







5.00 crédits	30.0 h + 30.0 h	Q2
--------------	-----------------	----

Enseignants	El Ghouch Anouar ;
Langue d'enseignement	Français
Lieu du cours	Louvain-la-Neuve
Préalables	Maîtrise de la langue française et des mathématiques, niveau dernier année du secondaire (math 6h/semaine). Connaissance passive de l'anglais.
Thèmes abordés	<p>Cette unité d'enseignement propose une introduction active aux méthodes exploratoires et aux principes fondamentaux de la modélisation probabiliste et statistique essentielle à l'analyse des données observationnelles et expérimentales. Des données réelles seront utilisées pour présenter les indices numériques et outils graphiques d'analyse descriptive et permettre à l'étudiant.e de découvrir et développer ses compétences opérationnelles en traitement de données à l'aide d'un logiciel spécialisé (par exemple, le logiciel libre R : https://www.r-project.org/).</p> <p>Des simulations de données, réalisées également sur logiciel, suivant des lois de probabilités de base permettront de développer une compréhension intuitive de la notion du hasard dans la modélisation. Une fois ces notions de probabilité (et de variable aléatoire) assimilées, des notions fondamentales de l'inférence statistique telles que l'échantillonnage, l'estimation ponctuelle et par intervalle seront présentées empiriquement via des simulations et appliquées à différents contextes réels. L'accent sera mis volontairement non pas sur la démonstration des résultats mathématiques mais sur la signification et l'interprétation des concepts. Le cours intégrera aussi un module sur l'analyse des incertitudes des mesures expérimentales, leurs sources, leur propagation et leur modélisation, ceci illustré sur des cas concrets dans des domaines tels que la physique ou la biologie.</p>
Acquis d'apprentissage	<p>A la fin de cette unité d'enseignement, l'étudiant est capable de :</p> <p>a. Contribution de l'activité au référentiel AA du programme</p> <p>Eu égard au référentiel AA du programme de Bachelier en sciences mathématiques, cette unité d'enseignement permet aux étudiant.e.s de maîtriser :</p> <ul style="list-style-type: none"> · de manière prioritaire les AA suivants : x.x, ; · de manière secondaire les AA suivants : x.x, <p>Eu égard au référentiel AA du programme de Bachelier en sciences de physiques, cette unité d'enseignement permet aux étudiants de maîtriser :</p> <ul style="list-style-type: none"> · de manière prioritaire les AA suivants : A1.2, A1.5, A2.3, A3.4 ; · de manière secondaire les AA suivants : A5.3, A6.5, A4.4. <p>b. Formulation spécifique pour cette activité des AA du programme</p> <p>Plus précisément, au terme de cette unité d'enseignement, l'étudiant.e sera capable de :</p> <ul style="list-style-type: none"> · sélectionner les outils de statistique descriptive (indices numériques et graphiques) adaptés pour résumer efficacement et répondre à des questions concernant un jeu de données ; · analyser une série de données, potentiellement de grande taille, via des outils descriptifs en utilisant un logiciel spécialisé ; · lire et interpréter les résultats d'une analyse descriptive de données en les formalisant dans le contexte de l'étude ; · savoir expliquer et définir les concepts de base de probabilité sur des événements et liés aux variables aléatoires univariées ; · réaliser du calcul de probabilité de base dans des situations variées ; · réaliser, à l'aide d'un logiciel, des simulations pour illustrer le comportement de variables aléatoires et les notions d'échantillonnage ; · expliquer les notations d'échantillonnage et les concepts fondamentaux de l'inférence statistique, et estimer les paramètres de base d'une variable aléatoire en quantifiant l'incertitude via des intervalles de confiance ; · identifier et quantifier l'incertitude expérimentale, et présenter les résultats d'une analyse ou d'un ensemble de mesures tout en indiquant le degré d'incertitude ; · expliquer les objectifs et notions générales de modélisation linéaire et non linéaire ; ajuster un modèle linéaire à une variable via une méthode des moindres carrés.

