

**CORSO DI LAUREA IN SCIENZE STATISTICHE**
**ANNO ACCADEMICO 2023 - 2024**
**STATISTICA MULTIVARIATA**

Principali informazioni sull'insegnamento	
Anno di corso	III
Periodo di erogazione	Primo semestre (11/09/2023 – 15/12/2023)
Crediti formativi universitari (CFU/ETCS):	10 CFU
SSD	Statistica, SECS-S/01
Lingua di erogazione	Italiano
Modalità di frequenza	Fortemente consigliata

Docente	
Nome e cognome	Alessio Pollice
Indirizzo mail	<a href="mailto:alessio.pollice@uniba.it">alessio.pollice@uniba.it</a>
Telefono	080 504 9267
Sede	Stanza n. 3, 5° piano
Sede virtuale	Canale MS Teams "Prof. A. Pollice - Ricevimento studenti", codice: y7zenm7
Ricevimento	martedì 9.30 – 11.30, venerdì 9.30 – 11.30, ovvero previo appuntamento

Organizzazione della didattica			
Ore			
Totali	Didattica in presenza	Pratica (laboratorio, campo, esercitazione, altro)	Studio individuale
250	46	24	180
CFU/ETCS			
10			

<b>Obiettivi formativi</b>	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.
<b>Prerequisiti</b>	<ul style="list-style-type: none"> <li>• Elementi di Analisi matematica;</li> <li>• Elementi di Algebra delle matrici;</li> <li>• Elementi di Calcolo delle probabilità;</li> <li>• Elementi di Inferenza statistica.</li> </ul>

<b>Metodi didattici</b>	<ul style="list-style-type: none"> <li>• 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.</li> <li>• Lezioni frontali su argomenti teorici (circa 46 ore) ed esercitazioni sugli stessi argomenti mediante il software R (circa 24 ore). In caso di necessità le lezioni e le esercitazioni potranno essere erogate in modalità di didattica a distanza.</li> <li>• Materiali didattici, test di autovalutazione ed esercitazioni su piattaforma e-learning.</li> <li>• 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.</li> </ul>
-------------------------	--

Risultati di apprendimento previsti	
DD1 Conoscenza e capacità di comprensione	Dal punto di vista disciplinare, i risultati di apprendimento attesi coincidono con l'acquisizione di competenze sui diversi capitoli del programma dettagliato: comprensione e conoscenza delle basi teoriche e degli sviluppi metodologici relativi all'inferenza multivariata, ai modelli lineari ed alle analisi di dati multidimensionali.
DD2 Conoscenza e capacità di comprensione applicate	Familiarità nell'applicazione dei metodi della statistica multivariata all'analisi di dati economici o provenienti da altri contesti applicativi con l'ausilio del software R.
DD3-5 Competenze trasversali	<ul style="list-style-type: none"> <li>• <i>DD3 - Autonomia di giudizio</i> Autonomia nella scelta delle metodologie della statistica multivariata e nella valutazione dei risultati dell'applicazione delle stesse con il software R ad esercizi e casi di studio economici o riferiti ad altri contesti applicativi;</li> <li>• <i>DD4 - Abilità comunicative</i> Abilità nel comunicare ad interlocutori specialisti e non specialisti i contenuti teorici della disciplina, le motivazioni delle scelte da effettuare per lo svolgimento di esercizi ed esempi e l'interpretazione dei risultati delle analisi condotte con il software R;</li> <li>• <i>DD5 - Capacità di apprendere in modo autonomo</i> Autonomia nell'apprendimento dei temi della statistica multivariata e nella pratica di utilizzo del software R allo scopo di valorizzare le abilità e le competenze maturate negli studi successivi e nell'attività lavorativa.</li> </ul>

Contenuti di insegnamento (Programma)	
	<p><u>Prima parte:</u> Inferenza multivariata e modelli lineari.</p> <ul style="list-style-type: none"> <li>• Variabili casuali multidimensionali discrete e continue. Indipendenza stocastica. Valori attesi. Matrici di varianze e covarianze. Matrici di covarianze. Funzioni generatrici dei momenti.</li> <li>• Distribuzione normale multidimensionale. Significato dei parametri. Normale bivariata. Standardizzazione. Funzione generatrice dei momenti. Proprietà della Normale multidimensionale. Distribuzioni di Wishart e di Hotelling.</li> <li>• 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.</li> <li>• 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.</li> <li>• 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 (cenni).</li> </ul> <p><u>Seconda parte:</u> Analisi di dati multidimensionali.</p> <ul style="list-style-type: none"> <li>• Analisi discriminante. Funzione discriminante lineare di Fisher. Funzione discriminante di massima verosimiglianza. Analisi discriminante bayesiana.</li> </ul>

