



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Scienze Economiche, Aziendali e Statistiche
ANNO ACCADEMICO OFFERTA	2023/2024
ANNO ACCADEMICO EROGAZIONE	2025/2026
CORSO DILAUREA	STATISTICA PER L'ANALISI DEI DATI
INSEGNAMENTO	STATISTICA SOCIALE 2
TIPO DI ATTIVITA'	B
AMBITO	50250-Statistico, statistico applicato, demografico
CODICE INSEGNAMENTO	06705
SETTORI SCIENTIFICO-DISCIPLINARI	SECS-S/05
DOCENTE RESPONSABILE	ATTANASIO MASSIMO Professore Ordinario Univ. di PALERMO
ALTRI DOCENTI	
CFU	9
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	141
NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA	84
PROPEDEUTICITA'	
MUTUAZIONI	
ANNO DI CORSO	3
PERIODO DELLE LEZIONI	1° semestre
MODALITA' DI FREQUENZA	Facoltativa
TIPO DI VALUTAZIONE	Voto in trentesimi
ORARIO DI RICEVIMENTO DEGLI STUDENTI	ATTANASIO MASSIMO Mercoledì 10:00 12:00 dseas ed 13 studio 104

PREREQUISITI	Il corso prevede la conoscenza della statistica descrittiva e inferenziale, di basi della matematica e del calcolo delle probabilità e di SAS. Il corso di Statistica Sociale 1 è propedeutico.
RISULTATI DI APPRENDIMENTO ATTESI	<p>Conoscenza e capacita' di comprensione</p> <p>Acquisizione:</p> <ol style="list-style-type: none"> 1. del linguaggio proprio dei dati longitudinali; 2. conoscenze di base per l'analisi dei dati di sopravvivenza. <p>Capacita' di applicare conoscenza e comprensione</p> <p>Essere in grado di:</p> <ol style="list-style-type: none"> 1. riconoscere la natura dei dati (individuale o aggregato, longitudinale o istantaneo, spaziale o puntuale, censurato o completo, ecc.) e individuare la possibilita' di acquisirli; 2. leggere studi sulla sopravvivenza (o di event history) e descrivere dati sanitari ed epidemiologici, sia attraverso indici, sia attraverso modelli statistici tipici della biostatistica e dell'epidemiologia. 3. essere in grado di usare SAS a un livello intermedio; riconoscimento tipologia di dati; saper usare le proc freq, univariate; sort; means; logistic; lifetest; phreg. <p>Autonomia di giudizio</p> <p>Essere in grado di:</p> <ol style="list-style-type: none"> 1. fornire una lettura critica dei risultati ottenuti in relazione alla tipologia di dati (dati da indagini o sperimentali; dati longitudinali o istantanei; ecc), e all'ambito disciplinare (sociale, medico-sanitario, educativo ecc.); 2. descrivere un data set di dati di sopravvivenza e di saper commentare in maniera elementare i risultati di un modello di Cox. 3. essere in grado di valutare le potenzialita' di SAS in problemi di analisi di dati reali <p>Abilita' comunicative</p> <p>Essere in grado di:</p> <ol style="list-style-type: none"> 1. conoscere gli elementi principali per redigere un rapporto scritto su uno studio di sopravvivenza; 2. esporre oralmente i risultati di un'analisi di sopravvivenza a seconda del destinatario dello studio, anche non esperto in statistica. 3. essere in grado di colloquiare con esperti non statistici su temi statistici di sopravvivenza <p>Capacita' d'apprendimento</p> <p>Essere in grado di:</p> <ol style="list-style-type: none"> 1. consultare la letteratura nazionale ed internazionale di base; 2. aumentare le conoscenze acquisite nel corso frequentando corsi di livello superiore; 3. distinguere i testi a seconda del loro contenuto statistico nei diversi ambiti di applicazione; 4. rielaborare quanto appreso attraverso l'adattamento alle condizioni e ai limiti imposti all'eventuale committente e dal tipo di problema da risolvere; 5. valutare l'uso degli strumenti informatici piu' comuni (excel, R e SAS) in termini comparativi
VALUTAZIONE DELL'APPRENDIMENTO	<p>La valutazione dell'apprendimento avviene attraverso una prova scritta e un colloquio orale, subordinato al superamento della prova scritta. La prova in itinere sara' stabilita a discrezione del docente e in accordo con gli studenti frequentanti.</p> <p>PROVA SCRITTA</p> <p>La prova scritta (anche in itinere) mira a rilevare le conoscenze e le abilita' possedute dallo studente, anche informatiche (Excel e SAS), e la sua capacita' di commentare i dati reali. La prova, della durata totale massima di 3 ore prevede da 4 a 5 quesiti di natura pratica e/o teorica, a risposta aperta. Per la prova in itinere il tempo sara' di 90 minuti e avra' la medesima struttura. La soglia di sufficienza (18-21) e' raggiunta se lo studente mostra un utilizzo sufficiente delle conoscenze di base in termini esecutivi. La soglia 21-24 e' raggiunta quando i quesiti sono svolti tutti ma presentano imprecisioni di calcolo e interpretazione. La soglia 24-27 viene assegnata quando il compito e' svolto in ogni sua parte, dimostrando una conoscenza approfondita (sa dimostrare proprieta' con un adeguato formalismo), mentre la votazione da 28 a 30 e lode viene raggiunta quando alle caratteristiche precedenti si aggiunge la capacita' di spaziare e generalizzare con altri ambiti della statistica ed e' in grado di leggere i dati, le formule e i risultati in termini statistico-matematici e in termini sostanziali.</p> <p>Nelle prove sara' richiesta la conoscenza di SAS, da saper applicare per l'analisi di dati reali</p> <p>PROVA ORALE</p> <p>La prova orale mira ad approfondire l'elaborato scritto e la valutazione dell'apprendimento dello studente. Questa consistera' in almeno due domande finalizzate a valutare la capacita' dello studente di formulare giudizi sia qualitativi che quantitativi sui concetti fondamentali del corso.</p>

