

Institut de statistique



STAT2411A Analyse multivariée des données.

[15h+7.5h exercices] 3 crédits

Enseignant(s): Léopold Simar
Langue d'enseignement : français
Niveau : Second cycle

Objectifs (en termes de compétences)

Objectifs Généraux.

Présenter les techniques modernes de l'analyse de grands ensembles 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 ensembles 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.

Objet de l'activité (principaux thèmes à aborder)

Contenu

- 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
- Analyse des corrélations canoniques
- Analyse factorielle discriminante
- Analyse factorielle des correspondances
- Introduction aux méthodes de classification
- L'analyse des données, en pratique

Autres informations (Pré-requis, Evaluation, Support, ...)

Prérequis:

L'étudiant doit être capable de manipuler et lire les expressions algébriques (calcul matriciel) ; comprendre et dominer les éléments de base de l'analyse statistique.

L'évaluation se fait :

- 1) par un travail sur données réelles selon les modalités précisées ci-dessous. Il s'agit de mettre en œuvre certaines des méthodes vues au cours dans un domaine d'application choisi par l'étudiant. Pour permettre aux étudiants de réaliser ce travail dans les meilleures conditions, le cours magistral sera concentré sur 10 semaines. Les étudiants travaillent, en principe, par paire. L'assistant du cours encadrera les étudiants pour ce travail (mise au courant du logiciel). Ce travail devrait prendre environ 12 heures de travail PAR étudiant (soit 24 h. pour la paire).
- 2) Par un examen écrit à livre fermé : il s'agira ici de voir si l'étudiant maîtrise les concepts abordés au cours, s'il comprend les méthodes utilisées (questions d'ordre général mais aussi commentaires sur des expressions matricielles importantes) et s'il peut interpréter correctement des résultats obtenus par les logiciels (du type de ceux présentés dans le syllabus).

Support:

Syllabus de L.SIMAR (2004) : " Multivariate Data Analysis ", 256 pages, Institut de Statistique, UCL.

Ce manuel est disponible à la DUC.

Programmes proposant cette activité**BIR2** Bio-ingénieur**Autres crédits de l'activité dans les programmes**

BIR22/0A	Deuxième année du programme conduisant au grade de bio-ingénieur: Sciences agronomiques (Technologies et gestion de l'information)	(2 crédits)	Obligatoire
BIR22/0C	Deuxième année du programme conduisant au grade de bio-ingénieur: chimie et bio-industries (Technologies & gestion de l'information)	(2 crédits)	Obligatoire
BIR22/0E	Deuxième année du programme conduisant au grade de bio-ingénieur: Sciences et technologies de l'environnement (Technologies et gestion de l'information)	(2 crédits)	Obligatoire
BIR22/4E	Deuxième année du programme conduisant au grade de bio-ingénieur : Sciences et technologie de l'environnement (Technologies environnementales: eau, sol, air)	(2 crédits)	Obligatoire
BIR22/5E	Deuxième année du programme conduisant au grade de bio-ingénieur : Sciences et technologie de l'environnement (Aménagement du territoire)	(2 crédits)	Obligatoire
BIR22/6E	Deuxième année du programme conduisant au grade de bio-ingénieur : Sciences et technologie de l'environnement (Nature, eau & forêts)	(2 crédits)	Obligatoire
BIR22/7E	Deuxième année du programme conduisant au grade de bio-ingénieur : Sciences et technologie de l'environnement (Ressources en eau et en sol)	(2 crédits)	Obligatoire