



Syllabus

Academic Year	2021/2022
Program	Data Science and Management
course	Advanced Statistics
Term	I semester
Year	1
SSD	SECS-S/01 Statistica
Credits	6

INSTRUCTIONAL GOALS

The course provides an overview of several advanced statistical methods for data science. The focus is on understanding advantages and limitations of each approach, interpretation, and main applications in various disciplines, particularly in economics, business, and management. This provides students with the basic ideas in modeling continuous and categorical outcomes, also with dependent data. Students in this course will not only gain a deep understanding of asymmetric modeling and dimension reduction for data science, but they will also acquire the practical skills necessary for their successful applications to problems in science and industry.

INTENDED LEARNING OUTCOMES

They describe what a learner is expected to know, understand and be able to demonstrate after completion of a learning path.

Knowledge and understanding:

The course will offer key statistical tools to investigate interrelationships between predictors and a continuous or categorical outcome. It will also discuss dimension reduction techniques and their use for visualization, scoring, and information retrieval in multivariate data settings. Strengths, weaknesses, use cases, and interpretation of the results of each method will be discussed in depth.

Applying knowledge and understanding:

On successful completion of this course students will be able to:

- Appreciate the different statistical methods for prediction of continuous and categorical outcomes
- Select, implement, and interpret the most appropriate statistical predictive tools in a range of real-world applications.
- Appreciate the perspectives offered by different conditional models
- Summarize multivariate data sets for information retrieval

Making judgements:

Students are expected to be able to choose the appropriate statistical method to pursue their aims with data analysis, taking into considerations data limitations and comparative performance. Students will demonstrate fluency with the software and interpretation of the results. Throughout the entire course, students will be stimulated to consider strengths and weaknesses of the different methods discussed in class.

Communications Skills:

This course will give the students the possibility to acquire the lexicon of statistical models and multivariate analysis. They will learn how to communicate effectively the results of their data analyses. A special emphasis will be given to writing concise and clear reports through the project group work.

Learning skills:

This course will empower students with the capability to learn several analytical tools for analyzing data, and to apply them to real-world problems in an independent and critical way. A strong emphasis



will be given to the application of the techniques and tools covered in the course to complex problems that are typical of today's data-driven companies.

Pre-requisites	Basic knowledge of descriptive statistics and statistical inference, including hypothesis testing and confidence interval estimation. Basic knowledge of the main probability distributions. Working knowledge of R is recommended.
Course content	The course will cover the following topics: <ul style="list-style-type: none">● Introduction to the R statistical software● Multivariate linear regression● Analysis of categorical data: multi-way tables, the generalized linear model. Logistic regression. Poisson regression.● Analysis of clustered and panel data: mixed models● Principal component analysis● Modern use cases: text mining, image processing
Reference Books	Lecture notes, R scripts, research papers and other course material will be made available on the e-learning platform. Recommended readings: Witten J.D., Hastie T., Tibshirani R. (2014). An Introduction to Statistical Learning with Applications in R. Springer, Springer Series in Statistics Chatfield, C. and Collins, A. J. (1981) Introduction to Multivariate Analysis, Chapman & Hall/CRC Press Everitt, B. S. and Hothorn, T. (2006) A Handbook of Statistical Analyses Using R. CRC Press. Available for free at: http://www.ecostat.unical.it/tarsitano/Didattica/LabStat2/Everitt.pdf
Teaching Methods	The course consists of lectures complemented by practical lab sessions and group project works.
Assessment	There will be a written midterm exam, a written final exam and a project. In the midterm and final exams students are required to demonstrate that: <ul style="list-style-type: none">● they have acquired a deep knowledge of the main advanced statistical methods for data science;● they have an in-depth understanding of the different statistical methods for prediction of continuous and categorical outcomes;● they are able to choose the appropriate statistical method for problems in data analysis. Midterm and final will count for 40% of the grade each. Students that will not take the midterm and final during the course are required to take an oral exam after the course, where they are required to demonstrate the same skills described above. In the project students are required to demonstrate that they are able to: <ul style="list-style-type: none">● apply analytical tools to real-world problems in an independent and critical way;● communicate effectively the results of their data analyses;● writing concise and clear technical reports about their work. The project will count for 20% of the grade. The overall assessment will take into account the level of knowledge and understanding of advanced statistical methods for data science acquired by the students; their capacity for thinking creatively, analytically, logically and critically; their ability to apply advanced statistical methods, such as asymmetric modeling and dimension reduction, to data science problems in science and industry; their



capacity to consider strengths and weaknesses of different advanced statistical methods for data science; and their capacity to present effectively findings and conclusions.
