



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Analisi Statistica Multivariata

2425-2-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 parti:

- **Modelli Statistici e R (9 CFU)** che fornisce un'introduzione al linguaggio R e tratta la specificazione, la stima e la verifica di modelli interpretativi dei dati.
- **Analisi esplorativa (6 CFU)** che presenta i metodi relativi all'esplorazione dei dati al fine di identificare strutture che consentano di ridurre la complessità preservando l'informazione originariamente presente nelle misurazioni.

Programma esteso

Prima parte: R per l'Analisi Statistica Multivariata (9 CFU)

- **Introduzione al linguaggio R:** utilizzo di R come calcolatrice scientifica; introduzione agli oggetti e alle loro

classi (vettori, booleani, matrici, data.frame, liste); strutture condizionali e cicli; funzioni.

- **Statistica descrittiva in R:** rivisitazione dei principali argomenti di statistica descrittiva univariata e bivariata applicata a dataset, incluse rappresentazioni grafiche e loro personalizzazioni.
- **Calcolo delle probabilità in R:** principali funzioni per lavorare con variabili casuali; metodi Monte Carlo per approssimare integrali e probabilità.
- **Inferenza statistica in R:** studio delle proprietà degli stimatori mediante simulazioni; metodi numerici per l'analisi di verosimiglianza.
- **Variabili casuali multidimensionali:** funzione di densità e di ripartizione congiunte; marginalizzazione; momenti; vettore delle medie e matrice di varianze e covarianze; v.c. Normale multivariata e sue proprietà.
- **Specificazione del modello:** fasi per la specificazione di un modello statistico; classificazione dei modelli.
- **Modello di regressione lineare semplice:** assunzioni; interpretazione dei parametri; stima dei parametri (minimi quadrati e massima verosimiglianza); proprietà degli stimatori; teorema di Gauss-Markov; indice di determinazione lineare R^2 .
- **Verifica ed utilizzo del modello:** verifiche di ipotesi sul valore di un singolo coefficiente; verifica di ipotesi sulla bontà del modello; utilizzo del modello per fare previsione puntuale ed intervallare.
- **Diagnostica del modello:** metodi per valutare le assunzioni relative alla struttura del modello, agli errori e all'assenza di osservazioni inusuali.
- **Modello di regressione lineare multiplo:** specificazione del modello in forma matriciale e sue assunzioni; interpretazione dei parametri; stima dei parametri (minimi quadrati e massima verosimiglianza); proprietà degli stimatori; teorema di Gauss-Markov; indice di determinazione multipla R^2 .
- **Variabili qualitative:** inserimento di variabili qualitative nel modello mediante l'inserimento di dummy; interazioni.
- **Verifica di un sistema di ipotesi lineare:** teoria generale e casi particolari.
- **Selezione del modello:** contributo assoluto e relativo di una variabile esplicativa; indice di determinazione parziale (IDP); selezione criterion-based delle variabili esplicative tramite approccio backward, forward e stepwise; AIC e BIC.

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

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.

In particolare:

- la parte di **Modelli Statistici e R (9 CFU)** prevede un totale di **73 ore** di lezioni frontali svolte in modalità erogativa in presenza, ciascuna formata da blocchi da 2 o 3 ore, molte delle quali saranno svolte in un'aula informatica. In aggiunta, saranno erogate delle attività di tutorato a supporto di studenti e studentesse.
- la parte di **Analisi Esplorativa (6 CFU)** prevede un totale di **42 ore** di lezioni frontali svolte in modalità erogativa in presenza, di cui 7 svolte in laboratorio informatico. In aggiunta, saranno erogate delle attività di tutorato a supporto di studenti e studentesse in modalità erogativa da remoto sincrona.

Modalità di verifica dell'apprendimento

Per superare l'insegnamento di Analisi Statistica Multivariata è necessario ottenere una valutazione maggiore o uguale a 18 in entrambe le parti che compongono l'insegnamento (Modelli Statistici e R (9 CFU) ed Analisi Esplorativa (6 CFU)). Il voto finale è determinato dalla media (ponderata con i rispettivi CFU) dei voti riportati nelle prove parziali.

Per la parte di **Modelli Statistici e R (9 CFU)**:

- la prova è scritta e articolata in 3 o 4 esercizi. Tali esercizi includono domande teoriche, esercizi di programmazione, classici esercizi scritti e/o analisi di dati reali e prevede anche l'uso di R.
- Sono previste due prove in itinere: la prima riguardante la programmazione in R e la seconda riguardante i modelli lineari.
- 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.
- Studenti e studentesse, così come il docente, possono richiedere una prova orale facoltativa (riguardante l'intero programma da 9 CFU).

La prova d'esame di **Analisi Esplorativa (6CFU)**

- è suddivisa in due parti: una *prima parte scritta* è articolata in 3 quesiti a risposta aperta che includono domande teoriche ed esercizi numerici da risolvere senza l'uso del calcolatore, la *seconda parte* consta di 2 esercizi di analisi dei dati da svolgere con R/RStudio accedendo alla piattaforma degli esami online.
- Studenti e studentesse, così come il docente, possono richiedere una prova orale facoltativa riguardante l'intero programma.
- 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, né di alcun supporto digitale.
- La valutazione delle due parti che compongono la prova di Analisi Esplorativa è proporzionale ai crediti dedicati durante il corso alla parte teorica e alla parte computazionale.

Testi di riferimento

Prima parte: Modelli Statistici e R (9 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](#).
- M. Grigoletto, F. Pauli, L. Ventura, Modello lineare, teoria e applicazioni con R. Giappichelli, 2017
- J. Fox. Applied regression analysis and generalized linear models, third edition. Sage.
- Piccolo, D. (2010), Statistica, Terza edizione, Il Mulino.

Seconda parte: Analisi Esplorativa (6 CFU)

- Appunti delle lezioni forniti dal docente
- Johnson, Wichern (2014) Applied Multivariate Statistical Analysis (6th Edition), Pearson Prentice Hall
- Everitt, Hothorn (2011) An Introduction to Applied Multivariate Analysis with R, Springer

Periodo di erogazione dell'insegnamento

- **Modelli Statistici e R (9 CFU)** : Il periodo del I semestre; I periodo del II semestre
- **Analisi esplorativa (6 CFU)** : I periodo del II semestre

Lingua di insegnamento

Italiano

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
