# Statistics (Statistical Learning)

## Prof. Silvia Angela Osmetti

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

Modern society is characterised by a growing amount of information and data (big data). The course aims to provide students with modern quantitative tools of Statistics and Data Science to model and understand complex data that can increasingly be found not only in the scientific field but also in economic-business areas. Special attention will be paid to computational aspects – through the use of the free and open-source R language – so as to develop concrete empirical analysis skills on real cases.

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

1. Knowledge of concepts, terms and methods of statistical learning techniques;
2. The ability to apply the statistical methods presented to real problems in the economic, financial and social fields, including through the use of suitable statistical software.
3. A capacity for critical thinking that enables them to interpret the results obtained, extrapolating qualitative information from quantitative data, to make independent judgements on the performance of the statistical models in their application to real problems and to select the most suitable model from those studied;
4. The ability to communicate the results of a statistical analysis clearly and accurately and to quantitatively justify the decisions that have been made to a specialist and non-specialist audience.
5. Good learning skills that enable students to identify links between the topics covered on the course and other topics within their own study plans; and the competences needed in professional positions related to managing data to a good level, rigorous reasoning skills and the ability to make data-driven decisions.

***COURSE CONTENT***

1. Introduction to Statistical learning and to R programming language.

2. Simple and multiple linear regression.

3. Classification.

4. Bootstrap methods.

5. Resampling methods and choice of model.

6. Statistical methods based on decision trees.

7. Unsupervised methods: Principal component analysis and cluster analysis.

***READING LIST***

J. Gareth-D. Witten-T. Hastie-R. Tibshirani, *An Introduction to Statistical Learning,* Springer, 2013 edition and following.

Course material on the *Blackboard* platform.

R Statistical Software *http://www.r-project.org/*

***TEACHING METHOD***

Lectures (60 hours) including description of statistical methods, analysis of data set and R codes and practical exercises.

***ASSESSMENT METHOD AND CRITERIA***

The exam is based on a two-hour written assessment and includes the following: open questions on the statistical methods presented in lectures (points 1-7 on the syllabus), interpreting programmes in *R*, and reading and interpreting results.

For attending students, the written exam may be substituted by two partial exams contributing equally to the final mark (the interim test carried out during reading week in semester 2 and the final exam of the summer session). The interim test covers points 1-4 of the syllabus. Students who pass the interim test must take a completion test which involves the development of a project. The project consists of the statistical analysis of a data set using the techniques presented in class and the use of the R software. Detailed instructions on exam procedures will be published on Blackboard.

The exam is designed to assess students’ reasoning and analytical skills on the course topics as well as their language and communication skills.

***NOTES AND PREREQUISITES***

Students are expected to have the same background knowledge of data analysis, probability and inference as obtained from courses of Statistics of a Faculty of Economics; for example, the topics covered in the courses of *Statistics* and *Applied Statistics*.

***OFFICE HOURS***

The teacher receives the students as reported on the personal web page and/or on the Blackboard page of the course.