

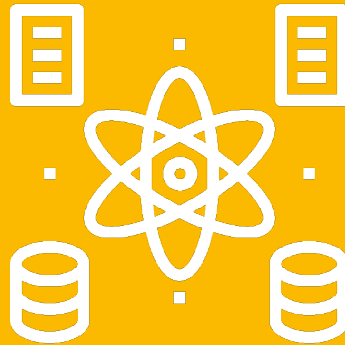
Master 2

Sciences, Technologies, Santé

2023-2024

Mathématiques et applications

Data Science



M2 DS

Apprentissage

Contrat Pro.



CONNAISSANCES
université
angers

SOMMAIRE

Contacts de la formation	03
Planning de la formation	04
Présentation de la formation	05
Volumes horaires et évaluations	09
Contenu des enseignements	
Semestre 3	11
Semestre 4	18
Calendriers détaillés	24

*Sommaire interactif
pour revenir
au sommaire
cliquer sur ►►*



CONTACTS DE LA FORMATION

— Sandrine TRAVIER : **Directrice adjointe à la Pédagogie**

Tél. : 02 41 73 50 01

sandrine.travier@univ-angers.fr

— Frédéric PROÏA : **M1 Responsable pédagogique et Président du Jury**

Tél. : 02 41 73 50 47

frederic.proia@univ-angers.fr

Bureau : bâtiment I, porte 212

— Fabien PANLOUP : **M2 Responsable pédagogique et Président du Jury**

Tél. : 02 41 73 50 29

fabien.panloup@univ-angers.fr

Bureau : bâtiment I, porte 221

— Sandrine HERGUAÏS : **Gestion de la scolarité et des examens**

Tél. : 02 41 73 54 85

sandrine.herguais@univ-angers.fr

— Charlotte BROSSET : **Alternance**

Tél. : 02 41 73 52 17

re.sciences@contact.univ-angers.fr

SCOLARITÉ – EXAMENS

Bâtiment A, Rez-de-chaussée

Horaires d'ouverture

9h00 – 12h30

13h30 – 17h00

Du lundi au vendredi



Fermé le mercredi après-midi



2023-2024

N.B. Calendriers détaillés pour les alternants et non-alternants p. 26/27

Semestre 1

Rentrée des Alternants	Vendredi 01 septembre 2023
Rentrée globale et début des cours	Lundi 11 septembre 2023
<i>Vacances d'automne</i> 	<i>Du samedi 28 octobre 2023 au dimanche 05 novembre 2023</i>
<i>Vacances de Fin d'année</i> 	<i>Samedi 23 décembre 2023 au dimanche 07 janvier 2024</i>
Fin des cours du 1 ^{er} semestre	Jeudi 11 janvier 2024
Examens 1 ^{er} semestre Session 1	Du lundi 15 janvier 2024 au vendredi 19 janvier 2024
Jury 1 ^{er} semestre Session 1	Lundi 05 février 2024

Semestre 2

Début des cours	Lundi 22 janvier 2024
Fin des cours du 2 ^{ème} semestre	Vendredi 01 mars 2024
Examens 2 ^{ème} semestre Session 1	Du lundi 04 mars 2024 au jeudi 07 mars 2024
Examens 1 ^{er} semestre Session 2	Du lundi 04 mars 2024 au jeudi 07 mars 2024
Soutenance de Projet Annuel	Semaine du 11 mars 2024 <i>pour les non alternants</i> A partir du mardi 12 mars <i>pour les alternants</i>
Stage non alternants	A partir du lundi 11 mars 2024
Soutenance de Stage / Alternance	Du mercredi 28 août 2024 au lundi 02 septembre 2024
Jury 1 ^{er} semestre Session 2	Vendredi 27 septembre 2024
Jury 2 ^{ème} semestre Session 1	

Planning susceptible de modifications



PRÉSENTATION DE LA FORMATION

Les mathématiques sont depuis longtemps partie prenante du monde socio-éco- nomique. «L'Étude de l'impact socio-économique des mathématiques en France » (Mai 2015) montre une contribution des mathématiques à hauteur de 15 % du PIB et 9 % des emplois en France. L'intégration de compétences mathématiques est désormais un choix stratégique du développement d'une entreprise, en France et dans le monde, dans tous les secteurs d'activité et particulièrement ceux à forte croissance.

La data Science désigne une discipline à l'interface entre modélisation mathématique, statistique et informatique, née de la nécessité croissante de traiter et d'exploiter les données du big data. La data science est désormais l'outil essentiel d'aide à la décision dans des domaines d'activités extrêmement variés : banque, finance, assurance, e-commerce et grande distribution, communication et marketing, santé, agro-alimentaire, aéronautique et défense, internet des objets et télécoms, etc..

Le **parcours Data Science (DS)** du master mention Mathématiques et Applications est une formation de niveau un (Bac+5). Il vise à former des cadres data-scientists capables de mettre en œuvre les techniques mathématiques et informatiques inhérentes à ce métier, possédant en outre des compétences métiers propres aux secteurs d'activités visés par le choix d'options.

OBJECTIFS, COMPÉTENCES ET CAPACITÉS VISÉS PAR LA FORMATION

Le master Data Science (DS) est une formation en deux ans qui débouche sur le diplôme national de Master mention Mathématiques et Applications, parcours Data Science. A l'issue de la formation, le ou la diplômé(e) possède :

Des compétences en mathématiques. Le diplômé maîtrise les méthodes de l'analyse statistique et algorithmique de l'information, qu'elles soient classiques (datamining, classification, ..) ou spécifiques au big data (machine learning, ..), ainsi que les méthodes associées d'optimisation. Le diplômé est en capacité de modéliser les processus aléatoires, de les simuler et de confronter le modèle aux données statistiques du problème.

Des compétences en informatique. Le diplômé du master dispose des compétences relatives aux outils classiques de data management et d'analyse des données (R, Python, SQL,..), et plus spécifiques aux données de grandes dimensions et/ou non structurées : plateforme Hadoop, HDFS, NoSQL, etc...

Des aptitudes professionnelles. Le diplômé du master sait définir un projet, ses objectifs et son contexte, les modalités de réalisation et d'évaluation des actions, leurs priorités et leur planning. Il sait utiliser les technologies de l'information et de la communication, effectuer une recherche d'information à partir de sources diverses, organiser une



veille technologique. Il a une bonne capacité de communication, une bonne maîtrise de l'anglais scientifique et technique (certification), et sait rédiger clairement une note ou un rapport selon des supports de communication adaptés.

