UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistica Multivariata

2627-1-F8206B002-F8206B002-1

Learning objectives

The aim of the course is to introduce multivariate statistical techniques for exploratory purposes and to apply these methodologies to real data through the use of R (R Studio).

By the end of the course, students will be able to apply the aforementioned techniques to real data sets collected in several contexts, e.g., economics, social and environmental fields, and to interpret the results of the corresponding models for the analysis of complex and high-dimensional phenomena.

The course contributes to strengthening knowledge in statistical methodologies and the ability to apply such knowledge, in line with the "Statistics" learning area of the Master's Degree in Statistics and Economics.

Contents

The main macro-topics of the course are listed below.

- Fundamentals of multivariate statistics: the multivariate normal distribution
- Classification methods
- Dimensionality reduction methods and latent variable models
- Distance-based methods

For further details, see the "Detailed Programme".

Detailed program

- **Fundamentals of multivariate statistics: the multivariate normal distribution**
    - Definition and introduction to the Mahalanobis distance
    - Properties
    - Maximum likelihood estimation
    - Distribution of the maximum likelihood estimators and Wishart distribution
    - Tools for assessing the normality assumption and multivariate Box-Cox transformation
    - Introduction to methods for identifying and handling missing data and outliers
- **Classification methods**
    - Classification for continuous variables
        - Linear Discriminant Analysis (LDA)
        - Quadratic Discriminant Analysis (QDA)
    - Classification for categorical (mixed) variables
        - Logistic Regression (LR)
    - Comparison between LDA, QDA and LR
- **Dimensionality reduction methods and latent variable models**
    - Continuous variables
        - Review of Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA)
        - Structural Equation Modelling (SEM)
    - Categorical variables
        - Introduction to Mutiple Correspondence Analysis (MCA)
        - Item Response Theory (IRT)
- **Distance-based methods**
    - Distances and dissimilarities
    - Multidimensional Scaling (MDS)
    - Relationships among MDS, PCA and MCA

Prerequisites

Students are expected to have a basic background in probability theory, statistical inference, programming, and exploratory data analysis, equivalent to that provided by an undergraduate degree in Statistical Sciences.

Teaching methods

The course is in Italian and consists of **47 hours** of **lectures** and **laboratory sessions** conducted in an interactive format.

The lectures will cover the introduction and in-depth study of the statistical methodologies addressed in the course, along with related exercises; the laboratory sessions will focus on the application of the aforementioned methodologies to real data sets using the R (R Studio) software.

Both lectures and laboratory sessions will be conducted in person. Some lectures may be conducted remotely, for up to 20% of the total hours.

Assessment methods

The exam is a written test consisting of theoretical questions, numerical exercises, and numerical exercises to be

completed using R.

There are no mid-term exams.

Textbooks and Reading Materials

- Rencher, A. C. , Christensen, W. F. (2012). Methods of Multivariate Analysis (3rd ed.). Wiley.
- Hastie, T., Tibshirani, R., Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Stanford, CA: Stanford University.
  Slides, note e appunti della docente.
- Adachi, K. (2020) Matrix-Based Introduction to Multivariate Data Analysis (2nd ed.). Springer.

Semester

The course is scheduled in the second term of the first semester.

Teaching language

Italian

Sustainable Development Goals

QUALITY EDUCATION