

Module 3 – MODÈLES DE DONNÉES LONGITUDINALES 2

I. OBJECTIFS DU COURS

Ce cours avancé en statistique appliqué aux données hiérarchiques longitudinales a pour but de permettre aux participants d'établir de meilleurs liens entre les données et les théories des sciences sociales ou de la santé (sociologie, psychologie, criminologie, démographie, médecine, etc.) en mettant en pratique les méthodes d'analyse de données longitudinales par trajectoires latentes de développement. Dans une perspective incluant les variables de contrôle, médiatrices et modératrices ainsi que les concepts relatifs à la causalité, les participants devront se familiariser avec les principes fondamentaux nécessaires à la compréhension des techniques d'analyse par regroupement (cluster analysis). Ce cours met l'accent sur les modèles statistiques dont la base est la régression linéaire généralisée, tout en incluant des variables latentes continues et catégorielles.

II. PRÉREQUIS

Ce cours nécessite une connaissance préalable de la statistique descriptive et inférentielle ainsi que des analyses multivariées (régression linéaire simple et multiple). Les connaissances introduites dans le module 1 sont également nécessaires. Plus précisément, la connaissance des modèles spécifiant des liens différentiels en fonction de sous-groupes dans la population (variable modératrice) ou spécifiant des mécanismes liant une variable dépendante à une variable indépendante (variable médiatrice). Il est également important de maîtriser un logiciel statistique comme SPSS, SAS, R, Mplus ou Stata. Les démonstrations de chacune des analyses sont effectuées à partir du logiciel SAS. Une macro SAS est également utilisée dans ce cours ([PROC TRAJ](#)).

Puisque le cours porte sur les concepts et méthodes d'analyse plutôt que sur l'usage d'un logiciel spécifique, il est possible de suivre le cours sans utiliser SAS. Les participants désireux d'utiliser un autre logiciel que SAS seront dirigés vers les ressources utiles pour reproduire les analyses présentées. Ils ne recevront toutefois pas de support en ce qui a trait à l'usage d'un autre logiciel que SAS.

III. DÉMARCHE PÉDAGOGIQUE

L'approche pédagogique de ce cours met l'emphasis sur l'écoute de capsules vidéo et la résolution de problèmes en ligne. Les chapitres suivent la séquence suivante: une lecture, une présentation théorique (vidéo), un mini-quiz, un exemple d'application (vidéo), des exercices pratiques et, enfin, un quiz final. Des rencontres synchrones auront lieu à quelques reprises. Le logiciel Adobe Connect sera utilisé lors de ces rencontres.

IV. CONTENU DES SÉANCES

Chapitre 1 : Modèles multiniveaux non paramétriques appliqués aux données longitudinales

- Introduction et historique du modèle avec variable latente catégorielle (classes latentes et modèle de mélange fini)
- La sélection des modèles (Bayesian Information Criterion BIC)
- Analyse des estimations et interprétations des probabilités postérieures
- Valeurs de départ, facteurs de risque et trajectoires jointes

Chapitre 2: Analyse des données manquantes

- Théorie sur les données manquantes (MCAR, MAR, MNAR)
- Les méthodes conventionnelles (listwise, pairwise, ajustement par régression)
- Les méthodes récentes (ML, pondération et imputation multiple) et leurs applications

Chapitre3 : Perspectives récentes dans les modèles multiniveaux non paramétriques

- L'estimation des intervalles de confiance autour des trajectoires
- L'ajout simultané de variables latentes continues et catégorielles (growth mixture model)
- L'estimation simultanée de groupes de trajectoires en incluant des données manquantes
- L'estimation de trajectoires jointes avec des variables prédictives
- L'estimation de trajectoires multiples
- Pondération échantillonnale et temps d'exposition