

5.00 crédits

30.0 h + 30.0 h

Q2

**Cette unité d'enseignement n'est pas accessible aux étudiants d'échange !**

Enseignants	Massart Estelle ;
Langue d'enseignement	Français
Lieu du cours	Charleroi
Préalables	Ce cours suppose acquises des notions de base en probabilités telles qu'enseignées dans les cours LSINC1211 , LEPL1108 ou LBIR1212 . <i>Le(s) prérequis de cette Unité d'enseignement (UE) sont précisés à la fin de cette fiche, en regard des programmes/formations qui proposent cette UE.</i>
Thèmes abordés	Le cours présente les concepts fondamentaux des statistiques dans un contexte d'ingénierie (analyse exploratoire, inférence, simulation) ainsi que les méthodes de base d'analyse de données multivariées (comme la régression linéaire, l'analyse en composantes principales ou la classification).
Acquis d'apprentissage	A la fin de cette unité d'enseignement, l'étudiant est capable de : • Explorer des données en petit et grand nombre, et en petite et grande dimension ; • Induire des caractéristiques d'une population de données à partir d'un échantillon en utilisant des techniques d'inférence, d'estimation, de calcul d'intervalles de confiance, et de tests statistiques ; • Relier l'approche déductive de la théorie des probabilités et l'approche inductive de la statistique, en identifiant clairement les modèles probabilistes qui seront soumis à l'inférence statistique ; • Reformuler l'énoncé textuel d'un problème d'inférence statistique dans un formalisme mathématique et statistique précis, en identifiant les modèles adéquats et les méthodes d'estimation correspondantes ; • Résoudre un problème appliqué en suivant une approche logique basée sur l'utilisation correcte des modèles et de l'inférence statistique ; • Utiliser des techniques de simulation Monte-Carlo, de validation croisée et de bootstrap à des fins d'estimation et de validation de résultats ; • Analyser des données multivariées avec des techniques fondamentales de régression linéaire, d'analyse en composantes principales et de classification/clustering ; • Utiliser des outils d'analyse statistique pour valider les conclusions de modèles (par exemple la régression linéaire) ; • Faire le lien entre les objectifs mathématiques d'une méthode d'analyse de données et ses objectifs applicatifs.
Modes d'évaluation des acquis des étudiants	Examen écrit individuel pour évaluer la compréhension des concepts et techniques (théorie et exercices, sous forme d'exercices à choix multiples et de questions ouvertes). Les hackathons interviennent pour 25% de la note finale. Les enseignants se réserve le droit d'interroger l'étudiant oralement tant sur les réponses à son examen que sur les hackathons.
Méthodes d'enseignement	(Remarque: Durant l'année académique 2021-2022, ce cours sera enseigné en français) Le cours est constitué de : - 9 cours magistraux sur les sujets énumérés dans le contenu du cours; - 7 séances pratiques, à la fois classiques et numériques (sur Python); - 4 hackathons, représentant 2 x 2 heures chacun, associés à de petits projets Python réalisés en groupe sur des sujets découverts en cours et au travers des séances pratique.
Contenu	- Analyse exploratoire et échantillonnage - Introduction à l'analyse de données multivariées - Estimation paramétrique (méthodes des moments et de maximum de vraisemblance) et propriétés des estimateurs (biais, variance, erreur quadratique moyenne) - Inférence statistique (intervalles de confiance et tests d'hypothèses) : comparaison de moyennes de deux et plusieurs populations normales, proportions, tests sur les variances - Régression linéaire, y compris l'analyse des coefficients et les tests d'hypothèses - Panorama des techniques d'apprentissage: apprentissage supervisé et non-supervisé - Lien entre les objectifs des méthodes d'analyse de données et leur formulation mathématique - Regression et méthode de classification (tels que modèles linéaires, k plus proches voisins, régression logistique)

	- Erreur d'apprentissage, de test et de généralisation, le compromis bias-variance, et éléments de théorie de la décision statistique - Resampling techniques for model selection/evaluation (e.g., ensemble de validation, principe de la validation croisée, bootstrap) - Apprentissage non-supervisé: réduction de dimension (analyse en composantes principales) et méthode de partitionnement (K-means).
Ressources en ligne	La totalité des supports de cours (slides et programmes) sont disponibles sur la plateforme Moodle via ce lien
Autres infos	Pour suivre ce cours, les étudiants doivent avoir une connaissance basique des probabilités, tels qu'enseignés dans le cours LEPL1108 ou LBIR1212. Le planning des cours est sujet à modification en raison des conditions sanitaires. Merci de consulter le site Moodle pour plus de détails.
Faculté ou entité en charge:	SINC

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Bachelier en sciences informatiques	SINC1BA	5		