Syllabus

Descrizione corso

Titolo insegnamento	Mathematics and Statistics for Data Science
Codice insegnamento	73065
Titolo aggiuntivo	
Settore Scientifico-Disciplinare	MATH-03/B
Lingua	Inglese
Corso di Studio	Corso di laurea magistrale in Informatica per la Data Science
Altri Corsi di Studio (mutuati)	
Docenti	dr. Paola Lecca, Paola.Lecca@unibz.it https://www.unibz.it/en/faculties/engineering/academic-staff/person/42241
Assistente	
Semestre	Primo semestre
Anno/i di corso	1
CFU	6
Ore didattica frontale	40
Ore di laboratorio	20
Ore di studio individuale	90
Ore di ricevimento previste	18
Sintesi contenuti	- Reminders of fundamentals of differential, integral calculus, and linear algebra. - Probability theory - Data distribution models: analysis and approximations - Confidence intervals and hypothesis tests - Joint probability distributions and analysis of independence - Regression analysis
Argomenti dell'insegnamento	The course consists of two parts.

	1) The first part of the course will cover the basic topics of probability, and will be preceded by a review of the concepts and mathematical tools of linear algebra and mathematical analysis necessary to understand the concepts and solve the problems of probability and statistics. In this part elements of Bayesian inference will be introduced, as it is a backbone of machine learning and AI methods, as well of computational procedure in medicine, life-sciences and decision-making processes in industry and manufacturing. 2) In the second part the course will deal with the basic topics of statistics, in particular inferential statistics whose core topics are confidence intervals, hypothesis testing, regression analysis and basic techniques for causality determination. The course aims to give a solid theoretical foundation of the concepts of probability and statistics and to provide the tools and methods for solving practical problems that probability theory and statistics can solve in their applications to data science. At the end of the course, the student will have - revised the foundations of mathematical calculus necessary to approach probability and statistics problems - acquired the foundations of mathematical calculus, probability and statistics that will allow him/her to solve the most common problems of statistical data processing and interpretation that are common to many scientific fields such as computer science and software engineering, artificial intelligence, and data processing in numerous applications of these fields (e.g., biology, medicine and social sciences, etc.) - acquired the ability to solve problems that frequently arise in data science, such as, but not limited to, calculating the probability of events, modelling events data using probability distributions, testing hypotheses and predictive modelling of data.
Parole chiave	Probability models; statistical data analysis; hypothesis testing; joint probabilities and correlation; causality; statistical data modelling.

Prerequisiti	Although the basic concepts and calculation techniques of integral and differential calculus and linear algebra are reviewed at the beginning of the course, it is necessary for the student to be familiar with the fundamental definitions, concepts and techniques of calculus and linear algebra.
Insegnamenti propedeutici	
Modalità di insegnamento	Frontal lectures and exercises to be carried out during the laboratories.
Obbligo di frequenza	Generally, attendance is not compulsory but recommended. Non-attending students must contact the lecturer at the start of the course to agree on the modalities of the independent study.
Obiettivi formativi specifici e risultati di apprendimento attesi	Knowledge and understanding D1.1 - Knowledge of the key concepts and technologies of data science disciplines D1.8 - Knowledge of the mathematical-statistical principles required for data analysis D1.11 - Knowledge of the main algorithms for data analysis, and of elements of the complexity theory Applying knowledge and understanding D2.1 - Practical application and evaluation of tools and techniques in the field of data science D2.2 - Ability to address and solve a problem using scientific methods D2.7 - Practical application of mathematical-statistical tools and methods from the field of data science Making judgements D3.2 - Ability to autonomously select the documentation (in the form of books, web, magazines, etc.) needed to keep up to date in a given sector Communication skills D4.1 - Ability to use English at an advanced level with particular reference to disciplinary terminology Learning skills D5.3 - Ability to deal with problems in a systematic and creative way and to appropriate problem solving techniques

Obiettivi formativi specifici e risultati di apprendimento attesi (ulteriori info.)	
Modalità di esame	Assessment is based on 1) the evaluation of a written report on an assigned theoretical topic 2) the evaluation of the final written exam consisting of exercises.
Criteri di valutazione	Written review report: 50% of the final mark. This part assesses the objectives: D1.1, D1.8, D3.2, D4.1 and D5.3. Written exam: 50% of the final mark This part assesses the objectives: D2.1, D2.2, D2.7, D5.3. The examination is deemed passed if the student scores at least 18/30 on both the review report and the written examination. If a mark of 18/30 or higher is obtained in both parts (review reports and written exam) of the exams, the final mark will be calculated as the average of the two marks obtained.
Bibliografia obbligatoria	The course includes topics from different disciplinary areas of mathematics that are unlikely to be contained in a single textbook. It is therefore advisable that the student follows the notes and the didactical material that the lecturer will make available at each lecture and laboratory. The notes provided during the course can be deepened by referring to textbooks, as, for example: • Howie, John M., Real Analysis, Springer, 2001 • Robert Hogg, Elliot Tanis , Dale Zimmerman, Probability and Statistical Inference, Pearson Education Inc. • Maurits Kaptein , Edwin van den Heuvel, Statistics for Data Scientists. An Introduction to Probability, Statistics, and Data

	Analysis, Springer 2022 • Frederik Michel Dekking, Cornelis Kraaikamp, Hendrik Paul Lopuhaä, Ludolf Erwin Meester, A Modern Introduction to Probability and Statistics, Understanding Why and How, Springer 2005 • James, E. Gentle, Matrix Algebra: Theory, Computations and Applications in Statistics (Springer Texts in Statistics) 2nd ed. 2017. Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it
Bibliografia facoltativa	Suggested by the lecturer during the course if needed.
Altre informazioni	Software used: The course does not include programming labs. However, example scripts in R (www.r-project.org) or in Python may be shown.
Obiettivi di Sviluppo Sostenibile (SDGs)	Istruzione di qualità