# Statistical modelling

## Prof. Guido Consonni; Prof. Federico Castelletti

***COURSE AIMS AND INTENDED LEARNING OUTCOMES***

The course presents two important aspects of current statistical modelling: i) statistical learning; ii) Bayesian statistics.

The former is based on recent advances in methods for applied statistical analysis also for big datasets; the latter concerns an approach to statistical theory which has become an indispensable resource for scientific and social science researchers. The teaching method will involve methods, computations and applications using the free and open-source language and environment R. The ultimate aim is to provide students with important tools to understand and perform statistical analyses.

At the end of the course the student is expected to:

have acquired a good knowledge of the principal aspects of statistical learning, such as: the trade-off between bias and variance, supervised and unsupervised learning, classification problems, cross-validation and regularization methods; foundations and main concepts of Bayesian statistics such as: prior, posterior and predictive distributions, computational methods (MCMC), hierarchical models. – *Knowledge and understanding*;

understand and interpret the assumptions behind models for empirical analyses, to evaluate the statistical properties of the methods, and to perform empirical analyses using the R language and environment.–*Applying knowledge and understanding*;

evaluate which models and techniques to use depending on the context and the applied problem, to evaluate the uncertainty associated to statistical inference. – *Making judgements*;

be able to describe with an appropriate statistical language the assumptions behind a model as well as its merits, and to communicate empirical findings. –  *Communication skills*;

acquire a conceptual framework and technical tools to advance his/her studies in the methods and techniques of Statistics and Data Science. – *Lifelong learning skills*.

***COURSE CONTENT***

The course is divided in two modules.

Module 1 (Statistical learning) - Federico Castelletti

* + The *R-programming language*
  + Statistical learning: main concepts and model accuracy
  + Cross-validation and the bootstrap
  + Model selection and regularization: ridge regression, lasso
  + Tree-based methods: regression and classification trees

## Module 2 (Bayesian statistics) - Guido Consonni

* + Bayes’ rule: prior, posterior and predictive distribution
  + Bayesian inference for binomial, Poisson and normal model using conjugate priors
  + Monte Carlo integration for Bayesian analysis
  + Markov Chain Monte Carlo methods
  + Bayesian hierarchical modelling

***READING LIST[[1]](#footnote-1)***

Module I

G. James, D. Wiiten, T. Hastie, R. Tibshirani. *An Inroduction to Statistical Learning.* Springer.2013. <http://faculty.marshall.usc.edu/gareth-james/ISL/>

Module II

P.D. Hoff, *A First Course in Bayesian Statistical Methods*, Springer*,* 2009.

Class notes*,* coding and further material will be posted on Blackboard

***TEACHING METHOD***

A blend of lectures, coding and data analysis (60 hours). Exercise and lab sessions (20 hours).

***ASSESSMENT METHOD AND CRITERIA***

Assignment (33%) on module I followed by a written paper with one question on module I and two on module II; alternatively written paper with two questions on module I (50%) and two on module II (50%). Questions involve methods, exercises, data analysis and R-code.

***NOTES AND PREREQUISITES***

Students enrolling in this course are expected to know data analysis, probability and frequentist inference, at the level of Statistics courses usually taught in a bachelor degree in Economics; see for instance the topics covered in ‘Statistica (analisi dei dati e probabilità) and ‘Statistica applicata’ (or ‘Statistics’ and ‘Applied Statistics’) at this University. These topics will be presented in a preliminary course in Statistics which will be offered in the week before lectures begin. Furthermore students are expected to know the theory and applications of the normal linear model and logistic regression as taught in an MSc. Course such as Empirical economics at this University.

In case the current Covid-19 health emergency does not allow frontal teaching, remote teaching will be carried out following procedures that will be promptly notified to students.

1. I testi indicati nella bibliografia sono acquistabili presso le librerie di Ateneo; è possibile acquistarli anche presso altri rivenditori. [↑](#footnote-ref-1)