

Analyse des données

Infos pratiques

- > ECTS : 3,0
- > Nombre d'heures : 30,0
- > Langue(s) d'enseignement : Français
- > Niveau d'étude : BAC +4
- > Période de l'année : Enseignement septième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral et Travaux dirigés
- > Ouvert aux étudiants en échange : Oui
- > Campus : Campus de Nanterre
- > Composante : Sciences économiques, gestion, mathématiques et informatique
- > Code ELP : 5E7DSADN

Présentation

Ce cours présente les principales méthodes statistiques d'analyse de données, parfois en grande dimension :

- Rappels de statistique descriptive univariée et bivariée
- Analyse en composantes principales
- Analyse factorielle des correspondances
- Méthodes de classification non supervisée

Les méthodes sont mises en oeuvre sur R. Les étudiants sont encouragés à présenter les exercices sous forme d'exposé oral.

Objectifs

- Maîtriser les bases théoriques de l'analyse des données.
- Maîtriser leur mise en oeuvre informatique.
- Interpréter les résultats à l'issue d'une étude statistique.
- Identifier des groupes d'observations ou de variables.
- Réduire la dimension des observations ou des variable.

Évaluation

Modalités : CC

SESSION 1 :

Contrôle Continu

- Type : Écrit, Oral
- Durée : TP notés
- Précisions : Le CC peut comporter une épreuve écrite; les travaux réalisés en TP sont évalués ainsi que les éventuelles présentations orales des étudiants de leurs travaux. L'ensemble constitue 50% de la note finale.

Partiel

- Type : Écrit
- Durée : 3h00
- Précisions : Le partiel effectué en fin de cours consiste en une épreuve sur ordinateur (logiciel R) avec pour objectif de répondre à des questions portant sur un jeu de données. Le rendu est un fichier qui doit détailler les résultats obtenus avec leur interprétation.

Régime Dérogatoire

- Type : Écrit
- Durée : 3h00
- Précisions : L'épreuve est réalisée sur ordinateur avec le logiciel R; elle a pour objectif de répondre à des questions portant sur un jeu de données. Le rendu est un fichier qui doit détailler les résultats obtenus avec leur interprétation.

SESSION 2 :

- Type : Écrit
- Durée : 3h00
- Précisions : L'épreuve est réalisée sur ordinateur avec le logiciel R; elle a pour objectif de répondre à des questions portant sur un jeu de données. Le rendu est un fichier qui doit détailler les résultats obtenus avec leur interprétation.

Utilisation de l'intelligence artificielle :

Dans le cadre de cet EC, l'usage de l'IA pour aider à la réalisation des travaux de contrôle continu de type devoir sur table *est interdit*.

Pré-requis nécessaires

Statistique descriptive et inférentielle, bases d'algèbre linéaire.

Compétences visées

A l'issue du cours, les étudiants sont capables de visualiser les données, et mettre en œuvre des méthodes pour les manipuler afin de mieux les comprendre, identifier des groupes d'observations ou de variables.

Bibliographie

- Analyse de données avec R (2e édition)
Auteurs : François Husson, Sébastien Lê, Jérôme Pagès
- Statistique exploratoire multidimensionnelle
Auteurs : Ludovic Lebart, Alain Morineau, Marie Piron
- An Introduction to Statistical Learning (with Applications in R)
Auteurs : Gareth James, Daniela Witten, Trevor Hastie et Robert Tibshirani.

Ressources pédagogiques

Diapos de cours ; modèles de codes R

Contact(s)

- > **Cecile Hardouin ceccantini**
Responsable pédagogique
hardouin@parisnanterre.fr