Sujets de mémoire en statistique proposés par Bernadette Govaerts

ISBA : Institut de statistique, biostatistique et sciences actuarielles
SMSC : Plate-forme technologique de support en méthodologie et calcul statistique

(Bernadette.govaerts@uclouvain.be)

Les sujets de mémoire/stage que je propose sont liés à mes sujets d'intérêt et de nombreuses collaborations de consultation statistique avec des chercheurs de l'UCL et des sociétés extérieures à l'UCL. Le mémoire ou le stage peut se faire en lien ou au sein du service de consultation et de calcul statistique SMCS (http://www.uclouvain.be/smcs.html). Les domaines d'application de ces consultations sont surtout les sciences expérimentales, les sciences du vivant et les méthodes analytiques.

Mes sujets statistiques d'intérêt sont:

- Les plans d'expériences et l'analyse de leurs résultats
- Les méthodes statistiques liées aux problèmes de traitement et de quantification d'incertitude de mesure en laboratoire d'analyse
- Le contrôle statistique de la qualité
- Les méthodes multivariées en chimiométrie
- Les méthodes –omics: métabonomique, génomique…
- La consultation et le calcul statistique

Les sujets que je peux proposer pour l'année académique 2015-2016 sont décrits ci-dessous. Ils peuvent être modulés selon le fait que vous recherchez un stage ou un mémoire.

Sujets dans le domaine de la métabolique, une domaine de pointe dans la recherche de biomarqueurs en médecine (en collaboration avec le SMCS et l'école de pharmacie de l'ULG et, potentiellement, le CraW à Gembloux).

- Test et optimisation d'une libraire R pour le pre-processing de spectres ^{1}H-NMR en métabolomique et amélioration des méthodes de correction de ligne de base et de réalignement de pics (par des algorithmes non paramétriques)
- Adaptation des méthodes ASCA, ANOVA-PCA et ComDim aux modèles linéaires mixtes. Actuellement ces méthodes permettent de réaliser d'appliquer décomposition de type ANOVA à des réponses multivariées de grande taille. Les facteurs sont considérés comme fixes et les plans balancés. Le but de mémoire est d'adapter les méthodes aux modèles comportant des facteurs aléatoires et aux designs non balancés. Applications en H-NMR et NIR
- Recherche de biomarqueurs à partir de données métaboliques spectrales récoltées au court du temps (en ^{1}H-NMR et MS)
- Préparation d'un petit manuel de traitement de données métabolomiques à l'aide des logiciels R et SIMCA à destination de chimistes, médecins et pharmaciens. Ceci implique une définition des étapes de la méthodologie multivariée (basée sur de la PCA, PLS-DA, O-PLS-DA, ICA, modèles GLM, recherche d'outliers multivariés…) et mise en œuvre sur des bases de données réelles et expérimentales

Plans expérimentaux

- Développement d'un package R pour évaluer les qualités statistiques d'un plan expérimental. Cet outil sera utile pour les consultations réalisées par le SMCS dans le domaine.

Statistique de la mesure - métrologie

- Le manuel GUM propose une méthodologie pour quantifier l'incertitude de mesures faites en laboratoire. La méthodologie, basée sur des simulations, proposée dans ce manuel pour calculer des intervalles de confiance sur des mesures composées reste très rudimentaire et des solutions plus élégantes peuvent être proposées. Le mémoire a pour but de revoir la méthodologie actuelle et comparer des nouvelles approches analytiques ou par simulations émanant de la littérature récente.
- La société Arlenda et le département R&D de la société GSK proposent régulièrement des mémoires et stages pour lesquels je peux être promoteur UCL. Me contacter si vous êtes intéressés.
- Mise en place d'un cadre bayesien pour le calcul statistique d'intervalles de tolérance et la quantification des incertitudes de mesures produites en laboratoire d'analyse : Application au dosage chimique et biologique. Mémoire en collaboration avec GSK, l'ULG et avec comme copromoteur P. Lambert. Sujet large pouvant mener potentiellement à une thèse.

Big data et data science

- Le monde statistique s'intéresse actuellement de plus en plus aux disciplines émergentes du big data et du data science. L'objectif du mémoire sera d'abord de tenter de cerner, par de la revue bibliographique, ce que ces deux domaines apportent de nouveau par rapport à la statistique classique et les compétences nouvelles que les experts doivent développer dans ces domaines. Le mémoire approfondira ensuite une ou deux techniques peu développées dans le master en statistique comme l'analyse de donnée de réseaux et testera des logiciels disponibles dans le domaine sur des données réelles.

Pédagogie de la statistique

- 3 cours de statistique ont été récemment créés en faculté de psychologie de l'UCL qui sont très fortement basées sur de l'e-learning. L'objectif est de construire sur cette base une formation à la statistique appliquée complètement basée sur de l'e-learning et que la LSBA pourrait offrir dans le cadre de son master en statistique. Le mémoire impliquera le choix d'une nouvelle structure du cours existant, le développement de podcast ainsi que d'exercices de simulation sous R visant à mieux comprendre les concepts clefs du cours (échantillonnage, tests, régression…). Les plateforme de travail seront selt et moodle.

Autres sujets possibles mais d'un intérêt moins prioritaire

- Comparaison des potentialités du logiciel ade4 et factor-miner pour faire de l'analyse multivariée en R. Travail intéressant le SMCS.
- Une étude et comparaison des logiciels et librairies existants pour la génération de plans d'expérience pour le traitement de problèmes en choix discret en marketing (par

exemple) et une revue de la littérature liée. (mémoire en lien avec la thèse de C. Taverne)

- Etude biblio et des logiciels pour l'analyse de données multi-tableaux dans le domaine des sciences dans le but de plus les utiliser dans le cadre des consultations réalisées par le SMCS. Les multi-tableaux sont des tableaux de données comprenant plus de deux dimensions (par exemple : individus, traitement, temps, répétitions…). Les domaines d'application sont par exemple l'analyse sensorielle et la chimiométrie en général.
- Une étude d'outils pour générer des plans factoriels fractionnaires pour facteurs à plus de deux niveaux et en particulier du nouveau package très prometteur PLANOR (http://w3.jouy.inra.fr/unites/miaj/public/logiciels/planor/) et la théorie sous-jacente. Cet outil sera utile pour les consultations et formations réalisées par le SMCS dans le domaine.

Chacun de ces sujets implique en général de la revue bibliographique et des éventuels nouveaux développements méthodologiques, de la programmation en R et/ou de l'utilisation de logiciels de statistique, de l'analyse de données et des contacts avec un client avec qui nous collaborons sur le sujet.

Pour les étudiants de l'EPL, un copromoteur, membre de votre commission de diplôme devra être choisi.

Vous êtes les bienvenus pour venir en parler !