

# Plan de cours : Science des données II : analyse et modélisation

année académique 2023-2024

## Enseignants

- Enseignant : Philippe Grosjean
- Assistant : Guyliann Engels

## Présentation du cours

La science des données s'intéresse à l'analyse des données sous toutes ses formes. Cette discipline requiert des compétences en informatique, en mathématiques et dans le domaine d'expertise. La Science fait actuellement face à de nombreux changements suite à l'introduction de l'Open Science, de l'Open Data, de l'augmentation exponentielle des données, ... Les biologistes doivent être capables d'évoluer dans ce contexte grâce à une bonne connaissance de la science des données biologiques.

Ce second cours s'inscrit dans une formation continue sur trois années successives. À la fin de ce second cours, les étudiants seront capables de modéliser, de produire des analyses multivariées et de présenter clairement leurs résultats de manière reproductible avec des outils informatiques et statistiques professionnels.

La fiche ECTS reprend toutes les informations essentielles : <http://applications.umons.ac.be/web/fr/pde/2023-2024/ue/US-B3-SCBIOL-006-M.htm>

## Organisation

Le cours de Science des données II : analyse et modélisation (UE) est une unité d'enseignement (UE) de cinq crédits. Il est divisé en deux activités d'apprentissages (AA) : Science des données II : modélisation (Q1) et Science des données II : analyse (Q2). Ce cours est constitué de dix modules. Chaque module représente une charge moyenne de travail de 12h à 15h dont 6h en présentiel et le reste en travail personnel à domicile. Vous pouvez retrouver à l'horaire les plages en présentiel. Le travail à distance est à réaliser entre ces sessions. Toutes les heures en présentiel sont assimilées à des heures de travaux pratiques encadrées avec **présence obligatoire**. Chaque absence doit être justifiée.

## Modules

Science des données II : modélisation (Q1)

- Régression linéaire I, régression linéaire simple, outils diagnostics
- Régression linéaire II, régression linéaire polynomiale, régression linéaire multiple, critère d'Akaike
- Modèle linéaire
- Modèle linéaire généralisé
- Régression non linéaire, courbes de croissance, courbe logistique, ...

Science des données II : analyse (Q2)

- Indices de distances, classification hiérarchique, K-moyennes, indices de diversité
- Analyse en composantes principales et analyse factorielle des correspondances
- Base de données I et analyse factorielle multiple
- Base de données II et positionnement multidimensionnel (MDS)
- Données ouvertes et carte auto-adaptatives (SOM)

## Méthode pédagogique

Ce cours est donné en classe **inversée** avec l'ensemble du matériel pédagogique disponible via le lien <http://bds.sciviews.org/>. Les étudiants seront actifs à chaque étape de l'apprentissage. Chaque module de cours comprend des exercices classés en quatre niveaux de difficulté croissante.

- niveau 1 : découverte des notions avec des exercices rapides et ciblés intégrés au sein du cours en ligne
- niveau 2 : application des notions dans des exercices sous la forme de tutoriel cadré
- niveau 3 : transfert de ces notions dans des projets individuels guidés et réalisés sur des données biologiques
- niveau 4 : réalisation d'analyses libres dans des projets de groupe et sur des données biologiques nouvelles

De plus, des challenges pratiques et des interrogations vont être réalisés durant les séances en présentiel. Néanmoins, la plus grande partie de votre temps sera consacrée à l'analyse de données biologiques au sein des exercices de niveau 3 et 4.

**L'apprentissage nécessite de poser régulièrement des questions.** Le canal privilégié pour ces échanges est constitué des “issues” dans GitHub, accessibles depuis le site du cours. Vous pouvez également nous contacter par mail à l'adresse [sdd@sciviews.org](mailto:sdd@sciviews.org), mais uniquement pour les questions à caractère privé par rapport aux autres étudiants, pas pour les questions sur la matière elle-même.

## Évaluation

L'activité dans les différents exercices est enregistrée pour chaque étudiant et est utilisée pour une évaluation continue. Il n'y a pas d'examen en session ni d'examen en seconde session pour cette matière. La pondération entre les AA du Q1 et du Q2 est de 50/50. Il faut avoir réalisé au minimum 80% des exercices de chaque module pour qu'il soit pris en compte dans l'évaluation.

La note de chaque AA est obtenue par addition des composantes suivantes :

- 25% pour les exercices, soit :
  - 5%, exercices N1 + N2
  - 20%, challenges
- 25% pour les projets, soit :
  - 5%, projets individuels N3
  - 20%, projets de groupe N4
- 25% pour une interrogation de milieu de parcours
- 25% pour une interrogation vers la fin du quadrimestre

## Équipement requis

- un ordinateur (en option)

Les heures en présentiel sont données dans des salles informatiques équipées d'ordinateurs. Il est conseillé d'utiliser son ordinateur personnel pour les étudiants qui en disposent. Cela simplifie le travail en distanciel.

- des écouteurs

Des vidéos sont présentes dans le cours en ligne et il faut pouvoir les écouter avec le son sans déranger ses voisins.

- un câble ethernet (en option)

Il s'agit d'un câble RJ-45 non croisé de catégorie 5E ou plus d'au moins 1m, avec adaptateur si nécessaire pour le connecter sur votre ordinateur (par exemple, USB -> RJ-45). La connexion filaire permet un accès à Internet plus stable et plus rapide que le Wifi.

L'UMONS met également des salles informatiques à disposition des étudiants qui ne disposent pas de leur propre PC. Le travail dans les salles informatiques doit se faire dans le respect de la **charte informatique de l'UMONS**, voir [https://go.sciviews.org/charte\\_info](https://go.sciviews.org/charte_info).

## Protection de la vie privée

La collecte des données personnelles des étudiants dans le cadre de ce cours se fait en respect du RGPD, voir <https://wp.sciviews.org/politique-de-confidentialite/>.

## Entraide versus plagiat

L'entraide entre les étudiants est encouragée, mais il faut citer ses sources. Le plagiat est par contre totalement **interdit** et sévèrement puni à l'UMONS. Les étudiants devront bien comprendre la différence entre les deux dans le cadre du cours.

## Ressources

- Site web du cours : <http://bds.sciviews.org/>
- Horaire des cours : <https://hplanning2023.umons.ac.be/invite>
- Ressources Moodle :
  - Science des données II : modélisation <https://moodle.umons.ac.be/course/view.php?id=1463>
  - Science des données II : analyse <https://moodle.umons.ac.be/course/view.php?id=3386>
- Fiche ECTS du cours : <http://applications.umons.ac.be/web/fr/pde/2023-2024/ue/US-B3-SCBIOL-006-M.htm>
- Définition du plagiat et les sanctions à l'UMONS : [https://alumniumonsac.sharepoint.com/sites/SAP\\_\\_Etudiants/SitePages/Plagiat\(1\).aspx](https://alumniumonsac.sharepoint.com/sites/SAP__Etudiants/SitePages/Plagiat(1).aspx)
- Je ne plagie pas ... enfin, je pense : [https://alumniumonsac.sharepoint.com/sites/SAP\\_\\_Etudiants/SitePages/Plagiat.aspx](https://alumniumonsac.sharepoint.com/sites/SAP__Etudiants/SitePages/Plagiat.aspx)
- Charte informatique de l'UMONS : [https://go.sciviews.org/charte\\_info](https://go.sciviews.org/charte_info)