UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Statistics III

2425-3-E4101B035

Obiettivi formativi

Il Corso si propone di approfondire le nozioni relative ai modelli statistici acquisite nel corso di Analisi Statistica Multivariata, offrendo una panoramica sui modelli non lineari.
Le competenze acquisite nel Corso mettono gli studenti in grado di:

1. specificare il modello statistico opportuno per il/i fenomeno/i oggetto di analisi
2. completare le fasi dell'inferenza relative al modello (stima, verifica, diagnostica)
3. utilizzare il modello (in termini sia interpretativi sia previsivi)
4. produrre ed interpretare l'output di analisi di dati reali effettuate tramite il software R

Il corso contribuisce al raggiungimento degli obiettivi formativi nell'area di apprendimento del CdS: "Statistica".

Contenuti sintetici

Il Corso tratta i modelli lineari generalizzati ed i modelli di regressione non parametrica.

Programma esteso

Modelli lineari generalizzati
- Fondamenti (famiglie di dispersione esponenziale)
- Inferenza basata sulla verosimiglianza
- Diagnostica ed utilizzo dei modelli
- Alcuni particolari modelli per risposta continua, binomiale, poisson

Regressione non parametrica

- Fondamenti
- Regressione: kernel e local polynomial

Prerequisiti

E’ vivamente consigliata la conoscenza delle nozioni impartite nell’insegnamento di Analisi Statistica Multivariata.

Metodi didattici

Il corso prevede lezioni frontali sia in aula sia in laboratorio informatico per un totale di 47 ore. Le lezioni in aula sono mirate all'approfondimento delle conoscenze teoriche dello studente sugli argomenti del Corso ed alla loro formalizzazione. Nelle lezioni svolte in laboratorio informatico si trattano gli aspetti di implementazione dei modelli su dati reali e simulati utilizzando il software R.

Tutte le lezioni sono svolte in modalità erogativa.
6 esercitazioni in laboratorio informatico della durata di 2 ore ciascuna saranno svolte in modalità erogativa da remoto.
Le rimanenti 35 ore saranno svolte in moduli da 2 o 3 ore in modalità erogativa in presenza.

Modalità di verifica dell'apprendimento

La modalità di verifica si basa su una prova scritta eseguita in laboratorio ed una prova orale (facoltativa). Nella prova scritta sono previste due sezioni: una costituita da domande teoriche che hanno l’obiettivo di verificare l'acquisizione dei concetti e della formalizzazione necessari per impostare correttamente un modello statistico; la seconda sezione richiede l’analisi di un data set tramite software R. Il voto finale della prova scritta è determinato dalla media dei voti riportati nelle due sezioni. Qualora lo studente (o il docente) richiedano la prova orale, il voto finale è la media dei due voti riportati nelle prove scritta ed orale, altrimenti coincide con l'esito della prova scritta.
Durante la prova non è ammesso l'uso di testi o altro materiale con l'esclusione dei codici che verranno messi a disposizione dal docente all'inizio della prova.
Durante la prova non è ammesso l'uso del cellulare.

Testi di riferimento

- Hardin J.W., Hilbe J.M., Generalized Linear Models and Extensions, Stata Press, 2007
- Dobson A.J., Barnett A.G., An introduction to Generalized Linear Models, CRC Press, 2008
- Azzalini A., Bowman A.W., Applied Smoothing Techniques for Data Analysis, Clarendon Press, 1997

Ulteriore materiale, in particolare i codici in R utilizzati a lezione nonché alcune simulazioni di prove d'esame, è reso disponibile a tutti gli studenti (frequentanti e non frequentanti) sulla pagina e-learning dedicata al Corso.

Periodo di erogazione dell'insegnamento

Il Corso viene erogato nel primo periodo (sei settimane) del primo semestre dell'Anno Accademico.

Lingua di insegnamento

Italiano

Sustainable Development Goals

ISTRUZIONE DI QUALITÁ