UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Analisi Statistica Multivariata

2122-2-E4101B037

Obiettivi formativi

L'obiettivo del corso è quello di fornire gli strumenti statistici necessari per l'analisi congiunta di più variabili misurate sul medesimo insieme di unità statistiche. Le competenze acquisite nel corso mettono gli studenti in grado di:

- esplorare e sintetizzare i dati;
- modellizzare i dati tramite regressione;
- produrre ed interpretare l'output di analisi di dati reali effettuate tramite il linguaggio R.

Contenuti sintetici

L'insegnamento (15 CFU) è articolato nelle seguenti tre parti:

- **R per l'Analisi Statistica Multivariata (3 CFU)** che fornisce un'introduzione al linguaggio R.
- **Analisi esplorativa (6 CFU)** che presenta i metodi relativi all'esplorazione dei dati al fine di identificare strutture che consentano di ridurne la complessità preservando l'informazione originariamente presente nelle misurazioni.
- **Modelli statistici (6 CFU)** che tratta la specificazione, la stima e la verifica di modelli interpretativi dei dati.

Programma esteso

Prima parte: R per l'Analisi Statistica Multivariata (3 CFU)

- Introduzione al linguaggio R (calcolo scientifico, elementi di programmazione, dataframes)
- Statistica descrittiva in R (statistica univariata e bivariata, rappresentazioni grafiche, analisi descrittiva dei dati)
- Calcolo delle probabilità in R (variabili aleatorie, metodi Monte Carlo)
- Inferenza statistica in R (metodi numerici per l'analisi di verosimiglianza, proprietà degli stimatori)

Seconda parte: Analisi Esplorativa (6 CFU)

- Rappresentazioni grafiche di due o più dimensioni
- Varianza totale e generalizzata
- Il teorema di decomposizione spettrale
- Analisi delle componenti principali
- Analisi dei gruppi: metodo delle K-medie e metodi gerarchici
- Analisi fattoriale

Terza parte: Modelli Statistici (6 CFU)

- Regressione lineare semplice e multipla
- Specificazione del modello
- Stima dei parametri (metodo dei minimi quadrati e di massima verosimiglianza)
- Verifica di ipotesi lineari
- Metodi diagnostici
- Tecniche per la selezione delle variabili
- Previsione

Prerequisiti

Questa attività formativa deve essere preceduta dal superamento degli esami di Algebra lineare, Analisi Matematica I, Calcolo delle Probabilità e Statistica I. Si consiglia inoltre la conoscenza degli argomenti trattati nel corso di Statistica II.

Metodi didattici

Il corso è erogato in italiano e prevede lezioni frontali sia in aula sia in laboratorio informatico.
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.

Modalità di verifica dell'apprendimento

La modalità di verifica si basa su tre prove parziali scritte con orale facoltativo, una per ciascuna parte del corso.

Ciascuna prova scritta prevede sia domande teoriche ed esercizi numerici, che hanno l'obiettivo di verificare l'acquisizione dei concetti e della loro formalizzazione; sia l'analisi di dati tramite il software R.

Il voto finale è determinato dalla media (ponderata con i rispettivi CFU) dei voti riportati nelle tre prove parziali. Qualora lo studente (oppure i docenti) richiedano la prova orale, il voto è una media dei voti di scritto e orale, altrimenti coincide con il voto dello scritto.

Testi di riferimento

Prima parte: R per l'Analisi Statistica Multivariata (3 CFU)

- Appunti delle lezioni forniti dal docente
- Albert, J. & M. Rizzo (2012). *R by Example*. Springer.
- Venables, W. N., Smith D. M. & the R Core Team (2021). *An Introduction to R*.

Seconda parte: Analisi Esplorativa (6 CFU)

- Appunti delle lezioni forniti dal docente
- Johnson, Wichern (2007) Applied Multivariate Statistical Analysis (6th Edition), Pearson Prentice Hall
- Everitt, Hothorn (2011) An Introduction to Applied Multivariate Analysis with R, Springer

Terza parte: Modelli Statistici (6 CFU)

- Appunti delle lezioni fornite dal docente
- M. Grigoletto, F. Pauli, L. Ventura, Modello lineare, teoria e applicazioni con R. Giappichelli, 2017

Periodo di erogazione dell’insegnamento

- **R per l'Analisi Statistica Multivariata (3 CFU) :** I periodo del I semestre
- **Analisi esplorativa (6 CFU) :** II periodo del I semestre
- **Modelli statistici (6 CFU)** : I periodo del II semestre

Lingua di insegnamento

Italiano