



## M2 - Statistique et Science des Données (SSD) - BIOSTATS



### Présentation

Ce M2 s'adresse à des étudiantes et étudiants titulaires du M1 Statistique Science des Données (SSD) ou tout autre M1 de mathématiques ou équivalent avec une spécialisation forte en probabilités et statistique.

Ce M2 est scindé en deux sous-parcours plus spécialisés, dont les enseignements restent partiellement mutualisés.

- La première de ces spécialisations est la Biostatistique, qui approfondit l'analyse et la modélisation des données du vivant.

- La seconde est le Management de l'information et de la décision (MIND), spécialisé dans l'analyse et la modélisation de la donnée économique ainsi que la gestion des décisions et risques associés.

\* L'ambition du parcours **SSD-Biostat** est de répondre aux attentes des étudiantes et étudiants du M1 SSD qui se sentent attiré-es par la modélisation des données du vivant ou de l'environnement. Les aspects statistiques abordés dans ce parcours vont de la modélisation du vivant jusqu'aux problématiques les plus théoriques de la statistique et de la modélisation stochastique. Les aspects numériques sont extrêmement présents dans ce parcours et requièrent un goût prononcé pour la programmation informatique.

Le parcours SSD-Biostat est une formation exigeante car elle met l'accent sur les concepts plutôt que sur les techniques. En effet, dans le domaine de la donnée au

sens large, les technologies numériques, avec l'avènement de l'intelligence artificielle évoluent vite et se périment encore plus vite. Les futur-es ingénieur-es statisticien-nes ou chercheurs et chercheuses qui devront traiter des données pourront, tout au long de leur vie professionnelle se former à de nouvelles technologies, d'autant mieux s'ils ou elles ont eu une formation initiale conceptuelle solide. La plus-value de la formation est justement d'apporter la compréhension théorique des concepts statistiques sous-jacents aux algorithmes automatiques. Les diplômé-es doivent aussi être capable d'assurer une veille technologique de qualité.

Le parcours SSD-Biostat reste en partie mutualisé en deuxième année avec le parcours management de l'information et de la décision (SSD-MIND). Toutefois, des UE de spécialisation pour la science des données du vivant ou de l'environnement et davantage tournées vers l'initiation à la recherche sont spécifiques au parcours SSD-BIOSTAT (deux UE par semestre du M2).

\* L'ambition du parcours **SSD-MIND** est de répondre aux attentes des étudiantes et étudiants du M1 SSD qui se sentent attiré-es par l'application de la science des données en entreprise. Compte tenu de la très grande diversité des entreprises et de leurs problématiques, ce M2 forme à une science des données généraliste, "tous terrains". Par surcroît, il fournit une formation plus spécifique au contexte de l'entreprise et à ses problématiques économique et managériale (information économique, gestion des risques financiers, clientèle, stratégie d'entreprise etc).



Le parcours SSD - MIND est une formation de type ingénierie mathématique, qui met au premier plan la méthodologie et la parfaite maîtrise des concepts et modèles statistiques. La ou le diplômé-e de ce parcours pourra faire face à tout type de données et de problématique, et concevoir une méthodologie complète et souvent originale servant cette problématique, partant de la gestion et de l'organisation des données, poursuivant avec leur exploration et leur réduction ciblée, puis avec la modélisation des phénomènes d'intérêt, et enfin synthétisant l'information extraite à des fins de décision. Elle ou il devra également savoir transmettre à l'entreprise la connaissance synthétisée à partir de l'information extraite des données. Chaque nouvel ensemble de données et chaque question posée le concernant est souvent un nouveau problème, et l'application à ces données d'une méthode standard est alors inadaptée. Il s'agit au contraire d'écrire un modèle mathématique adapté à ces données (au sens où il en traduit la complexité de façon satisfaisante) et de le rendre assimilable à une méthode d'estimation standard, ou de concevoir et programmer une méthode plus spécifique. L'accent mis par cette formation sur la maîtrise conceptuelle et mathématique des outils garantit aux diplômés de ce parcours la très grande capacité d'adaptation et d'auto-formation nécessitée par la rapide évolution de la science des données.

Le parcours SSD-MIND reste en partie mutualisé en deuxième année avec le parcours biostatistique (SSD-Biostat), plus spécialisé en analyse et modélisation des données du vivant ou de l'environnement. Le parcours SSD-MIND est une double formation en partenariat avec l'IAE (qui assure les enseignements d'économie et de management), donnant lieu à un double diplôme.

---

## Objectifs

Plusieurs objectifs sont visés par la formation.

- \* Rendre l'étudiant-e capable de faire face à tout type de données et de problématique, et de concevoir une méthodologie complète et souvent originale servant cette problématique.

