



3 crédits	15.0 h + 15.0 h	Q2
-----------	-----------------	----

Enseignants	El Ghouch Anouar ;
Langue d'enseignement	Français
Lieu du cours	Louvain-la-Neuve
Préalables	Maîtrise de la langue française niveau dernière année secondaire.
Thèmes abordés	Ce cours propose une introduction active aux méthodes exploratoires et aux principes fondamentaux de la statistique. Des bases de données réelles de plus ou moins grande taille sont utilisées pour présenter les indices numériques et outils graphiques d'analyse descriptive et permettre à l'étudiant de découvrir et développer ses compétences opérationnelles en traitement de données. Via des simulations, des notions fondamentales de l'inférence statistique telles que l'échantillonnage, l'estimation ponctuelle et par intervalle seront ensuite présentées empiriquement et appliquées à différents contextes réels.
Acquis d'apprentissage	<p style="text-align: center;">Au terme de ce cours, l'étudiant sera capable de</p> <ul style="list-style-type: none"> <li>• Structurer un ensemble de données, en évaluer la qualité et définir les variables en jeu.</li> <li>• Définir, calculer et interpréter les indices numériques et outils graphiques standards de statistique descriptive.</li> <li>• Ajuster un modèle linéaire par la méthode des moindres carrés.</li> <li>• Réaliser, à l'aide d'un logiciel de programmation statistique, une analyse exploratoire d'un grand ensemble de données pour répondre à des questions scientifiques d'un domaine d'application.</li> <li>• Écrire un rapport d'analyse statistique rigoureux, clair, bien argumenté et répondant aux objectifs d'une étude.</li> <li>• Expérimenter, sur des situations pratiques, le concept d'expérience aléatoire via la simulation de différentes lois probabilistes.</li> <li>• Expliquer les principes de l'échantillonnage, de l'estimation, de la distribution d'échantillonnage et des intervalles de confiance et les appliquer ou les démontrer via des simulations.</li> </ul> <p style="text-align: center;">-----</p> <p><i>La contribution de cette UE au développement et à la maîtrise des compétences et acquis du (des) programme(s) est accessible à la fin de cette fiche, dans la partie « Programmes/formations proposant cette unité d'enseignement (UE) ».</i></p>
Modes d'évaluation des acquis des étudiants	<ul style="list-style-type: none"> <li>• Devoirs à rendre chaque semaine (en préparation aux TPs ou post TPs)</li> <li>• Examen sur ordinateur : étude de cas - data management</li> <li>• Examen écrit.</li> </ul>
Méthodes d'enseignement	Le cours sera constitué - 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, - d'études de cas (projets) qui donneront l'occasion à l'étudiant d'appliquer les outils d'analyse statistique de données. L'approche pédagogique utilisée privilégiera l'apprentissage actif des étudiants et tentera de respecter les orientations pédagogiques proposées par la Faculté.
Contenu	<ul style="list-style-type: none"> <li>• Présentation de bases de données typiques dans différents domaines d'application utilisées pour illustrer le cours (données météo, résultats d'enquêtes, essais cliniques, données o-mics, données boursières, données de réseau (sociaux et autres) etc)</li> <li>• Structuration de données statistiques et nature des variables (quantitatives discrètes/continues, qualitatives nominales/ordinales)</li> <li>• 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.</li> <li>• 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</li> <li>• Notion de modèle et ajustement d'un modèle linéaire ou polynomial (régression linéaire) par méthode des moindres carrés</li> <li>• Notions d'expérience aléatoire. Loi uniforme discrète/continue, binomiale et normale. Algorithmes de génération de nombres liés.</li> <li>• Principe de l'échantillonnage aléatoire simple et de l'estimation.</li> <li>• Notion intuitive de distribution d'échantillonnage et d'intervalle de confiance</li> <li>• Application à la moyenne (et à la variance) en distribution normale.</li> </ul>

	<ul style="list-style-type: none"> <li>• Initiation à un langage de programmation statistique (comme R) et application de l'ensemble à la programmation en R (ou autre langage) et application de ces outils aux données concrètes via l'utilisation du logiciel "R" (approche par projet d'analyse des données)</li> </ul>
<p>Autres infos</p>	<p>COURS (7x2H)</p> <ul style="list-style-type: none"> <li>• C1 : Enjeux et Nature des données statistiques, structuration, problèmes et questions types de traitement de données (inférence en expérimentation/sondage, résumé de grands ensembles de données (météo-marketing), qui seront traités dans le cours. Mini Introduction au langage de programmation R</li> <li>• C2 : Indices et tables pour résumer une ou deux variables quantitatives ou catégorielles</li> <li>• C3 : Graphiques sur une et deux variables (et plus). Notions de distributions liées</li> <li>• C4 : Modélisation statistique : méthode des moindres carrés, régression simple, polynomiale.</li> <li>• C5 : Distributions uniforme, binomiale, normale. QQ plot. Simulation de distributions statistiques</li> <li>• C6 et C7 : Notion d'échantillonnage - estimation - distribution d'échantillonnage et intervalle de confiance - application à la distribution normale</li> </ul> <p>TPs (en salle informatique)</p> <ul style="list-style-type: none"> <li>• TP1 : Introduction au langage de programmation R (comme une calculatrice graphique puissante)</li> <li>• TP2, TP3 : Traitement de données avec R : lecture, écriture, mise en forme d'un tableau de données, résumés et graphiques de base, moindres carrés</li> <li>• TP4 Projet 1 : traitement de données (base de données d'observations météo)</li> <li>• TP5 et TP6 : Échantillonnage, simulation, intervalles de confiances...</li> <li>• TP7 Projet 2 : simulation et observation d'un procédé expérimental</li> </ul>
<p>Faculté ou entité en charge:</p>	<p>SC</p>

<b>Programmes / formations proposant cette unité d'enseignement (UE)</b>				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Bachelier en sciences mathématiques	MATH1BA	3		
Mineure d'accès en statistique et science des données	LSTAT100I	3		
Approfondissement en statistique et science des données	LSTAT100P	3		