



---

## **Guide pédagogique**

***Module « Sciences des données »  
Option IASD – 8.5 (4 crédits ECTS)***

---

### ***Place du module et enjeux***

A la croisée de plusieurs champs disciplinaires : mathématiques, statistiques, informatique, théorie de l'information et visualisation, les sciences des données mettent en œuvre différents outils d'analyse et de fouille de données afin d'extraire automatiquement des informations utiles, des connaissances, à partir de données potentiellement massives. Le but ultime est de rendre cette information plus facile à exploiter, la protéger et la valoriser. Elle pourra servir de base ensuite à des processus d'évaluation et d'aide à la décision. Ce module propose un enseignement approfondi sur les sciences des données : quels traitements sont les mieux adaptés en fonction de certains contextes, comment les mettre en œuvre. Il se termine par une étude de différents modes de visualisation qui permettent à l'opérateur humain de mieux appréhender les différentes caractéristiques des données analysées.

---

## **Teaching guide and syllabus**

***« Data Sciences » module  
DSAI option – 8.5 (4 ECTS credits)***

---

### ***Subject matter importance and associated issues***

At the crossroads of several disciplinary fields: mathematics, statistics, computer science, information theory and visualization, data science makes use of different tools for data analysis and data mining to automatically extract useful information and knowledge from potentially massive data. The aim is to make this information easier to use, protect and enhance. Based on it, evaluation and decision-making processes may be activated. This module offers in-depth teaching on data science: which treatments are best adapted to certain contexts, how to implement them. It ends with a study of different visualization models, that allow the human operator to better understand the particular features of the analyzed data.

**Responsable : Gérard Dray**  
Téléphone : 04 34 24 62 80  
Courriel : [gerard.dray@mines-ales.fr](mailto:gerard.dray@mines-ales.fr)



**IMT Mines Alès**  
École Mines-Télécom

ENSEIGNEMENTS ACADÉMIQUES	Volume horaire	Détail des coefficients	Crédits
<b>Sciences des données</b>	<b>50 h</b>		
○ Introduction à l'analyse de données	5		4
○ Statistiques et probabilités avancées	35	4	
○ Validation, visualisation, restitution	10	2	

**Matière 1**

<b>Titre de la matière : Introduction à l'analyse de données</b>	
<b>Code : 2IA-iasd-8.5.1</b>	<b>Titre du module : « Sciences des données »</b>
<b>Semestre : S8</b>	<b>Cursus de rattachement : Département 2IA, option IASD</b>

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
5	7	5					2	1	0.57

<b>Résumé</b>	L'option IASD du département 2IA forme des ingénieurs analystes de données. Cette introduction positionne l'analyse de données en regard de différentes problématiques. Une introduction au langage R est également proposée car c'est le langage de prédilection de ce domaine.
---------------	--

<b>Responsable</b>	Gérard Dray – LGI2P/IMT Mines Alès
<b>Équipe enseignante</b>	Gérard Dray – LGI2P/IMT Mines Alès Nicolas Sutton-Charani – LGI2P/IMT Mines Alès

<b>Mots-clés</b>	Analyse de données, langage R
<b>Prérequis</b>	Statistiques et probabilités du tronc commun

<b>Contexte et objectif général :</b> Les techniques actuellement mises en avant en Intelligence Artificielle reposent sur l'utilisation de données, nombreuses et hétérogènes, qui peuvent être croisées, analysées, afin de faire émerger des comportements récurrents, d'expliquer certains phénomènes, voire de prédire certains faits. Cette introduction permet de décrire les contextes propices à ce genre d'approches et donner un inventaire des techniques utilisées dans le domaine.
<b>Programme et contenu :</b> - 3h d'introduction à l'analyse de données - 2h d'introduction au langage R
<b>Méthode et organisation pédagogique :</b> Les enseignements sont prévus pour 35 élèves. Le découpage est prévu comme suit : - 5h de cours (illustration avec le langage R)
<b>Acquis d'apprentissage visés :</b> Etre conscient de la nécessité d'adapter les traitements d'analyse et de fouille de données au contexte et à la nature de ces données. Prise en main du langage R.
<b>Évaluation :</b> Le langage R sera évalué dans le cadre d'un projet qui fait intervenir la visualisation (matière 8.5.3)
<b>Retour sur l'évaluation fait à l'élève :</b> 3 semaines après la dernière séance.
<b>Support pédagogique et références :</b> 1 Polycopié