Des compétences métiers. Suivant le choix d'option, le diplômé possède des aptitudes spécifiques au traitement et à l'analyse des données de la biologie (données omiques notamment), ou celles des entreprises dédiées au management du risque ou/et à l'optimisation des actions.

DÉBOUCHÉS VISÉS ET RÉMUNÉRATION DES DIPLÔMÉS

Les diplômés du master Data Science (DS) occupent des emplois de cadre Data scientist (ou Data miner, Data analyst, etc..) couvrant un large spectre de secteurs d'activités, touchant aussi bien au domaine industriel qu'à celui du tertiaire. Sont plus particulièrement visés :

— Option données biologiques : secteurs de la santé, de la pharmacie, la bio-industrie, l'agro-industrie, les laboratoires de type Inserm, Cirad, Institut Pasteur, CEA, etc ..

— Option données numériques : tout secteur tel que sociétés de conseil en gestion des entreprises ou spécialisées en datamining, de la grande distribution, les grandes banques et compagnies d'assurances, etc..

L'enquête interne à 12 mois pour la promotion 2014-20151 montre une excellente insertion professionnelle (90 % en emploi CDI ou CDD) pour un salaire moyen brut mensuel de 2 600 Euros (hors primes et 13ème mois).

Selon l'APEC, la rémunération des

Data scientists se situe entre 32 et 40 keuros brut/an en début de carrière.

Le master est essentiellement à visée professionnelle. Des possibilités sont cependant offertes aux très bons étudiants d'une poursuite d'étude en doctorat par une thèse, en particulier dans le cadre d'une convention CIFRE avec une entreprise.



¹Enquête pour les diplômés du master Ingénierie Mathématique dont le master DS est une évolution. Exemples de stages ou d'emplois récents : Renault DI-RC, INRA, iAdvize, CPAM Maine et Loire, LNE, Sodifrance-Netapsys, Atmo Poitou Charentes, Airparif, DAS-MMA, SFR, Ineox, AVISIA, Metro Cash & Carry France, Tronico groupe Alcen, Imbriktion, Institut de Cancérologie de l'Ouest, CHU-Angers, Axens IFP Technologies, Coheris, BRED Banque populaire, Enza Zaden, Institut Gustave Roussy, Groupe PSA, CT Ingénierie, Harmonie Mutuelle, Banque de France, VYV Care IT, Capgemini, Sopra-Steria, Sigma Informatique, Socomec SAS,...

PUBLIC VISÉ, PRÉ-REQUIS, CONDITIONS D'ADMISSION, TARIFICATION

Le master Data Science est à capacité limitée et l'admission en première (M1-DS) et deuxième année (M2-DS) du master est sélective. En dehors des qualités académiques, le recrutement accorde une part importante à la motivation des candidats.

La validation du M1-DS entraîne l'admission de droit en M2-DS.

Formation initiale. Le M1-DS s'adresse en priorité aux étudiants diplômés d'une licence de Mathématiques, d'une licence de Mathématiques et informatique appliquées



aux sciences humaines et sociales (parcours Mathématiques appliquées), ou d'un niveau équivalent. Sous réserve d'un niveau suffisant en mathématique, les dossiers d'étudiants ayant eu un parcours plus dirigé vers l'informatique sont considérés avec intérêt. Les droits d'inscription sont fixés par arrêté ministériel (droits ministériels : 243 € en 2022-2023, en M1 et M2).

Formation continue. Les candidatures relevant de la formation continue sont les bienvenues et considérées avec la plus grande attention. Les tarifs sont variables et communiqués suite à l'étude du dossier. Un ensemble de financements est disponible (se renseigner auprès de la Direction de la Formation Continue).

Les pré-requis à l'entrée du M1-DS couvrent le programme classique d'une Licence de Mathématiques ou tout autre programme européen ou international équivalent. **Outre l'aptitude mathématique, une connaissance et une pratique de base d'outils de programmation sont recommandées** (Python, Scilab, C et/ou R par exemple). L'acquisition du certificat C2i-1 d'informatique est apprécié. Le diplôme de master visant des compétences en anglais certifiées de niveau B2 du CECRL, un niveau minimal B1 dans cette langue est recommandé à l'entrée du M1-DS. L'enseignement se faisant essentiellement en français, **un niveau minimal B2 certifié en français est exigé des étudiants internationaux à l'entrée du M1-DS.**

DES RAISONS DE CHOISIR LE MASTER DATA SCIENCE (DS) DE L'UR

La qualité de la recherche française en mathématiques fondamentales et appliquées est mondialement reconnue : suivant les critères choisis, elle se place du premier au troisième rang mondial. Dans ce cadre, la recherche en mathématique du laboratoire LAREMA de l'Université d'Angers (62 membres dont 40 permanents), associé au CNRS et partenaire du Labex Lebesgue, se distingue par son excellence. Aussi, **les étudiants du Master DS ont l'assurance de bénéficier des connaissances et des innovations les plus récentes**, celles à même d'intéresser les entreprises dans leur stratégie compétitive.

L'Université d'Angers est reconnue par ses résultats en matière de taux de réussite et d'insertion professionnelle de ses 22 700 étudiants : sur ces deux points, les enquêtes ministérielles la situent aux toutes premières places des universités françaises. **Les étudiants du Master DS ont l'assurance d'un accompagnement performant de leurs études jusqu'à leur insertion professionnelle.** Angers et son agglomération, ville étudiante par excellence (plus de 38 000 étudiants), apparaît dans les enquêtes nationales comme étant l'une des villes les plus attractives dans ce domaine : vitalité de la vie étudiante, qualité et modération des prix du logement étudiant, facilité de déplacement, cadre de vie, etc.. En résumé, **les étudiants du Master DS ont l'assurance d'une qualité de vie et d'étude particulièrement**



**propice à leur réussite et à leur
insertion professionnelle.**



VOLUMES HORAIRES – ÉVALUATIONS

SEMESTRE 3

30 ECTS

UE	Matières	Volumes horaires					ECTS	Coef.	Contrôle des connaissances			
		CM	TD	TP	Auto. Et.	Tot.			1 ^{ère} session		2 ^e session	Durée CT
									Assidus	D.A.		
1	Statistique en Grande Dimension et Apprentissage	24	16	16	-	56	6	6	CT-0,5 ¹ CC-0,5 ¹	CT ¹	CT ¹	3h
2	Introduction à l'analyse du signal en temps discret	16	12	8	-	36	4	4	CT-0,5 CC-0,5	CT	CT	2h30
3	Séries Chronologiques et Applications	24	12	16	-	52	6	6	CT-0,5 CC-0,5	CT	CT	2h30 - S1 2h -S2
4	Informatique pour le Big Data	24	0	32	-	56	6	6	CC	CT	CT	2h30
Culture d'Entreprise												
5	Anglais	0	0	16	-	16	2	2	CC	Oral	Oral	30mn
	Communication	0	0	6	-	6	1	1	Oral	Oral	Oral	1h
Métiers - Options au choix 1 parmi 2												
6	Option 1 : Stat. et Science des données pour la biologie I	16	0	20	-	36	5	5	CC	CT	CT	2h
	Option 2 : Économétrie de l'évaluation I	14	12	0	-	26	5	5	CC	CT	CT	2h

