



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

COURSE SYLLABUS

Statistical Models

2627-2-E4104B010-E4104B010-2

Obiettivi formativi

1. Conoscenza e capacità di comprensione

Al termine dell'insegnamento, lo studente avrà acquisito una solida conoscenza dei principali modelli statistici per l'analisi di dati multivariati, con particolare riferimento alla regressione lineare multipla, al modello lineare classico e alla regressione logistica. Sarà in grado di comprendere il ruolo del modello statistico nell'analisi dei fenomeni reali, distinguendolo dal modello matematico e dagli approcci di machine learning. Conoscerà le principali assunzioni alla base dei modelli studiati, le procedure di stima dei parametri, i criteri di valutazione della bontà di adattamento e le metodologie per l'interpretazione dei risultati. Acquisirà inoltre familiarità con la rappresentazione matriciale dei dati e dei modelli, nonché con i fondamenti teorici dell'inferenza statistica applicata ai modelli di regressione.

2. Conoscenza e capacità di comprensione applicate

Lo studente sarà in grado di applicare i modelli statistici studiati all'analisi di dati provenienti da differenti contesti applicativi, quali quelli economici, aziendali, sociali, biologici, medici e ambientali. Saprà specificare un modello statistico coerente con gli obiettivi dell'analisi, stimarne i parametri, verificarne le assunzioni e valutarne la capacità esplicativa e previsionale. Sarà inoltre in grado di utilizzare il software R per la gestione dei dati, la stima di modelli di regressione lineare e logistica, l'individuazione di problematiche quali multicollinearità e osservazioni anomale, e la produzione di analisi e report basati su dati reali e simulati.

3. Autonomia di giudizio

Durante il percorso formativo, lo studente svilupperà la capacità di valutare criticamente l'adeguatezza di un modello statistico rispetto al fenomeno oggetto di studio. Sarà in grado di analizzare la plausibilità delle assunzioni adottate, interpretare correttamente i risultati delle procedure inferenziali, confrontare modelli alternativi e valutarne punti di forza e limiti. Acquisirà inoltre la capacità di distinguere tra finalità esplicative e previsive dell'analisi statistica, formulando giudizi fondati sull'evidenza empirica e sulla corretta interpretazione dei risultati.

4. Abilità comunicative

Lo studente maturerà la capacità di comunicare in modo chiaro e rigoroso i risultati delle analisi statistiche, utilizzando il linguaggio tecnico proprio della modellistica statistica. Sarà in grado di presentare e discutere modelli

di regressione, risultati inferenziali, misure di bontà di adattamento e indicatori previsionali sia in forma scritta sia orale. Saprà inoltre predisporre report tecnici corredati da tabelle, grafici e commenti interpretativi, adattando il livello di approfondimento alle caratteristiche degli interlocutori.

5. Capacità di apprendere

Il corso mira a sviluppare nello studente la capacità di approfondire autonomamente metodologie statistiche avanzate e di affrontare problemi complessi di analisi dei dati. Le conoscenze teoriche e applicative acquisite costituiranno una base essenziale per insegnamenti successivi nell'ambito della statistica, della data science e dell'analisi quantitativa, nonché per l'utilizzo professionale dei modelli statistici in contesti decisionali e di ricerca. Lo studente sarà inoltre in grado di aggiornare autonomamente le proprie competenze rispetto all'evoluzione degli strumenti metodologici e informatici per l'analisi dei dati.

Contenuti sintetici

Il corso introduce i principali strumenti della modellistica statistica per l'analisi dei dati. Vengono affrontati il concetto di modello statistico e il processo di modellizzazione, i modelli di regressione lineare semplice e multipla e le principali estensioni per variabili risposta di diversa natura.

Il corso tratta le basi dell'inferenza statistica, la stima dei parametri, la valutazione e selezione dei modelli e gli elementi essenziali di diagnostica. Le metodologie sono applicate a dati reali e simulati mediante il software R.

Programma esteso

Il corso introduce i principali concetti e strumenti della modellistica statistica per l'analisi di dati multivariati. Vengono presentati il concetto di modello statistico e il suo ruolo nell'analisi dei fenomeni reali, le differenze rispetto al modello matematico e agli approcci di machine learning, nonché le principali fasi del processo di modellizzazione statistica, dalla formulazione delle ipotesi alla raccolta dei dati, dalla stima dei parametri alla valutazione della capacità esplicativa e previsionale dei modelli.

