UCLouvain

lsinc1109

2026

Statistics and data sciences

5.00 credits	30.0 h + 30.0 h	Q2

This learning unit is not open to incoming exchange students!

Teacher(s)	Massart Estelle ;
Language :	French
Place of the course	Charleroi
Prerequisites	To follow this course the student must have a basic knowledge of probabilities such as taught in courses **LSINC1211** , **LEPL1108** or **LBIR1212** . *The prerequisite(s) for this Teaching Unit (Unité d'enseignement – UE) for the programmes/courses that offer this Teaching Unit are specified at the end of this sheet.*
Main themes	This course presents the fundamental statistical concepts in an engineering context (exploratory analysis, inference, simulation) as well as basis method for analysing multivariate databases (like the linear regression, the principal component analysis and the classification).
Learning outcomes	**At the end of this learning unit, the student is able to :** · Explore datasets of small and big sizes with few or many dimensions; · Infer features of a population from a sample using techniques of inference, estimation, confidence intervals and statistical tests; · To connect the deductive approach from the probability theory to the statistical inductive approach, and to identify the probabilistic models used in statistical inference; · To translate the textual formulation of a problem of statistical inference into an accurate, statistical and mathematical formalism, while recognizing the adequate models and corresponding estimation methods; · To solve an applied problem by following a logical approach based on a correct use of models and statistical inference; · To use techniques of Monte-Carlo simulations, K-fold cross validation and bootstrapping in order to estimate models and validate results; · To analyse multivariate data with fundamental methods of linear regressions, of principal component analysis and of classification/clustering; · To use statistical tools to validate the conclusions from a model e.g. like the linear regression; · To make the link between the mathematical objectives of a method of data mining and its practical purposes.
Evaluation methods	Individual exam and activities realised during the term (hackathons). These hackathons are organised (and assessed) only once during the term, and represent 25% of the final mark for each session. Failure to comply with the methodological instructions communicated by the teacher, particularly with regard to the use of online resources or collaboration between students, in a hackathon will result in an overall mark of 0 for the hackathons. In particular, the use of generative AI tools and any collaboration is strictly prohibited for the hackathons. The distribution or exchange between students of (fragments of) code is not allowed by any means (GitHub, Facebook, Discord...), and this even after the deadline for submission of hackathons.
Teaching methods	The course is composed of: - lectures on the topics listed in the course content; - practical sessions, both classical and numerical; - 4 hackathons, representing 2 x 2 hours each, associated with small Python projects realized in group on subjects discovered both in the lectures and in the practical sessions.
Content	- Exploratory analysis and sampling - Introduction to multivariate data analysis - Parametric estimate (methods of moments and log-likelihood maximization) and properties of estimators (bias, variance, mean-squared error). - Statistical inference (confidence intervals and significance tests): comparison of means of two or several normal populations, proportions, variance testing. - Linear regression, including the analysis of coefficients and significance tests. - Panorama of learning techniques, supervised and unsupervised learning methods

	- Links between objectives of data analysis methods and their mathematical representation. - Regression and classification methods (such as linear models and least square, k-nearest neighbors, logistic regression) - Training, test error and generalization error, the Bias-Variance tradeoff, and elements of statistical decision theory - Resampling techniques for model selection/evaluation (e.g., validation set, K-fold cross validation, bootstrap) - Unsupervised learning: reduction of dimension (principal component analysis) and methods of clustering (K-means).
Inline resources	The totality of teaching material is available on the companion moodle website of the course. The schedule of course is subject to modification due to sanitary conditions, please consult the Moodle website of the course for additional information.
Other infos	To follow this course the student must have a basic knowledge of probabilities such as taught in courses LSINC1211, **LEPL1108** or **LBIR1212** . The schedule of course is subject to modifications. Please check the Moodle website for more details.
Faculty or entity in charge	SINC

Programmes containing this learning unit (UE)				
Program title	Acronym	Credits	Prerequisite	Learning outcomes
Bachelor in Computer Science	SINC1BA	5		🔍