**Applied statistics for business analytics**

Prof. Emilio Gregori; Prof. Giulia Cesari

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

This course will give you the fundamentals of probability, and methods for statistical inference, with focus on estimation and tests of hypotheses. Students will learn how to perform regression analysis for datasets with quantitative and qualitative variables, especially in the context of applied works in business analytics. More advanced statistical techniques based on the partition of the covariate space, and unsupervised learning methods for clustering will be presented.

The course deals with concepts, methods and techniques of multivariate analysis of data for estimating and predicting purposes. The focus is on analyzing real data, using the software R.

The following learning abilities are provided and expected to be achieved by participants at the end of the course:

1. Knowledge of concepts, terms and methods of statistical inference, regression models, tree-based methods and clustering techniques (DD1- Knowledge and understanding);
2. Ability to correctly apply methods of estimation and prediction through the use of the statistical software R. (DD2- Applying knowledge and understanding);
3. Quantitative thinking addressed to make independent evaluations, driven by estimations and predictions (DD3- Making judgements);
4. Ability to read and interpret data and communicate results, through the extraction of qualitative information from quantitative data (DD4-Communication);
5. Mastery of tools useful for quantitative analyses required in future careers involving management of data, rigorous reasoning and data-driven decision-making (DD5- Lifelong learning skills).

***COURSE CONTENT***

*Module I* (Prof. Emilio Gregori)

– Basics of probability theory: random variables, distribution theory.

– Point estimation for population mean, proportion and variance. Interval estimation for the population mean (in the normal case) and for the population proportion (binomial model). Extension to two sample problems.

– Hypothesis testing for population mean and proportion. Extension to two sample problems.

– Linear regression: inference and prediction, related t and F tests.

– Logistic regression: inference and prediction.

*Module II* (Prof. Giulia Cesari)

– Development of a data mining study: the CRISP-DM methodology.

– Tree-based methods: regression and classification trees.

– Clustering techniques: K-means clustering, hierarchical clustering.

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

The textbook of the course is:

A. Agresti, *Statistical Methods for the Social Sciences*, Pearson, 2018, 5th edition.

A selection of notes and other materials, available in the course reserve on Blackboard.

Suggested optional material for an introduction to R: W.J. Owen, *The R Guide*, CRAN, 2010.

***TEACHING METHOD***

The course involves lectures, exercise sessions using PC and discussion of use cases.

Should face-to-face lectures not be possible out of health considerations related to COVID-19, we will move classes online. Students will be notified in advance.

***ASSESSMENT METHOD AND CRITERIA***

Two alternative type of exams are foreseen:

*Option 1* - Two written partial exams, one mid-course exam and one final exam at the end of the course. Moreover, students can present a group-assignment. The first partial exam will cover approximately 60% of the topics of Module I, the second partial exam will cover the rest of Module 1 and the topics of Module II. The first partial exam will foresee close-ended questions about solving exercises on a real dataset through the use of the software R and the second partial exam will foresee close-ended questions about theoretical concepts and solving exercises on a real dataset through the use of the software R. The maximum grade for both the two partial exams is 28. The assignment will foresee an autonomous activity of analysis on a real dataset thorugh the software R, according to some basic instructions; the work is carried out in groups. The maximum grade for the assignment is 3 points.

The aim of the exams is to assess reasoning analytic abilities on the course subjects; language properties and presentation communication abilities are also assessed. The final grade is given by the weighted average between the grade of the first partial exam (50%) and the grade of the second partial exam (50%) plus the grade of the assignment.

*Option 2* - One general written exam, covering both the topics of Module I and the topics of Module II. The exam will foresee close-ended questions about theoretical concepts and about solving exercises on a real dataset through the use of the software R. The aim of the exam is to assess reasoning analytic abilities on the course subjects; language properties and presentation communication abilities are also assessed.

For the exam sessions other than the first, only Option 2 is applied.

Other details will be provided on Blackboard.

***NOTES AND PREREQUISITES***

Students enrolling in this course should ideally have a basic understanding of Statistics with regard to data analysis, probability and inference, at the level of the undergraduate courses taught at this University, namely Statistics or Statistica (analisi dei dati e probabilità).

Correspondence between students and teachers will be managed through Blackboard.

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