

SYLLABUS DEL CORSO

Analisi Statistica Multivariata

1920-2-E4101B037

Obiettivi formativi

Il Corso ha come scopo quello di fornire metodi statistici per lo studio di due o più fenomeni osservabili congiuntamente sulle unità statistiche di una popolazione.

Le competenze acquisite nel Corso mettono gli studenti in grado di:

- 1) esplorare e sintetizzare i dati
- 2) modellizzare i dati tramite regressione
- 3) 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 pervenire ad una loro "riduzione" che ne evidenzia e preserva le caratteristiche principali.
- **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
- Esplorazione dei dati con grafici
- Sintesi numeriche dei dati
- Associazione e tabelle di contingenza, Paradosso di Simpson
- La probabilità nella vita di tutti i giorni
- Distribuzioni campionarie

Seconda parte: Analisi Esplorativa (6 CFU)

- _____
- Rappresentazioni grafiche di due o più dimensioni
- Interpretazione geometrica della matrice dei dati
- Dati centrati e standardizzati
- Varianza totale e generalizzata
- Il teorema di decomposizione spettrale
- Analisi delle componenti principali
- La distanza di Mahalanobis
- 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 scritte/pratiche (prove parziali, una per ciascuna parte del corso) ed una prova orale finale facoltativa. Le tre prove parziali sono costituite da una prova scritta eseguita in laboratorio.

Al fine di accertare che il livello di preparazione dello studente sia sufficiente, la prova scritta è preceduta da un test a risposta multipla: l'accesso alla prova scritta è consentito solo se il tasso di risposte corrette è almeno pari al 60%.

Testi di riferimento

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

- Appunti delle lezioni forniti dal docente
- A. Agresti, C. Franklin (2016) "Statistica: l'arte e la scienza d'imparare dai dati". Ediz. mylab, Pearson Education Italia

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
- J. Fox, Applied regression analysis and generalized linear models, Sage, 2016
- D. Piccolo, Statistica (Parti III e IV), Il Mulino, Bologna, 2000.
- A. Azzalini, Inferenza Statistica: una Presentazione basata sul Concetto di Verosimiglianza. Milano: Springer-Italia, 2001. (2a edizione)

- 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 ciclo del I semestre
- **Analisi esplorativa (6 CFU)** : Il ciclo del I semestre
- **Modelli statistici (6 CFU)** : I ciclo del II semestre

Lingua di insegnamento

Italiano
