UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Generalized Linear Models in Epidemiology and Medicine

2122-3-E4102B059

Obiettivi formativi

Il corso permette di conoscere sotto il profilo teorico le principali tecniche di analisi dei dati esplorative sia a partire da dati quantitativi che qualitativi. Inoltre mette in grado di analizzare dati empirici mediante tali tecniche.

Conoscenza e comprensione. Mediante una indagine esplorativa su dati multivariati di elevata dimensione si fornisce conoscena e comprensione a riguardo delle relazioni fra variabili in un contesto multivariato secondo diverse ottiche quali la relazione fra insiemi di variabili quantitativi e qualitativi, le variabili che permettono di discrimniare fra gruppi di individui, la rappresentazioni di distanze fra individui in dimensioni ridotte.

Capacità di applicare conoscenza e comprensione. Attraverso le analisi dei dati esplorative più opportune nei diversi casi, si individuano, a partire da insiemi di variabili molto numerose , le variabili più rilevanti per costruire modelli lineari o non lineari di tipo multivariato, modelli logistici, modelli di analisi della varianza, rappresentazioni di individui in spazi di non elevate dimensioni.

Contenuti sintetici

Analisi della correlazione canonica

Analisi discriminante

Analisi delle corrispondenze

Multidimensional scaling

Programma esteso

Introduzione analisi dei dati

A Analisi della correlazione canonica

1 Scopo

2 Modo per ricavare 1° variabile canonica e successive

3 Lettura risultati

4 Esempi

5 Esercitazioni con SAS

B Analisi discriminante

1 Scopo

2 I dati

3 Modo per ricavare 1° variabili discriminante e successive

4 Lettura risultati

5 Modo alternativo per ricavare variabili discriminanti

6 Esempi

7 Esercitazioni con SAS

C Analisi corrispondenze

1 Scopo

2 Matrici profili riga e colonna

3 Analisi corrispondenze in ottica correlazione canonica e componenti principali

4 Lettura risultati: scomposizione ?2

5 Lettura altri risultati

6 Analisi delle corrispondenze multiple

7 Esempi

8 Esercitazioni con SAS

D Multidimensional scaling

1 Multidimensional scaling metrico con distanze euclidee

2 Teorema di Torgerson

3 Lettura risultati

4 Esempi

5 Esercitazioni con SAS

6 Multidimensional scaling con dissimilarità metriche

7 Multidimensional scaling con dissimilarità ordinali

8 Esempi

9 Esercitazioni con SAS

Prerequisiti

Nessuna propedeuticità formale.

Si richiede però la conoscenza dei contenuti degli insegnamenti di Calcolo delle probabilità, Statistica 1, Laboratorio di Informatica, Analisi statistica multivariata, Base di dati, Analisi dei dati

Metodi didattici

Il corso verrà erogato sia mediante lezioni frontali sia mediante esercitazioni pratiche in laboratorio informatico con il linguaggio SAS. Durante le lezioni verranno fornite le conoscenze teoriche dei metodi statistici oggetto del corso, mentre nelle esercitazioni si applicheranno queste tecniche ad alcuni casi reali soprattutto in medicina ed epidemiologia.

Modalità di verifica dell'apprendimento

L’esame è in forma scritta con orale facoltativo. La verifica scritta si compone di due domande di teoria sui metodi oggetto del corso e da una prova pratica in laboratorio, da svolgersi durante l'esame scritto con il software SAS, in analogia a quanto svolto durante le esercitazioni.
Le domande teoriche consentono di verificare la conoscenza degli scopi, dei metodi di risoluzione, del commento dei risultati dei metodi di analisi esplorativa dei dati. Si richiede la capacità di esprimersi con un adeguato linguaggio tecnico. La prova pratica vuole verificare la capacità di analizzare un data set di dati reali mediante gli strumenti di analisi dei dati più opportuni nei diversi casi commentando e interpretando adeguatamente i risultati.

Non sono previsti esami distinti per studenti frequentanti e studenti non frequentanti.

Non sono previste prove in itinere.

Testi di riferimento

Dispense e slides del docente.

Vitali O., Statistica per le scienze applicate. Vol. 2, Cacucci Editore, 1993 (per correlazione canonica e analisi discriminante).

Zani S., Cerioli A. Analisi dei dati e data mining per le decisioni aziendali. Giuffré, 2007 (per analisi delle corrispondenze e multidimensional scaling).

Periodo di erogazione dell’insegnamento

2 semestre a.a. 2021/22, secondo ciclo

Lingua di insegnamento

Italiano