

Nome docente	Alessio Pollice
Corso di laurea	Scienze Statistiche
Insegnamento	Statistica Multivariata
Anno accademico	2022-2023
Periodo di svolgimento	I semestre
Crediti formativi universitari (CFU)	10
Settore scientifico disciplinare	SECS/S-01
Pagina web docente	https://www.uniba.it/docenti/pollice-alessio/alessio-pollice

Pre-requisiti

- Elementi di Analisi matematica;
- Elementi di Algebra delle matrici;
- Elementi di Calcolo delle probabilità;
- Elementi di Inferenza statistica.

Conoscenze ed abilità da acquisire

- Comprensione e conoscenza delle basi teoriche e degli sviluppi metodologici relativi all'inferenza multivariata, ai modelli lineari ed alle analisi di dati multidimensionali.
- Familiarità ed autonomia nell'applicazione dei metodi suddetti con l'ausilio del software R.
- Dal punto di vista disciplinare, i risultati di apprendimento attesi coincidono con l'acquisizione di competenze sui diversi capitoli del programma dettagliato.

Programma dettagliato

Prima parte: Inferenza multivariata e modelli lineari.

- Variabili casuali multidimensionali discrete e continue. Indipendenza stocastica. Valori attesi. Matrici di varianze e covarianze. Matrici di covarianze. Funzioni generatrici dei momenti.
- Distribuzione normale multidimensionale. Significato dei parametri. Normale bivariata. Standardizzazione. Funzione generatrice dei momenti. Proprietà della Normale multidimensionale. Distribuzioni di Wishart e di Hotelling.
- Inferenza sui parametri della Normale multidimensionale. Stime di massima verosimiglianza. Distribuzioni campionarie degli stimatori di massima verosimiglianza. Teorema del limite centrale multivariato. Test multivariati: principio di unione-intersezione, rapporto di verosimiglianze generalizzato. Test di Hotelling. Verifica di ipotesi sulla matrice di varianze e covarianze.
- Modello lineare generale. Regressione lineare multipla. Stima dei parametri con il metodo dei minimi quadrati. Proprietà delle stime. Stima dei parametri con il metodo della massima verosimiglianza. Scomposizione della devianza e indice di determinazione. Verifica di ipotesi ed intervalli di confidenza per i coefficienti di regressione. Previsioni tramite il modello lineare. Rimozione delle assunzioni, analisi dei residui, modelli intrinsecamente lineari, scelta delle variabili antecedenti, minimi quadrati generalizzati (eteroschedasticità e autocorrelazione del 1° ordine), multicollinearità, stimatori ridge. Analisi della varianza. Analisi della covarianza.
- Modelli lineari generalizzati. Famiglia esponenziale, funzioni score e total score, famiglia esponenziale canonica. Definizione di MLG e generalità. Stima dei parametri dei MLG (metodi di Newton-Raphson e di scoring), distribuzione campionaria degli stimatori. Verifica dell'adeguatezza del modello. Quasi verosimiglianza.

Seconda parte: Analisi di dati multidimensionali.

- Analisi discriminante. Funzione discriminante lineare di Fisher. Funzione discriminante di massima verosimiglianza. Analisi discriminante bayesiana. Metodo del minimo costo di un'errata classificazione. Stima delle probabilità di classificazione errata (cenni).

- Analisi delle componenti principali. Determinazione delle componenti principali, proprietà campionarie. Problemi applicativi ed interpretazione delle componenti principali. Scelta del numero di componenti principali.
- Analisi della correlazione canonica. Determinazione delle componenti canoniche, proprietà campionarie. Verifica di ipotesi.
- Analisi Fattoriale. Modello fattoriale. Stima del modello: analisi dei fattori principali, metodo della massima verosimiglianza. Rotazione dei fattori. Stima dei punteggi fattoriali: Stimatori di Bartlett e di Thompson.
- Cluster analysis. Matrice di dissomiglianza. Tecniche gerarchiche aggregative e scissorie. Tecniche non gerarchiche. Determinazione del numero dei gruppi.
- Analisi delle Corrispondenze. Corrispondenze semplici, rappresentazione grafica. Cenni all'analisi delle corrispondenze multiple.

Riferimenti bibliografici e materiali didattici

- A. Pollice, Dispense del corso, disponibili in modalità e-learning.
- In modalità e-learning è possibile eseguire i test di autovalutazione e disporre delle dispense del corso, di ulteriori materiali didattici, delle tracce e dei dati utili allo svolgimento delle esercitazioni con il software R. L'indirizzo e la password del corso di Statistica Multivariata in modalità e-learning sono comunicati dal docente all'inizio del corso.

Organizzazione della didattica

- Cicli interni di lezione: No
- Corsi integrativi: No
- Esercitazioni: Si
- Seminari: No
- Attività di laboratorio: Si
- Project work: No
- Visite di studio: No

Modalità di erogazione delle attività formative

- Il corso di Statistica Multivariata fornisce un'introduzione alle tecniche statistiche che permettono lo studio di più caratteri, al fine di fare inferenza sulla loro distribuzione congiunta, studiare la relazione tra caratteri, la somiglianza tra unità statistiche e rappresentare le unità statistiche e/o le variabili in uno spazio di dimensione inferiore a quello originale. Le lezioni sono affiancate da esercitazioni in cui gli argomenti teorici affrontati sono applicati, utilizzando il software statistico R, ad esempi reali e casi di studio.
- Lezioni frontali su argomenti teorici (circa 45 ore) ed esercitazioni sugli stessi argomenti mediante il software R (circa 25 ore). In caso di necessità le lezioni e le esercitazioni potranno essere erogate in modalità di didattica a distanza.
- Materiali didattici, test di autovalutazione ed esercitazioni su piattaforma e-learning.
- I test di autovalutazione (domande a risposta multipla per ciascun capitolo del corso) vengono forniti allo scopo di familiarizzare con le modalità di svolgimento degli esoneri/esami e sono da affrontare individualmente nei limiti di tempo prestabilito. L'esito dei test di autovalutazione contribuisce a migliorare la valutazione complessiva dell'impegno degli iscritti al corso.

Modalità di accertamento delle conoscenze

- Per i frequentanti la valutazione delle attività formative è distribuita nell'arco del semestre e si conclude con la fine del corso. L'esito dei test di autovalutazione e delle attività di laboratorio contribuisce a migliorare la valutazione complessiva del profitto degli iscritti al corso. Concorrono alla valutazione una prova intermedia ed una finale riferite a parti distinte del programma ed entrambe basate su un test con domande a risposta multipla ed un'esercitazione in laboratorio con il software R. L'esonero intermedio e quello finale danno luogo ad una proposta di voto in trentesimi ottenuta come media delle due valutazioni. Alla fine di ciascun esonero segue un breve colloquio orale. Il risultato del colloquio orale andrà ad integrare il voto conseguito negli esoneri.

- I non frequentanti devono sostenere una prova riferita all'intero programma del corso e basata su un test con domande a risposta multipla ed un'esercitazione in laboratorio con il software R. Segue un colloquio orale.