UCLouvain

lstat2110

2018

Analyse des données

5 crédits	22.5 h + 7.5 h	Q1

Enseignants	Segers Johan ;
Langue d'enseignement	Français
Lieu du cours	Louvain-la-Neuve
Préalables	· calcul vectoriel et matriciel · géométrie euclidienne: points, espaces, orthogonalité, distances, angles · notions de base en statistique: moyenne, (co)variance, corrélation, matrice de covariance, probabilité conditionnelle, distribution normale, distribution khi-carré
Thèmes abordés	- Rappels d'algèbre et de géométrie utiles à l'analyse des données.. - Principes de base des méthodes factorielles. - Analyse en composantes principales et ses variations. - Analys edes corrélations canoniques. - Analyse factorielle discriminante. - Analyse factorielle des correspondances. - Introduction aux méthodes de classification. - L'analyse des données, en pratique.
Acquis d'apprentissage	1 A. Eu égard au référentiel AA du programme de master en statistique, orientation générale, cette activité contribue au développement et à l'acquisition des AA suivants, de manière prioritaire : 1.1, 1.3, 2.2, 3.3 Eu égard au référentiel AA du programme de master en statistique, orientation biostatistique, cette activité contribue au développement et à l'acquisition des AA suivants, de manière prioritaire : 1.1, 1.3, 2.2, 3.3 B. Objectifs généraux: Présenter les techniques modernes de l'analyse de grands ensemble de données et développer les outils de base du " data mining ". Objectifs spécifiques: A l'issue de ce cours, les étudiants doivent être capables de : - Traiter et décrire l'information contenue dans des grands ensemble de données ; - Comprendre les mécanismes qui justifient l'emploi de telle ou telle méthode ; - Interpréter correctement les graphiques et résultats fournis par les logiciels ; - Résoudre des problèmes avec données réelles. - - - - *La contribution de cette UE au développement et à la maîtrise des compétences et acquis du (des) programme(s) est accessible à la fin de cette fiche, dans la partie « Programmes/formations proposant cette unité d'enseignement (UE) ».*
Modes d'évaluation des acquis des étudiants	Tests lors des cours magistraux: • Test 1: Matrices de données et analyse en composantes principales • Test 2: Classification et Analyse discriminante linéaire Participation optionnelle. Au choix de l'étudiant, chaque test peut remplacer la partie de l'examen sur le même sujet. Examen (12/20): • par écrit, à livre fermé, à l'aide d'un formulaire et d'une calculatrice • exercices et questions de calcul, d'interprétation de sortie de logiciel, et de compréhension des formules et des résultats principaux Projet (8/20): • seul ou en binôme • application des méthodes sur une base de donnée apportée par l'étudiant lui-même • rapport écrit en R Markdown, à soumettre avant la session d'examens • consignes détaillés lors des TP et sur la page MoodleUCL du cours
Méthodes d'enseignement	Lors des cours magistraux, l'enseignant présente les différentes méthodes d'analyse, couvrant à la fois leur champ d'application, la théorie mathématique sous-jacante, et la programmation en R. Des devoirs sont donnés dont la solution est discuté lors des cours magistraux aussi. Les TP se déroulent en salle informatique et servent surtout pour permettre aux étudiants de s'entraîner à l'application des méthodes d'analyse en R et sur des vrais jeux de données.
Contenu	Français • Matrices de données • Analyse en composantes principales • Classification: moyennes mobiles et classification hiérarchique • Analyse discriminante linéaire • Analyse des correspondances simple et multiple

	Les méthodes sont réalisés dans le language R via le logiciel RStudio, et R Markdown est utilisé pour la réalisation de rapports comprenant à la fois du texte, des développements mathématiques, du code R et les résultats des analyses (tables, figures).
Ressources en ligne	Toutes les ressources nécessairs pour le cours sont fournies sur la page MoodleUCL du cours: les dias des cours magistraux et des TP, les script informatiques, les exercices. En plus, des liens vers des ressources électroniques externes sont fournis aussi: des cours en ligne, des vidéos, de la documentation du logiciel utilisé.
Bibliographie	• Escofier, B. et Pagès, J. (2016): Analyses factorielles simples et multiples, 5e édition, Dunod, Paris. • Lebart, L., Piron, M. et Morineau, A. (2006): Statistique exploratoire multidimensionnelle, 4e édition, Dunod, Paris. • Saporta, G. (2011): Probabilités, analyse des données et statistique, 3e édition révisée, Editions TECHNIP, Paris.
Autres infos	Prérequis: • calcul vectoriel et matriciel • géométrie euclidienne: points, espaces, orthogonalité, distances, angles • notions de base en statistique: moyenne, (co)variance, corrélation, matrice de covariance, probabilité conditionnelle, distribution normale, distribution khi-carré
Faculté ou entité en charge:	LSBA

Programmes / formations proposant cette unité d'enseignement (UE)

Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en sciences mathématiques	MATH2M	5		
Master [120] en statistique, orientation biostatistiques	BSTA2M	5		
Master [120] en sciences économiques, orientation générale	ECON2M	5		
Master [120] : ingénieur civil biomédical	GBIO2M	5		
Master [120] en statistique, orientation générale	STAT2M	5		
Master [120] : ingénieur civil en mathématiques appliquées	MAP2M	5		
Master [120] en science des données, orientation statistique	DATS2M	5		
Mineure d'accès en statistique et sciences des données	LSTAT100I	5		
Approfondissement en statistique et sciences des données	LSTAT100P	5		