



PROGRAMME *Health Management*
YEARS *2026-2027*
COURSE *Statistics and Big Data*

Principali informazioni sull'insegnamento	
Year	<i>I year II</i>
Period	<i>II semester</i>
CFU/ETCS:	<i>8</i>
SSD	<i>Statistics SECS-S/01</i>
Language	<i>English</i>
Attendance mode	<i>Virtual</i>

Teachers	
Name	<i>Niccolo Salvini and Giulio Giacomo Cantone</i>
Mail	<i>g.cantone@unicz.it</i>
Phone	
Room	<i>Room 28 West Wing</i>
Virtual room	<i>Provided upon prior email communication</i>
Office hours	<i>upon prior email communication</i>

Teaching			
Hours			
Total	In classroom	Labs	Individual study
<i>200</i>	<i>0</i>	<i>48</i>	<i>152</i>
CFU/ETCS			
<i>8</i>			



Prerequisites	<i>Basic mathematics. No prior programming experience is required.</i>
Expected Outcomes DD1 Knowledge DD2 Know-how DD3-5 Other	<i>I part</i> By the end of the module students will be able to (i) use R to import, describe and analyse health-related datasets; (ii) compute and interpret the main descriptive measures and the correlation between variables; (iii) run and interpret hypothesis tests on means; (iv) specify, estimate and interpret simple and multiple linear regression models, and understand why multivariable models are central to big-data analysis in healthcare. <i>II part</i> This will be a focus on the application of methods of multivariate regression within frameworks of causal inference, statistical matching, and statistical learning on large datasets Knowledge and understanding: The student will learn the principles of probabilistic reasoning and the role of quantitative methods in describing public health phenomena and taking decisions under uncertainty. Applying knowledge and understanding: The student will be able to collect, describe, present, and summarize data from. The student will know how to perform statistical hypothesis even in presence of highly multi-dimensional datasets Making judgements: Upon completion of the course, the student will be able to critically evaluate the appropriateness of statistical procedures applied to evidence-based policies. Communication skills: Upon completion of the course, the student will be able to provide descriptive graphical representations of various data types and use appropriate terminology to justify decisions based on statistical analysis. Learning skills: Upon completion of the course, the student will be able to correctly evaluate various general sources and materials related to data analysis.
Contents	1. Introduction to R and RStudio: objects, data frames, data import, basic plots. 2. Descriptive statistics: mean, variance, standard deviation and deviations from the mean, standardisation. 3. Covariance and correlation. 4. Hypothesis testing on means: one-sample, two-sample and paired t-tests. 5. Simple linear regression: estimation, interpretation, goodness of fit. 6. Multiple linear regression: interpretation of coefficients, confounding,



	model selection. Moving from two to many variables as the entry point to big data: all the previous topics are prerequisites for models with
References	- <i>Lecture slides, R scripts and datasets provided by the instructor on the e-learning platform.</i> - James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J. (2023). <i>An Introduction to Statistical Learning with Applications in R/Python</i>. Springer (free online edition), chapters 2–4. - Wickham, H., Çetinkaya-Rundel, M., Grolemund, G. (2023). <i>R for Data Science</i>, 2nd ed. O'Reilly (free online edition).
Materials	Formula sheets and additional teaching materials may be provided upon request during lectures or office hours.

Assessment	
Evaluation	Practical written exam in R (analysis and interpretation of a dataset), integrated with the assessment of the co-taught part of the course.
Other	
	.