



4BM019 (3 ECTS)

Biostatistique et Bioanalyse sous R

Initiez-vous à l'analyse de
données biologiques sous R
à travers des cours et des projets

1^{er} semestre 2025

Jeudi 4 septembre de 14h à 17h (CM)
Lundi 8 septembre de 9h à 12h (CM)
Mardi 9 septembre de 9h à 12h (CM)
Mercredi 10 septembre de 9h à 12h (CM)

Jeudi 11 septembre de 9h00 à 12h45 (TD)

Lundi 27 octobre de 9h à 18h (TP, projets)
Mardi 28 octobre de 9h à 18h (TP, projets)

Objectifs pédagogiques

Poussée par les révolutions biotechnologiques successives, l'analyse de données est aujourd'hui une compétence essentielle dans l'ensemble des métiers de la biologie moderne.

Cette unité d'enseignement a pour objectif de former les étudiants de 1^{ère} année du master BMC à l'analyse des données biologiques en utilisant la programmation en R, afin qu'ils soient capables d'interpréter des données de faible ou de grande dimension.

Les étudiants acquerront des compétences en biostatistiques (statistique descriptive et inférentielle) et en bioanalyse (analyses univariées, multivariées et intelligence artificielle), essentielles dans différents contextes des sciences du vivant. Ces compétences leur permettront de concevoir des stratégies d'analyse pour leurs projets académiques et leurs futurs métiers, d'être autonomes dans leur démarche scientifique, et de communiquer et discuter les résultats obtenus.

Programme pédagogique

Cette unité d'enseignement s'organise en quatre demi-journées de cours magistraux, une demi-journée de travaux dirigés, et deux jours de travaux pratiques sous forme de projets à réaliser en groupes d'étudiants.

Le programme pédagogique inclut les éléments suivants :

- Présentation de l'environnement de développement RStudio.
- Introduction au langage R, incluant la syntaxe, les variables et les fonctions.
- Utilisation de R pour manipuler, nettoyer et analyser des jeux de données biologiques.
- Introduction aux notions fondamentales de la statistique descriptive.
- Introduction aux notions fondamentales de la statistique inférentielle.
- Analyse de survie et génération de courbes de Kaplan-Meier.
- Comparaisons d'échantillons biologiques avec des outils d'analyse différentielle.
- Exploration des relations entre variables : corrélations et régressions linéaires.
- Analyses multivariées : ACP, clustering, et réduction de dimensionnalité.
- Création de visualisations graphiques avancées avec ggplot2.
- Présentation du logiciel GraphPad Prism.
- Introduction aux concepts de base en intelligence artificielle et premières applications en R.
- Initiation aux réseaux de neurones pour la classification de données.
- Introduction à l'analyse de données transcriptomiques.
- Communication des résultats et création de rapports automatisés avec R Markdown.

Organisation

Cette unité d'enseignement de 3 ECTS se déroulera au premier semestre de l'année universitaire.

Les modalités de contrôle des connaissances seront les suivantes :

- Un contrôle des connaissances sous forme de QCM (33 % de la note finale).
- Un projet d'analyse de jeux de données réalisé en groupe (CC2, 33 % de la note finale).
- Un contrôle des compétences sous forme de problèmes à résoudre (CC3, 34 % de la note finale).

Le projet consistera en l'analyse de trois jeux de données parmi un large choix et sera réalisé en équipes de trois étudiants. L'évaluation comprendra la restitution d'un rapport d'analyse ainsi que les scripts d'analyse. Les étudiants pourront également proposer leurs propres jeux de données à analyser et interpréter, sous réserve de validation par l'enseignant.