Use of Latent Variable Models in Editing and Imputation of Survey Data

Abstract

Latent variable models (LVMs) are a powerful tool for statistical modelling in a wide variety of situations. In the last years LVMs have been used at Istat in the context of editing and imputation (E&I) of survey data. In particular, in E&I LVMs are used for both classification and prediction. In the first case, the latent variable is discrete and its categories are typically associated with groups of units characterised by different error types [1],[3]. In the second case, the flexibility of LVMs in approximating a large class of probability distributions is exploited for prediction of incomplete data [2], or prediction of true values conditionally on observed values possibly affected by measurement errors [1]. Recently, models using LVMs have been proposed for imputation in the context of predictive mean matching (PMM). PMM is developed for continuous variables and it typically uses a regression model to compute the predictive mean of each unit. In the family of LVMs, the use of a factor model extends PMM to deal with target variables of any scale. The method is applied to the multipurpose sample survey data on enterprises gathered in the framework of the Business Census carried out by Istat in 2012.

References

