

Nome docente	Alessio Pollice
Corso di laurea	Scienze Statistiche
Anno accademico	2019-2020
Periodo di svolgimento	I semestre
Crediti formativi universitari (CFU)	10
Settore scientifico disciplinare	SECS/S-01

**Programma di Statistica Multivariata (a.a. 2019/2020)
(Prof. Alessio Pollice)**

Università degli Studi di Bari Aldo Moro
Corso di Laurea in Scienze Statistiche

Pre-requisiti

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

Conoscenze ed abilità da acquisire

- Comprensione e conoscenza delle basi teoriche e degli sviluppi metodologici relativi 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.

Programma

- Variabili casuali multidimensionali discrete e continue. Indipendenza stocastica. Valori attesi. Matrici di varianze e covarianze. Matrici di covarianze. Funzioni generatrici dei momenti. Valori attesi condizionati. Trasformazioni biunivoche di variabili casuali multidimensionali.
- 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 (metodo dei minimi quadrati). Proprietà delle stime. Stima dei coefficienti di regressione e della varianza dell'errore (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.
- 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. Metodi non parametrici (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. Dati qualitativi e variabili dicotomiche.
- 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.

Bibliografia

- 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, dei materiali didattici, delle tracce e dei dati utili allo svolgimento delle esercitazioni. 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 congiunto di più caratteri, al fine di rappresentare le unità statistiche e/o le variabili in uno spazio di dimensione inferiore a quello originale, studiare la relazione tra caratteri o la somiglianza tra unità statistiche. 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 ed esercitazioni sugli stessi argomenti mediante il software R.
- Materiali didattici e test di autovalutazione 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 valutazione delle attività formative

- 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, un'esercitazione in laboratorio. L'esonero intermedio e quello finale danno luogo ad una proposta di voto in trentesimi ottenuta come media delle due valutazioni. Questo voto può essere accettato così com'è o messo in discussione con una breve prova orale su tutto il programma del corso. Il risultato del colloquio orale andrà ad integrare il voto conseguito negli esoneri.
- I non frequentanti devono sostenere una prova scritta basata su esercizi e riferita all'intero programma del corso. Segue un colloquio orale.