



# UNIVERSITÀ DEGLI STUDI DI PALERMO

<b>DIPARTIMENTO</b>	Scienze Economiche, Aziendali e Statistiche
<b>ANNO ACCADEMICO OFFERTA</b>	2022/2023
<b>ANNO ACCADEMICO EROGAZIONE</b>	2023/2024
<b>CORSO DILAUREA</b>	STATISTICA PER L'ANALISI DEI DATI
<b>INSEGNAMENTO</b>	MODELLI LINEARI
<b>TIPO DI ATTIVITA'</b>	A
<b>AMBITO</b>	50244-Statistico - probabilistico
<b>CODICE INSEGNAMENTO</b>	20580
<b>SETTORI SCIENTIFICO-DISCIPLINARI</b>	SECS-S/01
<b>DOCENTE RESPONSABILE</b>	CHIODI MARCELLO      Professore Ordinario      Univ. di PALERMO
<b>ALTRI DOCENTI</b>	
<b>CFU</b>	9
<b>NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE</b>	141
<b>NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA</b>	84
<b>PROPEDEUTICITA'</b>	19596 - INFERENZA STATISTICA
<b>MUTUAZIONI</b>	
<b>ANNO DI CORSO</b>	2
<b>PERIODO DELLE LEZIONI</b>	2° semestre
<b>MODALITA' DI FREQUENZA</b>	Facoltativa
<b>TIPO DI VALUTAZIONE</b>	Voto in trentesimi
<b>ORARIO DI RICEVIMENTO DEGLI STUDENTI</b>	<b>CHIODI MARCELLO</b> Martedì    15:00    17:00    stanza del docente (edificio 13); eccezionalmente su teams Venerdì    12:00    13:00    stanza del docente (edificio 13); eccezionalmente su teams

DOCENTE: Prof. MARCELLO CHIODI

<b>PREREQUISITI</b>	inferenza statistica, calcolo delle probabilità, matematica e algebra lineare
<b>RISULTATI DI APPRENDIMENTO ATTESI</b>	<p>Conoscenza e capacità di comprensione</p> <p>Distinzione dei diversi tipi di correlazione (semplice parziale multipla);</p> <p>Conoscenza del linguaggio e della terminologia dei modelli di dipendenza.</p> <p>Comprensione della unicità di approccio all'analisi della dipendenza da variabili quantitative e qualitative</p> <p>Comprensione dell'analisi della verosimiglianza per la costruzione di test e stimatori e dei diversi tipi di scomposizione della devianza e del concetto di varianza "spiegata"</p> <p>Conoscenza delle tecniche di selezione dei modelli e di analisi dei residui</p> <p>Capacità di applicare conoscenza e comprensione</p> <p>Essere in grado di applicare la metodologia dei modelli di dipendenza a diversi casi di dati reali, provenienti da contesti diversi e con modalità diverse di acquisizione dei dati</p> <p>Capacità di distinzione fra diverse tipologie di matrici del disegno in funzione della natura dei dati e delle variabili.</p> <p>Capacità di impiego del software specialistico (R) per l'analisi dei modelli lineari e non</p> <p>Autonomia di giudizio</p> <p>Essere in grado di produrre risultati in relazione e al fenomeno sostantivo</p> <p>Giudicare l'utilità dei modelli impiegati per analizzare relazioni tra variabili e analizzare criticamente i risultati utilizzando tecniche di analisi dei residui.</p> <p>Capacità di effettuare un processo di selezione di un modello statistico di dipendenza, integrando con le informazioni provenienti dall'analisi dei residui</p> <p>Capacità di giudicare il software da utilizzare per l'implementazione di particolari tecniche</p> <p>Abilità comunicative</p> <p>Essere in grado di redigere un rapporto tecnico</p> <p>scegliere le rappresentazioni grafiche più opportune in relazione ai problemi;</p> <p>esporre verbalmente il contenuto del rapporto sia in forma sintetica che in forma estesa, in funzione delle conoscenze statistiche del destinatario</p> <p>Capacità d'apprendimento</p> <p>Essere in grado di consultare la letteratura scientifica sull'argomento; capacità di apprendere le estensioni dei modelli studiati a lezione; capacità di apprendimento di software statistico specialistico anche diverso da quello impiegato in aula.</p>
<b>VALUTAZIONE DELL'APPRENDIMENTO</b>	<p>Esame scritto con PC ed esame orale subordinato al superamento della prova scritta. La valutazione terrà conto di entrambe le prove</p> <p>La Commissione giudicatrice sarà formata secondo la normativa vigente; in funzione del calendario di esami può prevedere una prova in itinere, che viene eventualmente comunicata agli studenti almeno 3 settimane prima dello svolgimento dello stesso.</p> <p><b>PROVA SCRITTA</b></p> <p>La prova scritta (anche in itinere) mira a rilevare le conoscenze, le competenze, le abilità possedute dallo studente e la sua capacità di riportarle in un elaborato scritto che tenga conto anche della proprietà di linguaggio statistico posseduto.</p> <p>La prova, della durata massima di 2 ore, prevede usualmente 2 problemi ciascuno basato su un insieme di dati multivariato diverso. Almeno uno dei problemi riguarderà l'applicazione di un modello lineare o GLM.</p> <p>Per ciascuno dei problemi vengono posti dei quesiti specifici. In ogni caso lo studente è normalmente incoraggiato ad utilizzare tutti gli strumenti esplorativi e descrittivi a sua disposizione. Le situazioni poste sono di livello di difficoltà analoga a quelle affrontate in aula, ma normalmente non riguardano situazioni casi di studio veri e propri, ma piuttosto insiemi di dati tratti dalla realtà e riadattati ai fini dell'esame.</p> <p>Lo studente utilizzerà un Pc con il software R e scriverà il compito su carta (mantenendo comunque sul pc traccia delle proprie elaborazioni); nella stesura dell'elaborato lo studente dovrà commentare tutti i risultati principali in modo sintetico, ma con proprietà di linguaggio. In generale la semplice esecuzione dei comandi R per ottenere dei risultati non commentati non è considerata sufficiente, ma è semplicemente un prerequisito.</p> <p>Nei commenti ci si aspetta che lo studente descriva sinteticamente i metodi impiegati e giustifichi le scelte effettuate, senza addentrarsi in dimostrazioni o in approfondimenti teorici</p> <p>(L'eventuale prova in itinere sarà ovviamente calibrata sulla porzione di programma fino a quel momento svolta)</p> <p>La soglia di sufficienza per il superamento della prova scritta consiste nell'uso adeguato dei termini relativi ai concetti di base e nello svolgimento della maggior parte dei quesiti. Lo svolgimento di uno solo dei due problemi</p>

