







5.00 crédits	30.0 h + 7.5 h	Q1
--------------	----------------	----

Enseignants	Caelen Olivier ;
Langue d'enseignement	Français
Lieu du cours	Louvain-la-Neuve
Préalables	Concepts et outils équivalents à ceux enseignés dans les UE LSTAT2011 Eléments de mathématique pour la statistique LSTAT2014 Eléments de probabilités et de statistique mathématique
Thèmes abordés	<p>Le cours présente un survol des principaux outils de l'analyse exploratoire de données multivariées via les méthodes factorielles. Les données sont projetées sur un sous-espace de faible dimension tout en gardant un maximum d'information. Cette réduction de dimension facilite la visualisation et aide à la découverte de l'information et des tendances dans un tableau de données.</p> <ul style="list-style-type: none"> • Rappels d'algèbre et de géométrie utiles à l'analyse des données • Principes de base des méthodes factorielles • Analyse en composantes principales • Classification: moyennes mobiles et classification hiérarchique • Analyse discriminante linéaire • Analyse des correspondances simple et multiple • Régression sur composantes principales • Régression des moindres carrés partiels
Acquis d'apprentissage	<p>A la fin de cette unité d'enseignement, l'étudiant est capable de :</p> <p>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</p> <p>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</p> <p>1 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.</p>
Modes d'évaluation des acquis des étudiants	<p>Examen (12/20):</p> <ul style="list-style-type: none"> • 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 <p>Evaluation continue - tests lors des cours magistraux:</p> <ul style="list-style-type: none"> • Test 1: Matrices de données et analyse en composantes principales • Test 2: Classification et Analyse discriminante linéaire <p>Participation optionnelle. Au choix de l'étudiant, chaque test peut remplacer la partie de l'examen sur le même sujet.</p> <p>Evaluation continue - projet (8/20):</p> <ul style="list-style-type: none"> • seul ou en binôme • application des méthodes sur une base de donnée apportée par l'étudiant lui-même • rapport écrit, à soumettre avant une ou plusieurs dates spécifiées pendant le cours • consignes détaillés lors des TP et sur la page MoodleUCL du cours <p>La participation au projet est obligatoire pour obtenir une note d'examen. Le projet peut être refait lors d'une deuxième inscription à l'examen.</p>

Méthodes d'enseignement	<p>Lors des cours magistraux, l'enseignant présente les différentes méthodes d'analyse, couvrant à la fois leur champ d'application, la théorie mathématique sous-jacente, et la programmation en R. Des devoirs sont donnés dont la solution est discutée lors des cours magistraux aussi.</p> <p>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.</p>
Contenu	<ul style="list-style-type: none"> • Matrices de données • Analyse en composantes principales • Classification: moyennes mobiles et classification hiérarchique • Analyse discriminante linéaire • Analyse des correspondances simple et multiple • Régression sur composantes principales • Régression des moindres carrés partiels <p>Les méthodes sont réalisés dans le langage 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).</p>
Ressources en ligne	<p>Toutes les ressources nécessaires pour le cours sont fournies sur la page MoodleUCL du cours: les diapos 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.</p>
Bibliographie	<ul style="list-style-type: none"> • 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	<p>Préalables :</p> <ul style="list-style-type: none"> • 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 science des données, orientation statistique	DATS2M	5		
Master [120] : ingénieur civil biomédical	GBIO2M	5		
Master [120] en statistique, orientation biostatistiques	BSTA2M	5		
Master [120] en sciences mathématiques	MATH2M	5		
Master [120] en sciences actuarielles	ACTU2M	5		
Master [120] en statistique, orientation générale	STAT2M	5		
Master [120] : bioingénieur en chimie et bioindustries	BIRC2M	5		
Approfondissement en statistique et sciences des données	APPSTAT	5		
Master [120] : ingénieur civil en mathématiques appliquées	MAP2M	5		
Master [120] en sciences économiques, orientation générale	ECON2M	5		
Mineure en statistique, sciences actuarielles et science des données	MINSTAT	5		
Certificat d'université : Statistique et science des données (15/30 crédits)	STAT2FC	5		