

SYLLABUS DEL CORSO

Public Health Data Science

2627-2-FDS02Q043

Obiettivi

L'insegnamento vuole far apprendere a studenti e studentesse come analizzare dati medici (specialmente quelli di cartelle cliniche elettroniche) attraverso tecniche di statistica computazionale, di apprendimento automatico, e di analisi di sopravvivenza per scoprire nuova conoscenza sulle condizioni dei pazienti.

Il corso si propone di fornire i concetti di base dell'epidemiologia che sono alla base di un adeguato approccio metodologico a un progetto di ricerca nel campo della salute pubblica. Lo studente sarà in grado di gestire i big data nella salute pubblica, concentrandosi in particolare su diversi aspetti tra cui la progettazione, la gestione e l'analisi dei dati. Lo studente sarà in grado di implementare strategie di disegno adeguate ai registri, ai dati sanitari amministrativi, ai dati clinici ed omici.

1. Conoscenza e capacità di comprensione

Al termine del corso lo studente avrà acquisito:

Concetti core dell'epidemiologia e metodologia della ricerca clinica.

Logica dei diversi disegni di studio epidemiologici.

Caratteristiche, strutture e limiti dei Big Data sanitari.

2. Conoscenza e capacità di comprensione applicate

Lo studente sarà in grado di applicare le conoscenze statistiche e metodologiche per:

Progettazione, gestione e analisi statistica di Big Data in sanità pubblica.

Implementazione di strategie di disegno specifiche per dati amministrativi, clinici e omici.

Utilizzo dell'ambiente Bioconductor

3. Autonomia di giudizio

Lo studente svilupperà la capacità di:

Scegliere in maniera autonoma modello statistico e disegno di studio più idonei al quesito epidemiologico/clinico.

Valutare criticamente la qualità dei dati, identificando e correggendo ove possibile bias e confondimenti.

4. Abilità comunicative

Lo studente sarà in grado di:

Tradurre i risultati statistici complessi in report e indicatori chiari per un audience multidisciplinare
Lavorare e comunicare efficacemente in team multidisciplinari di sanità pubblica.

5. Capacità di apprendimento

Lo studente avrà sviluppato abilità di apprendimento autonomo che gli consentiranno di:

Aggiornarsi autonomamente su modelli statistici avanzati e nuove metodologie di gestione dei dati sanitari.

Contenuti sintetici

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

Epidemiologia della popolazione

Disegni di studio

Metodi statistici con applicazione a dati di natura diversa.

Integrazione di fonti dati diverse.

Programma esteso

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

Nozioni di base sull'epidemiologia di popolazione.

Disegni di studio per dati provenienti da fonti diverse (dati di registro, biomarcatori, biobanche, sondaggi).

Analisi della sopravvivenza: stima della funzione di sopravvivenza e modello di Cox.

Utilizzo dell'ambiente Bioconductor

Prerequisiti

Statistica descrittiva e inferenziale e basi dell'apprendimento automatico

Conoscenza di base di R o Python

Modalità didattica

Lezioni in presenza ed esercitazioni in presenza
3 lezioni di 2 ore condotte in modalità remota (asincrona)

Materiale didattico

Slides presentate a lezione ed articoli scientifici segnalati a lezione

Articoli scientifici:

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>

Libri di testo:

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.

Periodo di erogazione dell'insegnamento

Primo semestre

Modalità di verifica del profitto e valutazione

Progetto scientifico comprendente entrambe le unità didattiche per verificare le capacità dello studente nell'applicazione della metodologia di ricerca in sanità pubblica.

Consegna di una relazione e presentazione orale del lavoro svolto. (75% del voto finale).

Questionario a risposta chiusa per valutare la preparazione sul programma dell'unità didattica Statistical Modeling for Epidemiology (25% del voto finale).

Orario di ricevimento

Da concordare via email scrivendo a [davide.chicco\(AT\)unimib.it](mailto:davide.chicco(AT)unimib.it) o [paola.rebora\(AT\)unimib.it](mailto:paola.rebora(AT)unimib.it)

Sustainable Development Goals

SALUTE E BENESSERE