	<p>normalmente non e' giudicato sufficiente. La proprieta' di linguaggio sara' giudicata, nell'attribuzione generale del voto, insieme con l'adeguata spiegazione delle tecniche adottate</p> <p>La presenza di errori di calcolo (molto improbabili visto che si lavora con R e su dati gia' ampiamente controllati dal docente) non pregiudica in generale l'esito della prova, purché commenti e azioni siano coerenti</p> <p><b>PROVA ORALE</b>  la prova orale mira alla valutazione dell'apprendimento complessivo dello studente.  L'esame orale inizia sempre con un commento al compito e con qualche domanda di chiarimento dei docenti su quanto scritto dallo studente: Questa fase contribuisce alla valutazione complessiva.  La prova orale procede poi con almeno due domande, poste dai due componenti della commissione, su argomenti teorici 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.</p> <p>La soglia della sufficienza della prova orale sara' raggiunta quando lo studente abbia mostrato conoscenza e comprensione degli argomenti almeno nelle linee generali (definizione dei concetti) e abbia competenze applicative minime, consistenti nella esemplificazione di casi concreti semplici (tipicamente inerenti agli argomenti trattati nella parte iniziale). Quanto piu, invece, l'esaminando abbia brillantemente superato la prova scritta e abbia dato evidenza, nella prova orale, delle sue capacita' argomentative ed espositive, nonche' di proprieta' di linguaggio statistico, tanto piu' la valutazione sara' positiva.</p> <p>In ogni caso la valutazione tiene conto sia della prova scritta che della prova orale (e dell'eventuale prova in itinere). La soglia fra il superamento e il non superamento dell'esame complessivo tiene conto dell'acquisizione delle conoscenze minime previste fra gli obiettivi formativi</p>
<b>OBIETTIVI FORMATIVI</b>	<p>Obiettivi formativi del corso</p> <p>Il corso ha come obiettivo fondamentale di offrire allo studente gli strumenti fondamentali per la costruzione e l'analisi di modelli statistici di dipendenza e di relazioni fra piu' variabili utilizzando l'opportuno software statistico.</p> <p>L'attenzione sara' incentrata sui modelli di dipendenza lineare, sulla misura della dipendenza lineare fra piu' variabili, e sulle estensioni ai modelli lineari generalizzati</p>
<b>ORGANIZZAZIONE DELLA DIDATTICA</b>	<p>Lezioni frontali, Esercitazioni e laboratorio in aula di informatica</p> <p>Durante il corso, il docente condividera' con gli studenti un breve articolo, un capitolo di libro, o una sua parte in lingua inglese di carattere divulgativo, che sara' oggetto di dibattito con gli studenti.</p>
<b>TESTI CONSIGLIATI</b>	<p>Draper-Smith, Regression Analysis, John Wiley</p> <p>Per maggiori approfondimenti verranno indicate le fonti bibliografiche piu' utili durante il corso.</p> <p><b>Materiale didattico</b>  Il materiale didattico (appunti, software, casi di studio, dati e codice R) sara' a disposizione degli studenti sul web</p> <p>Verra' usato un software didattico ossia il package R MLANP che insieme a vari notebook didattici è disponibili sul sito del docente <a href="http://www.marcellochiodi.com">www.marcellochiodi.com</a> o sulle pagine istituzionali</p>

