



# UNIVERSITÀ DEGLI STUDI DI PALERMO

<b>DIPARTIMENTO</b>	Scienze Economiche, Aziendali e Statistiche
<b>ANNO ACCADEMICO OFFERTA</b>	2018/2019
<b>ANNO ACCADEMICO EROGAZIONE</b>	2019/2020
<b>CORSO DILAUREA</b>	STATISTICA PER L'ANALISI DEI DATI
<b>INSEGNAMENTO</b>	STATISTICA MULTIVARIATA
<b>TIPO DI ATTIVITA'</b>	A
<b>AMBITO</b>	50244-Statistico - probabilistico
<b>CODICE INSEGNAMENTO</b>	19595
<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>	147
<b>NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA</b>	78
<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

<b>PREREQUISITI</b>	statistica 2, matematica e algebra lineare
<b>RISULTATI DI APPRENDIMENTO ATTESI</b>	<p>Conoscenza e capacita' di comprensione  Distinzione dei diversi tipi di correlazione (semplice parziale multipla);  Conoscenza del linguaggio e della terminologia dei modelli di dipendenza.  Comprensione della unicità di approccio all'analisi della dipendenza da variabili quantitative e qualitative  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"  Conoscenza delle tecniche di selezione dei modelli e di analisi dei residui</p> <p>Capacita' di applicare conoscenza e comprensione  Essere in grado di applicare la metodologia dei modelli di dipendenza a diversi casi di dati reali, provenienti da contesti diversi e con modalita' diverse di acquisizione dei dati  Capacita' di distinzione fra diverse tipologie di matrici del disegno in funzione della natura dei dati e delle variabili.  Capacita' di impiego del software specialistico (R) per l'analisi dei modelli lineari e non</p> <p>Autonomia di giudizio  Essere in grado di produrre risultati in relazione e al fenomeno sostantivo  Giudicare l'utilita' dei modelli impiegati per analizzare relazioni tra variabili e analizzare criticamente i risultati utilizzando tecniche di analisi dei residui.  Capacita' di effettuare un processo di selezione di un modello statistico di dipendenza, integrando con le informazioni provenienti dall'analisi dei residui  Capacita' di giudicare il software da utilizzare per l'implementazione di particolari tecniche</p> <p>Abilita' comunicative  Essere in grado di redigere un rapporto tecnico  scegliere le rappresentazioni grafiche piu' opportune in relazione ai problemi;  esporre verbalmente il contenuto del rapporto sia in forma sintetica che in forma estesa, in funzione delle conoscenze statistiche del destinatario  Capacita' d'apprendimento  Essere in grado di consultare la letteratura scientifica sull' argomento; capacita' di apprendere le estensioni dei modelli studiati a lezione; capacita' 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 terra' conto di entrambe le prove</p> <p>La Commissione giudicatrice sara' formata secondo la normativa vigente; in funzione del calendario di esami puo' 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>  La prova scritta (anche in itinere) mira a rilevare le conoscenze, le competenze, le abilita' possedute dallo studente e la sua capacita' di riportarle in un elaborato scritto che tenga conto anche della proprieta' di linguaggio statistico posseduto. 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 riguardera' l'applicazione di un modello lineare o GLM.  Per ciascuno dei problemi vengono posti dei quesiti specifici. In ogni caso lo studente e' normalmente incoraggiato ad utilizzare tutti gli strumenti esplorativi e descrittivi a sua disposizione. Le situazioni poste sono di livello di difficolta' analoga a quelle affrontate in aula, ma normalmente non riguardano situazioni casi di studio veri e propri, ma piuttosto insiemi di dati tratti dalla realta' e riadattati ai fini dell'esame.  Lo studente utilizzerà un Pc con il software R e scrivera' il compito su carta (mantenendo comunque sul pc traccia delle proprie elaborazioni); nella stesura dell'elaborato lo studente dovra' commentare tutti i risultati principali in modo sintetico, ma con proprieta' di linguaggio. In generale la semplice esecuzione dei comandi R per ottenere dei risultati non commentati non e' considerata sufficiente, ma e' semplicemente un prerequisito.  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  (L'eventuale prova in itinere sara' 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, nonché la sua capacita' di trasmetterle con idoneo linguaggio statistico.  In base alla delibera del Consiglio di CdS L-41 del 28/09/2015, la prova orale prevede anche domande su uno di due testi in lingua inglese, atta a valutare la comprensione del testo e la capacita' di formulazione di una risposta. La prima vertera' sulla traduzione di una frase, la seconda sara' una domanda di carattere generale sul testo. I testi in lingua inglese oggetto di esame sono indicati nella sezione "Testi consigliati"  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, nonché di proprieta' di linguaggio statistico e di padronanza della lingua inglese, tanto piu' la valutazione sara' positiva.  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>
<b>ORGANIZZAZIONE DELLA DIDATTICA</b>	Lezioni frontali, Esercitazioni e laboratorio in aula di informatica
<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>Materiale didattico  Il materiale didattico (appunti, software, casi di studio, dati e codice R) sara' a disposizione degli studenti sul web <a href="http://dssm.unipa.it/chiodi">http://dssm.unipa.it/chiodi</a>.</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;
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.

## PROGRAMMA

<b>ORE</b>	<b>Lezioni</b>
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
<b>ORE</b>	<b>Esercitazioni</b>
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 $R$