

# Statistiques avancées pour Big Data

## Infos pratiques

---

- > ECTS : 3,0
- > Nombre d'heures : 21.0
- > Langue(s) d'enseignement : Français
- > Niveau d'étude : BAC +5
- > Période de l'année : Enseignement neuvième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral
- > Ouvert aux étudiants en échange : Oui
- > Composante : Sciences économiques, gestion, mathématiques et informatique
- > Code ELP : 4EgGASAI

## Présentation

---

Le cours a pour objectif de présenter les méthodes économétriques destinées au Big Data i.e. quand l'économètre possède plus de variables explicatives que d'observation. Il sera divisée en trois modules. Le premier module sera consacré aux régressions Ridge et Lasso. Le second présentera diverses méthodes d'apprentissage machine (Arbre de classification, forêts aléatoires, SVM, Réseaux de neurones). Le troisième sera consacré à l'analyse de données non numériques et des textes notamment.

L'évaluation se fera sur une examen écrit et un dossier portant sur l'un des trois modules.

## Objectifs

---

- Découverte des régression pénalisées dans le cadre de grandes bases de données
- Mise en application des logiciels statistiques pour des problèmes Big Data
- Utilisation des algorithmes d'apprentissage machine dans un but prédictif
- Connaître les limites des statistiques sur des données textuelles

## Évaluation

---

Session 1 : évaluation sur projet

Session 2 : écrit, oral ou dossier

## Pré-requis nécessaires

---

Économétrie paramétrique, Modélisation statistiques, Python niveau débutant

## Compétences visées

---

- Savoir quel modèle utiliser pour un problème donné
- Utiliser les logiciels statistiques pour mettre en oeuvre les différents modèles
- Savoir entraîner une algorithmme d'apprentissage machine

## Bibliographie

---

- Econometrics and Machine Learning, A. Charpentier, E. Flachaire and A. Ly, Economics and Statistics, 2018, 505, 147-169
- Python Data Science Handbook, Jake Vanderplas
- The Elements of Statistical Learning, J.H. Friedman, R.Tibshirani et T.Hastie, Springer 2009

## Ressources pédagogiques

---

Classe interactive et Travail sur ordinateur