UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Data Science: Longitudinal, Multilevel and Multivariate Analysis

2627-1-F8804N008

Aims

Knowledge and understanding: to acquire an in-depth knowledge of advanced data science techniques applicable to the analysis of sociological phenomena.

Applying knowledge and understanding: to provide the necessary skills to analyse microdata bases - including repeated cross-sectional, longitudinal (panel), and multidimensional data - in the sociological domain, in order to address questions of an associative, predictive, or causal nature.

Making judgements: to develop the ability to critically assess the methods and results of quantitative analyses in sociological research, recognising their theoretical assumptions, limitations, and practical implications.

Communication skills: to promote the ability to structure and effectively articulate quantitative findings within a coherent sociological theoretical framework.

Learning skills: to foster autonomy in study and research by encouraging the ability to critically and independently deepen the course content.

Contents

The course provides an advanced overview of analytical tools for the study of complex data structures in sociology, with a particular focus on hierarchical, longitudinal (panel), and multivariate data. Topics covered include multilevel regression techniques, econometric models for panel data, structural equation modeling (SEM), event history analysis, major dimensionality reduction techniques (such as cluster analysis and factor analysis), as well as selected supervised and unsupervised machine learning methods, including neural networks.

Detailed program

The first part of the course is dedicated to methods for analyzing hierarchical data characterized by complex variance structures. In particular, multilevel models will be introduced as an extension of traditional regression models applied to hierarchically structured data. These techniques are also applicable to longitudinal (panel) data, where the same outcome is observed at multiple time points within the same analytical units. This section concludes with an overview of key econometric techniques for panel data analysis, including fixed effects models, random effects models, and the Differences-in-Differences (DiD) estimator. The second part of the course focuses on multivariate data analysis. It covers the main dimensionality reduction techniques, such as Principal Component Analysis (PCA), Factor Analysis, and Cluster Analysis (both hierarchical and non-hierarchical). This is followed by an introduction to Structural Equation Models (SEM), which combine the logic of causal regression and factor analysis within a unified framework. The course also addresses Event History Analysis models, with specific reference to the Kaplan-Meier estimator and the Cox regression model. The course concludes with an in-depth exploration of selected machine learning techniques, particularly supervised neural networks (Multilayer Perceptron) and unsupervised neural networks (Self-Organizing Map

Prerequisites

Prerequisites for the course include knowledge of linear and logistic regression models, as well as a basic theoretical and methodological background in social research.

Teaching form

The course consists of a total of 56 hours of in-person teaching, structured into sessions that combine lecture-based instruction with interactive activities. Each session includes a first part devoted to the presentation of theoretical and methodological content (lecture format), followed by a second part focused on active student participation through individual or group exercises, presentations, and class discussions. Overall, approximately 70% of the course time is allocated to lecture-based teaching, while the remaining 30% is dedicated to interactive and hands-on activities. The course is taught in Italian, and the exercises will be carried out using the statistical software Stata.

Textbook and teaching resource

Slides and reading materials will be provided by the instructor during the course.

Kreft, I. G. G., & de Leeuw, J. (1998). Introducing multilevel modeling. Thousand Oaks, CA: Sage Publications. (optional)

Longhi, S., & Nandi, A. (2015). A practical guide to using panel data. London: SAGE Publications Ltd. (optional)

Singer, J.D. & Willett, J.B. (2003), *Applied Longitudinal Data Analysis* (*ALDA*), Oxford University Press. (optional)

De Lillo, A., Argentin, G., Lucchini, M., Sarti, S., & Terraneo, M. (2007). L'analisi multivariata per le scienze sociali (Cap. 7–8). Milano: Pearson Education. (optional)

Semester

february 2027 - may 2027

Assessment method

Students may choose to take an oral examination, based on the materials provided by the instructor and listed in the course bibliography. The exam will consist of six questions, each relating to topics covered during the course. Each answer will be evaluated on a scale from 0 to 6. The final grade will be the sum of the scores awarded for each response. If the total score is below 18, the exam will be considered failed. If the score is equal to or greater than 31, the final grade will be 30 with honors (30 e lode). Answers will be evaluated according to three main criteria: accuracy, completeness, and clarity of exposition.
Alternatively, students may take a written examination in class, using their own personal computer and the Stata software. The test will require the implementation of four data analysis models presented during the course. Each answer will be graded on a scale from 0 to 8. The final grade will be calculated as the sum of the individual scores. If the total score is below 18, the exam will be considered failed; if the score is 31 or above, the final grade will be 30 with honors.
Evaluation will be based on three key criteria: accuracy of the analysis, completeness of the response, and clarity of exposition. During the exam, the instructor will provide the datasets required for the analyses. The time allowed for the written test is 120 minutes

Office hours

Wednesday (11.00-12.00)

Sustainable Development Goals

QUALITY EDUCATION