**Matière 2**

<b>Titre de la matière : Statistiques et probabilités avancées</b>	
<b>Code : 2IA-iasd-8.5.2</b>	<b>Titre du module : « Sciences des données »</b>
<b>Semestre : S8</b>	<b>Cursus de rattachement : Département 2IA, option IASD</b>

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
35	50	14	18			3	15	4	2.28

<b>Résumé</b>	L'option IASD du département forme des ingénieurs analystes de données. Cet enseignement a pour objectif de renforcer les compétences théoriques avec un approfondissement des statistiques et de la théorie des probabilités.
---------------	--

<b>Responsable</b>	Nicolas Sutton Charani – LGI2P/IMT Mines Alès
<b>Équipe enseignante</b>	Nicolas Sutton Charani – LGI2P/IMT Mines Alès Abdelhak Imoussaten – LGI2P/IMT Mines Alès Stefan Janaqi – LGI2P/IMT Mines Alès

<b>Mots-clés</b>	Statistiques, probabilités, statistiques bayésiennes, estimation et vraisemblance, réduction de dimension, tests
<b>Prérequis</b>	Notions de base des statistiques et probabilités, langage R.

<b>Contexte et objectif général :</b>	Les enseignements de l'option IASD (Intelligence Artificielle et Science des Données) et en particulier toutes les méthodes d'apprentissage automatique et de <i>business intelligence</i> , reposent sur des calculs statistiques et/ou des modèles probabilistes. Cet enseignement en fournit les fondements théoriques et présente des outils et des techniques pour leur mise en œuvre.
<b>Programme et contenu :</b>	Cet enseignement détaillera les chapitres suivants : <ul style="list-style-type: none"> <li>- Estimation et vraisemblance (3h de cours et 3h de TD)</li> <li>- Mélange de lois normales et algorithme EM – Espérance-maximisation (3h de cours et 3h de TD)</li> <li>- Statistiques Bayésiennes (3h de cours et 6h de TD)</li> <li>- Réduction de dimensionnalité (4h de cours et 4h de TD)</li> <li>- Tests (1h de cours 2h de TD)</li> </ul>
<b>Méthode et organisation pédagogique :</b>	Les enseignements sont prévus pour 35 élèves en promotion complète, avec la possibilité d'un enseignant supplémentaire pour certaines séances de mise en pratique. Le découpage est prévu comme suit : <ul style="list-style-type: none"> <li>- 14h de cours (illustration avec le langage R)</li> <li>- 18h de TD</li> <li>- 3h de contrôle sur table</li> </ul>
<b>Acquis d'apprentissage visés :</b>	Mettre en œuvre des méthodes statistiques et probabilistes pour l'analyse de données. Appliquer des techniques qui permettent le passage à l'échelle.
<b>Évaluation :</b>	Épreuve sur table (3h, coef. 4)
<b>Retour sur l'évaluation fait à l'élève :</b>	3 semaines après la dernière séance.
<b>Support pédagogique et références :</b>	1 Polycopié

**Matière 3**

<b>Titre de la matière : Validation, visualisation, restitution</b>	
<b>Code : 2IA-iasd-8.5.3</b>	<b>Titre du module : « Sciences des données »</b>
<b>Semestre : S8</b>	<b>Cursus de rattachement : Département 2IA, option IASD</b>

Heures présentiel	Heures total	Cours	TD	TP	Projet	Contrôles	Travail personnel	Coef /module	ECTS
10	15	2		8			5	2	1.14

<b>Résumé</b>	Un pan important de l'analyse de données concerne la restitution à l'utilisateur des différents résultats obtenus après traitement. Ce cours permet de présenter les différents modes d'affichage de ces résultats et d'outiller les étudiants avec la présentation de bibliothèques logicielles dédiées.
---------------	---

<b>Responsable</b>	Gérard Dray – LGI2P/IMT Mines Alès
<b>Équipe enseignante</b>	Yan Holtz – Datadog Doctorant en appui pour l'encadrement de TP

<b>Mots-clés</b>	Visualisation, Restitution.
<b>Prérequis</b>	Connaissance basique de R.

