Handling Missing Rows in Multi-Omics Data Integration: Multiple Imputation in Multiple Factor Analysis Framework

Valentin Voillet[*1], Philippe Besse[2], Laurence Liaubet[1], Magali Sancristobal[*1], and Ignacio Gonzalez[*3]

[1]Génétique, Physiologie et Systèmes d'Élevage – Institut national de la recherche agronomique (INRA) : UMR1388 – France
[2]Institut de Mathématiques – Université de Toulouse Paul Sabatier – France
[3]Mathématiques et Informatiques Appliquées – Institut national de la recherche agronomique (INRA) : UMR875 – France

Résumé

In omics data integration studies, it is common, for a variety of reasons, that some individuals are not present in all data tables. Missing row values are challenging to deal with because most statistical methods cannot be directly applied to incomplete datasets. To overcome this issue, we propose a multiple imputation (MI) approach in a multiple factor analysis (MFA) framework. MI involves filling the missing rows with plausible values, resulting in m completed datasets. MFA is then applied to each completed dataset leading to m different component configurations. Finally, the m configurations are combined to yield one consensus solution. We showed that MI-MFA configurations were closer to the true configuration (obtained from the original data) even when a significant number of individuals were missing, thus providing improved results.

---

[*]Intervenant