

Biostatistique III

Traitement des données en statistique

- Cours (CM) : 17h (25,5 heures ETD)
- Travaux dirigés (TD)/Travaux pratiques (TP) : 11,5 heures
- Total : 37 heures ETD

Description du contenu de l'enseignement

Cette UE figure dans les UE « obligatoires » du troisième semestre du parcours *Biomatériaux pour la Santé* et dans les UE « obligatoires à choix » du troisième semestre du parcours *Recherche en Biomédecine*. Des connaissances préalables en statistique de base et notamment en inférence bayésienne sont très fortement souhaitables.

Cette UE vise en effet à apporter des connaissances théoriques et pratiques à l'étudiant en situation d'analyse de données et de recherche. Ces données peuvent issues de plans expérimentaux ou de l'épidémiologie. L'enseignement sera donc fortement axé sur la pratique, après l'acquisition des connaissances théoriques minimales, donc largement constitué de TD/TP.

Compétences à acquérir

A l'issue de ce cours l'étudiant est capable de :

- comprendre le principe d'un test d'hypothèse, le test d'hypothèse nulle selon la procédure de Neyman-Pearson, selon la procédure de Fisher et l'approche bayésienne des tests ;
- savoir caractériser les données issues d'expérimentation, de recherche clinique, d'épidémiologie observationnelle, ... et en tirer les conséquences en termes de modélisation ;
- maîtriser les analyses bayésiennes quant à leur technique d'estimation, vérification de la convergence des algorithmes, vérification de l'adéquation des modèles ;
- pouvoir estimer les paramètres des modèles habituellement rencontrés dans les données de santé (modèle linéaire généralisé, régression Beta, survie) ;
- savoir gérer différents types de données (longitudinales, spatialisées, ...), ainsi que les données manquantes.

Bibliographie, lectures recommandées

Aucune

Modalités d'organisation et de suivi

Méthodes pédagogiques :

Les supports de cours seront mis en ligne avant les séances. Le cours magistral est constitué des diapositives de cours ainsi que, le cas échéant de matériel complémentaire (par exemple articles scientifiques). Pour les séances de TD/TP des fichiers de données et des fichiers de code seront mis à disposition des étudiants.

Travail personnel recommandé :

Il est très fortement recommandé de revoir régulièrement les diapositives des cours ainsi que le matériel complémentaire fourni. De même, il est impératif de pouvoir reproduire des analyses de données en partant des codes logiciels mis à disposition et qui auront été explicités en TD/TP.

Plan de l'enseignement :

- 1 Inférence fréquentiste et tests d'hypothèses, inférence bayésienne
 1. Tests d'hypothèses : procédure Neyman-Pearson, procédure de Fisher (CM)
 2. Des lois à l'inférence bayésienne (CM)
 3. Typologie des variables et des situations (longitudinal, univarié, ...) (0,5 heure, TD)
 - 2 Techniques bayésiennes (CM)
 - 3 Bases de modélisation
 1. Introduction à la modélisation (CM)
 2. Régression linéaire simple et multiple, ANOVA (2 heures, CM)
 3. GLM (régressions logistique, de Poisson) (CM)
 4. GLM (régression gamma), régression Beta (CM)
 5. Adéquation des modèles (CM)
 6. Survie (TD)
 7. Modèles mixtes (2 heures, CM)
 - 4 Schémas d'étude, situations particulières
 1. Données manquantes (2 heures, CM)
 2. Données longitudinales (2 heures, TD)
 3. Plans expérimentaux, essais thérapeutiques, pilotage, nombre nécessaire de sujets, analyses séquentielles (3 heures, CM)
 4. Modèles temporels, spatiaux (2 heures, TD)
- TD Modélisation (6 heures, TD/TP)

Responsables

Pr. Erik-André Sauleau

Pr. Nicolas Meyer

Laboratoire de Biostatistique et Informatique médicale

Laboratoire ICube, UMR CNRS 7357

Faculté de médecine

4, rue Kirschleger

F-67085 Strasbourg cédex

ea.sauleau@unistra.fr