- \* Rendre l'étudiant-e capable de s'intégrer très vite à tout type d'entreprise en saisissant rapidement les problématiques.
- \* Amener les étudiant-es qui le souhaitent à un niveau théorique leur permettant de faire une thèse de doctorat en statistique.
- \* Former de futur-es chercheuses et chercheurs ou enseignants-chercheurs et enseignantes-chercheuses dans le domaine des mathématiques de l'aléatoire : probabilités ou statistique théorique ou appliquée. Elles ou ils pourront ensuite intégrer, après un doctorat, des laboratoires dans les universités, les grandes écoles d'ingénieurs ou les organismes de recherche, tels le CNRS, l'INRAE, Inria, le CIRAD, l'INSERM, ... Il est aussi possible d'intégrer, directement après le M2, une entreprise ou un laboratoire de recherche dans l'industrie.
- \* Former des spécialistes du traitement statistique des données de haut niveau pour des organismes de recherche ou des entreprises pour lesquels la statistique est maintenant un outil indispensable comme les laboratoires pharmaceutiques ou des instituts de veille épidémiologique ou de surveillance de la qualité de l'air, des eaux, les entreprises agro-alimentaires, les entreprises de biotechnologies, les entreprises de la filière santé (aide au diagnostic, médecine personnalisée), etc

---

## Savoir faire et compétences

- \* Etre capable d'extraire les données pertinentes
- \* Réaliser le pré-traitement des données (nettoyage et mise en forme éventuels)
- \* Mener l'analyse exploratoire des données en utilisant des outils de visualisation et de réduction de dimension.
- \* Modéliser un problème : maîtriser les méthodes usuelles en science des données moderne et savoir proposer la/ les méthode(s) adéquate(s) pour la résolution du problème posé, écrire un / plusieurs modèle(s) mathématique(s) adapté(s) au service de la problématique, et le(s) mettre sous forme idoine au traitement par les méthodes usuelles de la science des données moderne.
- \* Mettre en œuvre la méthode d'un point de vue computationnel et être capable de proposer des stratégies de choix de modèles



- \* Programmer efficacement dans au moins un langage (python, R)
- \* Savoir analyser et interpréter les résultats, c'est-à-dire produire de la connaissance à partir de l'information extraite.
- \* Relier la connaissance produite à la décision de sorte à éclairer et optimiser cette dernière.
- \* Être capable de communiquer les résultats à l'écrit et à l'oral

## Admission

---

### Public cible

Public cible\* : Ce parcours s'adresse à des étudiant-es titulaires du M1 Maths - Statistique Science des Données (SSD) ou tout autre M1 de mathématiques équivalent avec une spécialisation forte en probabilités et statistique.

### Pré-requis nécessaires

M1 Maths - Statistique Science des Données (SSD)

### Pré-requis recommandés

M1 Maths - Statistique Science des Données (SSD)

## Et après

---

### Poursuites d'études

Doctorat possible à l'issue du M2.

### Insertion professionnelle

Métiers : statisticien-ne, biostatisticien-ne, data-scientist, analyste de données, tous de niveau ingénieur.

Tous secteurs d'activité : industrie, recherche et développement, santé, agronomie, banques et assurances, commerce, etc.

## Infos pratiques

---

### Contacts

#### Responsable pédagogique

Elodie Brunel-piccinini

+33 4 67 14 41 64

elodie.brunel-piccinini@umontpellier.fr

#### Responsable pédagogique

Xavier Bry

+33 4 67 14 35 78

xavier.bry@umontpellier.fr

### Etablissement(s) partenaire(s)

Université de Sherbrooke (double diplôme)

IAE Montpellier (double diplôme en M2)

<https://iae.umontpellier.fr/>

### Laboratoire(s) partenaire(s)

Institut Montpelliérain Alexander Grothendieck  
IMAG

<https://imag.edu.umontpellier.fr/Montpellier>

### Lieu(x)

Montpellier - Triolet



# Programme

## Organisation

M2 ouvert à l'alternance avec des périodes longues d'enseignement et en entreprise.

Première période : 7 semaines d'enseignement de septembre à fin octobre.

Deuxième période : 7 semaines en entreprise pour les alternant-es ou projet tutoré long en laboratoire pour les non-alternant-es de novembre à mi-janvier.

Troisième période : 7 semaines d'enseignements de mi-janvier à mi-mars.

Quatrième période : alternance en entreprise de mars à août ou stage de 4 à 6 mois pour les non alternant-es.

Stages et projets tutorés : Un projet tutoré de 7 semaines pour les non-alternant-es faisant l'objet d'un rapport et d'une soutenance orale.

Stage de 4 à 6 mois en fin de M2.

### M2S9 - Statistiques et Science des Données (SSD) - BIOSTATS

Analyse multivariée	5 crédits	21h
Projet ou Soutenance alternance	3 crédits	
Estimation non-paramétrique	5 crédits	21h
Statistique Bayésienne	5 crédits	21h
Anglais	2 crédits	18h
Apprentissage statistique	5 crédits	21h
Modèles linéaires généralisés	5 crédits	21h

### M2S10 - Statistiques et Science des Données (SSD) - BIOSTATS

Analyse des durées de vie	4 crédits	18h
Modèles à variables latentes	4 crédits	21h
Complément 1	4 crédits	18h
Complément 2	4 crédits	18h
Stage	14 crédits	120h