**Quantitative methods for corporate and banking decisions**

## Prof. Enrico Fabrizi

***COURSE AIMS AND EXPECTED LEARNING RESULTS***

The aim of the course is to provide students with a set of basic statistical tools for data analysis and modeling. Students should also be able to orient themselves in the scientific literature when using regression models as well as super-supervised and unsupervised classification methods.

At the end of the course students will be able to:

- perform routine statistical analyzes (descriptive statistics, graphical representations, hypothesis testing) using the R software;

- perform basic database management operations in order to prepare them for statistical analysis;

- estimate linear and logistic / probit regression models using the R software, interpret the results and diagnostics, implement alternative estimation strategies in case of failure of the standard assumptions;

- know the theoretical foundations and apply statistical learning techniques related to supervised and unsupervised classification (classification trees, random forests, Gradient Boosting, k-means clustering) with R and know how to apply them in data analysis.

***COURSE PROGRAM***

1. Review of descriptive and inferential statistics: averages, variability measures, confidence intervals, t test for one mean, t test for the comparison of two means, one-way ANOVA.

2. Introduction to the R environment for statistical data analysis. Data analysis and graphical exploration of data with R. Data matrices and relationships between variables: association and correlation measures and related tests;

4. The multiple linear regression model for the analysis of cross section data; inference on model parameters; diagnostics. Introduction to quantile regression.

5. Introduction to the problem of classification in statistics: basic definitions, classification accuracy tools.

6. The logistic and the probit regression models. Their use as a classification tool. Elements of multinomial logistic regression.

7. Other statistical techniques of supervised classification: regression trees, random forests, gradient boosting. Introduction to statistical techniques of unsupervised classification (k-means clustering).

***REFERENCES***

James G., Witten D., Hastie T., Tibshirani R., *An introducton to statistical learning with applications in R*, Springer Verlag, 2013.

Notes and supplementary material provided by the teacher

***DIDACTICS***

The lectures in the classroom (which will also be accessible remotely through streaming) will be supplemented by online material and videos posted by the teacher on the blackboard page of the course. The active participation of students in the lessons, that may also take place in the computer lab, is required.

***EVALUATION METHODS***

Laboratory test in which students will be asked to perform simple data analysis with the R software and comment on the results. The exercises can be supplemented by more theoretical questions. Specifically, the test will be divided into 7 questions / problems, each being evaluated on a 0 to 4.3 points scale.

During the course exercises (based on the use of the R software) will be assigned to students for home working. This assignments will allow students to accumulate some points (max 3) which will add up to the mark obtained in the laboratory test.

***PREREQUISITES***

##### Active attendance to lessons, although not mandatory, is strongly recommended. Prerequisite of the course is the knowledge of the main techniques of descriptive and inferential statistics, or the contents of the volume Alan Agresti: Statistical Methods for the Social Sciences, 5th Edition (chapters 1-8).

.

***OFFICE HOURS***

# Reception hours are available online on the teacher's personal page, which can be consulted at http://docenti.unicatt.it/. In any case it is possible to contact the teacher via email for any need. The reception of the students is normally held at the teacher's office (or electronically via Teams / Skype), Faculty of Economics and Law (Economics building, III floor).