Successivamente vengono approfonditi i modelli di regressione per lo studio delle relazioni tra variabili quantitative. Sono illustrati i metodi di stima dei parametri, i criteri per la valutazione della bontà di adattamento, le procedure di selezione dei modelli e gli strumenti per la diagnosi delle principali criticità che possono emergere nell'analisi dei dati. Vengono inoltre introdotti elementi di rappresentazione matriciale dei dati e dei modelli statistici.

Particolare attenzione è dedicata agli aspetti inferenziali della modellistica statistica. Si analizzano le proprietà degli stimatori, le ipotesi alla base del modello lineare classico, la costruzione e l'interpretazione di intervalli di confidenza e test di ipotesi, nonché i criteri per il confronto tra modelli alternativi e per la valutazione del loro utilizzo a fini esplicativi e previsivi.

Il corso affronta inoltre alcune estensioni dei modelli lineari, con particolare riferimento all'inclusione di variabili qualitative, alle trasformazioni delle variabili e ai modelli per variabili risposta non continue. Vengono presentati i fondamenti della regressione logistica e l'interpretazione dei relativi parametri e indicatori.

Gli argomenti teorici sono costantemente affiancati da applicazioni pratiche mediante il software R, utilizzando dati reali e simulati per sviluppare competenze operative nell'analisi statistica e nell'interpretazione dei risultati.

Prerequisiti

Si richiede di aver superato gli esami degli insegnamenti propedeutici: **Statistica I, Analisi Matematica I, Algebra Lineare, Calcolo delle Probabilità.**

Per una più agevole comprensione dei contenuti del corso è fortemente consigliato conoscere le nozioni di inferenza statistica impartite al corso di **Statistica II.**

Metodi didattici

Il corso prevede 49 ore di lezione e 12 ore di esercitazioni svolte in presenza. Le attività didattiche sono affiancate da un tutor.

Le lezioni frontali affrontano gli aspetti teorici della modellistica statistica, con il supporto di strumenti informatici e di esempi applicativi sviluppati mediante il software R. Le esercitazioni si svolgono in laboratorio informatico e sono dedicate all'applicazione delle tecniche di analisi statistica su dati reali e simulati.

Durante il corso gli studenti sono guidati nello sviluppo di analisi statistiche, nella scelta delle metodologie più appropriate e nell'interpretazione dei risultati. Il materiale didattico viene reso disponibile sulla piattaforma e-learning prima delle lezioni.

Modalità di verifica dell'apprendimento

L'esame prevede una prova scritta e una prova orale. Non sono previste prove intermedie.

La prova scritta si svolge mediante l'utilizzo del software R e consiste nell'analisi di un insieme di dati forniti dal docente. Lo studente è chiamato a svolgere un'analisi statistica completa, includendo l'applicazione dei modelli studiati, la produzione di grafici e tabelle e la redazione di un report in formato R Markdown. Il report deve contenere l'interpretazione dei risultati e il commento delle principali evidenze emerse dall'analisi.

La prova orale verte sugli argomenti del corso e ha lo scopo di verificare la comprensione teorica dei metodi utilizzati e la capacità di interpretazione critica dei risultati ottenuti.

È inoltre prevista una giornata di presentazione di lavori di gruppo, durante la quale gli studenti illustrano un'analisi svolta autonomamente su dati reali reperiti in modo indipendente, mediante una presentazione in aula. Tale attività contribuisce alla valutazione complessiva del percorso formativo.

Testi di riferimento

I principali testi di riferimento sono:

- J. J. Faraway: *Linear Models with R*, 2nd ed., Chapman & Hall/CRC, 2014.
- G. James, D. Witten, T. Hastie, R. Tibshirani: *An Introduction to Statistical Learning*, 2nd ed., Springer, 2021 (con particolare riferimento ai capitoli dedicati alla regressione lineare e alla regressione logistica, con applicazioni in R).

- Carter Hill, William E. Griffiths, Guay C. Lim: *Principles of Econometrics*, 5th ed., Wiley, 2018.
- D. Nolan, D. T. Lang: *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving*, Chapman & Hall/CRC, 2015.
- Dispense fornite dal docente durante lo svolgimento del corso.

Periodo di erogazione dell'insegnamento

II Semestre, III Ciclo

Lingua di insegnamento

Italiana

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
