UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Modelli Statistici

2627-2-E4104B010-E4104B010-2

Learning objectives

1. Knowledge and Understanding

By the end of the course, students will have acquired a solid understanding of the main statistical models for the analysis of multivariate data, with particular emphasis on multiple linear regression, the classical linear model, and logistic regression. They will be able to understand the role of statistical models in the analysis of real-world phenomena, distinguishing them from mathematical models and machine learning approaches. Students will be familiar with the main assumptions underlying the models studied, parameter estimation procedures, model fit evaluation criteria, and methods for interpreting results. They will also gain familiarity with the matrix representation of data and statistical models, as well as with the theoretical foundations of statistical inference applied to regression models.

2. Applied Knowledge and Understanding

Students will be able to apply the statistical models studied to the analysis of data from different application contexts, such as economics, business, social sciences, biology, medicine, and environmental studies. They will be able to specify a statistical model consistent with the objectives of the analysis, estimate its parameters, verify its assumptions, and evaluate its explanatory and predictive performance. They will also be able to use the R software environment for data management, estimation of linear and logistic regression models, identification of issues such as multicollinearity and outliers, and the production of analyses and reports based on real and simulated data.

3. Autonomy of Judgment

Throughout the course, students will develop the ability to critically assess the adequacy of a statistical model with respect to the phenomenon under study. They will be able to evaluate the plausibility of the underlying assumptions, correctly interpret the results of inferential procedures, compare alternative models, and assess their strengths and limitations. They will also acquire the ability to distinguish between explanatory and predictive purposes of statistical analysis, formulating judgments based on empirical evidence and correct interpretation of results.

4. Communication Skills

Students will develop the ability to clearly and rigorously communicate the results of statistical analyses using the technical language of statistical modeling. They will be able to present and discuss regression models, inferential results, goodness-of-fit measures, and predictive indicators both in written and oral form. They will also be able to prepare technical reports supported by tables, graphs, and interpretative comments, adapting the level of detail to the characteristics of the audience.

5. Learning Skills

The course aims to develop students' ability to independently deepen advanced statistical methodologies and to tackle complex data analysis problems. The theoretical and applied knowledge acquired will provide a solid foundation for subsequent courses in statistics, data science, and quantitative analysis, as well as for the professional use of statistical models in decision-making and research contexts. Students will also be able to independently update their skills in response to the evolution of methodological and computational tools for data analysis.

Contents

The course introduces the main tools of statistical modeling for data analysis. It covers the concept of a statistical model and the modeling process, simple and multiple linear regression models, and the main extensions for response variables of different types.

The course addresses the foundations of statistical inference, parameter estimation, model evaluation and selection, and essential diagnostic tools. The methodologies are applied to real and simulated datasets using the R software environment.

Detailed program

The course introduces the main concepts and tools of statistical modeling for the analysis of multivariate data. It presents the concept of a statistical model and its role in the analysis of real-world phenomena, the differences between statistical and mathematical models as well as machine learning approaches, and the main stages of the statistical modeling process, from hypothesis formulation and data collection to parameter estimation and the assessment of the explanatory and predictive performance of models.

Subsequently, regression models for studying relationships between quantitative variables are introduced. The course covers parameter estimation methods, criteria for assessing model fit, model selection procedures, and diagnostic tools for identifying potential issues in data analysis. Elements of matrix representation of data and statistical models are also introduced.

Particular attention is devoted to the inferential aspects of statistical modeling. The properties of estimators are analyzed, along with the assumptions underlying the classical linear model, the construction and interpretation of confidence intervals and hypothesis tests, as well as criteria for comparing alternative models and evaluating their use for explanatory and predictive purposes.

The course also addresses extensions of linear models, with particular reference to the inclusion of categorical variables, variable transformations, and models for non-continuous response variables. The fundamentals of logistic regression and the interpretation of its parameters and related indicators are introduced.

Theoretical topics are consistently complemented by practical applications using the R software environment, employing both real and simulated datasets to develop operational skills in statistical analysis and interpretation of results.

Prerequisites

Students are required to have successfully completed the prerequisite courses: **Statistica I**, **Analisi Matematica I**, **Algebra Lineare**, and **Calcolo delle Probabilità**.

For a smoother understanding of the course contents, a solid knowledge of the statistical inference topics covered in **Statistica II** is strongly recommended.

Teaching methods

The course consists of 49 hours of lectures and 12 hours of in-person practical sessions. Teaching activities are supported by a tutor assistant.

The lectures cover the theoretical foundations of statistical modeling, complemented by computer-based demonstrations and applied examples developed using the R software environment. The practical sessions take place in a computer laboratory and focus on the application of statistical analysis techniques to both real and simulated datasets.

Throughout the course, students are guided in developing statistical analyses, selecting appropriate methodologies, and interpreting the results. Course materials are made available on the e-learning platform before each lecture.

Assessment methods

The assessment consists of a written examination and an oral examination. No midterm assessments are scheduled.

The written examination is conducted using the R software environment and consists of the analysis of a dataset provided by the instructor. Students are required to perform a comprehensive statistical analysis, including the application of the models covered during the course, the production of graphs and tables, and the preparation of a report in R Markdown format. The report must include the interpretation of the results and a discussion of the main findings.

The oral examination covers the topics addressed during the course and is designed to assess students' theoretical understanding of the methods employed as well as their ability to critically interpret the results obtained.

In addition, a group project presentation session is scheduled, during which students present an independent statistical analysis based on real-world data that they have identified and collected on their own. The project is presented in class and contributes to the overall course assessment.

Textbooks and Reading Materials

Main Reference Texts

- J. J. Faraway: *Linear Models with R*, 2nd ed., Chapman & Hall/CRC, 2014.
- G. James, D. Witten, T. Hastie, R. Tibshirani: *An Introduction to Statistical Learning*, 2nd ed., Springer, 2021 (with particular emphasis on the chapters covering linear regression and logistic regression, including applications in R).
- Carter Hill, William E. Griffiths, Guay C. Lim: *Principles of Econometrics*, 5th ed., Wiley, 2018.
- D. Nolan, D. T. Lang: *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving*, Chapman & Hall/CRC, 2015.
- Lecture notes and supplementary materials provided by the instructor during the course.

Semester

II Semester, III Cycle

Teaching language

Italian

Sustainable Development Goals

QUALITY EDUCATION