	<p>Metodo del minimo costo di un'errata classificazione. Stima delle probabilità di classificazione errata (cenni).</p> <ul style="list-style-type: none"> <li>• Analisi delle componenti principali. Determinazione delle componenti principali, proprietà campionarie. Problemi applicativi ed interpretazione delle componenti principali. Scelta del numero di componenti principali.</li> <li>• Analisi della correlazione canonica. Determinazione delle componenti canoniche, proprietà campionarie. Verifica di ipotesi.</li> <li>• 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.</li> <li>• Cluster analysis. Matrice di dissomiglianza. Tecniche gerarchiche aggregative e scissorie. Tecniche non gerarchiche. Determinazione del numero dei gruppi.</li> <li>• Analisi delle Corrispondenze. Corrispondenze semplici, rappresentazione grafica. Cenni all'analisi delle corrispondenze multiple.</li> </ul>
Testi di riferimento	<p>A. Pollice, Dispense del corso, disponibili in modalità e-learning. G. James, D. Witten, T. Hastie, R. Tibshirani (2020) <u>Introduzione all'apprendimento statistico con applicazioni in R</u>, Piccin Nuova Libreria Editore.</p>
Note ai testi di riferimento	<p>L'utilizzo di slides o appunti personali è <b>FORTEMENTE SCONSIGLIATO</b> e ritenuto insufficiente per la preparazione.</p>
Materiali didattici	<p>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 all'inizio del corso.</p>

<b>Valutazione</b>	
Modalità di verifica dell'apprendimento	<p>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 ad integrare la valutazione complessiva del profitto. Concorrono alla valutazione un esonero intermedio ed uno finale riferiti a parti distinte del programma ed entrambi basati su un test con domande a risposta multipla ed un'esercitazione in laboratorio con il software R. Alla fine di ciascun esonero segue un breve colloquio orale.</p> <p>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.</p>
Criteri di valutazione	<ul style="list-style-type: none"> <li>• <i>Conoscenza e capacità di comprensione</i> La prima parte degli esoneri riferiti alle due parti del corso consiste in un test con 20 domande a risposta multipla da svolgere in 30 minuti. Per i non frequentanti il test riguarda l'intero programma del corso e comprende 40 domande a risposta multipla da svolgere in 50 minuti.</li> <li>• <i>Conoscenza e capacità di comprensione applicate</i> La seconda parte degli esoneri prevede lo svolgimento di un'esercitazione in cui si richiede di sviluppare in 2 ore una traccia riferita all'analisi di un insieme di dati con il software R. Analogamente, si richiede ai non frequentanti di svolgere in 3 ore l'analisi di un insieme di dati con il software R con riferimento alle metodologie dell'intero programma del corso.</li> <li>• <i>Autonomia di giudizio</i> Allo scopo di valutare l'autonomia di giudizio dei/le candidati/e verranno valutati i commenti contenuti nello svolgimento delle esercitazioni e riferiti <ul style="list-style-type: none"> <li>• alle motivazioni delle scelte effettuate;</li> <li>• all'interpretazione dei risultati dell'applicazione delle metodologie oggetto del programma del corso con il software R.</li> </ul> </li> <li>• <i>Abilità comunicative</i></li> </ul>

	<p>Dopo aver effettuato la consegna dello svolgimento dell'esercizio con R, i/le candidati/e verranno chiamati/e individualmente per una breve discussione pubblica orale basata sulle risposte alle domande a risposta multipla.</p> <p>Allo scopo di valutare le abilità comunicative dei/le candidati/e verranno inoltre considerati i commenti contenuti nello svolgimento delle esercitazioni e riferiti</p> <ul style="list-style-type: none"> <li>• alle motivazioni delle scelte effettuate;</li> <li>• all'interpretazione dei risultati dell'applicazione delle metodologie oggetto del programma del corso con il software R.</li> </ul> <ul style="list-style-type: none"> <li>• <i>Capacità di apprendere in modo autonomo</i> Allo scopo di valutare se i/le candidati/e hanno sviluppato le capacità di apprendimento necessarie per intraprendere studi successivi con un alto grado di autonomia gli/le stessi/e sono interrogati/e individualmente per una breve discussione pubblica orale basata sulle risposte alle domande a risposta multipla.</li> </ul>
<p>Criteria di misurazione dell'apprendimento e di attribuzione del voto finale</p>	<p>La correzione e la valutazione degli svolgimenti dell'esercitazione con R sono effettuate nei giorni successivi allo svolgimento, prestando molta attenzione all'autenticità dei contenuti. Sono annullati automaticamente gli esoneri dei/delle candidati/e i cui svolgimenti riportano frasi o espressioni identiche nell'interpretazione dei risultati o gli stessi errori nei comandi di R. L'annullamento dell'esonero comporta l'obbligo di sostenere l'esame sull'intero programma del corso.</p> <p>Ciascun test dà luogo ad una valutazione in 100esimi. Anche le esercitazioni con R danno luogo a valutazioni in 100esimi. Il risultato del colloquio orale va ad migliorare o peggiorare il punteggio conseguito nelle due prove precedenti. La valutazione dell'esonero intermedio e di quello finale danno luogo ad una proposta di voto in trentesimi ottenuta come sintesi delle due valutazioni. Per i non frequentanti la proposta di voto è formulata sulla base delle valutazioni del test, dell'esercitazione con R e del colloquio orale riferiti all'intero programma del corso.</p>
<p>Altro</p>	