

Analyse des données

Infos pratiques

- > ECTS : 3.0
- > Nombre d'heures : 36.0
- > Période de l'année : Enseignement huitième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral et Travaux dirigés
- > Ouvert aux étudiants en échange : Oui
- > Composante : Sciences économiques, gestion, mathématiques et informatique

Présentation

L'objectif de ce cours/TP est d'introduire les principales méthodes pour l'analyse de données. Il sera présenté trois méthodes factorielles et deux méthodes de classification non supervisée. L'objectif est que les étudiants soient capables de mener une analyse descriptive des données : manipulation, nettoyage, visualisation et analyse des données avec la méthode appropriée. Le logiciel R et l'environnement RStudio seront utilisés.

Plan du cours :

- * Introduction à Rmarkdown et à tidyverse
- * Rappel sur l'analyse univariée et bivariate
- * Analyse en composantes principales (ACP)
- * Classification non-supervisée (CAH et K-means)
- * Analyse factorielle des correspondances (ACM)
- * Analyse factorielle discriminante (AFD)

Objectifs

- * Acquérir les méthodes d'analyse des données.
- * Être capable de faire une analyse descriptive des données, d'identifier des questions concrètes et savoir interpréter les résultats fournis par le logiciel R.

- * Réaliser des rapports automatisés avec Rmarkdown et apprendre les récentes extensions (dplyr, tidyverse, ggplot2).

Évaluation

- * Un projet (50%)
- * Un examen (50%)

Bibliographie

- * Philippe Besse. Statistique exploratoire multidimensionnelle (<https://www.math.univ-toulouse.fr/~besse/enseignement.html>)
- * François Husson et al. *R pour la statistique et la science des données*