¹ Note plancher de 8 à effectuer sur l'UE



Conditions de validation du semestre 3 :
Admis-e si moyenne générale > ou= 10 et
UE1 > ou= à 8 et
UE 4 > ou= à 8



Pas de DA et pas de Session 2

CT = Contrôle Terminal
 CC = Contrôle Continu

P = Validation en Présentiel
 DA = Dispensé d'Assiduité

Auto. Et. = Autonomie Étudiant



SEMESTRE 4

30 ECTS

UE	Matières	Volumes horaires					ECTS	Coeff.	Contrôle des connaissances			
		CM	TD	TP	Auto. Et.	Tot.			1 ^{re} session		2 ^e session	Durée CT
									Assidus	D.A.		
Culture d'Entreprise												
1	Conférences professionnelles et création d'entreprise	23	0	0	-	23	0	0	P	P	P	-
	3-PE	0	8	0	4	12	0	0	P	P	P	-
	Projet Annuel Tutoré	0	3	0	53	56	7	7	CC	-	-	-
Métiers 1 EC obligatoire + 1 EC d'Option au choix												
2	Obligatoire : Traitement d'images	12	0	12	-	24	3	3	CC	-	-	-
	Option 1 Bio : Stat. et Science des données en biologie II	8	0	8	-	16	2	2	CC	-	-	-
	Option 2 Eco : Économétrie de l'évaluation 2	14	12	0	-	26	2	2	CC	-	-	-
Choix de 1 parmi 2 entre stage ou alternance												
3	Stage	0	0	0	-	0	18	18	Rapport + soutenance	-	-	-
	Alternance											

Conditions de validation du semestre 4 :
Admis-e si moyenne générale > ou=10 et
Si Conférences professionnelles validées par la présence

Conditions de validation de l'année :
Admis-e si (Semestre 3 + Semestre 4) / 2 > ou= 10
et conditions de validation des semestres 3 et 4 validées

Pas de DA et pas de Session 2

CT = Contrôle Terminal
 CC = Contrôle Continu

P = Validation en Présentiel
 DA = Dispensé d'Assiduité

Auto. Et. = Autonomie Étudiant



CONTENUS DES ENSEIGNEMENTS

SEMESTRE 3

Les noms des responsables et intervenants sont donnés sous réserve de modifications.

BLOC MATHÉMATIQUES – UE OBLIGATOIRES

UE1

STATISTIQUE EN GRANDE DIMENSION ET APPRENTISSAGE

Statistics for Big Data and Machine Learning

Responsable **Fabien Panloup**
fabien.panloup@univ-angers.fr

Pré-requis

Notions et contenus : Algèbre linéaire et analyse (licence mathématiques L3). Statistique inférentielle (S1-UE3-DS). Optimisation non linéaire (S1-UE2-DS). Datamining et classification (S2-UE1-DS). Manipulation standard de R et de Python. Compétences et capacités :

— Maîtriser les compétences enseignées dans les modules de statistique inférentielle et de datamining et classification du M1, particulièrement tout ce qui concerne les modèles de régression linéaire et les méthodes de classification.

— Maîtriser les bases de l’algèbre linéaire, du calcul matriciel, de l’analyse et de l’optimisation, en particulier : recherche des valeurs propres et singulières, la notion de produit scalaire (sur l’exemple de \mathbb{R}^n et de L^2), les propriétés des espaces de Hilbert de fonctions ou encore l’optimisation sous contraintes.

— Maîtriser les notions de loi conditionnelle/espérance conditionnelle.

— Maîtriser les bases de l’optimisation déterministe (descente de gradient et extensions).

— Avoir une connaissance minimale des langages R et Python (syntaxe, manipulation élémentaire, calcul matriciel, gestion des graphiques, etc.).

Contenus

Bases mathématiques de l’apprentissage statistique. Bases de la grande dimension (« fléau » de la dimension...). Régression en grande dimension (motivation par l’exemple, théorie du Lasso, seuillage, sparsité, inégalités classiques, régressions basées sur les composantes principales, PLS). Méthodes de pénalisation (généralisation du Lasso (Elastic-Net, Group-Lasso), sélection de variables, Ridge, etc.). Classification et apprentissage en grande dimension (SVM, Logistic-Lasso, forêts aléatoires, k plus proches voisins, ACP à noyaux, Boosting, Gradient Boosting). Introduction aux réseaux de neurones et au deep learning. Bases de l’optimisation stochastique. Large mise en pratique avec R ou Python.

Compétences

— Avoir un recul professionnel sur la notion de grande dimension. Comprendre par l’exemple les raisons théoriques mais aussi les raisons techniques qui différencient les méthodes statistiques classiques des méthodes statistiques en grande dimension.

— Comprendre la théorie du Lasso et d’une manière générale la méthodologie consistant à pénaliser une estimation par des contraintes sur les coefficients de la régression. Avoir une vision claire des objectifs visés par les méthodes de pénalisation (estimation ou sélection de variables).

— Savoir réduire la dimension des observations par des méthodes classiques d’analyse des données (par exemple l’ACP) et savoir exploiter cette réduction de dimension pour les régressions en grande dimension ou possédant de la



multicolinéarité.

— Savoir ce qui relie et ce qui différencie le Lasso des méthodes de pénalisation basées sur des normes différentes (Elastic-Net, Ridge, etc.). Connaître les points forts et points faibles de chacune des approches pour en permettre une mise en pratique pertinente sur des jeux de données réelles.

— Connaître les méthodes simples d'apprentissage automatique appliquées à la classification (comme les k plus proches voisins) et les méthodes plus sophistiquées (comme les forêts aléatoires). Avoir une vision d'ensemble de la théorie des SVM appliqués à la classification.

— Connaître les bases du boosting (Adaboost/Gradient Boosting, XGBoost).

— Comprendre le fonctionnement et savoir mettre en oeuvre un algorithme d'apprentissage par deep learning (perceptron et réseaux de neurones convolutionnels).

— Être capable de mobiliser et exploiter les méthodes étudiées dans ce module dans des cas pratiques sous R ou Python, en particulier les méthodes de régression et de sélection de variables en grande dimension.

