

Multilevel analysis of student ratings with missing level-two covariates: a comparison of imputation techniques

Analisi multilivello dell'opinione degli studenti universitari in presenza di valori mancanti delle variabili di secondo livello: un confronto tra metodi di imputazione

Maria Francesca Marino e Carla Rampichini

Abstract We analyse the relationship between student ratings of university courses and several characteristics of the student, the course and the teacher. In particular, we exploit data from a survey collecting information about teacher beliefs and practices at the University of Padua in a.y. 2012/13. Student ratings are nested into classes, calling for multilevel modelling. However, due to survey non-response, the information about beliefs and practices is missing for about half of the teachers, posing a serious issue of missing data at level 2. To avoid listwise deletion, we make multiple imputation via fully conditional specification, exploiting information at all hierarchical levels. The proposed approach turns out to be effective. We found that some of the teacher beliefs and practices are significantly related to student ratings.

Abstract *Il presente lavoro analizza la relazione tra le valutazioni dei corsi universitari fornite dagli studenti e le caratteristiche di docenti e studenti. I dati provengono da un'indagine ad hoc svolta dall'Università di Padova nell'a.a. 2012/13 su convinzioni e pratiche dei docenti universitari. La struttura gerarchica dei dati, con valutazioni raggruppate per insegnamento, richiede l'uso di modelli multilivello. L'indagine presenta un alto tasso di non-risposta, ponendo un serio problema di dati mancanti a livello 2. Si considera un'imputazione multipla basata sull'approccio fully conditional, sfruttando anche l'informazione proveniente dalle unità di livello 1. I risultati ottenuti evidenziano l'efficacia del metodo utilizzato. Inoltre, risulta che alcune delle convinzioni e delle pratiche adottate dai docenti sono significativamente correlate alle valutazioni degli studenti.*

Key words: fully conditional specification, multiple imputation, multivariate mixed model

Maria Francesca Marino

Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence - Italy, e-mail: mariafrancesca.marino@unifi.it

Carla Rampichini

Department of Statistics, Computer Science, Applications "G. Parenti", University of Florence - Italy, e-mail: carla.rampichini@unifi.it

1 Introduction

We analyse student satisfaction, as measured by student evaluation of teaching (SET). The peculiarity of the study lies in the availability of many variables about teacher characteristics and beliefs, and teaching practices. This work exploits data from the University of Padua for academic year 2012/13 about bachelor degree courses. The data set is obtained by merging three different sources: (i) the traditional SET survey with 18 items, measured on a ten-point scale (1: low, 10: high); (ii) administrative data on students, teachers and didactic activities; (iii) an online survey carried out by the PRODID project on teacher beliefs and practices.

Data have a two-level hierarchical structure, with 56,775 student ratings at level 1 and 1,016 classes at level 2. The average class size is 79 (min 5, max 442). We are interested in student satisfaction about two key aspects of teaching, i.e. teacher ability to involve students (item D06 of the SET questionnaire) and teacher clarity (item D07).

The analysis is based on the following bivariate two-level linear mixed model for item m (m : 1 for D06, 2 for D07) recorded on student i in class j :

$$Y_{mij} = \alpha_m + \boldsymbol{\beta}'_m \mathbf{x}_{ij} + \boldsymbol{\gamma}'_m \mathbf{w}_j + u_{mj} + e_{mij} \quad (1)$$

where \mathbf{x}_{ij} is the vector of student covariates (level 1), and \mathbf{w}_j is the vector of teacher and class covariates (level 2). Level 1 errors e_{mij} are assumed to be independent across students; level 2 errors (random effects) u_{mj} are assumed to be independent across classes and independent from level 1 errors. We make standard assumptions for the distributions of the model errors, including homoscedasticity (within each outcome) and normality. Therefore, the response vector $\mathbf{Y}_{ij} = (Y_{1ij}, Y_{2ij})'$ has residual variance-covariance matrix $\text{Var}(\mathbf{Y}_{ij}) = \boldsymbol{\Sigma}_u + \boldsymbol{\Sigma}_e$, where $\boldsymbol{\Sigma}_u$ and $\boldsymbol{\Sigma}_e$ are the covariance matrices of the errors at level 2 and level 1, respectively.

The survey on teacher beliefs and practices has about fifty percent of missing questionnaires, posing a serious issue of missing data at level 2. An analysis based on listwise deletion would discard the entire set of student ratings for non responding teachers, causing two main problems: (i) a dramatic reduction of the sample size, and thus of the statistical power, and (ii) possibly biased estimates if the missing mechanism is not MCAR.

2 Multiple imputation of level 2 covariates

In multilevel models, the treatment of missing data requires special techniques. In fact, the data have a hierarchical structure and, thus, missing values can be at any level of the hierarchy. Moreover, missing values can alter the variance components and the correlations. Multiple imputation (MI) is the most flexible approach to missing data. MI has been extended to the multilevel setting to deal with these special issues, following two main approaches (Mistler and Enders, 2017; Grund *et*

al., 2018): fully conditional specification, also known as multivariate imputation by chained equations, and joint modelling.

In our case study, the substantive model (1) is multilevel, however missing data are only at level 2. This feature makes the imputation simpler than in the general multilevel setting. Indeed, we can apply standard MI techniques to level 2 data and, then, merge level 1 and level 2 data to obtain complete datasets.

Besides this simplification, the MI step for the data under investigation remains a challenging matter, since we have to deal with a large number of categorical variables having a high percentage of missingness. As stated before, about 50% of the teachers did not respond to the whole questionnaire, thus producing missing values on 10 binary items (teacher practices) and 20 ordinal items (teacher beliefs on a seven-point scale). In particular, the imputation model at level 2 should include all the level 2 covariates and information on level 1 variables, especially on the response variables.

