Modelling Longitudinal Health-Related Constructs: A Latent Variable Approach

Authors:

Niccolò Cao[1*†], Silvia Cagnone[1], Livio Finos[2]
Marilina Amabile[3]

[1] Department of Statistical Sciences "Paolo Fortunati", University of Bologna
[2] Department of Statistical Sciences, University of Padova
[3] IRCCS Istituto Ortopedico Rizzoli

[*] Corresponding author [†] Presenter

Contact: niccolo.cao@unibo.it

Keywords:

Underlying Variable Approach, Ordinal data, Full information maximum likelihood, Limited information estimation methods

Abstract:

Questionnaires are increasingly used tools in clinical and medical research to collect self-reported data on patients' perceived health outcomes. In particular, Patient-Reported Outcome Measures (PROMs) have been specifically designed to assess health-related constructs - such as health status and treatment effects - based on patients' responses to ordinal items. When questionnaires are administered at multiple time points, latent variable models (LVMs) offer a variety of methods to gain insights from the data. For instance, first-order latent growth models are commonly employed to analyse changes across measurement occasions, treating each item separately. Alternatively, longitudinal confirmatory factor analysis (CFA) models offer a powerful framework for studying the measurement properties of a set of items over time, such as assessing whether the items consistently measure the same latent construct across time (i.e., measurement invariance). Within this framework, a common approach to dealing with ordinal items is the Underlying Variable Approach (UVA), which assumes continuous latent variables underlying the observed categorical responses. A longitudinal CFA model for ordinal data, under the UVA, incorporates the following structures: (i) an auxiliary model, linking the ordinal response to an underlying continuous response; and (ii) a measurement model, relating the underlying continuous responses to latent factors at different occasions. CFA can be extended by including a structural model to account for the temporal dynamics of the latent variables, for example through a second-order latent growth model with random effects. A standard approach for fitting LVMs with categorical items is full-information maximum likelihood (FIML), which is theoretically robust and effectively handles data affected by different types of missing data mechanisms. However, FIML can be computationally unfeasible, as it requires the evaluation of high-dimensional integrals, especially in models with many items and/or many latent variables. In contrast, limited information methods, such as diagonally weighted least squares (DWLS), rely on first- and second-order sample statistics and represent a computationally faster solution. While limited information methods can handle missing completely at random data, they need to be integrated with additional methodologies (e.g., multiple imputation or estimator corrections) to appropriately address data that are missing at random. Methodological advancements involving longitudinal latent variable modelling will be presented, along with an example based on longitudinal PROMs data.