

Ph.D. in “Life Course Research” – Socio-demographic curriculum

Academic Year 2023-2024

Linear mixed models, M-quantile and regression, machine learning models and its application to Small Area Estimation

Prof. Salvati Nicola

University of Pisa

nicola.salvati@unipi.it

Objectives

The course offers a review of the main Small Area Estimation Methods. In particular, the course focuses on SAE based on Linear Mixed models, M-quantile regression models, machine learning models and teaches how to apply them to European survey data to have a local monitoring of the Sustainable Development Goals and to estimate deprivation and inequality indicators. R codes for the application of the SAE will be presented.

- 1) Analysis of the collected data for estimation and testing for the phenomenon under study; definition of planned and unplanned domains.
- 2) Direct and indirect estimates for unplanned domains; R codes for the application of the SAE estimators (EURAREA and SAMPLE project libraries)
- 3) quality issues in SAE and usage of SAE in European Statistical System.

At the end of the module student will be able to deal with small area estimation both at the theoretical and empirical level and to apply aggregation methods to indicators from the European survey data.

Students will learn the fundamental small area methods and what might be the problems that arise in the application of them and in the definition of their statistical quality.

Program

Monday morning (Instructor: G. Bertarelli):

- Introduction to small area estimation
- Area level small area estimation
- Introduction to small area estimation with R

Tuesday afternoon (Instructor: F. Schirripa-Spagnolo):

- Unit level small area estimation based on mixed models and M-quantile regression models
- Unit level small area models in R

Wednesday morning (Instructor: N. Tzavidis):

- Unit level small area estimation for general parameters with unit level models
- The ELL and the empirical best predictor
- Data driven transformations
- Real world applications

- Examples in R: The emdi package
- Introduction to machine learning (Random forests) for small area estimation
- Extensions of random forests for small area estimation
- Real world applications

Suggested lecture

Pratesi and Chamabers (2016) Analysis of Poverty Data by Small Area Estimation. New York, Wiley.
Introductory Chapters.
Rao, J.N.K. and Molina, I. (2015) Small Area Estimation. New York, Wiley

Requirements

Correlation and multiple linear regression. Some acquaintance with linear mixed models and with the R software is useful.