Supports ou références

— *Christophe Giraud, Introduction to High-Dimensional Statistics (2015).*

UE2

INTRODUCTION À L'ANALYSE DU SIGNAL EN TEMPS DISCRETS

Introduction to Discrete-Time Signal Processing

Responsable **David Rousseau**
david.rousseau@univ-angers.fr

Pré-requis

Notions et contenus : espace euclidien, produit scalaire (licence mathématiques L3); séries de Fourier (licence mathématiques L2).

Compétences et capacités : maîtriser le calcul des coefficients d'une série de Fourier, la décomposition d'une fonction sur

une base orthogonale ; connaître et savoir appliquer le théorème de Parseval et les inégalités de Bessel ; maîtriser le langage de programmation Python.

Contenus

Analyse du signal, motivations, exemples
Séries de Fourier, Transformée de Fourier
Filtres FIR (Réponse Impulsionnelle Finie) et IIR (Réponse Impulsionnelle Infinie) – Impulsion – convolution

Transformation en Z – Application à la stabilité des filtres

Problèmes d'aliasing

Application à des signaux temporels de natures diverses

Compétences

— Maîtriser les calculs de transformation de Fourier et de Fourier Inverse.

— Savoir tester la stabilité d'un filtre.

— Savoir concevoir un filtre correspondant à des objectifs donnés

— Savoir utiliser Python et des bibliothèques de type `scipy.signal` pour mettre en pratique la conception des filtres.

Supports ou références

— *Tisserand, É., Pautex, J. F., & Schweitzer, P. (2009). Analyse et traitement des signaux-2e éd.: Méthodes et applications au son et à l'image. Dunod.*

UE3

SÉRIES CHRONOLOGIQUES ET APPLICATIONS

Time Series and Applications

Responsable **Frédéric Proïa**
frederic.proia@univ-angers.fr

Pré-requis

Notions et contenus :

— Bagage mathématique (algèbre, probabilités, etc.) de Licence et de Master 1.

— Modules de Statistique et de Modèles de régression du M1.

— Manipulation standard de R et de Python.

Compétences et capacités :



- Savoir manipuler les opérateurs du calcul des probabilités (espérance, variance, covariance, etc.) appliqué aux variables et aux vecteurs aléatoires.
- Maîtriser les notions de statistique inférentielle du S1 et de régression linéaire du S2.
- Maîtriser les bases de l'algèbre linéaire et de l'analyse concernant les espaces vectoriels, les produits scalaires et la projection orthogonale (sur l'exemple de \mathbb{R}^n et de L^2).
- Avoir une connaissance minimale des langages R et Python (syntaxe, manipulation élémentaire, gestion des graphiques, etc.).

Contenus

Décomposition des séries chronologiques (modèles additif et multiplicatif, filtrage linéaire). Processus aléatoires discrets (bruit blanc, loi d'un processus, espace L^2). Stationnarité (sens fort, sens faible). Processus linéaires (opérateurs chronologiques, bruit d'innovation, causalité, ACF/PACF). Étude détaillée des processus ARMA. Inférence dans les processus ARMA. Introduction à la non-stationnarité (ARIMA-SARIMA) et à l'hétéroscédasticité (ARCH-GARCH). Application à la prévision (meilleur prédicteur linéaire, lissages exponentiels). Introduction à la statistique spatiale (krigeage). Large mise en pratique avec R et Python.

Compétences

- Être capable de reconnaître visuellement une série à composantes additives ou multiplicatives, avec tendance et/ou périodicité, et d'éliminer les tendances par un filtrage approprié.
- Savoir étudier la stationnarité d'un processus discret et savoir la tester en pratique. Comprendre la différence entre stationnarité forte et stationnarité au second ordre.
- Savoir calculer la fonction d'autocorrélation d'un processus stationnaire. Savoir construire son autocorrélation partielle par un algorithme approprié.
- Acquérir une approche théorique solide d'un processus ARMA. Savoir établir des conditions sur ses paramètres per-

mettant d'étudier ses propriétés.

- Savoir estimer l'ACF et la PACF d'une série chronologique stationnaire, en proposer une modélisation ARMA adaptée par comparaison avec le comportement théorique des autocorrélations.
- Savoir étendre la définition des processus ARMA aux processus ARIMA, SARIMA, ARCH et GARCH ainsi que leur mise en pratique.
- Connaître les principaux tests d'hypothèses adaptés au cadre ARMA/ARIMA/SARIMA (tests de stationnarité, de significativité des paramètres, d'autocorrélation, etc.).
- Avoir un recul professionnel sur la modélisation d'une série, savoir comparer plusieurs approches sur la base de critères adaptés aux objectifs de l'étude.
- Être capable de mobiliser toutes ces compétences dans des cas pratiques avec R et Python.
- Connaître les rudiments de la statistique spatiale par l'intermédiaire du krigeage.

Supports ou références

- *P. Brockwell, R. Davis. Time Series: Theory and Methods. Seconde édition. Springer Series in Statistics, 1991.*



UE4

INFORMATIQUE POUR LE BIG DATA

Big Data

Responsable **Jacquelin Charbonnel**
jacquelin.charbonnel@univ-angers.fr

intervenants **Marie Badreau | Mohamed Benzarga | Jacquelin Charbonnel | David Rousseau**

Pré-requis

Notions et contenus : Bases de données relationnelles (S2-UE4-DS). Programmation (S1-UE5-DS). Statistique (S2-UE1-DS).

Compétences et capacités : Connaissances de base des modèles relationnels. Connaissances en statistique et des langages de programmation.

Contenus

Concepts et les enjeux du Big Data

Technologies du Big Data :

- Architecture et composants de la plateforme Hadoop.
- Modes de stockage (NoSQL, HDFS).
- Principes de fonctionnement de MapReduce.
- Programmation concurrente (générateurs, multi-threading, multi-processing, asynchronisme).
- Web services.
- Nasse de data visualisation.

Gérer les données structurées et non structurées :

- Principes de fonctionnement HDFS.
- Importer des données externes vers HDFS.
- Réaliser des requêtes SQL avec HIVE.
- Utiliser PIG pour traiter la donnée.

Les méthodes d'analyse des données pour le Big Data (en liaison avec S3-UE1-DS) :

- Les méthodes d'exploration.
- Segmentation et classification.
- Estimation et prédiction.
- Implémentation des modèles.
- Méthodes de réduction de dimensions.

Introduction au MLOps : Mise en production ou déploiement d'un modèle de Machine Learning.

