UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

# COURSE SYLLABUS

# Data Science: Longitudinal, Multilevel and Multivariate Analysis

**2425-1-F8802N057**

## Aims

This course covers the theory and application of hierarchical, longitudinal and multivariate statistical models. By the end of the course students will be able to recognize and implement study designs that would be appropriate for multilevel and multivariate analyses.

## Contents

This course introduces analytical tools for hierarchical, longitudinal and multivariate data. Topics to be covered include multilevel regression techniques, structural equation models (SEM) and cluster analysis.

## Detailed program

The first part of the course is designed to provide students the methods for the analysis of hierarchical data with a complex variance structure. Multilevel models are powerful regression models used for data that are clustered within units (i.e. children within classrooms, patients within hospitals, employees within companies). These kind of techniques can also be used for longitudinal data - that's to say repeated measures of an outcome over time within units - which are quite common in the field of social sciences.

The second part of the course will be devoted to the multivariate data. More precisely, topics to be covered include structural equation models (SEM) and cluster analysis. The strength of SEM approach is that it combines the logic of causal regression models with the logic of factor analysis.Finally, the course will cover cluster analysis techniques that are data exploration (mining) tools for dividing a multivariate dataset into homogeneous groups.

## Prerequisites

Students should have familiarity with ordinary least- squares (OLS) linear regression models.

## Teaching form

The lessons will be carried out in presence.
Exercises will be carried out using Stata.
The course comprises 56 hours, of which approximately 80% is delivered through traditional lectures (lectures with slide presentations) and 20% through interactive teaching (exercises)

## Textbook and teaching resource

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling* Thousand Oaks, CA: Sage Publications, Ltd

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

De Lillo, A., Argentin, G., Lucchini, M., Sarti, S., & Terraneo, M. (2007). *L'analisi multivariata per le scienze sociali*. Milano: Pearson Education. (cap.8 cap.9)

## Semester

february 2025 - may 2025

## Assessment method

The student can choose whether to perform an oral exam, which will focus on the materials listed in the bibliography or alternatively take a written exam with the software STATA. In this specific case, the implementation of multilevel, longitudinal and multivariate models presented during the course will be requested. The teacher will make the databases available for the analysis. The work carried out will be evaluated and discussed in an oral exam session in presence or via Webex device.

## Office hours

Wednesday (11.00-12.00)

## Sustainable Development Goals