UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

R Laboratory For Biostatistics

2122-1-F8203B044

Obiettivi formativi

Il corso si propone di fornire le basi teoriche e le conoscenze di R utili alla gestione e analisi di dati raccolti mediante un disegno epidemiologico sperimentale o osservazionale. Verranno inoltre forniti esempi di utilizzo di tecniche di machine learning in questo contesto.

Conoscenza e comprensione

Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- la gestione dei dati con R
- l'uso di diversi modelli e tecniche biostatistiche con R
- l'applicazione di tecniche di machine learning di classificazione e di selezione delle variabili negli studi sperimentali ed osservazionali

Capacità di applicare conoscenza e comprensione

Alla fine dell'insegnamento gli studenti saranno in grado di:

- gestire con il linguaggio R il database di partenza per una buona sintesi e visualizzazione dei dati
- analizzare con il linguaggio R i dati provenienti da uno studio sperimentale o osservazionale
- applicare tecniche di classificazione e di selezione delle variabili nella costruzione dei modelli statistici

Contenuti sintetici

Gestione dei dati con R

Metodi statistici per l’analisi dei principali disegni sperimentali e osservazionali in R

Metodi di machine learning in R per la classificazione e la selezione delle variabili

Programma esteso

Gestione dei dati con R: i) Introduzione al linguaggio R, ii) metodi per la pulizia dei dati, iii) metodi per la creazione di report e la visualizzazione dei dati

Metodi statistici per l’analisi dei principali disegni sperimentali e osservazionali in R: i) analisi di uno studio trasversale, ii) analisi di uno studio di coorte, iii) analisi di uno studio caso-controllo, iv) analisi di uno studio sperimentale

Metodi di machine learning in R per la: i) verifica della capacità discriminatoria di un test diagnostico, ii) selezione delle variabili in un modello multivariato

Prerequisiti

Nessuna propedeuticità formale. Si richiede però la conoscenza dei contenuti degli insegnamenti di Modelli statistici I

Metodi didattici

Lezioni frontali alternate a esercitazioni pratiche su dati reali erogate in laboratorio informatico.

In caso di emergenza COVID-19, le attività didattiche si svolgeranno da remoto secondo le modalità indicate sulla piattaforma e-learning.

Modalità di verifica dell'apprendimento

Modalità prova finale

Lavoro di gruppo di analisi su dati reali da consegnare prima della prova orale che consiste nell'esposizione del lavoro svolto. Tale modalità d'esame consentirà di valutare nei candidati la capacità di elaborazione critica dei metodi applicati e dei risultati ottenuti.

Non sono previste prove in itinere.

La modalità d'esame è la stessa per frequentanti e non frequentanti

Le modalità sopra descritte potranno subire variazioni in base all'andamento dell'emergenza Covid-19. Se saranno necessarie delle variazioni verranno tempestivamente rese disponibili in questa pagina.

Testi di riferimento

Slide delle lezioni e materiale integrativo disponibili sulla piattaforma della didattica online http://elearning.unimib.it/.

Periodo di erogazione dell’insegnamento

I semestre, II periodo (approssimativamente novembre-gennaio).

Lingua di insegnamento

L'insegnamento è completamente erogato in italiano. La maggior parte del materiale di riferimento è in inglese.