<p><b>Contexte et objectif général :</b> Les méthodes d'Intelligence Artificielle et de Science des Données sont toutes basées sur une compréhension de la donnée. Utiliser des méthodes de visualisations appropriées est une compétence indispensable pour mener ce travail à bien.</p>
<p><b>Programme et contenu :</b></p> <ul style="list-style-type: none"> <li>- Introduction à la visualisation de données (2h de cours)</li> <li>- Introduction à ggplot2 (3h de TP)</li> <li>- Application à un jeu de données du choix des étudiants (2h de TP)</li> <li>- Construction et partage d'un rapport reproductible avec R Markdown et GitHub (3h de TP)</li> </ul>
<p><b>Méthode et organisation pédagogique :</b> Les enseignements sont prévus pour 35 élèves en promo complète. Les TP seront réalisés sur les ordinateurs personnels de ces derniers. Le découpage est prévu comme suit :</p> <ul style="list-style-type: none"> <li>- 2h de cours</li> <li>- 8h de mise en application en TP.</li> <li>- Évaluation : étude du dépôt de projet dans GitHub et du rapport en ligne. Les critères de notation seront en particulier : l'originalité, le choix des graphiques, la qualité du rapport et la clarté du code.</li> </ul>
<p><b>Acquis d'apprentissage visés :</b> Connaitre les bonnes pratiques de visualisation de données et les différents types de graphique. Savoir les construire à l'aide de R et ggplot2. Être capable de les intégrer dans un rapport reproductible et les partager grâce à des outils de contrôle de version.</p>
<p><b>Évaluation :</b> TP notés (2)</p>
<p><b>Retour sur l'évaluation fait à l'élève :</b> 3 semaines après la dernière séance.</p>
<p><b>Support pédagogique et références :</b> Supports de cours et site Internet dédié.</p>

## Méthode et organisation pédagogique

Il s'agit d'un enseignement relativement classique avec une partie réalisée en cours magistral et une partie appliquée au travers de TP et Projets.

## Modalité d'évaluation

Le niveau d'acquisition des compétences sera évalué selon les exigences suivantes :

N° indicateur	Indicateur
1	connaître les savoirs formels et pratiques du socle des fondamentaux
2	Exploiter les savoirs théoriques et pratiques
3	Analyser, interpréter, modéliser, émettre des hypothèses, et résoudre

## Répartition

Matière	Contrôle	Coefficients	Type de notation	Indicateurs évalués	Chapitres
Statistiques et probabilités avancées	Devoir sur table	4	Individuelle	1 et 2	Tous
Validation, visualisation, restitution	TP noté	2	En binôme	3	Tous

## Engagement de l'étudiant, éthique et professionnalisme

La démarche éthique est définie dans le règlement intérieur de l'établissement. Chaque étudiant s'engage à en prendre connaissance et à la respecter.

**Obligation des cours** : Présence obligatoire pour tous à chaque séance

**Nombre d'heures estimées de travail personnel** : pour acquérir les compétences demandées, il est nécessaire que l'étudiant consacre minimum 45 min de travail personnel de compréhension et d'approfondissement par séance de cours.

**Nombre d'heures estimées de préparation aux travaux dirigés (TD)** : 1 à 2h.

### Pénalité pour retard

Tout travail remis en retard sans motif valable peut être pénalisé de 1 point par jour de retard (notation effectuée sur 20).

## Équipe enseignante

Nom	Domaine d'expertise	Courriel/Téléphone
Gérard DRAY	Analyse et fouille de données	<a href="mailto:gerard.dray@mines-ales.fr">gerard.dray@mines-ales.fr</a> / 04 34 24 62 80
Abdelhak IMOUSSATEN	Analyse multicritère, Apprentissage des préférences, Théorie de l'incertain, Fusion de données, Opérateurs d'agrégation	<a href="mailto:abelhack.imoussaten@mines-ales.fr">abelhack.imoussaten@mines-ales.fr</a> 04 34 24 62 64
Stefan JANAQI	Data Science, Mathématiques appliquées	<a href="mailto:stefan.janaqi@mines-ales.fr">stefan.janaqi@mines-ales.fr</a> 04 34 24 62 19
Nicolas SUTTON-CHARANI	Science de la Donnée, Apprentissage Automatique et Théories de l'Incertain	<a href="mailto:nicolas.sutton-charani@mines-ales.fr">nicolas.sutton-charani@mines-ales.fr</a> 04 34 24 62 67
Yann HOLTZ	Visualisation de données, Data science.	<a href="mailto:yan.holtz.data@gmail.com">yan.holtz.data@gmail.com</a>

ACADEMIC TEACHING	Teaching hours	Coefficients	Credits
<b>Data Sciences</b>	<b>50 h</b>		
○ Introduction to data analysis	5		
○ Advanced Statistics and Probability	35	4	4
○ Validation, Visualisation, Reporting	10	2	

