

# Data Science



## Présentation

**Code interne :** EI7IF252

### Informations complémentaires

Dans le domaine de la science des données, des chercheurs perçoivent l'émergence de trois communautés professionnelles: (i) la gestion de bases de données, (ii) les statistiques et le machine learning convertissant les données en connaissances et (iii) les systèmes informatiques permettant le traitement efficaces de ces données.

Ce cours se concentrera sur le point (ii) d'un point de vue d'informaticien en s'appuyant sur des hypothèses de confort pour les points (i) et (iii). Seront abordés avec différents niveaux de détails:

les démarches scientifiques, dont la modélisation et la conception d'expérience,

sous l'angle de l'analyse d'algorithmes probabilistes : les statistiques descriptives, les lois de probabilités classiques, les estimateurs, l'inférence statistique : les paradigmes fréquentiste ou bayésien, les tests statistiques,

l'inférence causale, incluant une discussion sur cause et corrélation, le Structural Causal Model (SCM) de Pearl,

la visualisation (sommaire) de données et (si le temps le permet) un peu d'analyse topologique de données,

l'éthique de l'analyse de données (biais, condition d'expérimentation,...).

### Modalités de contrôle des connaissances

#### Évaluation initiale / Session principale - Épreuves

Type d'évaluation	Nature de l'épreuve	Durée (en minutes)	Nombre d'épreuves	Coefficient de l'épreuve	Note éliminatoire de l'épreuve	Remarques
Contrôle Continu Intégral	Contrôle Continu			1		