- Aspects techniques d'un déploiement.
- Mise en pratique : déploiement d'une API Web avec Flask.
- Maintenance post-déploiement : logging, model drift, qualité des données.
- Prérequis d'un déploiement réussi : versionnage, tests unitaires, bonnes pratiques de code, etc.
- Vue d'ensemble des métiers et logiciels liés au MLOps.

Compétences

- Comprendre les concepts et l'apport du Big Data par rapport aux enjeux métiers.
- Comprendre l'écosystème technologique nécessaire pour réaliser un projet Big Data.
- Acquérir les compétences techniques pour gérer des flux de données complexes, non structurés et massifs.
- Visualisation d'espaces latents, implémenter des modèles d'analyses statistiques pour répondre aux besoins métiers.
- Comprendre l'intérêt du déploiement d'un modèle de Machine Learning
- Avoir une vue d'ensemble des points d'attention d'une mise en production, avant, pendant et après sa réalisation
- Acquérir une compétence technique de base pour créer une API Web avec Flask ; mettre en pratique quelques aspects du MLOps.

UE5

ANGLAIS

English

Responsable [Philippe Torrès](mailto:philippe.torres@univ-angers.fr)
philippe.torres@univ-angers.fr
Intervenant [Bruno Menan](mailto:bruno.menan@univ-angers.fr)
bruno.menan@univ-angers.fr

Pré-requis

Notions et contenus : Les bases de la langue anglaise.

Compétences et capacités : Maîtriser le niveau B1 du CECRL (dit « d'utilisateur indépendant »).

Contenus

—Permettre aux étudiants de continuer à travailler les cinq compétences en langue - compréhension écrite et orale, expression écrite et orale, et interaction orale - à travers des supports authentiques (articles, documentaires, documents audio et vidéo d'internet, graphiques...) et des activités variées (exercices de compréhension, d'expression écrite, jeux de rôle, débats, présentations orales...).

— Etoffer les connaissances lexicales.

— Améliorer la prononciation et revoir certains points de langue le cas échéant.

—Valider à l'issue du 3ème semestre un niveau B2 du CECRL (par le biais d'une certification).

Compétences

On vise, à l'issue du 2ème voire 3ème semestre du Master, le niveau de compétence B2 du Cadre Européen Commun de Référence pour les Langues (CECRL) qui est résumé comme suit : « Peut comprendre le contenu essentiel de sujets concrets ou abstraits dans un texte complexe, y compris une discussion technique dans sa spécialité. Peut communiquer avec un degré de spontanéité et d'aisance tel qu'une conversation avec un locuteur natif ne comporte de tension ni pour l'un ni pour l'autre. Peut s'exprimer de façon

claire et détaillée sur une grande gamme de sujets, émettre un avis sur un sujet d'actualité et exposer les avantages et les inconvénients de différentes possibilités. »

UE5

COMMUNICATION – PRÉPARATION À L'INSERTION PROFESSIONNELLE

Communication – Professional Integration Program

Responsable [Fabien Panloup](mailto:fabien.panloup@univ-angers.fr)
fabien.panloup@univ-angers.fr
Intervenants [Intervenants du SUIO-IP](#)
[et/ou autres intervenants professionnels](#)

Pré-requis

Compétences de l'UE S2-UE5 communication.

Contenus

—Techniques de communication - Préparation au travail en groupe : Les fonctions de l'animateur ; Conduite de réunions - Techniques de rédaction des documents — Techniques de présentation orale - La recherche d'emploi : Lettres de candidature ; Curriculum vitae ; Entretien de candidature.



UE6

OPTION 1

STATISTIQUE ET SCIENCE DES DONNÉES POUR LA BIOLOGIE I

Responsable **Mikael Escobar-Bach**
mikael.escobar-bach@univ-angers.fr
Intervenants **Jérémy Clotault** | **Mikael Escobar-Bach** | **Frédéric Proia**
jeremy.clotault@univ-angers.fr
frederic.proia@univ-angers.fr

Pré-requis

Notions et contenus :

- Cours de Introduction à la génomique, Statistique et Modèles de régression du M1.
- Compétences en R et/ou Python.

Compétences et capacités : Maîtriser les concepts généraux des probabilités, des statistiques et des modèles de régression vus en M1, ainsi que l'introduction à la génomique.

Contenus

Cet enseignement, divisé en trois parties respectivement intitulée Introduction à l'Analyse de Survie, Introduction aux plans d'expériences et Modèles de reconstruction de généalogie, a pour objectif de présenter quelques thèmes importants de Statistique et Modélisation pour la biologie. Ci-dessous une description plus précise.

Introduction à l'Analyse de Survie

Introduction to Survival Data Analysis

Responsable **Mikael Escobar-Bach**
mikael.escobar-bach@univ-angers.fr

Le problème des données manquantes est fréquemment rencontré dans un grand nombre de domaines d'applications statistiques. Dans le cadre de la biologie, il n'est pas rare que l'acquisition de données de suivi soit en partie incomplète : pour un essai thérapeutique, le temps de

rémission puis de guérison d'un patient peut être trop long et donc jamais observé, on parle alors de données censurées. Dans ce cours, nous introduirons la méthodologie (fonction de survie, risque instantané, ...) et les outils statistiques (estimateur de Kaplan-Meier, Nelson-Aalen, ...) qui permettent d'appréhender ce genre de problèmes. Une implémentation sur R sera aussi proposée.

Introduction aux plans d'expérience

Introduction to Design of Experiments

Responsable **Mikael Escobar-Bach**
mikael.escobar-bach@univ-angers.fr

Les plans d'expérience forment un outil essentiel en statistique expérimentale, ils permettent d'esquisser les contours d'une méthodologie et d'organiser au mieux la suite des mesures à effectuer afin de répondre à une problématique statistique donnée.

Nous décrivons de manière succincte quelques plans d'expérience classiquement utilisés dans les sciences du vivant.

Modèles de reconstruction de généalogie

Probabilistic reconstruction of genealogies

Responsable **Jérémy Clotault**
jeremy.clotault@univ-angers.fr

Ce module est dédié à l'étude d'un cas pratique de reconstruction généalogique dans une population végétale. Après quelques concepts génétiques généraux liés aux niveaux de ploïdie, à la production des gamètes et aux croisements, nous proposerons un modèle permettant de reconstruire, à partir de marqueurs génétiques, un ensemble d'arbres généalogiques probabilisés dans une population végétale soumise à des lois de reproduction spécifiques. Nous étudierons des reconstructions dans des populations diploïdes ou tétraploïdes, et un travail pratique sera demandé dans le cas diploïde.