**Class 1**

<b>Class title: Introduction to data analysis</b>	
<b>Code : 2IA-iasd-8.5.1</b>	<b>Module title: « Data Sciences »</b>
<b>Semester: S8</b>	<b>Classification : 2IA Department, DSAI option</b>

Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
5	7	5					2	1	0.57

<b>Summary</b>	The DSAI option of the CSAI department trains data analyst engineers. This introduction positions data analysis in different contexts and for dedicated issues. An introduction to the R language is also provided since it is the preferred language in this domain
----------------	--

<b>Head</b>	G�rard Dray – LGI2P/IMT Mines Al�s
<b>Teaching team</b>	G�rard Dray – LGI2P/IMT Mines Al�s Nicolas Sutton-Charani – LGI2P/IMT Mines Al�s

<b>Key words</b>	Data analysis, R programming language
<b>Prerequisites</b>	Statistics and probabilities.

<b>Context and general objective:</b>
Currently, Artificial Intelligence techniques rely mostly on huge amount of heterogeneous data that can be crossed and analysed in order to reveal recurrent behaviours, explain certain phenomena or even predict certain facts. This introduction describes different contexts where such approaches may be used and provides an inventory of the techniques available in this field.
<b>Programme and contents:</b>
The course covers the following topics: <ul style="list-style-type: none"> <li>- introduction to data analysis (3h)</li> <li>- introduction to R programming language (2h)</li> </ul>
<b>Method and pedagogic organisation:</b>
The course will be organized as follows for 35 students <ul style="list-style-type: none"> <li>- 5h of lectures</li> </ul>
<b>Targeted skills or knowledge :</b>
To better apprehend the outline of data analysis and have a general idea of techniques that can be used in particular contexts. R language programming.
<b>Evaluation:</b> Other courseworks of this module need R language (in particular 8.5.3) and evaluation will be done in these project context.
<b>Feedback made to the student :</b>
Exam copies within three weeks
<b>Teaching material and references :</b>
Lecture slides and reference textbooks

Class 2

<b>Class title: Advanced statistics and probability</b>	
<b>Code : 2IA-iasd-8.5.2</b>	<b>Module title: « Data Sciences »</b>
<b>Semester: S8</b>	<b>Classification : 2IA Department, DSAI option</b>

Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
35	50	14	18			3	15	4	2.28

<b>Summary</b>	DSAI option of the CSAI department trains data analyst engineers. The objective of this course is to strengthen theoretical skills with a deepening of statistics and probability theory.
----------------	---

<b>Head</b>	Nicolas Sutton Charani – LGI2P/IMT Mines Alès
<b>Teaching team</b>	Nicolas Sutton Charani – LGI2P/IMT Mines Alès Abdelhak Imoussaten – LGI2P/IMT Mines Alès Stefan Janaqi – LGI2P/IMT Mines Alès

<b>Key words</b>	Statistics, probability, Bayesian statistics, likelihood estimation, Dimensionality reduction, tests.
<b>Prerequisites</b>	Basics of statistics and probabilities, R language

**Context and general objective:**

The lessons of the DSAI option (Data Science and Artificial Intelligence) and in particular machine learning and business intelligence methods, are based on statistics and/or probabilistic models. This teaching provides the theoretical foundations, tools and techniques for their implementation.

**Programme and contents:**

The course covers the following topics:

- Maximum likelihood estimation (3h lectures and 3h tutorial)
- GMM (Gaussian Mixture Model) and Expectation–maximization algorithm (3h lectures and 3h tutorial)
- Bayesian Statistics (3h lectures and 6h de TP)
- Reduction of dimensionality (4h lectures and 4h tutorial)
- Testing (3h)

**Method and pedagogic organisation:**

The course will be organized as follows for 35 students

- 16h of lectures
- 6h of labs
- 3h written exam

One teaching assistant for some labs sessions

**Targeted skills or knowledge:**

Apply statistics and probability technics to data analysis. Ensure scaling up.

