

Robustesse aux données manquantes : comparaison entre modèle mixte et analyse en composantes principales fonctionnelles

jeudi 16 novembre 2023 16:15 (25 minutes)

En recherche médicale, la collecte de données longitudinales est très fréquente. L'analyse de ces données permet de décrire l'évolution temporelle d'un processus biologique tout en prenant en compte l'importante variabilité individuelle. Par exemple, les cohortes qui étudient l'évolution cognitive du sujet âgé intègrent les résultats des patients à différents tests psychométriques collectés sur des intervalles de temps réguliers. L'étude rétrospective de ces trajectoires permet de mieux décrire l'évolution cognitive du sujet âgé et constitue un outil précieux dans la compréhension de l'histoire naturelle de la démence.

Le modèle statistique le plus couramment utilisé dans l'étude de données longitudinales dans ce contexte est le modèle mixte. Il permet de modéliser avec une grande flexibilité des trajectoires variées, hétérogènes d'un sujet à l'autre avec souvent peu de mesures par sujet et la présence de données manquantes. Ce modèle est connu pour être, dans certains cas, robuste à la présence de données manquantes, fréquentes dans les données longitudinales. Néanmoins, il s'agit d'un modèle paramétrique et des hypothèses souvent fortes sont nécessaires pour estimer le modèle.

Si l'on considère ces données longitudinales comme des données fonctionnelles, c'est-à-dire comme des réalisations aléatoires d'une fonction inconnue sous-jacente, il devient envisageable d'appliquer d'autres outils statistiques. En particulier, l'analyse en composantes principales fonctionnelles (ACPF) est un outil qui permet de décrire des trajectoires temporelles, sans hypothèses paramétriques. Néanmoins, si cette méthode a été éprouvée sur des données denses et régulières, son utilisation sur des données longitudinales avec, en particulier, la présence de données manquantes reste à explorer.

Nous avons réalisé une étude empirique qui montre que l'ACPF se comporte bien malgré la présence de données manquantes et peut se comparer avec l'approche par modèle mixte. Une application de l'ACPF sur des données issues de la cohorte 3C qui étudie l'évolution du sujet âgé permet d'illustrer son intérêt.

Auteurs principaux: SÉGALAS, Corentin (Université de Bordeaux / Bordeaux Population Health); PROUST-LIMA, Cécile (Université de Bordeaux / Bordeaux Population Health); GENUER, Robin (Université de Bordeaux / Bordeaux Population Health)

Orateur: SÉGALAS, Corentin (Université de Bordeaux / Bordeaux Population Health)

Classification de Session: Données longitudinales