

Introduzione all'analisi statistica multivariata

Il corso si propone di fornire i concetti introduttivi sottostanti all'analisi statistica multivariata. Ogni partecipante avrà la possibilità di acquisire gli strumenti teorici, ma soprattutto pratici, per condurre autonomamente analisi statistiche di tipo multidimensionali. Più in dettaglio, il fruitore del corso potrà sviluppare le competenze necessarie per a) la formalizzazione di problemi statistici caratterizzati da non elementari gradi di complessità; b) l'estrazione della quota di informazione rilevante presente nei dati; c) la costruzione di modelli uniequazionali per la soluzione di problemi specifici (ad esempio analisi dei dati di indagine, individuazione di valori anomali, qualità del dato).

Particolarmente incoraggiate sono le risorse provenienti dall'area informatica o che, più in generale, siano in possesso di un background non prettamente statistico (ad esempio ingegneri, matematici, economisti, sociologi). Per agevolare la fruizione del corso, sono stati introdotti dei moduli di ripasso di alcuni fondamentali concetti statistici. Infine, il corso ben si adatta a coloro che, a diverso titolo, sono impegnati in problemi inerenti l'analisi e l'interpretazione dei dati di vario tipo, ad esempio di matrice economica o sociale oppure di fonte non convenzionale (Big Data).



Data: 2-4, 9-11 febbraio

Orario: 9.30-13.00

Luogo: Aula Virtuale

Docente: Livio Fenga

Info: Stefania Brandetti

PROGRAMMA

1° Giornata

9.30 – 13.00

- Richiami di base per la corretta fruizione del corso (Indici di sintesi e dispersione, regressione semplice, stimatori e metodi di stima, intervalli di confidenza degli stimatori, distribuzioni statistiche)
- Vettori, matrici, concetto di ortogonalità e metriche
- Teoria dell'informazione applicata alla scelta del modello multivariato
- Indipendenza e correlazione
- Analisi multivariata e regressione multipla

2° giornata

9.30 – 13.00

- Multidimensionalità ed incertezza associata al modello statistico multivariato
- Multicollinearità e selezione dei regressori (selezione forward e backward)
- Modello lineare multiplo: stima dei parametri e test di specificazione.
- L'analisi dei residui e la distribuzione degli errori

3° giornata

9.30 – 13.00

- Identificazione e trattamento di valori anomali.
- Concetto di leverage e dato influente
- Stabilizzazione della varianza e trasformata Box Cox
- Modelli Lineari Generalizzati, aspetti pratici ed applicazioni

4° giornata

9.30 – 13.00

- Cenni all'analisi multivariata per dati indipendenti
- Il problema della riduzione della dimensionalità
- Analisi di due gruppi in caso di violazione della normalità nella distribuzione dei dati
- L'uso dei criteri di informazione per la scelta del modello multivariato

5° giornata

9.30 – 13.00

- Problemi di estrazione dell'informazione con il metodo della regressione
- L'estrazione dell'informazione con grandi matrici di dati
- Criteri Lasso, Ridge, Elastic Net per la data reduction

6° giornata

9.30 – 13.00

- Analisi multivariata assistita da metodi di ricampionamento
- Il problema di data set con gruppi sbilanciati: il package R "ROSE"