

Proposition de stage de Master 2 en Biostatistique – 13/11/2019

Titre : Apport de la régression linéaire généralisée sur composantes supervisées dans l'étude de l'association entre déficit visuel et morphologie cérébrale.

Lieu : Inserm, U1061, Hôpital La Colombière, Montpellier, France

Durée : 4 à 6 mois

Encadrant : Isabelle Carrière, IR1, statistique et épidémiologie, INSERM UMR_S 1061.

Co-encadrants :

- Xavier Bry, Enseignant Chercheur, mathématique et statistique, IMAG, Université Montpellier
- Vincent Daïen, PU-PH, ophtalmologie, CHU Montpellier Gui de Chauliac, Inserm UMR_S 1061
- Catherine Trottier, Enseignant Chercheur, mathématique et statistique, IMAG, Université Montpellier.

Objectif :

L'objectif du projet est d'explorer l'utilisation d'une nouvelle méthode statistique de régularisation de modèles, de développer un complément à cette méthode et de l'appliquer à l'étude des associations entre une acuité visuelle basse et les données d'imagerie cérébrale de 310 participants de la cohorte montpelliéraine des Trois-cités (<http://www.three-city-study.com/>).

Contexte :

Les données déjà acquises de morphologie cérébrale consistent en 1035 variables évaluant notamment pour les deux hémisphères et pour 111 régions cérébrales, leur volume, épaisseur corticale, surface et courbure.

La méthode d'analyse des volumes cérébraux actuellement employée (ou des aires ou épaisseurs ou courbures) est de tester une à une ces régions en ajustant le seuil de significativité pour contrôler la multiplicité des tests. Une autre méthode traditionnelle qui consiste à rechercher la combinaison linéaire des volumes maximisant la vraisemblance d'un modèle linéaire généralisé (GLM) a pour limite majeure une instabilité des estimations due à la forte collinéarité des régresseurs. Ces deux méthodes ne permettent pas d'identifier les structures sous-jacentes des volumes. Nous proposons d'utiliser une extension du GLM appelée "régression linéaire généralisée sur composantes supervisées" (SCGLR) permettant de construire des composantes orthogonales sur les variables explicatives (i.e. volumes cérébraux) afin de régulariser la régression et d'extraire des informations structurelles sur les données cérébrales.

Mission du stage :

Il s'agira d'abord d'explorer les très nombreuses variables potentiellement explicatives indépendamment de la réponse en utilisant la classification automatique de variables afin d'identifier leurs partitionnements possibles en classes de variables redondantes. Dans un deuxième temps, en tenant compte des résultats obtenus, on mettra les variables explicatives en relation avec la réponse à l'aide de la méthode THEME-SCGLR déjà développée sous R. Une recherche du réglage optimal de la méthode devra être faite, tant pour le partitionnement des variables explicatives que pour les paramètres de réglage fin. Enfin, le stagiaire développera la construction des intervalles de confiance pour les coefficients estimés par la méthode du bootstrap.

Logiciel : R

Contact: Envoyer CV et lettre de motivation à Isabelle Carrière, Responsable de la plateforme Modélisation statistique, à l'adresse isabelle.carriere@inserm.fr