

Programmes des UEs du master ISADS (à partir de 2022)

Les UEs dotées d'une astérisque sont mutualisées avec les étudiant.e.s du master Mathématiques et Apprentissage Statistique.

Programme du M1

Méthodes stochastiques et statistiques - 81h - 9 ECTS [Agnès Grimaud, Mariane Pelletier]

Rappels de probabilités, fondements de la statistique: Estimation, Intervalles et régions de confiance, tests statistiques.

Traitement statistique des données (*) - 54h - 6 ECTS [Ester Mariucci, Charles Tillier] :

Le but de ce cours est de prédire quels types d'algorithmes de machine learning existants seront les plus appropriés pour quels types de tâches, sur la base de propriétés formelles et de résultats expérimentaux. Et également d'évaluer et analyser les algorithmes d'apprentissage existants. Dans un premier temps on s'intéressera à l'analyse des données, à comment extraire ou synthétiser l'information contenue dans un jeu de données. Nous nous intéresserons à plusieurs classes de méthodes et tout particulièrement aux méthodes d'analyse factorielle (ACP, AFC et ACM). Ensuite, on se focalisera sur les méthodes d'apprentissage supervisé et on abordera plusieurs algorithmes : K-plus proches voisins, régression logistique, SVM, et arbres de décision.

Outils mathématiques - 54h - 6 ECTS [Catherine Donati, Julien Worms, Ester Mariucci]

UE en 3 parties : (i) rappels sur les matrices symétriques et orthogonales, sur les espaces euclidiens, l'Orthogonalité, et les projections. Application en statistique à la méthode des moindres carrés. (ii) Vecteurs aléatoires, matrice de variance-covariance et ses propriétés, loi gaussienne multivariée, vecteurs gaussiens, propriétés des vecteurs gaussiens, Théorème de Cochran, Théorème Central Limite multivarié. (iii) Introduction aux concepts de risque empirique et processus empiriques, inégalités de concentration. .

Bases de données (*) - 45h - 5 ECTS [Département d'informatique]

Généralités sur les bases de données ; Type de données ; Dictionnaire de données ; Le modèle relationnel ; Le langage SQL ; Méthodes de conception de BD relationnelles, graphe des dépendances fonctionnelles, entités-associations

Logiciels statistiques (*) - 72h - 8 ECTS (R (15h), SAS (27h), Python(30h)) [Agnès Grimaud, Charles Tillier, Julien Worms] :

Partie R : programmation R, manipulation de différents objets, boucles, fonctions. Applications en statistique

Partie Python : introduction générale sur Python et la programmation object; utilisation et manipulation des bibliothèques de bases (NumPy, Math, Scipy); initiation à l'apprentissage statistique avec Scikit-Learn.

Partie SAS : concepts généraux de la programmation SAS (étapes, librefs, formats, options de noms de tables), fonctionnement fin de l'étape data, jointures, manipulation de données textuelles, arrays, macro-variables et macro-langage.

Statistique linéaire (*) - 54h - 6 ECTS [Julien Worms]

Modèle linéaire général et cas particulier de la régression simple ou multiple, traitement par la

méthode des moindres carrés. Propriétés des estimateurs avec ou sans normalité. Décomposition de la variance. Ratios de Student et tests basés sur les matrices de contraste. Intervalles de prédiction. Test de Fisher général pour modèles emboîtés. Approche par vraisemblance et lien entre le LRT et le test de Fisher. Diagnostics (résidus studentisés, QQ-plots, autres indicateurs usuels). Introduction à la pénalisation de la dimension. Une attention particulière sera portée à la gestion des régresseurs catégoriels (interprétation des coefficients et tests associés)

Outils professionnels (Excel (10h), VBA (20h), actuariat (15h)) - 45h - 5 ECTS [Intervenant extérieur, Mariane Pelletier]

Programme de la partie actuariat : Introduction à l'actuariat. Méthodes classiques déterministes de provisionnement (méthode chain-ladder, london chain, london pivot).

Stage – 15 ECTS (non compensable)

Programme M2

Techniques de prévision (*) – 36h – 4 ECTS [Agnès Grimaud]

Le but est d'effectuer des prévisions à partir d'une série chronologique, les différentes étapes sont : modélisation par un processus, estimation du modèle, validation et prévisions. Programme : Décomposition d'une série chronologique : tendance, saisonnalité et bruit. Estimation des différentes composantes. Etude du bruit : processus stationnaires, ARMA. Savoir les identifier. Processus ARIMA et SARIMA. Mise en pratique sur des exemples avec le logiciel R.

Statistique non paramétrique (*) – 36h – 4 ECTS [Ester Mariucci]

Le module est une introduction à la statistique non-paramétrique. Les sujets traités seront l'estimation de la fonction de répartition, la construction des estimateurs à noyau et des estimateurs par projection de la fonction de densité, estimateurs adaptatifs, modèles de régression, régions de confiance non-paramétriques. Les objectifs majeurs de cette UE sont la compréhension des hypothèses sous-jacentes aux méthodes non paramétriques et l'acquisition d'une part de la capacité à appliquer les méthodes non paramétriques aux problèmes modernes d'analyse de données et d'autre part d'une expérience pratique dans la mise en œuvre des méthodes et de l'utilisation des paquets R existants.

Statistique bayésienne (*) – 24h – 3 ECTS [Mariane Pelletier] ---

Projet de master 30h - 3 ECTS

Apprentissage non supervisé (*) - 30h – 3 ECTS [Charles Tillier]

Introduction à l'intelligence artificielle et au concept d'apprentissage; algorithme K-means et clustering; classification hiérarchique; réduction de la dimensionnalité ; introduction à la détection de nouveautés/anomalies, one-class-classification.

Méthodes de régression (*) - 30h – 3 ECTS [Charles Tillier]

Rappel sur le modèle linéaire; principe de pénalisation; régression LASSO, Ridge et Elasticnet; régression PLS; introduction aux régressions non-linéaires.

Classification supervisée et analyse des données catégorielles – 30h – 3 ECTS [Intervenant extérieur]

Rappels sur l'analyse bivariée de données catégorielles. Modèle de régression logistique et

logistique multinomiale, discussion de l'hypothèse des risques proportionnels.

Applications marketing – 30h – 3 ECTS [Julien Worms et Intervenante extérieure]

Problématiques statistiques en marketing, notions de score, suivi d'un score. Scoring RFM. Scoring basé sur un modèle de classification supervisée ou non-supervisée. Introduction rapide d'outils associés (Classification non-supervisée, ACM, arbres de décision, forêts aléatoires,...). Construction et sélection d'indicateurs. Mise en oeuvre sous R, et dans le cadre d'un projet sous Python.

Assurance vie 45h - 5 ECTS [Mariane Pelletier] ---

Assurance non vie 45h - 5 ECTS [Mariane Pelletier] ---

Analyse des risques– 30h – 3 ECTS [Mariane Pelletier] ---

Alternance – 18 ECTS (non compensable)