UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Data Science: Analisi Longitudinale, Multilivello e Multivariata

2324-1-F8802N057

Obiettivi

Il presente corso offre una presentazione teorica ed applicativa di modelli statistici applicati a dati gerarchici, longitudinali e multivariati. Al termine del corso gli studenti saranno in grado di riconoscere e implementare disegni di studio appropriati per analisi multilivello ed analisi multivariate.

Contenuti sintetici

Il corso in questione offre una panoramica di strumenti per l'analisi di dati gerarchici, longitudinali e multivariati. Gli argomenti affrontati includono tecniche di regressione multilivello, modelli di equazioni strutturali e analisi dei gruppi.

Programma esteso

La prima parte del corso è focalizzata sui metodi per l'analisi di dati gerarchici che presentano una struttura di varianza complessa. Più nello specifico verranno presentati i modelli multilivello che rappresentano un'estensione dei tradizionali modelli di regressione applicati a dati gerarchicamente strutturati, ovvero raggruppati entro unità di livello superiore (come per esempio gli studenti classificati entro le classi scolastiche, i pazienti raggruppati entro ospedali, i dipendenti accorpati entro le aziende). Questo tipo di tecniche è utilizzabile anche per dati longitudinali - nella forma di occasioni ripetute di un dato outcome nel tempo entro particolari unità - che sono ampiamente utilizzati nell'ambito delle scienze sociali.

La seconda parte del corso sarà dedicata ai dati multivariati. Verranno presentati i modelli di equazioni strutturali (SEM) e l'analisi dei gruppi. La forza dei modelli SEM sta nel fatto di inglobare sia la logica di regressione causale che la logica di analisi fattoriale entro un unico approccio.

Il corso terminerà con un approfondimento dedicato alla cluster analysis che rappresenta un prezioso strumento di esplorazione per dividere una base dati multivariata entro raggruppamenti omogenei.

Prerequisiti

Gli studenti dovranno avere una certa familiarità con i modelli di regressione lineare

Modalità didattica

Le lezioni saranno svolte in presenza
Gli esercizi verranno svolti usando il software statistico Stata
Il corso si compone di 56 ore, di cui indicativamente il'80% con didattica erogativa (lezioni frontali con utilizzo di slides) e il 20% con didattica interattiva (esercitazioni)

Materiale didattico

Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling* Thousand Oaks, CA: Sage Publications, Ltd

Singer, J.D. & Willett, J.B. (2003), *Applied Longitudinal Data Analysis* (*ALDA*), Oxford University Press. (opzionale)

De Lillo, A., Argentin, G., Lucchini, M., Sarti, S., & Terraneo, M. (2007). *L'analisi multivariata per le scienze sociali*. Milano : Pearson Education. (cap.7 cap.8)

Periodo di erogazione dell'insegnamento

febbraio 2025 - maggio 2025

Modalità di verifica del profitto e valutazione

Lo studente potrà scegliere se effettuare un esame orale, che verterà sui materiali riportati in bibliografia o in alternativa sostenere un esame scritto in aula con il proprio pc e il software stata. In questo caso specifico, verrà richiesta l'implementazione di modelli di analisi multilivello, longitudinale e multivariata presentati durante il corso. Il docente metterà a disposizione le basi dati su cui effettuare le analisi. Il lavoro svolto verrà valutato e discusso in una sessione di esame orale in presenza oppure in remoto tramite dispositivo Webex.

Orario di ricevimento

Mercoledi' 11.00-12.00

Sustainable Development Goals