

<b>CORSO DI STUDIO</b>	<b>LAUREA MAGISTRALE IN MATEMATICA (LM-40)</b>
<b>ANNO ACCADEMICO</b>	<b>2023-2024</b>
<b>INSEGNAMENTO</b>	<b>STATISTICA MATEMATICA</b>

Principali informazioni sull'insegnamento	
Anno di corso	Secondo
Periodo di erogazione	Primo semestre (25 settembre 2023 – 22 dicembre 2023)
Crediti formativi universitari (CFU)	7
Settore scientifico disciplinare (SSD)	MAT/06 – Probabilità e Statistica Matematica
Lingua di erogazione	Italiano
Modalità di frequenza	Facoltativa

Docente	
Nome e cognome	Marcello De Giosa
Indirizzo mail	marcello.degiosa@uniba.it
Telefono	+39 080 544 2707
Sede	Dipartimento di Matematica, piano 4, stanza 12.
Sede virtuale	
Pagina web	<a href="https://www.dm.uniba.it/it/members/degiosa">https://www.dm.uniba.it/it/members/degiosa</a>
Ricevimento	Mercoledì ore 15:30-17:30 e su appuntamento.

Organizzazione della didattica				
	Totali	Didattica frontale	Pratica (esercitazioni)	Studio individuale
<b>Ore</b>	175	40	30	105
<b>CFU</b>	7	5	2	

Obiettivi formativi	
	Acquisire conoscenza dei principali metodi della Statistica Matematica, della loro applicazione per risolvere problemi reali, e dei relativi moderni codici del software R.

Prerequisiti	
	Calcolo, calcolo multivariato, algebra lineare, probabilità elementare, come offerti usualmente in un corso di studi in Matematica della classe L-35.

Syllabus	
Contenuti dell'insegnamento (Programma)	<p><b>Parte I. Statistica moderna con uso intensivo del computer.</b></p> <p><b>Introduzione ai dati.</b> Campionamenti: principi e strategie. Esperimenti. Studi osservazionali.</p> <p><b>Analisi dei data esplorativa.</b> Tecniche grafiche per dati categorici e codici R. Tecniche grafiche per dati numerici e codici R. Sommarî numerici. Caso studio con R.</p> <p><b>Modello di regressione.</b> Regressione lineare con un predittore. Regressione lineare con più predittori. Regressione logistica. Laboratorio con R.</p> <p><b>Fondamenti di Inferenza.</b> Verifica di ipotesi con randomizzazione. Intervalli di fiducia con il bootstrap. Inferenza con modelli matematici. Errori di decisione. Caso studio con R: vaccino della malaria.</p>



**Inferenza Statistica.** Stima di una proporzione, test con bootstrap e modelli matematici. Confronto di due proporzioni, test con randomizzazione, intervalli di fiducia bootstrap, modelli matematici. Inferenza per tabelle a due vie, test di indipendenza con randomizzazione e modelli matematici. Stima di una media, intervalli di fiducia con bootstrap e modelli matematici. Confronto di due medie: test di randomizzazione per la differenza, intervalli di fiducia bootstrap per la differenza, modelli matematici. Confronto di medie appaiate: test di randomizzazione per la differenza, intervalli di fiducia bootstrap per la differenza, modelli matematici. Confronto di più medie: test di randomizzazione e modelli matematici. Caso studio e tutorial con codici R.

**Inferenza for regressione lineare con un predittore.** Caso studio. Test di randomizzazione e intervalli di fiducia bootstrap per la pendenza. Modello matematico per test di ipotesi e intervalli di fiducia per la pendenza. Verifica delle ipotesi del modello.

**Inferenza for regressione lineare con più predittori.** L'output del software, la multicollinearità, e le cross-validation per l'errore di previsione.

**Inferenza for regressione logistica.** Diagnostica per ipotesi del modello. L'output del software, le cross-validation per l'errore di previsione.

#### **Parte II. Approfondimenti matematici dell'approccio modellistico.**

**Rivisitazione del modello lineare.** I concetti chiave nella stima dell'età dell'universo: stima puntuale ai minimi quadrati, intervalli di fiducia e test di ipotesi, per i parametri. Il modello lineare in generale. Codici R. Tre utili teoremi fondamentali: distribuzioni chi-quadro, t-student, Fisher.

**La Teoria del modello lineare.** Stima ai minimi quadrati con decomposizione QR. Distribuzione degli stimatori. Stima della varianza. Dalla decomposizione QR ai test di ipotesi ed intervalli di fiducia per i parametri. La distribuzione di Fisher attraverso la decomposizione di Cholesky per test di ipotesi sui parametri.

Dalla decomposizione QR alla distribuzione di Fisher nell'Analisi della Varianza (Anova) sequenziale per confrontare modelli annidati.

Dalla decomposizione QR alla matrice di influenza. Distribuzioni dei vettori dei valori adattati e dei residui.

I risultati in termini di matrice del modello.

Esempio di uso dei risultati con codici R su dati reali (Galapagos).

Il teorema di Gauss-Markov.

Interpretazione geometrica di stima ai minimi quadrati, della decomposizione QR, e di modelli annidati.

**Modelli lineari in pratica con codici R.** I comandi base, plot diagnostici, R-quadro, selezione del modello, l'Akaike Information Criteria (AIC), previsioni, collinearità.

