

## Stage Master II en biostatistique

# **Implémentation d'un modèle IRT dynamique sous R et application à l'étude de la dépendance fonctionnelle chez les personnes âgées**

Direction : Cécile Proust-Lima (CR INSERM)  
INSERM, Centre INSERM U1219, Equipe Biostatistiques

Les données de réponse aux items sont de plus en plus courantes dans les études de santé, notamment par l'utilisation croissante de questionnaires pour évaluer l'état de santé des individus et leurs comportements.

L'analyse des données de réponse aux items nécessite l'utilisation de modèles statistiques appropriés à la nature multivariée (plusieurs items), corrélée (plusieurs items mesurant la même quantité sous-jacente) et possiblement catégorielle (items binaires ou ordinaux) des items issus de questionnaires. La méthodologie appelée « IRT » pour Item Response Theory (Baker et Kim, 2004) apporte un cadre d'analyse bien adapté à ces caractéristiques. Néanmoins, dans les études épidémiologiques et notamment celles basées sur des données de cohorte, les questionnaires sont mesurés de façon répétée au cours des suivis et leur analyse doit donc aussi inclure l'aspect dynamique.

Ce stage vise à implémenter sous R la méthodologie IRT étendue au cas dynamique pour permettre d'analyser les trajectoires de phénomènes de santé mesurés par le biais d'items répétés au cours du temps. L'approche combinera pour cela la méthodologie IRT pour lier les items à la quantité d'intérêt qu'ils mesurent, et la méthodologie des modèles mixtes pour décrire la quantité d'intérêt au cours du temps en tenant compte de la corrélation entre les mesures répétées d'un même individu. Cette méthodologie a été implémentée sous R pour des items continus (Proust-Lima, et al., JSS 2017). Le stage vise à étendre cette implémentation pour des items binaires et ordinaux. On pourra s'appuyer pour cela sur la méthodologie statistique proposée par Proust-Lima et al. (BJMSP 2013).

Le modèle IRT dynamique, une fois implémenté sous R, sera appliqué à l'étude de la dépendance fonctionnelle chez les personnes âgées, mesurée à partir des items d'activité de la vie quotidienne (e.g., autonomie dans la toilette, l'alimentation, l'habillement) et activités instrumentales de la vie quotidienne (e.g., autonomie dans les courses, la gestion des finances, des médicaments). Suivant une analyse précédemment réalisée dans le centre de recherche (Edjolo et al., 2015), les trajectoires de dépendance fonctionnelle seront analysées en amont du décès pour capter l'histoire naturelle de la dépendance.

L'objectif final de cette application est d'utiliser la méthodologie IRT pour calculer a posteriori des indicateurs cruciaux pour la Santé Publique et la gestion des besoins de santé comme la proportion de personnes dépendantes à une activité spécifique au moment du décès ou X années avant le décès ou encore le nombre d'années passées avec une dépendance à une activité spécifique de la vie quotidienne (e.g., la gestion des médicaments ou la toilette).

La partie application du stage reposera sur les données de la cohorte PAQUID, cohorte populationnelle sur le vieillissement cérébral avec plus de 30 ans de suivi.

Spécifiquement, l'étudiant aura pour missions de :

- s'approprier la méthodologie des modèles IRT et modèles mixtes à processus latents à partir de la bibliographie et de l'expertise de l'équipe de recherche ;

**Inserm U1219**

Bordeaux Population Health Research Center

- se familiariser avec l'implémentation du package R lcmm qui implémente cette méthodologie pour des données continues ;
- implémenter le modèle IRT dynamique (i.e., modèle mixte à processus latent pour des marqueurs ordinaux répétés) sous R ;
- appliquer la méthodologie à l'analyse des données de dépendance fonctionnelle chez les personnes âgées à partir des données de la cohorte PAQUID ;
- implémenter le calcul a posteriori d'indicateurs de santé et l'appliquer sur les données de la cohorte PAQUID.

Suivant le déroulement du stage et les souhaits du stagiaire, ce stage pourra déboucher sur une thèse en Biostatistique.

Compétences requises :

Bonnes connaissances en biostatistique  
Maîtrise du logiciel statistique R  
Aptitudes à la programmation

Structure d'accueil :

Centre de Recherche Bordeaux Population Health  
Equipe de Biostatistique  
146, rue Léo-Saignat  
33076 Bordeaux Cedex

Durée du stage : 4-6 mois

Période du stage : à partir de début 2019

Gratification de stage : selon les grilles prévues par l'Inserm

Contacts :

Merci d'envoyer une lettre de motivation et un CV détaillé à Cécile Proust-Lima, [cecile.proust-lima@inserm.fr](mailto:cecile.proust-lima@inserm.fr).

Références :

- Baker FB, Kim SH. Item Response Theory. Parameter Estimation Techniques. 2nd ed. New York: Marcel Dekker; 2004. 495 p. (Statistics: Textbooks & Monographs).
- Edjolo A, Proust-Lima C, Delva F, Dartigues J-F, Pérès K. Natural History of Dependency in the Elderly: A 24-Year Population-Based Study Using a Longitudinal Item Response Theory Model. Am J Epidemiol. 2016 Feb 15;183(4):277–85.
- Proust-Lima C, Amieva H, Jacqmin-Gadda H. Analysis of multivariate mixed longitudinal data: a flexible latent process approach. Br J Math Stat Psychol. 2013 Nov;66(3):470–87.
- Proust-Lima C, Philipps V, Liquet B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. Journal of Statistical Software, Articles. 2017;78(2):1–56.