UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Public Health Data Science

2627-2-FDS02Q043

Aims

This module aims at teaching students how to analyze medical data (especially, data of electronic health records) through computational statistics and machine learning techniques to infer new knowledge about the conditions of patients.
This course aims to provide the basic concepts of epidemiology that are at the basis of a proper methodological approach to a research project in public health. The student will be able to deal with big data in public health particularly focusing on several aspects including design, data managment and analysis. The student will be able to implement optimal design strategies on registries, administrative health data, clinical and omics data.

1. Knowledge and Understanding
   By the end of the course, students will have acquired:
   Core concepts of epidemiology and clinical research methodology.
   The rationale behind various epidemiological study designs.
   Characteristics, structures, and limitations of health-related Big Data.

2. Applied Knowledge and Understanding
   Students will be able to apply statistical and methodological knowledge to:
   Design, manage, and perform statistical analysis of Big Data in public health.
   Implement specific design strategies for administrative, clinical, and omics data.
   Use the Bioconductor environment

3. Independent Judgment
   Students will develop the ability to:
   Independently select the study design and statistical tool that best suit to the epidemiological/clinical question.
   Critically evaluate data quality, identifying and correcting bias and confounders, where possible.

4. Communication Skills
   Students will be able to:

Translate complex statistical results into clear reports and indicators for multidisciplinary audience
Work and communicate effectively in multidisciplinary public health teams.

5. Learning Skills
   Students will have developed independent learning skills that will enable them to:
   Keep up to date with advanced statistical models and new methods for managing health data.

Contents

Dataset search and retrieval
Data preparation and data cleaning
Exploratory data analysis
Unsupervised machine learning
Supervised machine learning
Feature ranking
Result understanding and validation
R and Python programming languages
Survival analysis
Population epidemiology
Study designs
Statistical methods with application to various types of health data.
Integration of different data sources.

Detailed program

Dataset search and retrieval
Data preparation and data cleaning
Exploratory data analysis
Unsupervised machine learning
Supervised machine learning
Feature ranking
Result understanding and validation
R and Python programming languages
Basics in population epidemiology.
Advanced study designs to combine data from different sources (registry data, biomarkers, biobanks, surveys).
Survival analysis: survival estimate and Cox model regression.
Use of the Bioconductor environment.

Prerequisites

Descriptive and inferential statistics and basic machine learning
Basic knowledge of R o Python

Teaching form

In-person theory classes and practice exercise classes
3 2-hour lectures conducted in a remote (asynchronous) delivery mode

Textbook and teaching resource

Classes slides and scientific papers mentioned during classes

Articles:
Davide Chicco, Vasco Coelho (2025) "A teaching proposal for a short course on biomedical data science", PLOS Computational Biology 21(4): e1012946. https://doi.org/10.1371/journal.pcbi.1012946

Textbooks:
Kenneth J. Rothman Sander Greenland, Timothy L. Lash . Modern Epidemiology. Lippincott Williams & Wilkins; 3rd ed.
Eric Vittinghoff, David V. Glidden, Stephen C. Shiboski, Charles E. McCulloch. Regression Methods in Biostatistics Linear, Logistic, Survival, and Repeated Measures Models. Statistics for Biology and Health book series. Springer; 2nd edition (March 6, 2012)
Marie Reilly "Beyond classic epidemiological designs" https://www.routledge.com/Controlled-Epidemiological-Studies/Reilly/p/book/9780367186784 Chapman & Hall/CRC Biostatistics Series 2023
Modern Statistics for Modern Biology, by Susan Holmes and Wolfgang Huber, Cambridge University Press (2019). ISBN: 9781108705295.

Semester

First semester

Assessment method

Scientific project including both teaching units to test the ability of the student in the application of research methodology in public health. Delivery of a report, and oral presentation of the work done (75% of the overall grade).
Questionnaire with closed answer to evaluate the preparation on the program of the teaching unit Statistical Modeling for Epidemiology (25% of the overall grade).

Office hours

To define via email by writing to davide.chicco(AT)unimib.it or paola.rebora(AT)unimib.it

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING