variety of situations; ' Using software, simulations to illustrate the behaviour of random variables and the notions of sampling; ' explain the sampling notations and fundamental concepts of statistical inference, and estimate the basic parameters of a random variable by quantifying uncertainty via confidence intervals; ' identify and quantify the experimental uncertainty, and present the results of an analysis or a set of measurements while indicating the degree of uncertainty; ' explain the general objectives and concepts of linear and nonlinear modeling; adjust a linear model to a variable via a least squares method. - - - - *The contribution of this Teaching Unit to the development and command of the skills and learning outcomes of the programme(s) can be accessed at the end of this sheet, in the section entitled "Programmes/courses offering this Teaching Unit".*
Evaluation methods	· During the practicals: assignments in the form of mini-projects (case studies) to be solved in groups of two students using specialized software. · During the exam session: computer-based written exam.

Teaching methods	The teaching unit will consist of
	· lectures that will present the subject on the basis of examples,
	· exercises sessions to systematically put into practice the different notions seen in the course on well-targeted cases and using specialized software,
	· projects that will give the student the opportunity to integrate the different tools in fields of application of mathematics and physics.
	The pedagogical approach will favor the active learning of students and will try to respect the pedagogical orientations proposed by the Faculty of Sciences.
Content	· Presentation of databases from different fields of application in science and technology that are used to illustrate the course (results of laboratory experiments, weather data, survey results, clinical trials, stock market data, network data (social and other), etc.
	· Structuring of statistical data and nature of the variables (discrete / continuous quantitative, nominal / ordinal quality).
	· Numeric tools to summarize data by nature: frequency tables, position indices (mode, median, mean), dispersion indices (range, interquartile range, standard deviation, variance, coefficient of variation), percentiles, correlation coefficient.
	· Graphical tools to summarize data: histogram, empirical distribution function, bar graph, box-plot, time graph, X-Y graph (simple or matrix), Q-Q-plot.
	· Random experimentation and basic notions of probability theory: the definition of a probability and its elementary properties and the calculation of probability on events.
	· Random variables (univariate) and their properties (moments, distribution ...). Introduction of the most frequently used distributions in the analysis of data encountered in applications seen in the course: uniform law, binomial, Fish and normal.
	· Simple algorithms for generating random numbers according to these introduced probability laws.
	· Principle of random sampling and estimation.
	· Intuitive concept of sampling distribution and confidence interval.
	· Application to mean (and variance) in normal distribution.
	· Notions of error and measurement uncertainty.
	· Quantification and expression of uncertainty on a simple or repeated measurement (including uncertainty widened by confidence interval).
	· Calculation of composite uncertainties in the case of sums, products and nonlinear transformations of independent or correlated measurements. Applications to laboratory measurements in various fields (physics, biology, ...).
	· Adjustment of a linear or polynomial model (simple linear regression) by least squares method.
	· Initiation to a statistical programming language (for example, R). Approach through management projects and data analysis and simulations.
Faculty or entity in charge	SC

Programmes containing this learning unit (UE)

Program title	Acronym	Credits	Prerequisite	Aims
Bachelor in Physics	PHYS1BA	5		🔍
Bachelor in Mathematics	MATH1BA	5		🔍
Minor in Statistics and data sciences	LSTAT100I	5		🔍
	LSTAT100P	5		🔍
	LOSTA100I	5		🔍