



UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Statistica III

2526-3-E4101B035

Obiettivi formativi

Il corso si propone di ampliare e consolidare le conoscenze sui **modelli statistici** introdotte in Analisi Statistica Multivariata, offrendo una trattazione sistematica dei **modelli lineari generalizzati** (GLMs). In particolare, gli obiettivi formativi includono:

- i. fornire una comprensione avanzata della teoria e delle proprietà dei GLMs;
- ii. introdurre le principali tecniche per l'analisi di dati con distribuzioni non gaussiane (es. binomiale, Poisson);
- iii. applicare i modelli mediante il software **R** attraverso esercitazioni pratiche e laboratori;
- iv. discutere problematiche reali quali la errata specificazione del modello, l'inferenza basata sulla verosimiglianza e l'uso delle quasi-verosimiglianze.

Il corso contribuisce al raggiungimento degli obiettivi formativi nell'area di apprendimento del CdS: "**Statistica**". In particolare, fornisce le conoscenze e competenze per individuare e implementare opportuni modelli statistico-probabilistici e trarre le appropriate conclusioni inferenziali.

Contenuti sintetici

Statistica III è un corso monografico sui modelli lineari generalizzati (*Generalized Linear Models*, GLMs). Il programma dettagliato è disponibile nella [pagina web del corso](#). Gli argomenti principali sono:

- Modelli lineari: complementi (recap, errata specificazione, robustezza)
- Modelli lineari generalizzati (teoria e metodi)
- Modelli notevoli (regressione binaria e binomiale, regressione di Poisson)
- Quasi-verosimiglianza.

Programma esteso

- **Modelli lineari ed errata specificazione**
 - Il modello lineare: ripasso e notazione
 - Robustezza dello stimatore ai minimi quadrati
 - Stimatore *sandwich* della varianza
 - Minimi quadrati pesati
 - Trasformazione di Box-Cox, trasformazioni che stabilizzano la varianza
- **Modelli lineari generalizzati (GLM)**
 - Famiglia di dispersione esponenziale
 - Inferenza basata sulla verosimiglianza: stima e verifica d'ipotesi
 - Algoritmo *IRLS*: minimi quadrati pesati ed iterati
 - Devianza, diagnostica, analisi dei residui
 - Tecniche di selezione del modello
 - Implementazione in **R**
- **Modelli notevoli**
 - Regression per dati binari e regressione binomiale
 - Regressione di Poisson per dati di conteggio
- **Quasi-verosimiglianze**
 - Ipotesi al second'ordine
 - Equazioni di stima non distorte
 - Funzione di quasi-verosimiglianza
 - Sovradispersione

Prerequisiti

Questo è un corso di livello triennale, ma ci sono alcuni prerequisiti: si assumano siano già acquisite conoscenze solide dei seguenti argomenti:

- Regressione lineare semplice e statistica descrittiva, ad esempio da [Statistica I](#);
- Statistica inferenziale, ad esempio da *Statistica II*;
- Modelli lineari, ad esempio da *Analisi Statistica Multivariata e Econometria*;
- Software R, ad esempio da [Analisi Statistica Multivariata](#).

Metodi didattici

Le lezioni si svolgono sia in aula che in laboratorio, integrando aspetti di carattere teorico con quelli pratico-applicativi di analisi dei dati e di programmazione in R. Le 47 ore di didattica saranno così suddivise:

- 35 ore di lezione svolte in modalità erogativa in presenza;
- 12 ore di attività di laboratorio.

Modalità di verifica dell'apprendimento

La modalità di verifica si basa su una **prova scritta eseguita in laboratorio** ed una prova orale (facoltativa).

Nella prova scritta sono previste due sezioni: una costituita da **domande teoriche** che hanno l'obiettivo di verificare l'acquisizione dei concetti e della formalizzazione necessari per impostare correttamente un modello statistico; la seconda sezione richiede **l'analisi di un data set tramite software R**. Il voto finale della prova scritta è determinato dalla media dei voti riportati nelle due sezioni. Qualora lo studente (o il docente) richiedano la prova orale, il voto finale è la media dei due voti riportati nelle prove scritta ed orale, altrimenti coincide con l'esito della prova scritta.

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.

Testi di riferimento

Riferimenti principali (in italiano)

- Salvan, A., Sartori, N. and L. Pace. 2020. [Modelli Lineari Generalizzati](#). Springer.
- Azzalini, A. 2008. [Inferenza Statistica: Una Presentazione Basata sul Concetto di Verosimiglianza](#). Springer

Approfondimenti (in Inglese)

- Agresti, A. 2015. [Foundations of Linear and Generalized Linear Models](#). Wiley.
- McCullagh, P., and J. A. Nelder. 1989. [Generalized Linear Models](#). Second Edition. Springer.

Ulteriore materiale didattico verrà messo a disposizione nella [pagina web del corso](#).

Periodo di erogazione dell'insegnamento

Il Corso viene erogato nel primo periodo del primo semestre dell'Anno Accademico.

Lingua di insegnamento

Italiano

Sustainable Development Goals

ISTRUZIONE DI QUALITÀ
