

COURSE OF STUDY *Strategie di Impresa e Management (SIM)*

ACADEMIC YEAR 2023/24

ACADEMIC SUBJECT *Multivariate Data Analysis*

General information	
Year of the course	2nd year
Academic calendar (starting and ending date)	11/09 – 22/12 2023
Credits (CFU/ETCS):	8
SSD	Statistics (SECS-S/01)
Language	Italian (English on demand)
Mode of attendance	Classroom attendance (not mandatory)

Professor/ Lecturer	
Name and Surname	Massimo Bilancia
E-mail	massimo.bilancia@uniba.it
Telephone	N/A
Department and address	Ionic Department - Taranto
Virtual room	Microsoft teams (code: a3ixcpa)
Office Hours	By appointment

Work schedule			
Hours			
Total	Lectures	Hands-on (laboratory, workshops, working groups, seminars, field trips)	Out-of-class study hours/ Self-study hours
200	64		136
CFU/ETCS			
8	8		

Learning Objectives	The course provides the theoretical and practical foundations for Data Science techniques and their applications in the field of business management
Course prerequisites	Basic knowledge of statistics and mathematics.

Teaching strategies	Lectures + PC-based lab instruction by R
Expected learning outcomes in terms of	
Knowledge and understanding on:	The course aims to teach the basic elements of Data Science techniques
Applying knowledge and understanding on:	The student will learn to appreciate and apply in practice the models learned in the theoretical part of the course, using common data analysis software
Soft skills	<ul style="list-style-type: none"> • <i>Making informed judgments and choices</i> <ul style="list-style-type: none"> ○ The student will be able to decide on the most appropriate model to gain new insights in the different intended application areas (business domain applications, business intelligence and marketing) • <i>Communicating knowledge and understanding</i> <ul style="list-style-type: none"> ○ The student will be able to produce reports on the new knowledge extracted by Data Mining techniques. • <i>Capacities to continue learning</i>

	<ul style="list-style-type: none"> ○ The course aims to provide the basic elements of Data Science techniques, knowledge on which is essential for the possibility of attending courses of a more advanced nature that prepare for the figure of Data Scientist
Syllabus	
Content knowledge	<p>Part I – Basic probability and statistical inference</p> <ol style="list-style-type: none"> 1. Introduction 2. Elementary probability 3. Discrete and continuous random variables 4. Double and multivariate random variables 5. The elements of statistical inference 6. Data matrices <p>Part II – Data Mining and knowledge discovery</p> <ol style="list-style-type: none"> 1. Supervised and unsupervised learning 2. Association rules and Market Basket Analysis 3. Naïve Bayes classification 4. Decision trees 5. Clustering I: hierarchical clustering 6. Clustering II: k-means clustering <p>Part III – Lab: Introduction to R and Data Mining libraries</p>
Texts and readings	M. Bilancia (2020) <i>Dispense per il Corso di Metodi Statistici Multivariati – Versione 1.2 Febbraio 2020</i> . Freely available under Creative Commons 4.0 CC BY-NC-ND Licence
Notes, additional materials	N/A
Repository	Microsoft™ Teams classroom

Assessment	
Assessment methods	<ol style="list-style-type: none"> 1. PRACTICAL TEST AT THE CALCULATOR (duration 1 hour), consisting of the analysis of a real data set using the methods explained in the laboratory lectures: The candidate must prepare a Microsoft™ Word™ document (or using Libre Office word processor) containing all R commands used, all textual and graphical output generated, and explanatory comments. You may refer to lab handouts 4-5-6-7, which are printed on paper, to complete the practical exam: You will have available for this purpose the handouts of these lectures, reproduced in a more practical format of 4 slides per sheet. These slides must be printed as you received them, i.e., you may not annotate them or add your notes to them (also, you may not use the PDF file of the handouts: Violation of these requirements will result in your inability to use them during the exam). Consultation of other texts or documents (in written or electronic form) is not allowed and will result in immediate exclusion from the exam 2. WRITTEN EXAM TEST (duration 1 hour), focused on the preparation of a paper on one open question (free form) chosen from all the theoretical topics covered in class. No texts, documents or computer/telephone assistance of any kind may be used to complete the paper, as this will result in immediate disqualification from the exam
Assessment criteria	<ul style="list-style-type: none"> • <i>Knowledge and understanding</i> <ul style="list-style-type: none"> ○ Ability to accrue knowledge in the field of probability theory and multivariate data analysis, with a view toward applications in economics and business.

	<ul style="list-style-type: none"> • <i>Applying knowledge and understanding</i> <ul style="list-style-type: none"> ○ Ability to apply learned theoretical skills in practice using specialized data analysis software. • <i>Autonomy of judgment</i> <ul style="list-style-type: none"> ○ Ability to select the appropriate data analysis technique for the nature of the proposed problem. • <i>Communicating knowledge and understanding</i> <ul style="list-style-type: none"> ○ Demonstrated ability to effectively explain the skills learned. • <i>Communication skills</i> <ul style="list-style-type: none"> ○ Ability to produce a meaningful report on the proposed problem.
Final exam and grading criteria	<p>The laboratory test is graded YES / NO. The written test is graded in 30/30. The grade given to the candidate who passes both tests is the grade of the written test.</p> <p>A candidate who fails one of the two tests may retake it in one of the subsequent calls.</p> <p>Each candidate may selectively reject only the grade of the written test and save the laboratory test for later calls</p>
Further information	N/A
	.