Une partie de ce travail sera dédiée à la programmation effective de la reconstruction généalogique en R ou en Python. La visualisation des résultats sera réalisée à l'aide du package `igraph` de R ou de son équivalent en Python.

Compétences

Logit-Probit-Tobit-Modèles de Heckman. Choix discrets/ Les modèles de choix binaire (probit)/modèle Logit/modèle Logit multinomial/modèle Logit conditionnel/modèles à choix ordonnés/modèles de comptage/modèles à variables limitées (tronquées)/modèles d'autosélection/ Arbres de décision et Nested Logit. Visualisation de graphes. Arbres probabilisés. Principe des plans d'expérience.

Supports ou références

— *Analysis of Survival Data* by D.R. Cox, David Oakes, 1984 by Chapman and Hall/CRC.
— *Le modèle linéaire par l'exemple*, JM.

UE6

OPTION 2

ÉCONOMÉTRIE DE L'ÉVALUATION 1

Econometric Techniques 1

Responsable [Philippe Compaire](mailto:philippe.compaire@univ-angers.fr)
philippe.compaire@univ-angers.fr

Pré-requis

Cours d'économétrie niveau L3

Notions et contenus :

- 1- Introduction
- 2- Chapitre I : Les modèles de choix binaires (probit)
- 3- Chapitre II : Le modèle logit
- 4- Chapitre III : Le modèle logit multinomial
- 5- Chapitre IV : Le modèle logit conditionnel
- 6- Chapitre V : Les modèles à choix ordonnés
- 7- Chapitre VI : Les modèles de comptage
- 8- Chapitre VII : Les modèles à variables limitées (tronquées)

9- Chapitre VIII : Arbres de décision et Nested logit

Compétences :

On expose les méthodes d'estimations d'un modèle non linéaire dans la variable endogène. Cas de choix discrets dichotomiques et polytomiques, modèles de comptage et des observations censurées.

Contenus

Modèles Logit-Probit-Tobit. Arbres de décision et Nested Logit.

Compétences

Applications sur données réelles macroéconomiques et microéconomiques.

Supports ou références

— *Greene-Maddala-Train-Thomas*



SEMESTRE 4

BLOC CULTURE D'ENTREPRISE – UE OBLIGATOIRE

UE1

CRÉATION D'ENTREPRISE - PRÉPARATION À L'INSERTION PROFESSIONNELLE

Responsable **Fabien Panloup**
fabien.panloup@univ-angers.fr
Intervenants **SUIO-IP**

Contenus

L'enseignement dédié à la création d'entreprise vise à fournir aux étudiants les éléments méthodologiques nécessaires. Il s'appuiera sur l'intervention de professionnels et sollicitera un travail personnel des étudiants sous forme de capsule.

Conférences Professionnelles

Responsable **Fabien Panloup**
fabien.panloup@univ-angers.fr
Intervenants **Extérieurs**

Contenus

Des exposés et des mini-cours seront proposés par des professionnels à la fois pour présenter le métier de data scientist mais aussi pour compléter la formation par des introductions à des thèmes de développement souvent récents.

Projet Annuel Tutoré

Annual Training Project

Responsable **Fabien Panloup**
fabien.panloup@univ-angers.fr

Contenus

Les étudiants réalisent en groupe un projet annuel entre septembre et février avec une soutenance à mi-projet, sous l'encadrement d'un tuteur enseignant. La réalisation du projet s'efforcera de calquer le déroulement de la gestion de projets dans le monde socio-économique. Dans une première phase, après avoir circonscrit le problème, un cahier des charges sera

réalisé et un calendrier sera défini pour la réalisation. La seconde phase du projet consistera au traitement du problème en utilisant les concepts, méthodes et outils appropriés et satisfaisants au cahier des charges. Le projet annuel met en oeuvre l'ensemble des compétences théoriques et pratiques du master, sur un thème en lien avec l'option métier.

Compétences

- Autonomie.
- Capacité à travailler en équipe.
- Capacité à mettre en oeuvre les contenus du master.
- Capacité à rédiger un rapport.
- Capacité à réaliser une soutenance.
- Capacité créative et de recherche.



UE2

TRAITEMENT D'IMAGES

Responsable [David Rousseau](mailto:david.rousseau@univ-angers.fr)
david.rousseau@univ-angers.fr

Pré-requis

Notions et contenus :

- Contenus des cours de -S3-M2 DS
« Statistique en Grande Dimension et
Apprentissage » et « Bases de Données
Avancées-Big Data ».

Compétences :

- Traitement des données avec Python.
- Bases sur les réseaux de neurones.

Contenus

Cet enseignement a pour objectif d'introduire les concepts principaux du traitement d'images pour les mettre en oeuvre via des outils logiciels libres standards (Scikit image Python, imageJ, Illastik, Knime, ...) sur des questions variées. A l'issue du cours, les étudiants devront être capables de construire un «pipeline» de traitement d'images pour réaliser de l'apprentissage machine basé expert ou par réseau de neurones profonds.

Supports ou références

— Peters, J., Janzing, D., & Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms* (p. 288). The MIT Press.

— Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). *Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more*. O'Reilly Media.



UE2

OPTION 1 Bio

STATISTIQUE ET SCIENCES DES DONNÉES POUR LA BIOLOGIE II

Responsable [Olivier Goudet](mailto:olivier.goudet@univ-angers.fr)
olivier.goudet@univ-angers.fr

Pré-requis

Notions et contenus :

- Contenu du cours S2-M1 DS « Introduction à la génomique ».
- Contenus des cours de S3-M2 DS « Statistique en Grande Dimension et Apprentissage » et « Bases de Données Avancées-Big Data ».

Compétences :

- Traitement des données avec Python.
- Modèles de régression multivariée.
- Modèles graphiques.

Contenus

Cet enseignement se focalise principalement sur l'inférence de réseaux et la transcriptomique. Il a pour objectif de renforcer les compétences en analyse de données génomiques et d'introduire les réseaux biologiques.

Cet enseignement sera dispensé sous forme de CM et de TP de mise en pratique sur des données réelles. La réalisation d'un projet de traitement de données transcriptomiques et d'inférence de réseaux sera proposée à l'issue du cours.

Compétences

Introduction aux réseaux biologiques | Méthodes d'inférences de réseaux | Causalité | Méthode de validation.

Supports ou références

- *Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: foundations and learning algorithms (p. 288). The MIT Press.*
- *Ramsundar, B., Eastman, P., Walters, P., & Pande, V. (2019). Deep learning for*

the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more. O'Reilly Media.

