

SYLLABUS DEL CORSO

Analisi Esplorativa

2122-2-E4102B084-E4102B084M

Obiettivi formativi

Il modulo di Analisi Esplorativa introduce i principali metodi statistici descrittivi per lo studio di due o più fenomeni osservabili congiuntamente su un insieme di unità statistiche. Si tratta di metodi finalizzati all'esplorazione dei dati multivariati per individuarne la struttura soggiacente e ridurre la dimensionalità in modo da preservare le caratteristiche principali osservate. Dal punto di vista applicativo l'analisi dei dati viene affrontata con il ricorso al software R in ambiente RStudio.

Conoscenza e comprensione. Questo insegnamento fornirà conoscenze e capacità di comprensione relativamente a:

- Principali metodologie esplorative di base dell'analisi statistica multivariata finalizzate al problema della classificazione delle unità statistiche e alla sintesi delle variabili osservate in un numero ridotto di indicatori
- Applicazione dei metodi nella pratica mediante esercizi numerici svolti con la calcolatrice (ossia, senza l'ausilio del software statistico)
- Logica e funzionamento alla base del linguaggio R e suo utilizzo nell'ambito dell'applicazione delle principali analisi statistiche per dati multidimensionali e delle relative rappresentazioni grafiche
- Lettura e interpretazione degli output delle analisi prodotte con R.

Capacità di applicare conoscenza e comprensione. Alla fine dell'insegnamento gli studenti saranno in grado di:

- Scegliere le metodologie esplorative di base dell'analisi statistica multivariata più adeguate in base agli scopi delle analisi e alla natura dei dati a disposizione
- Ridurre la dimensionalità di un dataset aggregando le unità statistiche in gruppi e/o costruendo indicatori di sintesi delle variabili osservate
- Interpretare e confrontare i risultati delle analisi ottenute con metodi diversi per stabilire quale approccio sia

da ritenersi più opportuno in base a specifici criteri fissati a priori

- Importare in R file di dati esterni di varia provenienza e formato e utilizzare in modo autonomo la sintassi di base del linguaggio R.

L'insegnamento consente allo studente di acquisire solide basi teoriche e applicative relativamente ai principali metodi esplorativi dell'analisi di dati multidimensionali necessarie in qualsiasi contesto lavorativo in cui si utilizzino file di dati e che rappresentano una base imprescindibile per il proseguimento del percorso universitario.

Contenuti sintetici

Introduzione all'analisi statistica multivariata, matrici di dati quantitativi, qualitativi e misti, rappresentazioni grafiche per dati multidimensionali. Cluster Analysis: metodi di raggruppamento gerarchici e non gerarchici. Analisi delle componenti principali. Analisi discriminante lineare. Uso integrato dei metodi esplorativi di analisi multivariata. Applicazioni a dati reali con il software R in ambiente RStudio.

Programma esteso

- Introduzione all'analisi statistica multivariata: scuola francese e scuola anglosassone, classificazione delle metodologie di analisi multivariata
- Matrici di dati quantitativi, qualitativi e misti. Principali sintesi e trasformazioni. Rappresentazione dei dati, spazio degli individui e spazio delle variabili. Dissimilarità e distanze fra unità, distanze fra variabili
- Cluster Analysis: metodi di raggruppamento gerarchici e non gerarchici, bontà della classificazione, applicazione a variabili quantitative e qualitative
- Analisi delle componenti principali: estrazione delle componenti principali, criteri di arresto, valutazione della variabilità riprodotta, interpretazione delle componenti principali, applicazioni
- Analisi discriminante lineare: determinazione delle funzioni discriminanti lineari nel caso di due o più popolazioni, metodo alternativo per ricavare le funzioni discriminanti, regole decisionali e valutazione dei risultati, applicazioni
- Uso integrato delle tecniche esplorative di analisi multivariata
- Analisi di casi empirici con RStudio

Prerequisiti

Superamento degli esami degli insegnamenti propedeutici di I anno: Statistica I, Analisi Matematica I, Algebra Lineare, Calcolo delle Probabilità

Metodi didattici

Lezioni teoriche in aula ed esercitazioni pratiche in laboratorio statistico-informatico con il software R in ambiente RStudio.

Modalità di verifica dell'apprendimento

L'esame consiste in una prova scritta (durata complessiva: 2 ore) con tre quesiti (articolati in più punti) che riguardano gli aspetti sia teorici sia applicativi delle metodologie trattate durante il corso. Un quarto quesito è facoltativo e riguarda la programmazione con il software R. _____

La prova orale è facoltativa (su richiesta del docente o dello studente) e riguarda argomenti sia teorici sia pratici. L'accesso alla prova orale è subordinato al superamento della prova scritta con un esito di almeno 18/30. Si fa presente che la prova orale può comportare sia l'aumento, sia il mantenimento, che la diminuzione della valutazione conseguita alla prova scritta.

Testi di riferimento

- Materiale didattico della docente pubblicato sul sito e-learning del corso (ad accesso riservato con password)
- Frosini, B.V. (2014). Complementi di analisi statistica multivariata, EDUCatt, Milano
- Zani, S., Cerioli, A. (2007). Analisi dei dati e data mining per le decisioni aziendali, Giuffrè Editore, Milano
- Gherghi, M., Lauro, C. (2004). Appunti di analisi dei dati multidimensionali. Metodologia ed esempi, RCE Edizioni, Napoli
- Bolasco, S. (1999). Analisi multidimensionale dei dati: strategie e criteri di interpretazione, Carocci, Roma
- Dillon, W.R., Goldstein, M. (1984). Multivariate Analysis, J. Wiley, New York
- Everitt, B.S., Hothorn, T. (2011). An Introduction to Applied Multivariate Analysis with R, Springer, Berlin

Periodo di erogazione dell'insegnamento

I Semestre, II periodo

Lingua di insegnamento

Italiano