## PROGRAMMA

ORE	Lezioni
2	Introduzione al corso e problemi reali introduttivi;
6	variabili statistiche multiple; definizione della matrice dei dati; momenti primi e secondi (multivariati) di una variabile statistica multipla e di combinazioni lineari di variabili statistiche multipla Correlazione fra piu' variabili: Correlazione parziale Analisi delle componenti principali
4	la distribuzione normale bivariata; funzioni di regressione; distribuzione normale multivariata (cenni); combinazioni lineari di variabili normali; risultati principali sulle distribuzioni condizionate;
4	Funzioni di regressione e dipendenza in media: modello lineare come approssimazione del problema generale della dipendenza in media; regressione non parametrica Modelli lineari: versatilita' del modello lineare generale; componente sistematica e accidentale; la matrice delle x; osservazioni ripetute; disegni fattoriali; regressione multipla; regressione polinomiale, regressori del tipo 0/1, modelli con interazioni;

## PROGRAMMA

ORE	Lezioni
6	assunzioni di base nel modello lineare; la verosimiglianza del modello lineare: costruzione del test LR e degli stimatori ML; minimi quadrati ordinari; distribuzione campionaria di $b$ , scomposizione della devianza nel modello lineare e verifica di ipotesi. test F per la verifica di ipotesi nel modello lineare; configurazioni della matrice X e di $X'X$ modello lineare: verifica di ipotesi generali; intervalli di confidenza per $E(Y)$ ; errori di previsione.
6	regressione multipla; scomposizione della devianza empirica e coefficiente di determinazione lineare multipla $R^2$ ; prova di ipotesi particolari; test per l'eliminazione di $q$ regressori; la multicollinearità Scelta delle variabili; confronto fra modelli; AIC; errori di previsione e cross validation.
4	Violazione delle assunzioni di base nella regressione e nel modello lineare: analisi dei residui; proprietà dei residui empirici; grafici impiegati per l'analisi dei residui.
6	stima dei parametri del modello lineare con fattori qualitativi; analisi della varianza ad una via; assunzioni di base; divergenza dalla linearità per fattori quantitativi, differenza fra i test di omogeneità; analisi della varianza a due vie; significato delle interazioni; analisi della varianza con variabili concomitanti: l'analisi della covarianza. allontanamento dalle assunzioni di base nel modello lineare e nell'analisi della varianza; analisi dei residui nell'analisi della varianza.
10	Modelli lineari generalizzati: funzioni legame e famiglia di distribuzioni esponenziale. Regressione logistica, Regressione di Poisson e regressione gamma. Stima dei parametri e scelta del modello
ORE	Esercitazioni
8	variabili statistiche multiple; definizione della matrice dei dati; momenti primi e secondi (multivariati) di una variabile statistica multipla e di combinazioni lineari di variabili statistiche multipla Correlazione fra più variabili: Correlazione parziale Analisi delle componenti principali
2	la distribuzione normale bivariata; funzioni di regressione; distribuzione normale multivariata (cenni); combinazioni lineari di variabili normali; risultati principali sulle distribuzioni condizionate;
14	Modelli lineari: impostazioni, stima parametri, inferenza, scelta variabili con R
6	regressione logistica di poisson regressione gamma con R