**Modelli lineari in pratica con variabili fattore.** Il problema dell'identificabilità. Fattori multipli ed interazioni.

**Specifiche in R di modelli lineari.**

**Approfondimenti di teoria dei modelli lineari.** Vincoli, contrasti e fattori. Verosimiglianza. AIC e statistica di Mallows.

**Pratica con i modelli lineari e codici R.** Il modello lineare semplice. Esempi su dati reali. Stima della media e previsione per una nuova osservazione. Il modello multiplo. Collinearità ed Inflazione della varianza. Selezione del modello: AIC, test di ipotesi, test di Fisher. Interpretazione dei coefficienti. Un esempio di analisi di soddisfazione dei clienti di un parco divertimenti.

	<p><b>Rivisitazione della Regressione Logistica.</b> Distribuzione Binomiale e Bernoulli. Odds, logit, logistica, odds-ratio. Predittore binario. Predittore continuo. Interpretazione dei parametri. Selezione del modello. Rappresentazione grafica del modello.</p> <p><b>Regressione Logistica multipla.</b> Massima verosimiglianza e minimi quadrati iterativamente ripesati (IRLS). La devianza. AIC. Test di rapporto di verosimiglianza (LRT). Residui. Stima del parametro di dispersione.</p> <p><b>Modelli Lineari Generalizzati (GLM).</b> Due esempi introduttivi: epidemie e animali predatori. La famiglia esponenziale di distribuzioni. Media e varianza di distribuzioni esponenziali. Formulazione generale di GLM. Stima dei parametri e distribuzioni asintotiche degli stimatori. Paragonare modelli con AIC e devianza. Residui.</p> <p><b>GLM in pratica con R.</b> Un modello binomiale per probabilità di attacco cardiaco. Regressione Poisson come modello di epidemie.</p>
Testi di riferimento	<p>S.N. Wood - <i>Generalized Additive Models</i> - 2017 - CRC Press</p> <p>M. Cetinkaya-Rundel, J. Hardin - <i>Introduction to Modern Statistics</i> - 2023 - OpenIntro</p> <p>R.S. Kenett, S. Zacks, P. Gedeck - <i>Modern Statistics. A computer-based approach with Python</i> - 2022 - Birkhauser</p> <p>A. Agresti, C.A. Franklin, B. Klingenberg - <i>Statistics. The art of learning from data.</i> - 2023 - Pearson</p> <p>A. Agresti, M. Kateri - <i>Foundations of Statistics for Data Scientists</i> - 2021 - CRC Press</p> <p>G. Casella, R.L. Berger – <i>Statistical Inference</i> – 2002 - Duxbury</p>
Note ai testi di riferimento	
Materiali didattici	

<b>Risultati di apprendimento previsti (secondo i Descrittori di Dublino)</b>	
DD1 Conoscenza e capacità di comprensione	Acquisire la comprensione dei concetti chiave, delle tecniche e dei codici R per la Statistica Matematica
DD2 Conoscenza e capacità di comprensione applicate	Abilità nell'applicare i risultati teorici ed i codici R per risolvere problemi reali con la Statistica Matematica.
DD3-5 Competenze trasversali	<i>DD3 Autonomia di giudizio:</i> Essere capaci di scegliere gli strumenti della Statistica Matematica ed i giusti codici R.
	<i>DD4 Abilità comunicative:</i> Essere capaci nel documentare e spiegare i metodi della Statistica Matematica ed i risultati delle analisi statistiche.
	<i>DD5 Capacità di apprendere:</i> Essere capaci di utilizzare i metodi di apprendimento per migliorare la conoscenza nel campo della Statistica Matematica.

<b>Metodi didattici</b>	
	Lezioni teoriche frontali e laboratori R con uso di diapositive.

<b>Valutazione</b>	
Modalità di verifica dell'apprendimento	L'esame finale si compone di una parte orale e di una parte di laboratorio R.
Criteri di valutazione	<ul style="list-style-type: none"> <li><i>Conoscenza e capacità di comprensione:</i> adeguata padronanza degli aspetti teorici ed applicativi della materia, quanto dei codici R presentati a lezione.</li> <li><i>Conoscenza e capacità di comprensione applicate:</i> essere capaci nell'applicare i metodi della Statistica Matematica ed i relativi codici R per risolvere problemi pratici reali.</li> </ul>



	<ul style="list-style-type: none"><li>• <i>Autonomia di giudizio</i>: autonomia nella scelta degli strumenti teorici e pratici per la soluzione di problemi reali e nella scelta dei codici R adeguati.</li><li>• <i>Abilità comunicative</i>: abilità nella presentazione dei fondamenti teorici dei metodi applicati e dei risultati di un'analisi statistica.</li><li>• <i>Capacità di apprendere</i>: Organizzare la conoscenza acquisita ed essere autonomi nel trovare fonti di approfondimento.</li></ul>
Criteria di misurazione dell'apprendimento e di attribuzione del voto finale	L'esame consiste di una parte orale sui metodi visti a lezione, e di una parte di discussione di alcuni codici R, scelti a caso tra quelli presentati a lezione. Il voto finale, in trentesimi, dipenderà dalla padronanza degli aspetti teorici ed applicativi quanto dalla capacità di interpretare codici R, ma anche dalla capacità di spiegare chiaramente le conclusioni..

Ulteriori informazioni	