



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistica Multivariata

2526-1-F8206B002-F8206B002-1

Obiettivi formativi

Il corso si pone come obiettivo quello di introdurre tecniche di analisi statistica multivariata con finalità esplorativa e di applicare tali metodologie a dati reali attraverso l'utilizzo del **R (R Studio)**.

Al termine del corso, le studentesse e gli studenti saranno in grado di applicare le tecniche di analisi a dati raccolti in molteplici contesti, quali ad esempio quello economico e sociale, e di interpretare i risultati dei suddetti modelli per l'analisi di fenomeni complessi e ad elevata dimensionalità.

Il corso contribuisce a consolidare le conoscenze nell'ambito delle metodologie statistiche e la capacità di applicare tali conoscenze, in coerenza con l'area di apprendimento di "Statistica" del corso di laurea magistrale in Scienze Statistiche ed Economiche.

Contenuti sintetici

I macro-argomenti principali del corso sono riportati di seguito.

- Fondamenti di statistica multivariata: la normale multivariata
- Metodi di classificazione
- Metodi di riduzione dimensionale e modelli a variabili latenti
- Metodi basati sulle distanze

Per maggiori dettagli si rimanda al "Programma esteso".

Programma esteso

- **Fondamenti di statistica multivariata: la normale multivariata**
 - Definizione e introduzione alla distanza di Mahalanobis
 - Proprietà
 - Stima di massima verosimiglianza
 - Distribuzione degli stimatori di massima verosimiglianza e distribuzione di Wishart
 - Strumenti per la verifica dell'assunzione di normalità multivariata e trasformazione di Box-Cox multivariata
- **Metodi di classificazione**
 - Classificazione per variabili continue
 - Analisi Discriminante Lineare (ADL)
 - Cenni ad estensioni
 - Classificazione per variabili categoriali (miste)
 - Regressione logistica (formulazione, stima, estensione multinomiale)
 - Confronto ADL e regressione logistica
 - Cenni ad estensioni
- **Metodi di riduzione dimensionale e modelli a variabili latenti**
 - Variabili latenti continue
 - Analisi delle Componenti Principali (ACP) e Analisi Fattoriale (AF)
 - Structural Equation Modelling (SEM)
 - Variabili latenti categoriali
 - Analisi delle Corrispondenze Multiple (ACM)
 - Item Response Theory (IRT)
- **Metodi basati sulle distanze**
 - Distanze e dissimilarità
 - Multidimensional Scaling (MDS)
 - Relazione tra MDS, ACP e ACM

Prerequisiti

Analisi esplorativa, modelli statistici, calcolo delle probabilità, inferenza statistica, programmazione.

Metodi didattici

Il corso è erogato in lingua italiana e prevede **48 ore** tra **lezioni frontali** e **sessioni di laboratorio** svolte in modalità interattiva.

Le lezioni frontali tratteranno l'introduzione e l'approfondimento delle metodologie statistiche oggetto del corso e relativi esercizi; le sessioni di laboratorio si focalizzeranno invece sulla applicazione delle suddette metodologie a dati reali mediante l'utilizzo del software R (R Studio).

Le lezioni frontali e le sessioni di laboratorio saranno erogate in presenza. Alcune lezioni potranno essere erogate a distanza (da remoto) fino ad un massimo del 20% delle ore.

Modalità di verifica dell'apprendimento

L'esame consiste in una prova scritta che si compone di domande di teoria, esercizi numerici da svolgere in forma

cartacea e tramite l'utilizzo di R Studio.

Non sono previste prove in itinere.

Testi di riferimento

TBD

Periodo di erogazione dell'insegnamento

Il corso è erogato nel primo periodo del secondo semestre.

Lingua di insegnamento

Italiano

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
