UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Statistics III

2627-3-E4101B035

Learning objectives

The course aims to broaden and consolidate the knowledge of **statistical models** introduced in Multivariate Statistical Analysis, providing a systematic treatment of **Generalized Linear Models** (GLMs). In particular, the learning objectives include:

i. providing an advanced understanding of the theory and properties of GLMs;
ii. introducing the main techniques for the analysis of data with non-Gaussian distributions (e.g., binomial, Poisson);
iii. applying these models using the **R** software through practical exercises and labs;
iv. discussing real-world issues such as model misspecification, likelihood-based inference, and the use of quasi-likelihoods.

The course contributes to the achievement of the learning objectives in the study program area of "**Statistics**." In particular, it provides the knowledge and skills needed to identify and implement appropriate statistical-probabilistic models and to draw sound inferential conclusions.

Contents

Statistics III is a monographic course on Generalized Linear Models (GLMs). The detailed syllabus is available on the course webpage. The main topics are:

- Linear models: additional topics (recap, misspecification, robustness)
- Generalized linear models: theory and methods
- Important models: binary and binomial regression, Poisson regression
- Quasi-likelihood

Detailed program

- **Linear models and misspecification**
  - Linear model: review and notation
  - Robustness of the least squares estimator
  - Sandwich variance estimator
  - Weighted least squares
  - Box-Cox transformation, variance-stabilizing transformations
- **Generalized linear models (GLM)**
  - Exponential dispersion family
  - Likelihood-based inference: estimation and hypothesis testing
  - IRLS algorithm: iteratively reweighted least squares
  - Deviance, diagnostics, residual analysis
  - Model selection techniques
  - Implementation in **R**
- **Notable models**
  - Regression for binary and binomial data
  - Poisson regression for count data
- **Quasi-likelihood**
  - Second-order assumptions
  - Unbiased estimating equations
  - Quasi-likelihood function
  - Overdispersion

Prerequisites

This is a B.Sc.-level course, but there are some prerequisites: it is assumed that you have already been exposed to:

- Simple linear regression and descriptive statistics, e.g. from *Statistica I*;
- Inferential statistics, e.g. from *Statistica II*;
- Linear models, e.g. from *Analisi Statistica Multivariata* and *Econometria*;
- R software, e.g. from *Analisi Statistica Multivariata*.

Teaching methods

Classes take place both in the classroom and in the computer lab, combining theoretical aspects with practical applications in data analysis and R programming. The 47 hours of teaching are organized as follows:

- 35 hours of lectures delivered in person;
- 12 hours of laboratory activities.

Assessment methods

The assessment consists of a **written exam carried out in the computer lab** and an oral exam (optional).

The written exam is divided into two sections: the first consists of **theoretical questions** aimed at verifying the acquisition of concepts and formalization skills necessary to correctly set up a statistical model; the second requires the **analysis of a dataset using the R software**. The final grade of the written exam is determined by the average of the two section grades. If either the student or the instructor requests an oral exam, the final grade will be the average of the written and oral exam grades; otherwise, it will coincide with the written exam grade.

During the exam, the use of textbooks or other material is not permitted, with the exception of the R scripts made available by the instructor at the beginning of the exam. The use of mobile phones during the exam is strictly prohibited.

Textbooks and Reading Materials

Required (In italian)

- Salvan, A., Sartori, N. and L. Pace. 2020. *Modelli Lineari Generalizzati*. Springer.
- Azzalini, A. 2008. *Inferenza Statistica: Una Presentazione Basata sul Concetto di Verosimiglianza*. Springer

Optional

- Agresti, A. 2015. *Foundations of Linear and Generalized Linear Models*. Wiley.
- McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*. Second Edition. Springer.

Additional teaching material is made available in the course website.

Semester

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

Teaching language

Italian

Sustainable Development Goals

QUALITY EDUCATION