

Statistical Models – GLM

Data Analytics for Business and Finance – a.a. 2025/26

Fabio Demaria

Course objectives

This course introduces students to the concept of causality, extending their knowledge beyond the simple linear regression models previously studied. The curriculum covers multiple linear regression and logistic regression in greater depth, focusing on both theoretical nuances and practical applications. All statistical techniques are implemented in R through hands-on, in-class tutorials.

Prerequisites

Undergraduate level mathematical and statistical knowledge. Nevertheless, the initial lectures will include a review of key concepts such as vectors and matrices, as well as linear regression.

Course content

0. Introduction to R and RStudio

- 0.1 The R syntax
- 0.2 Functions in R
- 0.3 Data wrangling
- 0.4 Descriptive statistics
- 0.5 Data visualization

1. Linear regression

- 1.1 Simple linear regression
- 1.2 Multiple linear regression
- 1.3 Model visualization
- 1.4 Model selection and evaluation
- 1.5 Export results

2. Logistic regression

- 2.1 The logistic model
- 2.2 Multinomial logistic regression
- 2.3 Ordered logistic regression

3. Poisson regression

- 3.1 Poisson regression for count data
- 3.2 Poisson regression for incidence rates
- 3.3 Overdispersion and extensions

Teaching methods

Teaching is in-person and delivered in Italian. The course utilizes a hybrid lecture-laboratory approach, combining theoretical lessons with applied sessions. Students are supported by comprehensive teaching materials (slides, exercises, etc.) and lecture recordings, with a strong emphasis on data analysis practice using the R programming language.

Materials and books

- [ENG] An Introduction to Generalized Linear Models, A. J. Dobson
- [ENG] Applied Regression Analysis and generalized linear models, J. Fox
- [ITA] Complementi di Statistica Economica, T. Di Fonzo, F. Lisi
- Slides and exercises from MS Teams/Moodle.

Final examination

Students may choose between two assessment methods: a final written examination or the submission and presentation of a group project developed in R.

Expected results

1. Knowledge and understanding: Mastery of the principles of causality and the theoretical framework of generalized linear models, including multiple, logistic, and Poisson regressions.
2. Applying knowledge and understanding: Proficiency in using R for data wrangling, visualization, and the practical implementation of advanced statistical models on real-world datasets.
3. Making judgments: Ability to critically select and evaluate the most appropriate model based on data distribution, diagnostic checks, and overdispersion.
4. Communication skills: Competence in translating R-based outputs into professional reports and visualizations that effectively communicate causal relationships.
5. Learning ability: Development of a coding-based analytical mindset that enables the independent acquisition of new statistical techniques and R packages.

Sustainable Development Goals

This course contributes to the achievement of the UN 2030 Agenda for Sustainable Development goals.

- 4 – Quality Education
- 9 – Industry, Innovation, and Infrastructure