

Atelier Statistique

Clustering :
**Une vision unifiée pour
une utilisation éclairée**

PARIS (I.H.P.)

15 et 16 juin 2023
9h00-12h30 et 14h00-17h30

Tél : (33) 01 44 27 66 60

E-mail : sfds.gestion@ihp.fr
Site internet : <http://www.sfds.asso.fr>

Depuis 2001, la SFdS organise des Ateliers en **science des données** destinés avant tout aux ingénieurs, chercheurs, doctorants et post-doctorants. Ces Ateliers sont originaux dans le sens où ils abordent des sujets porteurs et/ou nouveaux.

Le second Atelier de l'année 2023 portera sur la **classification non supervisée et les modèles de mélange**. La statistique est la science de la synthèse des données : il s'agit de fournir une meilleure compréhension des données et de conduire à des décisions éclairées. Le *clustering* (ou classification non supervisée) est l'un de ses principaux paradigmes, **combinant une réduction extrême des données à une extraction de sens**. Dans le contexte contemporain d'abondance et de complexité croissante des données, cet atelier donnera les clefs pour un **usage raisonné et utile** de méthodes de *clustering* modernes et performantes.

L'Atelier débutera par une journée portant sur les **méthodes exploratoires** ainsi que sur l'**évaluation d'une méthode de clustering**. La deuxième journée sera consacrée à la **formalisation par modèles de mélange** et au traitement de la **grande dimension**.

Chaque partie du cours sera illustrée par **des travaux pratiques sur données réelles** avec le logiciel R et la présentation des packages R comme Mixmod, Mixtcomp, blockcluster.

L'atelier sera animé par **Christophe Biernacki** (INRIA)

Jeudi 15 Juin 2023

Méthodes exploratoires en clustering

- ✓ Genèse et contemporanéité en *data science*
- ✓ Kmeans, classification hiérarchique (CAH)
- ✓ Autres méthodes : noyaux, spectral,...
- ✓ Mise en œuvre sur données réelles avec R

Evaluation d'une méthode de clustering

- ✓ Importance des données et des dissimilarités
- ✓ Effet des algorithmes et du nombre de classes
- ✓ Interprétabilité par l'utilisateur
- ✓ Mise en œuvre sur données réelles avec R

Vendredi 16 Juin 2023

Formalisation par modèles de mélange

- ✓ Lien entre Kmeans/CAH et mélanges de lois
- ✓ Estimation et sélection en mélanges de lois
- ✓ Ouverture vers des pistes de recherche
- ✓ Mise en œuvre R : Mixmod, Mixtcomp

Traitement de la grande dimension (HD)

- ✓ Bénédiction/malédiction de la HD en *clustering*
- ✓ *Clustering(s)* HD : réduction dim., parsimonie
- ✓ Focus sur le *co-clustering*
- ✓ Mise en œuvre R : ACP/Kmeans, Blockcluster

Droits d'inscription

Tarif général (ex : établissements privés, EPIC)

2 jours de formation au choix 700 €*

Salarié(e)s des universités et des établissements publics d'enseignement ou de recherche

2 jours de formation au choix 400€*

Doctorant(e)s

2 jours de formation au choix 350€*

Étudiant(e)s non-salarié(e)s (ex : master)

2 jours de formation au choix 150€*

Les droits d'inscription au cours comprennent :

- la participation à la formation
- les documents de cours
- les pauses café
- les repas du midi

Modalités d'inscription

L'inscription aux Ateliers Statistiques SFdS se fait désormais **uniquement en ligne.**

Pour vous inscrire à cet Atelier, merci de compléter impérativement le formulaire suivant :

<https://framaforms.org/2023-juin-ateliers-statistiques-1673511312>

A la suite de votre inscription, vous recevrez une confirmation par voie électronique ainsi qu'une facture pour le règlement et une convention de formation à signer.

Important

Pour des raisons juridiques, les ateliers doivent s'adresser uniquement aux membres de la SFdS. Les non-membres de l'association doivent s'acquitter de leur adhésion pour l'année en cours pour pouvoir participer aux formations. Toutes les informations utiles pour adhérer à la SFdS sont disponibles ici :

https://www.sfds.asso.fr/fr/722-adherer_a_la_sfds/

Annulation

Le remboursement des droits d'inscription sera possible sur demande écrite transmise **avant le 8 juin 2023** et donnera lieu à une retenue de 150 € pour frais. Aucun remboursement ne sera effectué après cette date.

Organisation

Le cours se déroulera à l'IHP, 11 rue Pierre et Marie Curie 75005 Paris.

Accès : RER ligne B station : Luxembourg

Bus : 21, 27, 38, 84, 85, 89

Afin d'assurer une plus grande convivialité à la formation et de favoriser les échanges entre les participants, **le nombre de places est limité à 25.**

Date limite d'inscription : 8 juin 2023