

5.00 crédits

22.5 h + 10.0 h

Q2

Enseignants	François Thomas ;
Langue d'enseignement	Anglais > Facilités pour suivre le cours en français
Lieu du cours	Louvain-la-Neuve
Thèmes abordés	L'accès à ce cours est réservé à des étudiants ayant déjà suivi avec succès un cours d'introduction aux statistiques, par exemple LFIAL2260. Le cours abordera les thèmes suivants : • Concept et structure d'un modèle statistique (linéaire ou non linéaire) et exemples de questions de recherche linguistiques typiques • Modèles de régression : introduction à différents modèles de régression (ex. régression linéaire, logistique, etc.), algorithmes de sélection des variables, notion de multicollinéarité, entraînement des modèles, interprétation des paramètres, qualité des prédictions • Analyse de la variance : présentation de différentes techniques d'analyse de la variance (données paramétriques et non paramétriques; indépendante et à mesures répétées; à un ou plusieurs critères de classification, etc.), logique sous-jacente au test F, comparaison multiple de moyennes (tests post-hoc) • Modèles linéaires mixtes : notion de modèle linéaire généralisé, effets aléatoires, modèles hiérarchiques • Méthodes exploratoires : méthodes d'exploration des données typiques de la stylométrie, par exemple analyse en composantes principales, analyse factorielle, etc. • Méthodes de classification : introduction aux modèles de classification, par exemple les arbres de décision • Validation des modèles : mesures de la qualité de l'ajustement, analyse des résidus, test d'homogénéité de variance et de sphéricité, détection d'outliers ou de points influents, transformation de variable
Acquis d'apprentissage	A la fin de cette unité d'enseignement, l'étudiant est capable de : 1 Traduire une problématique linguistique de recherche en une série de questions statistiques, choisir les méthodes adaptées, les appliquer et présenter l'ensemble des résultats dans un rapport. 2 Comprendre et expliquer les concepts statistiques sous-jacents aux différentes méthodes utilisées dans le cours. 3 Appliquer les différentes méthodes statistiques couvertes dans le cours à des données textuelles à l'aide du logiciel R Cette unité d'enseignement contribue au développement et à la maîtrise des compétences et acquis des programmes de l'Ecole de langues et lettres suivants#(cf. Tableau des AA ELAL) : 1.4 ; 2.3 ; 2.6 ; 3.1 ; 3.2 ; 3.3 ; 3.5 ; 4.5 ; 5.1 ; 5.2 ; 5.3
Modes d'évaluation des acquis des étudiants	L'évaluation se déroulera selon trois axes : • Évaluation continue (tests, participation au cours, travaux) (30 % de la note) • Examen écrit (30 % de la note) • Travail personnel écrit (40 % de la note): il s'agira soit d'un travail personnel sur un problème linguistique abordé via les outils statistiques, soit de la participation à un projet collectif visant à la production d'un article scientifique. Pour la session de septembre, l'évaluation est adaptée comme suit : • Examen écrit (50% de la note) • Travail personnel (50% de la note) NB : Les intelligences artificielles (IA) génératives doivent être utilisées de manière responsable et conformément aux pratiques de l'intégrité académique (cf. balises des bonnes pratiques publiées par l'UCLouvain). L'intégrité scientifique impliquant que l'on cite ses sources, l'usage d'une IA doit toujours être signalée. L'utilisation des intelligences artificielles pour les tâches où celles-ci sont explicitement interdites sera considérée comme un cas de tricherie.

Méthodes d'enseignement	cours oraux + lectures + exercices (au cours et dans le cadre de travaux pratiques)
Contenu	Après une introduction sur le rôle central des méthodes statistiques pour la linguistique, le cours abordera divers thèmes : • Notions de modélisation statistique • ANOVA I - Analyse de la variance à un critère de classification : Modèle classique, comparaisons post-hoc et ANOVA de Kruskal-Wallis • ANOVA II - Analyse de la variance à deux critères de classification • ANOVA pour mesures répétées : Modèle classique et ANOVA de Friedman • Modèles de régression linéaire simple et multiple et analyse des résidus • Modèle de régression logistique simple et multiple • GLM - Modèle linéaire général et modèles mixtes - Analyses statistiques multivariées exploratoires : Analyse en composantes principales et analyse factorielle • Méthodes de classification : arbres de décision
Bibliographie	Field, A. (2013). <i>Discovering statistics using IBM SPSS statistics</i> . Sage. Howell, D. (2008). <i>Méthodes statistiques en sciences humaines</i> , Paris, De Boeck Université. Muller, Charles (1992). <i>Initiation aux méthodes de la statistique linguistique</i> , Champion. Rasinger, S.M. (2008). <i>Quantitative Research in Linguistics</i> . New York, Continuum International Publishing Group
Autres infos	Support de cours (sur Moodle) : • slides ; • articles ou chapitres de livres ; • exercices complémentaires.
Faculté ou entité en charge:	ELAL

Programmes / formations proposant cette unité d'enseignement (UE)				
Intitulé du programme	Sigle	Crédits	Prérequis	Acquis d'apprentissage
Master [120] en linguistique	LING2M	5		
Master [120] en langues et lettres modernes, orientation germaniques	GERM2M	5		
Master [120] en langues et lettres modernes, orientation générale	ROGE2M	5		