

# Données, Big Data et analyse économique empirique

## Infos pratiques

---

- > ECTS : 3,0
- > Nombre d'heures : 24,0
- > Langue(s) d'enseignement : Français
- > Période de l'année : Enseignement dixième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral
- > Ouvert aux étudiants en échange : Oui
- > Composante : Sciences économiques, gestion, mathématiques et informatique

## Présentation

---

La recherche empirique en économie est devenue centrale, la part des articles théoriques a décliné drastiquement. La grande majorité des travaux qui circulent aujourd'hui s'appuie sur des bases de données, qu'elles soient déjà existantes (recensements, bases de données administratives, enquêtes ménages des services statistiques, « big data ») ou construites par le chercheur (analyses textuelles, enquêtes et sondages, expérimentations en laboratoire, expérimentations sur le terrain, ...). Dans le déroulement d'une recherche, le temps pour identifier les sources répondant à la problématique, se les procurer, les mettre en forme, construire les variables pertinentes, avant de passer à l'analyse économétrique proprement dite est une étape essentielle, lourde et qui conditionne la qualité de l'analyse économétrique. Les « Big data », par rapport à l'analyse économétrique standard, mettent davantage l'accent sur la prédiction à partir de modèles non paramétriques que sur l'identification des mécanismes causaux.

Plan du cours :

### 1. A la base, les bases

L'individu de base n'est pas toujours humain :  
l'architecture des bases de données

Des données aux variables : la construction des catégories du raisonnement – sous contraintes

Tout est-il accessible à tous ?

Accéder aux données : Le coût des données, la recherche de financements, le coût de la recherche de financement

Faire ses bases de données : l'expérimentation, les enquêtes à partir des réseaux sociaux, l'éthique dans les expérimentations

### 1. *Les big data : l'ancien et le nouveau.*

Quelles méthodes pour quelles questions et pour quelles données ? Quel logiciel choisir ?

Qu'entend-on par big data ? Qu'apportent-elles par rapport aux grandes bases classiques ?

Classifier et prédire – Neural network, ACM, Decision and regression trees, Random Forest

Des traitements économétriques paramétriques aux traitements non paramétriques

Quantifier l'écrit : utiliser les réseaux et les textes comme données

### 1. *La restitution : comment le chercheur peut-il sortir de sa tour sans passer au pilori ?*

Revenir devant les producteurs de données

La visualisation des résultats : ce n'est pas un détail

L'article scientifique, l'abstract, la diffusion dans le grand public, l'interview

## Objectifs

---

Ce cours propose une introduction aux différentes techniques d'analyse des bases de données, des difficultés, apports et limites dans leur appropriation par les recherches empiriques actuelles. L'accent est mis sur la présentation des grandes bases de données de référence, françaises, européennes et internationales, et des choix des techniques économétriques appropriées en fonction des données et des questions de recherche. Il s'appuie sur l'étude d'articles de recherche en économie mettant en œuvre un ensemble varié de techniques.

## Évaluation

---

Réalisation d'un dossier avec exploitation d'une base de données

## Pré-requis nécessaires

---

Avoir des notions statistiques de base acquises en M1 SES ou équivalent

## Compétences visées

---

Savoir construire/exploiter une base de données ; identifier les techniques appropriées à la question de recherche

## Bibliographie

---

Hastie, T., & Tibshirani, R. & Friedman, J.(2008). *The Elements of Statistical Learning; Data Mining, Inference and Prediction*. New York: Springer

Dean, J. (2014). *Big data, data mining, and machine learning: value creation for business leaders and practitioners*. John Wiley & Sons.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, pp. 3-7). New York: Springer.

## Ressources pédagogiques

---

Le cours et les articles étudiés à chaque séance sont disponibles sur cours en ligne