UE2

OPTION 2 Eco

ÉCONOMÉTRIE DE L'ÉVALUATION 2

Econometric Techniques 2

Responsable [Christophe Daniel](mailto:christophe.daniel@univ-angers.fr)
christophe.daniel@univ-angers.fr

Pré-requis

Notions, contenus et compétences :

- Modèles micro-économiques du travail-Econométrie des données transversales et de panel.

Contenus

1 Théories microéconomiques :

- Des inégalités et des discriminations (hommes-femmes).
- De l'éducation (rendements de l'éducation ou de l'expérience), négociations collectives.
- Du choix des conditions de travail (théories des différences compensatrices, de la segmentation, et des salaires hédoniques).
- De la satisfaction au travail.

2 Passage des modèles théoriques aux modèles économétriques : fonctions de gains (Mincer, élargies aux conditions de travail...), équation de choix des conditions de travail, biais d'endogénéité et de sélection, méthodes de correction (Variables instrumentales et méthode d'Heckman de l'Inverse du Ratio de Mill), modèle à deux régimes, décomposition de Blinder-Oaxacca, de Machado-Mata, indices de satisfaction...

3 Applications sur données transversales : estimations et tests

Méthodes d'estimation : MCO, Moindres Carrés Pondérés, Moindres Carrés Généralisés, probit et logit (odd-ratios, modèles simples ou bivariés), Variables instrumentales, Heckman, estimation par



quantiles, Principal Component Regression et Partial Least Squares.

4 Évaluations contrefactuelles des politiques publiques de l'emploi :

Approche expérimentale (groupe test, groupe de contrôle), différences-en-différences, score de propension, bootstrap...

UE3

STAGE OU ALTERNANCE

Stage de 5 à 6 mois en entreprise ou alternance sur l'année

L'étudiant devra accomplir un stage en entreprise ou dans un laboratoire de recherche sur un sujet en adéquation avec les objectifs de la formation. Pour les alternants, la période de stage s'effectue sur les périodes de l'année scolaire définies dans le calendrier ci-joint. Pour les étudiants en formation initiale, il s'agira d'un stage de 22 à 26 semaines environ, à partir de la mi-mars. Les étudiants sont suivis par un tuteur entreprise et par un tuteur université.

Ce stage est finalisé par la rédaction d'un rapport d'une trentaine de pages et par une soutenance qui se déroule début septembre. La soutenance se déroule devant un jury constitué de membres de l'équipe pédagogique du master, en présence d'un ou plusieurs membres de la structure dans laquelle s'est déroulé le stage (possibilité de visioconférence pour ces derniers). L'évaluation prend en compte l'avis des tuteurs entreprise, la qualité scientifique et rédactionnelle du rapport ainsi que la prestation de l'étudiant lors de la soutenance, tant dans la présentation que dans les réponses aux questions.

N.B. Les stages peuvent se prolonger jusqu'à la veille de la date de jury si nécessaire.



CALENDRIERS

Semaines	#	Début	Fin	Formation initiale	Alternance
36	1	04/09/2023	08/09/2023		Entreprise
37-43	7	11/09/2023	27/10/2023	Enseignement S3	Enseignement S3
44-46	3	30/10/2023	17/11/2023	Projet annuel	Entreprise
47-50	4	20/11/2023	15/12/2023	Enseignement S3	Enseignement S3
51	1	18/12/2023	22/12/2023	Projet annuel	Entreprise
52-1	2	26/12/2023	05/01/2024	Vacances	Entreprise
2	1	08/01/2024	12/01/2024	Enseignement S3	Enseignement S3
3	1	15/01/2024	19/01/2024	Examens S3	Examens S3 <i>Selon convocation sinon entreprise</i>
4-5-6	3	22/01/2024	09/02/2024	Enseignement S4	Enseignement S4
7-8	2	12/02/2024	23/02/2024	Projet annuel	Entreprise
9	1	26/02/2024	01/03/2024	Enseignement S4	Enseignement S4
10	1	04/03/2024	08/03/2024	S4 Examen S3 Session 2	S4 Examen - S1 Session 2 <i>Selon convocation sinon entreprise</i>
11	1	11/03/2024	15/03/2024	Soutenance Projet annuel et/ou début stage	Entreprise
12-18	7	18/03/2024	3/05/2024	Stage	Entreprise
19-21	1	06/05/2024	24/05/2024	Stage	Projet annuel
22-35	14	27/05/2024	30/08/2024	Stage	Entreprise
35	0,5	02/09/2024	03/09/2024	Soutenance Stage	Soutenance Stage



CALENDRIER ALTERNANCE (CONTRAT PRO OU APPRENTISSAGE)

