Abstract

REAL-world data often contain grouping structures that, if improperly modeled, can lead to inaccurate assumptions about data dependencies and produce misleading results. Properly accounting for these structures improves model accuracy and provides insights that can inform policy decisions at the group level. This thesis aims to develop and apply novel models and methods for effectively handling grouped data with a particular focus on the fields of learning and healthcare analytics, where grouping typically takes the form of hierarchical structures and recurrent events. Specifically, this thesis builds on mixed-effects and time-to-event models to enhance both the institutional effectiveness and the assessment of treatment efficacy, which are critical challenges in the learning and healthcare context. Key novelties include advancements in model-based clustering methods, propensity-score techniques, management of time-dependent group-level variability in time-to-event models, and the assessment of treatment effects in recurrent events setting. These contributions aim to enhance existing methodologies and introduce new strategies for analyzing grouped data structures in both domains.

Keywords: linear mixed-effects models; survival analysis; nonparametric estimation; hierarchical data; recurrent events