

# Analyse de Données



## Présentation

---

### Description

**Programme :**

- I. Analyse factorielle
  - 1. Analyse en composantes principales
  - 2. Analyse factorielle de correspondance - Analyse de correspondances multiples
  - 3. Décomposition en valeurs singulières
- II. Régression linéaire
  - 1. Régression simple
  - 2. Régression multiple
  - 3. ANOVA (Analyse de la Variance)
- III. Classification
  - 1. Classification paramétrique (classification hiérarchique)
  - 2. Classification non-paramétrique (k-means, arbres de décision)

**Logiciels :**

Les travaux pratiques de ce module seront réalisés avec Python ou R.

---

### Objectifs

- 1. Entreprendre une analyse factorielle complète
- 2. Appliquer les méthodes de réduction de la dimension de l'information statistique
- 3. Analyser et décrire les relations entre plusieurs variables statistiques mesurées sur un grand nombre d'individus
- 4. Classifier des individus statistiques
- 5. Calibrer un modèle statistique et évaluer l'ajustement par validation croisée

## Pré-requis obligatoires

Module SM1, licence de mathématiques plus particulièrement les modules d'algèbre linéaire (calcul matriciel).

---

## Bibliographie

1. Probabilités, analyse des données et Statistique. Gilbert Saporta. Troisième édition, Technip, 2011.
2. Analyse de données avec R. François Husson, Sébastien Lê, Jérôme Pagès. Presses Universitaires de Rennes (PUR), 2016.
3. <https://scikit-learn.org/stable/Documentation> de scikit-learn