Several strategies to summarize information from level 1 variables can be adopted (Erler *et al.*, 2016; Grund *et al.* 2017). Here, we choose to consider the cluster means as they are rather effective and easy to be implemented in cases where level 1 variables (including the response) are completely observed, as in our case.

We perform multivariate imputation by chained equations using the `mi` command of Stata (StataCorp, 2017). In our case, we use a binary logit as imputation model for the 10 binary items (teacher practices), and a cumulative logit for the 20 ordinal items (teacher beliefs). For all imputation models, we consider the following covariates: the fully observed class and teacher characteristics, the cluster means of level 1 variables (covariates and outcomes), and the cluster size.

3 Results

The bivariate two-level model (1) is fitted by maximum likelihood on 10 imputed data sets, and the results are combined with the standard MI rules. The analysis is conducted using the `gsem` and `mi` commands of Stata (StataCorp, 2017).

We first fit the model without covariates in order to explore the correlation structure of the two outcomes, i.e. teacher ability to involve students and teacher clarity. We find out that the ICC is about 30% for both outcomes. The two outcomes are highly correlated (0.83), especially at level 2 (0.933).

Then, we add the available covariates in the model. These include 6 student characteristics, centered on class averages (gender, age, high school grade, year of enrollment, regular student, number of exams passed in 2012), 8 fully observed covariates at level 2, including teacher and course characteristics (gender, age, role, and involvement of the teacher in the course; credits, class size, school and compulsoriness of the course) and 30 further level 2 covariates summarizing teacher practices and beliefs. To ensure parsimony, we select a subset of the 30 practices and beliefs using *leaps and bounds* method proposed by Furnival and Wilson (1974) and implemented in the `gvselect` command of Stata (Lindsey and Sheather, 2010). Specifically, we

apply the procedure to each equation of model (1) separately and retain in the final model only those covariates significant for at least one of the two equations according to the AIC criterion. Such a procedure leads to the selection of the following binary indicators of teacher practices: Q01 – *practicals*, Q02 – *exploiting contribution from experts*, Q07 – *using multimedia resources*. As regards teacher beliefs, the following ordinal indicators are retained into the final model: Q12 – *passion for teaching*, Q14 – *usefulness of practicals*, Q17 – *usefulness of student opinions*, Q23 – *usefulness of student-oriented teaching*, Q25 – *need for teaching support*, Q27 – *usefulness of sharing teaching experiences with colleagues*.

To quantify the influence of missing data on the sampling variance of the parameter estimates, we can consider the fraction of missing information (FMI), i.e. the ratio between the sampling variance among imputations and the total sampling variance. For the imputed covariates, the FMI ranges from 0.15 to 0.68, with a median value equal to 0.44. The fraction of missing data is about 0.5; therefore, for the majority of imputed covariates, the trade-off between the sampling error inflation due to MI and its reduction due to sample size increase is favourable.

The main substantive finding is that teacher practices and beliefs from the PRODID survey are significantly related with the SET ratings. In particular, the item *practicals* is negatively related to the outcomes, while *contribution from experts* is positively related. As for teacher beliefs, *passion for teaching* and *usefulness of student opinions* are positively associated with the ratings, while *usefulness of practicals* and *need for teaching support* are negatively associated with the ratings.

An alternative imputation method is joint modelling (e.g. Grund *et al.*, 2017), which allows to impute binary and ordinal data through latent normal variables. We implement this approach using the R package `jomo` (Quartagno and Carpenter, 2017). However, in our case with 10 binary and 20 ordinal variables this approach is not feasible due to the computational burden, unless we treat the ordinal variables as continuous or we previously select a subset of the ordinal variables. The second option is preferable to preserve the original scale of measurement. A thorough comparison between chained equations and joint modelling is the object of our future research.

Acknowledgements This work has been supported by the research project "Advances in Multilevel and Longitudinal Modelling" (BIRD162088/2016) of the University of Padua.

References

1. Eler, N.S., Rizopoulos, D., van Rosmalen, J., Jaddoe, V.W.V., Franco, O.H., Lesaffre, E.M.E.H: Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full Bayesian. *Statistics in Medicine*. **35**, 2955–2974 (2016)
2. Furnival, G. M., Wilson, R. W.: Regression by leaps and bounds. *Technometrics*. **16**, 499511 (1974)
3. Grund, S., Lüdtke, O., Robitzsch, A.: Multiple Imputation of Missing Data for Multilevel Models Simulations and Recommendations. *Organizational Research Methods*. **21**, 111–149

- (2018)
4. Grund S., Lüdtke O., Robitzsch A.: Multiple Imputation of Missing Data at Level 2: A Comparison of Fully Conditional and Joint Modeling in Multilevel Designs, *Journal of Educational and Behavioral Statistics*. (2017)
 5. Lindsey, C., Sheather, S.: Variable selection in linear regression. *The Stata Journal*. **10**, 650–669 (2010)
 6. Mistler, S. A., Enders, C. K.: A comparison of joint model and fully conditional specification imputation for multilevel missing data. *Journal of Educational and Behavioral Statistics*. **42**, 432–466 (2017)
 7. Quartagno, M., Carpenter, J.: jomo: A package for Multilevel Joint Modelling Multiple Imputation. <https://CRAN.R-project.org/package=jomo> (2017)
 8. StataCorp.: Stata, Release 15. Statistical Software. StataCorp LLC, College Station, TX (2017)