UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Multivariate Statistical Analysis

2021-2-E4101B037

Learning objectives

The course aims at introducing multivariate statistical techniques both from the methodological and from the applicative point of view.

Contents

The course is composed of three parts: R for the Multivariate Statistical Analysis (first part, 3 CFU), Exploratory data analysis (second part, 6 CFU) and Statistical models (third part, 6 CFU).

- **R for the Multivariate Statistical Analysis (3 CFU)** provides “hands-on” training for learning how to analyse data in the R statistical software package. It covers data input/output, data management and manipulation, and how to make useful and informative graphics.
- **Exploratory Analysis (6 CFU)** offers an introduction to the statistical analysis of multivariate observations with the goal of dimensionality reduction thereby facilitating the understanding of the data.
- **Statistical models (6 CFU)** offers an introduction to linear regression models.

Detailed program

First part: R for the Multivariate Statistical Analysis (3 CFU)

- Reading data
- Recoding and manipulating data
- Making exploratory plots
- Multiway contingency tables and Simpson's paradox
- Performing basic statistical analysis with R

Second part: Exploratory Analysis (6 CFU)

- Graphical representation of multivariate data
- Total and generalized variance
- Spectral decomposition theorem
- Principal components analysis
- Cluster analysis: K-means and hierarchical methods
- Factorial analysis

Third part: Statistical Models (6 CFU)

- Simple and multiple linear regression

- Model specification
- Parameter estimation
- Linear hypotheses tests
- Diagnostics
- Variable selection
- Prediction

Prerequisites

Knowledge of the notions given in the courses "Statistics I", "Probability", "Matrix Algebra", and "Statistical inference (Statistics II)" is required.

Teaching methods

Class lectures and lab sessions.

Assessment methods

Students are supposed to pass three written exams (one for each part of the course) and an oral exam (optional). Each written exam consists of questions about theory, numerical exercises and analysis of data sets. The overall mark in the written exam is obtained by averaging the marks obtained in each part. The final mark is an average between written and oral ones.

Textbooks and Reading Materials

First part: R for the Multivariate Statistical Analysis (3 CFU)

- Lecture notes from the instructor
- Agresti, A. and Franklin, C. (2016) “The Art and Science of Learning from Data ”, Pearson

Second part: Exploratory Analysis (6 CFU)

- Lecture notes from the instructor
- Johnson, Wichern (2007) Applied Multivariate Statistical Analysis (6th Edition), Pearson Prentice Hall
- Everitt, Hothorn (2011) An Introduction to Applied Multivariate Analysis with R, Springer
- Zani, Cerioli (2007) Analisi dei dati e data mining per le decisioni aziendali, Giuffré Editore

Third part: Statistical models (6 CFU)

- Lecture notes from the instructor
- M. Grigoletto, F. Pauli, L. Ventura, Modello lineare, teoria e applicazioni con R. Giappichelli, 2017

Semester

The course is scheduled in the first semester (first and second part) and in the first six weeks of the second semester (third part).

Teaching language

Italian