

Metodi statistici per l'analisi di dati gerarchici

SECS-S/03

Conoscenze e abilità da acquisire:

Il corso si propone di introdurre metodologie e tecniche avanzate per il trattamento e l'analisi di dati aventi una struttura gerarchica.

Contenuti:

Parte I. Introduzione

- Richiami dei principali concetti e strumenti per lo studio della variabilità intra e tra gruppi e delle principali caratteristiche dei modelli multilivello.

Parte II. Approfondimenti pratici

- Definizione e caratterizzazione della struttura gerarchica in funzione dell'obiettivo dell'analisi.
- Centratura delle variabili in un contesto multilivello (definizioni, effetti, criticità).

Parte III: Approfondimenti metodologici

- Effetti delle numerosità campionarie dei diversi livelli sull'accuratezza delle stime.
- Estensione in ambito multilivello di modelli e metodi statistici classici, quali Generalized Linear Models, Structural Equation Modeling, Latent Class Analysis, modelli di selezione, etc. La scelta di quali modelli e approcci approfondire sarà valutata anche sulla base del background degli studenti.

Tutte le diverse sezioni prevedono la visione e la discussione collegiale di numerose applicazioni per favorire la comprensione e l'approfondimento dei vari temi trattati.

Attività di apprendimento previste e metodologie di insegnamento:

Sono previste sia lezioni frontali, sia un'intensa attività laboratoriale. Il corso ha una connotazione pratica molto importante, per cui viene richiesta una partecipazione attiva da parte dello studente durante tutte le settimane di lezione. L'approccio didattico prevede infatti attività di lavoro e di approfondimento sui temi del corso in gruppo.

Dati multi-fonte e analisi territoriali

SECS-S/05

Sempre più spesso, le basi di dati statistiche nascono dall'integrazione di archivi diversi, costruiti con dati provenienti da fonti sia di tipo tradizionale (censimenti, dati amministrativi, dati da indagini campionarie) che moderno (web, social network, droni-telecamere).

È quindi opportuno che gli studenti della laurea magistrale di scienze statistiche o anche in vista del loro impegno lavorativo acquisiscano le competenze necessarie per costruire le basi di dati multifonte.

La prima parte di questo corso sarà dedicata alle procedure di linkage, di gestione della selezione, di identificazione e correzione degli errori, di gestione degli outliers. Tali procedure statistiche (deterministiche e probabilistiche) verranno apprese attraverso un'intensa attività laboratoriale, grazie all'analisi di basi di dati reali.

Gran parte delle basi di dati multi-fonte interessa singoli o aggregati dislocati in un territorio. La seconda parte del corso sarà dedicata alla rappresentazione dei dati territoriali, sia mediante

l'acquisizione delle tecnologie GIS, sia attraverso tecniche di cartografia statisticamente determinate. Anche questa parte verrà sviluppata mediante un'intensa attività laboratoriale.

Metodi statistici per dati con elevata dimensionalità

SECS-S/01

- 1) Verifica di ipotesi su larga scala e test multipli: Bonferroni, FWER e FDR e loro controllo, metodi bayesiani empirici e FDR locale
- 2) Modelli di regressione per dati con elevata dimensionalità
 - 2a) dati con numerosità elevata (tall data): algoritmi incrementali a memoria limitata, discesa stocastica del gradiente, inferenza
 - 2b) dati con un numero elevato di variabili: sparsità, penalizzazioni che inducono sparsità, Lasso, inferenza post-selezione
- 3) Stima di matrici di varianza e covarianza ad elevata dimensionalità
 - 3a) modelli fattoriali, modelli che inducono sparsità nella matrice di precisione/concentrazione
 - 3b) grafi di concentrazione, algoritmi di apprendimento basati su calcoli locali
 - 3c) applicazioni in ambito finanziario e bioinformatico

Metodi statistici per dati funzionali

SECS-S/01

- 1) Metodi esplorativi
- 2) Modelli frequentisti
 - lisciamento di funzioni e funzioni vincolate
 - registrazione di curve
 - metodi di riduzione della dimensionalità (componenti principali di funzioni e relative regolarizzazioni)
 - modelli con esplicative funzionali
 - modelli per risposta funzionale ed esplicative multivariate
 - modelli per risposta funzionale ed esplicative funzionali e modelli concorrenti
 - modelli per equazioni differenziali
- 3) Modelli bayesiani
 - - modelli parametrici
 - - modelli non parametrici (BNP)