**Evaluation:** Three hour written examination (coef 4).

**Feedback made to the student:** annotated exam copies within three weeks

**Teaching material and references:**

Lecture slides and reference textbooks

## Class 3

<b>Class title: Validation, Visualisation, Reporting</b>									
<b>Code : 2IA-iasd-8.5.3</b>					<b>Module title: « Data Sciences »</b>				
<b>Semester: S8</b>					<b>Classification : 2IA Department, DSAI option</b>				
Hours of presence	Total hours	Lectures	Work shop	Labs	Project	Testing	Personal work	Coef /module	ECTS
10	15	2		8			5	2	1.14
<b>Summary</b>	An important part of data analysis concerns the feedback to the user after processing. This course presents the different ways of displaying these results and equips students with the presentation of dedicated software libraries.								
<b>Head</b>	G�rard Dray – LGI2P/IMT Mines Al�s								
<b>Teaching team</b>	Yan Holtz – Datadog Teaching assistant								
<b>Key words</b>	Data visualization, Reproducibility, Reporting								
<b>Prerequisites</b>	Basic knowledge on R								
<b>Context and general objective:</b> Artificial intelligence and data science in general is always based on a previous step of data exploration. Visualizing the data the right way and creating a reproducible report is a crucial step in the pipeline.									
<b>Programme and contents:</b> The course covers the following topics: <ul style="list-style-type: none"> <li>- Introduction to data visualization (2h)</li> <li>- Introduction to ggplot2 (3h)</li> <li>- Application to a dataset of your choice (2h)</li> <li>- Building and sharing report with R markdown and Github (3h)</li> </ul>									
<b>Method and pedagogic organisation:</b> The course will be organized as follows for 35 students <ul style="list-style-type: none"> <li>- 2h of lectures:</li> <li>- 8h of labs to learn the basis of ggplot2 and learn how to create and share a report</li> <li>- Evaluation</li> </ul>									
<b>Targeted skills or knowledge:</b> Knowledge of different types of chart, Overview of dataviz best practices, Capacity to build basic charts with R and ggplot2, and to share them thanks to R Markdown and Github.									
<b>Evaluation:</b> On project: analysis of the report shared online, including clarity of code.									
<b>Feedback made to the student:</b> 3 weeks after end of Module.									
<b>Teaching material and references:</b> Slides will be provided and dedicated web site.									

## Method and teaching organisation

Teachings will consist of lectures, seminars and labs.

## Testing procedures

The student's level of knowledge acquisition will be evaluated according to the following points:

N° Indicator	Indicator
1	To know the formal and practical knowledge constituting the foundation of a given field
2	Exploit theoretical and practical knowledge
3	Analyse, interpret, model, hypothesize and solve problems

### Grading scheme:

Class	Exam	Coefficients	Administration mode	Evaluated Indicators	Chapters
Advanced Statistics and Probability	Exam	4	Individual	1, 2	All
Validation, Visualisation, Reporting	Project	2	Group	3	All

## Student commitments, ethics and professionalism

Expectations concerning ethics are defined in the establishment's code of conduct. Each student is expected to know and respect the code of conduct.

**Obligatory presence in classes:** Students must attend all courses, seminars and labs.

**Estimated hours of personal study:** *in order to acquire the required learning level, the student is expected (must) to spend a minimum of 45min of personal study time per hour spent in class.*

**Estimated hours of preparation required for labs/Work Shop:**

1 to 2 hours

**Late penalties:** Late works are subject to penalties as follows: 3 points per day (ratings are between 0 and 20).

## Teaching team

Name	Expertise	Email/Phone
G�rard DRAY	Data analysis, Data mining	<a href="mailto:gerard.dray@mines-ales.fr">gerard.dray@mines-ales.fr</a> / 04 34 24 62 80
Abdelhak IMOUSSATEN	Multiple criteria decision analysis, Preference learning, Uncertainty theory, data merging, aggregation operators	<a href="mailto:abelhack.imoussaten@mines-ales.fr">abelhack.imoussaten@mines-ales.fr</a> 04 34 24 62 64
Stefan JANAQI	Data Science, applied mathematics	<a href="mailto:stefan.janaqi@mines-ales.fr">stefan.janaqi@mines-ales.fr</a> 04 34 24 62 19
Nicolas SUTTON-CHARANI	Data Science, Machine Learning and Uncertainty theory	<a href="mailto:nicolas.sutton-charani@mines-ales.fr">nicolas.sutton-charani@mines-ales.fr</a> 04 34 24 62 67
Yan HOLTZ	Data visualization, data science	<a href="mailto:Yan.holtz.data@gmail.com">Yan.holtz.data@gmail.com</a>

## Approbation

Ce guide p dagogique entre en vigueur   compter du 07 janvier 2019

Il est port    la connaissance des  l ves par une publication sur le site de l' cole

R�daction	V�rification	Validation
L'enseignant responsable du module : G�rard DRAY	Le responsable d'UE / de d�partement : Sylvie RANWEZ	Le directeur de l'�cole, Pour le directeur et par d�l�gation, Le directeur de la DFA / de la DE : Michel Ferlut