

Stage de 6 mois (M2): Intégration de données hétérogènes pour l'étude de la malnutrition chronique chez l'enfant en Afrique

Lieu de stage: Hub de bioinformatique / biostatistique, Institut Pasteur, Paris, France

Contexte de l'étude

Mondialement, un enfant sur quatre de moins de cinq ans souffre d'un retard de croissance. La cause majeure du retard de croissance est la sous-nutrition chronique, qui constitue un terrain favorable à la survenue d'infections et représente donc un frein considérable au développement normal de ces enfants et peut entraîner des séquelles jusqu'à l'âge adulte. Pour le moment, aucun traitement efficace n'existe. Les causes du caractère chronique de cette sous-nutrition sont nombreuses et incluent le microbiote intestinal ainsi que différents métabolites. Ce projet de stage, qui s'intègre dans un projet clinique de grande envergure, AFRIBIOTA [1, 2, 3], vise à intégrer des données cliniques et microbiologiques grâce à des méthodes statistiques afin de mieux comprendre l'interaction entre ces différentes entités et, à terme, aider à développer des traitements pour ce syndrome.

Les données qui seront analysées incluent des données de microbiote intestinal, de métabolites (sels biliaires et dérivés) ainsi que des paramètres socio-économiques et cliniques. Dans le contexte de ce stage, les données d'environ 250 enfants seront analysées.

Le stage se déroulera à l'Institut Pasteur, dans les locaux du C3BI (<https://c3bi.pasteur.fr/>).

Méthodes

RGCCA [3] est une méthode d'analyse de données structurées en blocs qui a pour but, sur la base d'un "design" représentant les liens entre les différents blocs, d'extraire des composantes qui permettent à la fois d'expliquer la variabilité de chacun des blocs et les liens décrits par le design. Le stage aura pour but d'adapter RGCCA à la nature particulière des données de questionnaire et de microbiote et de l'appliquer aux données, mais également de comparer ses performances à d'autres méthodes utilisées par ailleurs pour des analyses similaires.

Profil recherché

Statisticien/ne, data analyst ou data scientist, le/la candidat/e doit avoir de solides compétences en analyse de données multivariées et de bonnes compétences en programmation R et maîtriser l'utilisation de R Markdown.

Le/la candidat/e doit également avoir un intérêt pour l'application des méthodes statistiques à l'analyse de données biologiques, être autonome et enthousiaste de travailler, sous la supervision d'un statisticien, dans une équipe interdisciplinaire.

Contact

Envoyez CV et lettre de motivation à vincent.guillemot@pasteur.fr

Références

[1] Vonaesch, P., Morien, E., Andrianonimiadana, L., et al. (2018), Stunted childhood growth is associated with decompartmentalization of the gastrointestinal tract and overgrowth of oropharyngeal taxa, *Proc Natl Acad Sci USA*. doi: 10.1073/pnas.1806573115

[2] Vonaesch, P., Rendremanana, R., Gody, J.-C., et al. (2018). Identifying the etiology and pathophysiology underlying stunting and environmental enteropathy: study protocol of the AFRIBIOTA project, *BMC Pediatr*. doi: 10.1186/s12887-018-1189-5.

[3] Tenenhaus, A., & Tenenhaus, M. (2014). Regularized generalized canonical correlation analysis for multiblock or multigroup data analysis. *European Journal of Operational Research*. doi: 10.1016/j.ejor.2014.01.008

Page web

<https://www.pasteur.fr/fr/institut-pasteur/institut-pasteur-monde/programmes-recherche-internationaux/malnutrition-infantile>