COURSE DESCRIPTION – ACADEMIC YEAR 2019/2020

Course title	**Statistics for Data Science**
Course code	73004
Scientific sector	MAT/06
Degree	Master in Computational Data Science (LM-18)
Semester	1
Year	1
Credits	6
Modular	No

Total lecturing hours	40
Total lab hours	20
Attendance	NB – Generally, attendance is not compulsory, but non-attending students have to contact the lecturer at the start of the course to agree on the modalities of the independent study. The exam modalities are the same both for attending and non-attending students (see Assessment).
Prerequisites	None
Course page	https://ole.unibz.it/

Specific educational objectives	The course belongs to the type "affini o integrative – formazione affine" in the curriculum "Data Analytics". The course is designed for acquiring professional skills and knowledge in statistical analysis of data common to different applicative domains. The educational objectives are: 1) to introduce the students to the main concepts of probability and statistics; 2) to provide the students with the theoretical foundations, the methodologies, the practical techniques, and the software tools related to probabilistic reasoning, regression, descriptive and inferential statistics.

Lecturer	Paola Lecca
Contact	Piazza Domenicani 3, Room 3.11, Paola.Lecca@unibz.it, +39 0471 016162
Scientific sector of lecturer	INF/01
Teaching language	English
Office hours	Tuesday 10:15 - 12:15, arrange by email.
Lecturing Assistant (if any)	Fabiola Del Greco
Contact LA	Piazza Domenicani 3, Room 1.04, fabiola.delgreco@eurac.edu
Office hours LA	Wednesday 18:00-19:00, arrange by email.
List of topics	• Introduction to probability. Probability diagrams, conditional probability. • Hypothesis testing and ANOVA • Test of independence • Goodness of fit tests. • Correlation • Linear and Logistic regression with one and multiple variables • Time series • Probabilistic models (EM)

Teaching format	Frontal lectures, theoretical exercises and exercises on computer with software R.

Learning outcomes	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 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 judgments • 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.

Assessment	This course foresees a PASS/FAIL exam. Assessment: written exam both for attending and non-attending students.
Assessment language	English
Assessment Typology	Monocratic
Evaluation criteria and criteria for awarding marks	This course foresees a PASS/FAIL exam. Assessment: written exam both for attending and non-attending students. The written exam is evaluated with a score expressed in thirtieths. The minimum threshold for passing the exam is 18/30. Example: The exam will consists of a set of exercises and questions whose scores sum to 30/30. The correct solution as well as the correct answer to a question score positively, whereas wrong solutions of exercise score zero, and wrong answer to question score negatively.

Required readings	Text books: 1) Christian Heumann, Michael Schomaker, Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R, Springer 2017. Reading suggestions:

	1) Peter J. Brockwell, Introduction to Time Series and Forecasting, Springer 2016. Subject Librarian: David Gebhardi, David.Gebhardi@unibz.it
Supplementary readings	Supplementary readings will be suggested during the course.
Software used	The software used in this course is R (The R project for statistical computing: https://www.r-project.org/).