



UNIVERSITÀ  
DEGLI STUDI DI MILANO-BICOCCA

## COURSE SYLLABUS

### Statistics for Sustainability

2627-1-F7603Q032-F7603Q03201

---

#### Aims

The aim of the course is to introduce multivariate statistical modeling and supervised/unsupervised statistical learning, making students able to identify the most suitable techniques for describing and predicting a phenomenon, in relation to the type of data available.

Students are invited to consult the syllabus of the entire course for details regarding learning- and skill-related objectives.

#### Contents

- Brief recall of the basic notion of statistical inference (point and interval estimation, hypothesis tests).
- Linear regression.
- Logistic regression.
- Principal component analysis.
- Linear discriminant analysis.
- K-means clustering.
- Hierarchical clustering.
- Use of the R software environment on data related to sustainability.

#### Detailed program

- Review of basic probability theory and statistical inference to recall the notion of confidence interval and hypothesis test.
- Simple linear regression: estimating the coefficients and assessing the accuracy of the coefficient estimates.

- Multiple linear regression: estimating the coefficients and assessing the accuracy of the coefficient estimates, linear model selection, prediction.
- Logistic regression: estimating the coefficients and assessing the accuracy of the coefficient estimates, variable selection, prediction.
- Cross-validation to evaluate the performance of a statistical model.
- Dimensionality reduction with the principal components analysis: estimation of principal components and their interpretation, biplot.
- Linear discriminant analysis: estimating discriminant functions, cluster prediction based on linear discriminant scores.
- K-means clustering and hierarchical clustering.

## **Prerequisites**

- Basis of probability theory and knowledge of the most relevant continuous and discrete random variables.
- Basic notions of statistical inference: point estimation, confidence interval hypothesis testing.
- Linear algebra.
- matrix theory.
- Optimization of functions of several variables.

## **Teaching form**

3 CFUs of mixed theoretical and interactive lessons in the classroom (24 hours):

- 8 two-hour lectures, in person, Delivered Didactics;
- 4 two-hour lectures, in person, discussing problems/exercises, Mixed Didactics.

Attendance to lectures and interactive exercises is highly recommended.

## **Textbook and teaching resource**

- James G., Witten D., Hastie T. and Tibshirani R. (2021). An Introduction to Statistical Learning, with applications in R (2nd edition). Springer Verlag.
- Hastie T. , Tibshirani R., Friedman J. (2021). The Elements of Statistical Learning (2nd edition). Springer Verlag.
- Material provided by the lecturer.

## **Semester**

I Semester (December - January)

## **Assessment method**

The exam at the end of the module consists in a written test with open-ended questions to assess the student's knowledge, jointly with the development of an application in R on real data.  
An oral discussion can be requested by the lecturer.

The final score will be between 18/30 and 30/30 *cum laude*, based on the overall assessment considering the following criteria:

- (1) knowledge and understanding;
- (2) ability to connect different concepts;
- (3) autonomy of analysis and judgment;
- (4) ability to correctly use scientific language.

## **Office hours**

Always, after scheduling an appointment *via* e-mail.

## **Sustainable Development Goals**

QUALITY EDUCATION

---