MASTER MATHÉMATIQUES Parcours Data Science

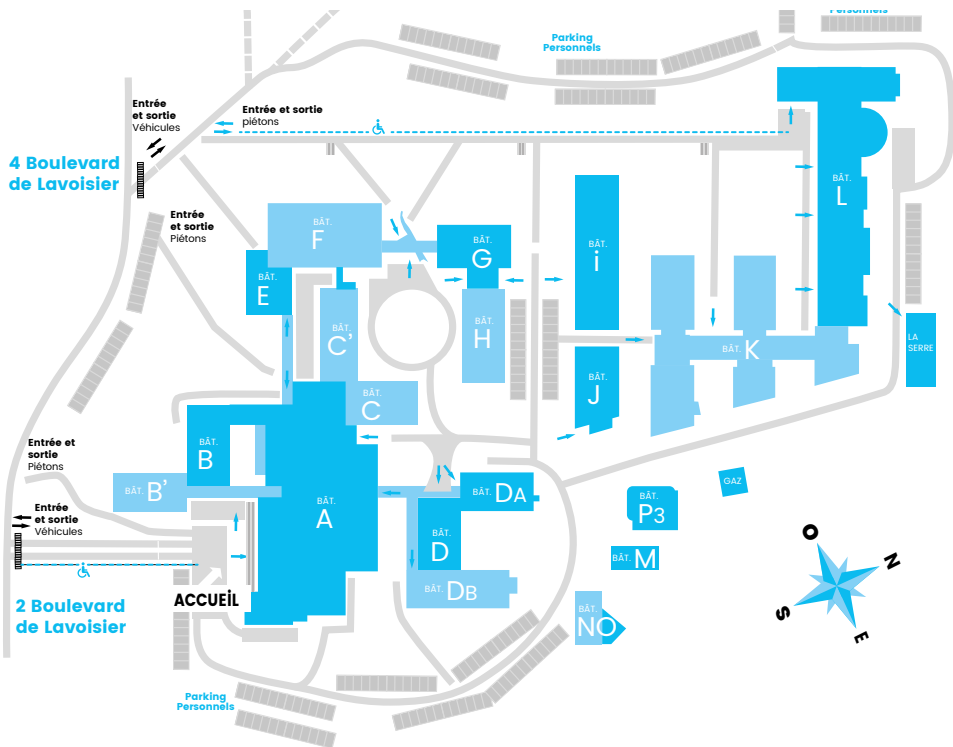
		2023					2024							
		Sept.	Oct.	Nov.	Déc.	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Sept.
		M A	M A	M A	M A	M A	M A	M A	M A	M A	M A	M A	M A	M A
1 M		1 V	1 D	1 M	1 V	1 L	1 J	1 V	1 L	1 M	1 S	1 L	1 J	1 D
2 M		2 S	2 L	2 J	2 S	2 M	2 V	2 S	2 M	2 J	2 D	2 M	2 V	2 L
3 J		3 D	3 M	3 V	3 D	3 M	3 S	3 D	3 M	3 V	3 L	3 M	3 S	3 M
31		4 V	4 M	4 J	4 S	4 L	4 D	4 L	4 J	4 S	4 M	4 J	4 D	4 M
5 S		5 M	5 D	5 V	5 M	5 V	5 L	5 M	5 V	5 D	5 M	5 V	5 L	5 J
6 D		6 M	6 V	6 L	6 M	6 S	6 M	6 M	6 S	6 L	6 J	6 S	6 M	6 V
7 L	36	7 J	7 S	7 M	7 J	7 D	7 M	10	7 D	7 M	7 V	7 D	7 M	7 S
8 M		8 V	8 D	8 M	8 V	8 L	8 J	8 V	8 L	8 M	8 S	8 L	8 J	8 D
9 M		9 S	9 L	9 M	9 S	9 M	9 V	9 S	9 M	9 J	9 D	9 M	9 V	9 L
32	10	10 D	10 M	10 V	10 D	10 M	10 S	10 D	10 M	10 V	10 L	10 M	10 S	10 M
11 V		11 L	11 M	11 S	11 L	11 J	11 D	11 L	11 S	11 V	11 M	11 J	11 D	11 M
12 S		12 M	12 V	12 J	12 S	12 V	12 L	12 M	12 V	12 D	12 M	12 V	12 L	12 J
13 D		13 M	13 V	13 L	13 M	13 S	13 M	13 M	13 S	13 L	13 J	13 S	13 M	13 V
14 L		14 J	14 S	14 M	14 J	14 D	14 M	11	14 J	14 D	14 V	14 D	14 M	14 S
15 M		15 V	15 D	15 M	15 V	15 L	15 J	15 V	15 L	15 M	15 S	15 L	15 J	15 D
16 M		16 S	16 L	16 J	16 S	16 M	16 V	16 S	16 M	16 J	16 D	16 M	16 V	16 L
33	17	17 D	17 M	17 V	17 D	17 M	17 S	17 D	17 M	17 V	17 L	17 M	17 S	17 M
18 V		18 L	18 M	18 S	18 L	18 J	18 D	18 L	18 J	18 S	18 M	18 J	18 D	18 M
19 S		19 M	19 V	19 D	19 M	19 V	19 L	19 M	19 V	19 D	19 M	19 V	19 L	19 J
20 D		20 M	20 V	20 L	20 M	20 S	20 M	20 M	20 S	20 L	20 J	20 S	20 M	20 V
21 L		21 S	21 M	21 V	21 D	21 M	21 L	21 J	21 D	21 M	21 V	21 D	21 M	21 S
22 M		22 D	22 M	22 V	22 L	22 V	22 L	22 V	22 L	22 M	22 S	22 L	22 M	22 D
23 M		23 S	23 L	23 V	23 S	23 M	23 V	23 S	23 M	23 J	23 D	23 M	23 V	23 L
34	24	24 D	24 M	24 V	24 L	24 M	24 D	24 D	24 M	24 V	24 L	24 M	24 S	24 M
25 V		25 L	25 M	25 S	25 L	25 J	25 D	25 L	25 J	25 S	25 M	25 J	25 D	25 M
26 S		26 M	26 V	26 D	26 M	26 V	26 L	26 M	26 V	26 D	26 M	26 V	26 L	26 J
27 D		27 M	27 V	27 L	27 M	27 S	27 M	27 M	27 V	27 L	27 J	27 S	27 M	27 V
28 L		28 S	28 M	28 V	28 D	28 M	28 J	28 D	28 S	28 M	28 V	28 D	28 M	28 S
29 M		29 V	29 D	29 M	29 V	29 L	29 J	29 V	29 L	29 M	29 S	29 L	29 J	29 D
35	30	30 M	30 V	30 L	30 S	30 M	30 M	30 S	30 M	30 J	30 D	30 L	30 V	30 L
31 J		31 M	31 M	31 J	31 D	31 M	31 M	31 D	31 M	31 V	31 D	31 M	31 M	31 L

Formation à l'Université
 périodes en entreprise
 Week-end
 Projet annuel (en entreprise si lien avec la mission, sinon université)

Soutenances (selon convocation sinon entreprise)
 Exams (selon convocation sinon entreprise)
 Jours fériés

Volume heures de cours & examens : 403h
 Volume des heures de travail en autonomie (non éméragé) : 262h
 Volume horaire total de la formation : 665h
 Reentrée universitaire : Lundi 4 septembre 2023





- A** Administration | Scolarité | Enseignement (Amphi A à E)
- B** Biologie végétale | Physiologie végétale | Travaux pratiques biologie
- B'** Travaux pratiques biologie
- C** Travaux pratiques chimie
- C'** Département de Géologie | Recherche environnement (LETG -LEESA) | Recherche géologie (LPGN-BI&F)
- D** Travaux pratiques physique
- Da** Enseignement | Travaux pratiques physique
- Db** Département de Physique | Recherche physique (LPHIA)
- E** Travaux pratiques biologie
- F** Département de Biologie | Recherche neurophysiologie (SIFCIR) | Travaux pratiques biologie, géologie
- GH** Département informatique | Recherche informatique (LERIA) | Travaux pratiques géologie
- I** Département Mathématiques | Recherche Mathématiques (LAREMA)
- J** Chimie enseignement | Travaux pratiques
- K** Département de Chimie | Recherche Chimie (MOLTECH Anjou)
- L** Espace multimédia | Enseignement (Amphi L001 à L006) | Salle d'examen rez-de-jardin

Impression Service Reprographie UA

Ua
**FACULTÉ
 DES SCIENCES**
 UNIVERSITÉ D'ANGERS

2, Boulevard Lavoisier
 49045 ANGERS CEDEX 01
 T.0241735353
 www.univ-angers.fr



LE TRI + FACILE