



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Scienze Economiche, Aziendali e Statistiche
ANNO ACCADEMICO OFFERTA	2021/2022
ANNO ACCADEMICO EROGAZIONE	2021/2022
CORSO DILAUREA MAGISTRALE	STATISTICA E DATA SCIENCE
INSEGNAMENTO	PROGRAMMAZIONE AVANZATA IN R PER DATI AD ALTA DIMENSIONALITÀ
TIPO DI ATTIVITA'	C
AMBITO	21031-Attività formative affini o integrative
CODICE INSEGNAMENTO	21977
SETTORI SCIENTIFICO-DISCIPLINARI	SECS-S/01
DOCENTE RESPONSABILE	AUGUGLIARO LUIGI Professore Ordinario Univ. di PALERMO
ALTRI DOCENTI	
CFU	6
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	108
NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA	42
PROPEDEUTICITA'	
MUTUAZIONI	
ANNO DI CORSO	1
PERIODO DELLE LEZIONI	2° semestre
MODALITA' DI FREQUENZA	Facoltativa
TIPO DI VALUTAZIONE	Voto in trentesimi
ORARIO DI RICEVIMENTO DEGLI STUDENTI	AUGUGLIARO LUIGI Martedì 10:00 12:00 Stanza n. 201 - secondo piano

DOCENTE: Prof. LUIGI AUGUGLIARO

PREREQUISITI	Lo studente deve possedere conoscenza dei concetti della Statistica Inferenziale. Deve inoltre possedere conoscenza del linguaggio di programmazione R.
RISULTATI DI APPRENDIMENTO ATTESI	<p>Conoscenza e capacita' di comprensione:</p> <ul style="list-style-type: none">- della programmazione funzionale in R- degli algoritmi per l'analisi dei dataset ad alta dimensionalità;- delle procedure e degli algoritmi propri della teoria dei test multipli;- della teoria e degli algoritmi propri dell'inferenza sparsa; <p>Capacita' di applicare conoscenza e comprensione</p> <ul style="list-style-type: none">- Capacità di utilizzo del linguaggio di programmazione R per l'applicazione delle metodologie acquisite durante il corso;- Capacità di implementazione degli algoritmi proposti nelle riviste scientifiche internazionali;- Capacita' di affrontare problemi concreti con i metodi acquisiti durante le lezioni frontali; <p>Autonomia di giudizio</p> <p>Lo studente deve essere in grado di utilizzare criticamente i metodi visti durante il corso. Lo studente deve inoltre saper implementare, nel linguaggio di programmazione R, algoritmi di ottimizzazione per l'analisi dei dati ad alta dimensionalità.</p> <p>Abilita' comunicative</p> <p>Essere in grado di esporre oralmente quanto imparato durante il corso modulando il linguaggio e il formalismo a seconda del destinatario.</p> <p>Capacita' di apprendimento</p> <p>Essere in grado di consultare la letteratura scientifica nazionale e internazionale; distinguere i testi a seconda dell'ambito applicativo e del loro contenuto statistico, rielaborare quanto appreso attraverso l'adattamento alle condizioni e ai limiti imposti dal tipo di problema da risolvere.</p>
VALUTAZIONE DELL'APPRENDIMENTO	<p>L'esame finale consisterà nella preparazione di un rapporto tecnico, individuale o di gruppo (al più 3 studenti), basato sull'implementazione di una metodologia statistica per l'analisi di un dataset con il linguaggio di programmazione R, cui segue un colloquio orale per presentare e commentare i risultati ottenuti. La prova orale mira ad accertare l'apprendimento dello studente. Questa consisterà in almeno due domande finalizzate a graduare la valutazione delle conoscenze, competenze, abilita' e trasversalita' con gli argomenti dei corsi seguiti precedentemente, possedute dallo studente, nonche' la sua capacita' di trasmetterle con idoneo linguaggio statistico. La soglia della sufficienza sarà raggiunta quando lo studente avrà mostrato conoscenza e comprensione degli argomenti almeno nelle linee generali (definizione dei concetti). Quanto più, invece, l'esaminando avrà dato evidenza delle sue capacita' argomentative ed espositive, nonche' di proprieta' di linguaggio statistico, tanto più la valutazione sarà positiva.</p> <p>METODO DI VALUTAZIONE FINALE</p> <p>La valutazione finale dell'esame prenderà in considerazione tre aspetti: i) la padronanza degli argomenti; ii) la capacita' di applicazione delle conoscenze e iii) la proprieta' di linguaggio. La Commissione graderà ciascun aspetto secondo le modalità "Assente", "Scarso", "Sufficiente", "Buono", e "Ottimo". Pertanto il metodo di valutazione sarà:</p> <p>Insufficiente: se almeno due "Assente" e nessun "Ottimo" 18 - 20: se almeno due "Sufficiente" e nessun "Ottimo" 21 - 24: se almeno due "Buono" e nessun "Ottimo" 25 - 27: se 1 "Ottimo" 28 - 30: se 2 "Ottimo" 30 e lode: se tre "Ottimo"</p> <p>Il range dei voti consentirà al docente di tenere conto dei fattori di contesto dell'esame (come ad esempio la partecipazione attiva durante le lezioni e le esercitazioni, oppure la presenza di qualche disabilità).</p>
OBIETTIVI FORMATIVI	Il corso è finalizzato a fornire allo studente le conoscenze dei metodi di base per l'analisi dei dati ad alta dimensionalità. Enfasi verrà posta sugli algoritmi derivanti dalla teoria dei test multipli e sugli algoritmi derivanti dalla teoria dell'inferenza sparsa.
ORGANIZZAZIONE DELLA DIDATTICA	Lezioni frontali, esercitazioni in laboratorio informatico.
TESTI CONSIGLIATI	<p>Advanced R, Second Edition. H. Wickham (2019). CRC Press.</p> <p>Simultaneous Statistical Inference: With Applications in the Life Sciences. T. Dickhaus (2014). Springer-Verlag.</p>

Statistical Learning with Sparsity: The Lasso and Generalizations. T. Hastie, R. Tibshirani and M. Wainwright (2015). CRC Press.

PROGRAMMA

ORE	Lezioni
14	Programmazione avanzata in R: gli oggetti di base, funzioni, programmazione orientata agli oggetti, programmazione funzionale, funzionali.
4	Teoria e algoritmi delle procedure per i test multipli applicati ai dati ad alta dimensionalità: misura e controllo degli errori (FWER e FDR), teoria del p-value, procedure marginali per i test multipli, procedure step-down/up per il controllo dei FWER e FDR.
6	Inferenza statistica per modelli sparsi: relazione tra teoria dei test multipli e i criteri di informazione, penalizzazione lasso per modelli lineari generalizzati, metodi di ottimizzazione, selezione del parametro di tuning.

ORE	Esercitazioni
6	Programmazione avanzata in R: applicazione della programmazione funzionale in R
6	Teoria dei test multipli per dati ad alta dimensionalità: applicazione in R
6	Inferenza statistica per modelli sparsi: applicazione in R