<p>Modes d'évaluation des acquis des étudiants</p>	<ul style="list-style-type: none"> • Pendant les TPs : devoirs sous forme de mini projets (études de cas) à résoudre en groupe de 2 étudiants à l'aide d'un logiciel spécialisé. • Pendant la session des examens : examen écrit assisté par ordinateur.
<p>Méthodes d'enseignement</p>	<p>Le cours sera constitué</p> <ul style="list-style-type: none"> • d'exposés qui présenteront la matière sur base d'exemples, • de séances d'exercices visant à mettre systématiquement en pratique les différentes notions vues au cours sur des cas bien ciblés et à l'aide d'un logiciel spécialisé, • des projets qui donneront l'occasion à l'étudiant.e d'intégrer les différents outils dans des domaines d'application de la mathématique et de la physique. <p>L'approche pédagogique utilisée privilégiera l'apprentissage actif des étudiants et tentera de respecter les orientations pédagogiques proposées par la Faculté des sciences.</p>
<p>Contenu</p>	<ul style="list-style-type: none"> • Présentation de bases de données provenant des différents domaines d'application en sciences et technologie qui sont utilisées pour illustrer le cours (résultats d'expériences de laboratoire, données météo, résultats d'enquêtes, essais cliniques, données boursières, données de réseau (sociaux et autres), etc.). • Structuration de données statistiques et nature des variables (quantitatives discrètes/continues, qualitatives nominales/ordinales). • Outils numériques pour résumer des données en fonction de leur nature : tableaux de fréquence, indices de position (mode, médiane, moyenne), indices de dispersion (étendue, écart interquartile, écart-type, variance, coefficient de variation), percentiles, coefficient de corrélation. • Outils graphiques pour résumer des données : histogramme, fonction de distribution empirique, diagramme en barres, box-plot, graphique temporel, graphe X-Y (simple ou matriciel), Q-Q-plot. • Expérience aléatoire et notions de base de la théorie des probabilités : la définition d'une probabilité et ses propriétés élémentaires et le calcul de probabilité sur des événements. • Variables aléatoires (univariées) et leurs propriétés (moments, distribution). Introduction des distributions les plus fréquemment utilisées dans l'analyse des données rencontrées dans applications vues au cours : loi uniforme, binomiale, poisson et normale. • Algorithmes simples de génération de nombres aléatoires suivant ces lois de probabilité introduites • Principe de l'échantillonnage aléatoire et de l'estimation. • Notion intuitive de distribution d'échantillonnage et d'intervalle de confiance. • Application à la moyenne (et à la variance) en distribution normale. • Notions d'erreur et d'incertitude de mesure. • Quantification et expression de l'incertitude sur une mesure simple ou répétée (dont l'incertitude élargie par intervalle de confiance). • Calcul d'incertitudes composées dans le cas de sommes, produits et transformations non linéaires de mesures indépendantes ou corrélées. Applications à des mesures de laboratoire dans des domaines divers (physique, biologie,'). • Ajustement d'un modèle linéaire ou polynomial (régression linéaire simple) par méthode des moindres carrés. • Initiation à un langage de programmation statistique (par exemple, R). Approche par des projets de gestion et analyse des données et par simulations.
<p>Faculté ou entité en charge:</p>	<p>SC</p>

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Mineure en statistique et science des données	MINDATA	5		
Bachelier en sciences physiques	PHYS1BA	5		
Bachelier en sciences mathématiques	MATH1BA	5		
Approfondissement en statistique et sciences des données	APPSTAT	5		
Mineure en statistique, sciences actuarielles et science des données	MINSTAT	5		
Certificat d'université : Statistique et sciences des données (15/30 crédits)	STAT2FC	5		
Master [120] en science des données, orientation statistique	DATS2M	5		