# Applied statistics and big data (Business intelligence)

## Prof. Enrico Lazzaretto; Prof. Maria Lucia Pace

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

The aim of the first part of the course is to provide students with the fundamental statistics tools required for analysing data supported by R software for R quantitative analysis.

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

* Apply the fundamental methods of inferential statistics (point estimation and interval estimation and hypothesis testing);
* Manage and organise data using R software.
* Descriptive and bivariate analysis using R.

The aim of the second part of the course is to teach students how to use statistics tools to face and solve real problems, as well as explore certain topics which can be applied to the context of Big Data.

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

* estimate linear regression models and logistics models using R software
* carry out residue analysis so as to assess the value of the regression model which was applied
* use classic methods data reduction summary and data simplification (analysis of the main components of factor analysis).

***COURSE CONTENT***

The course is organised into 6 modules and covers the following topics:

1. **Outline of inference and probability**: Fundamental methods of inferential statistics: random variables; estimates and statistics tests
2. **Introduction to R as an environment and tool for the analysis and processing of data**
3. **Simple linear regression:** Introduction to statistics techniques used to shape the dependence between two variables and residue tests
4. **Multiple linear regression:** Introduction to statistics techniques used to shape the dependence between multiple variables
5. **Advanced R with a focus on multiple regressions**
6. **Regression logistics:** Understand the specification of models which use qualitative variables as a response variable

***READING LIST***

Espa, G. Micciolo, R. *Analisi esplorativa dei dati* con R Apogeo, 2012.

S. Borra-A. Di Ciaccio, Statistica 3/ed - *Metodologie per le scienze economiche e sociali*, McGraw-Hill, 2014.

D.M. Levine-T.C. Krehbiel-M.L. Berenson, Statistica, Pearson Italia, 7a edizione, Con Mylab E Etext, Milan, 2018. (Chapters 3, 5-9).

***TEACHING METHOD***

The course alternates frontal lectures with individual work and/or online group work.

Students will be required to study topics presented in the video-lessons and material for further study remotely. The webinars (practice and feedback) are an important part of the learning process.

***ASSESSMENT METHOD AND CRITERIA***

The course offers the opportunity to take an ongoing assessment: 50% of the mark will consist of two written tests, one of which is practical and requires the use of R software and the remaining 50% is an oral test on theory. To take the ongoing assessment, the first written test must be passed with a minimum mark of 18, otherwise students will change to the final summative assessment format. The final summative exam can be taken in all official exam sessions, both in the winter and in the summer sessions. The final summative exam will consist of a written and an oral part.

***NOTES AND PREREQUISITES***

Students should have prior knowledge of the basic notions of statistics (description of probability data and theory). The following text is recommended:

- D.M. Levine-T.C. Krehbiel-M.L. Berenson, Statistica, Pearson Italia, 7a edizione, Con Mylab E Etext, Milan, 2018. (Chapter 2-4)

- S. Borra - A. Di Ciaccio, *Statistica. Metodologie per le scienze economiche e sociali*, Mc-Graw-Hill, Milan, 2014, 3a ed. [Chapters 1-4, 6, 16]

The introductory lectures will consolidate certain fundamental concepts.

Further information can be found on the lecturer's webpage at http://docenti.unicatt.it/web/searchByName.do?language=ENG or on the Faculty notice board.