	<p>La prova orale consta anche di almeno due domande sul linguaggio SAS relativo alle procedure e funzioni previste dal programma, nonché del linguaggio di programmazione base di SAS.</p> <p>L'insegnamento si considera superato se lo studente abbia anche mostrato conoscenze, capacità e abilità utili almeno a impostare la risoluzione dei quesiti posti e con non rilevanti lacune relative alla sintassi SAS</p> <p>METODO DI VALUTAZIONE FINALE La valutazione finale dell'esame prenderà in considerazione tre aspetti: i) la padronanza degli argomenti; ii) la capacità di applicazione delle conoscenze e iii) la proprietà di linguaggio, valutati nel complesso della prova scritta e orale. Il docente avrà anche la possibilità 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) ai fini della determinazione dell'esito della prova. Entrambe le prove (scritta e orale) sono valutate in trentesimi e si ritengono superate con una votazione minima di 18/30.</p>
OBIETTIVI FORMATIVI	<p>Lo studente alla fine del corso dovrebbe essere in grado di: 1. riconoscere e descrivere le caratteristiche essenziali della regressione logistica a risposta dicotomica e polinomica; 2. riconoscere e descrivere le caratteristiche essenziali dei dati di sopravvivenza con gli "strumenti statistici di sopravvivenza più comuni"; 3. riconoscere e interpretare in termini matematici, probabilistici e statistici gli oggetti più comuni (funzione hazard, funzione di sopravvivenza e hazard cumulato); 4. elaborare e interpretare in maniera elementare un modello di regressione di Cox. Inoltre deve essere in grado di utilizzare fogli elettronici e il software statistico SAS.</p>
ORGANIZZAZIONE DELLA DIDATTICA	<p>lezioni frontali ed esercitazioni. Attività di laboratorio anche con il supporto del software SAS per l'analisi di dati reali.</p> <p>Durante il corso, il docente condividerà con gli studenti un breve articolo, un capitolo di libro, o una sua parte in lingua inglese di carattere divulgativo, che sarà oggetto di dibattito con gli studenti.</p>
TESTI CONSIGLIATI	<p>Agresti A. (1990) An Introduction to Categorical Data Analysis, Wiley (pp.1 e 2.1)</p> <p>Collett D. (1999), Modelling binary data. Chapman & Hall. (capp 3 e 7)</p> <p>Stokes M., Davis C.S., Koch G.G (2012) Categorical Data Analysis Using SAS®, Third Edition,(cap.9) SAS Institute</p> <p>Collett D. (1994), Modelling survival data in medical research, Chapman & Hall. (capp.1,2,3 e 5.1 e 5.2)</p> <p>Kleinbaum (1996) Survival Analysis – A Self learning text, Springer (capp 1,2,3 e 4)</p> <p>Attanasio M. Dispense sull'aggregazione di categorie e collapsamento di variabili.</p> <p>Dispense di Statistica Sanitaria (a cura di F.Aiello, M.Attanasio e L.Librizzi). Università di Palermo.</p>

PROGRAMMA

ORE	Lezioni
2	Introduzione al corso. Illustrazione del programma: obiettivi
2	Introduzione alle tipologie di dati (prospettivo contro retrospettivo, aggregato contro individuale, survey contro sperimentale, longitudinali contro puntuali), Principali misure di associazione in dati rappresentati in tabelle di contingenza. Data analysis
4	La rappresentazione dei dati multidimensionali. Richiami modello di regressione logistica.
2	Costruzione del dato attraverso un'indagine e utilizzo di dati rilevati da altri (indagine fondamentali in campo medico, sociale, demografico)
2	Data analysis e tecniche di categorizzazione
2	Interazione: definizione, forme funzionali e analisi grafica.
2	Bontà di adattamento di un modello di regressione logistica: calcolo della deviance. Applicazioni. Software: SAS
7	Modelli multinomiali. Proportional odds. Partial proportional Odds. Applicazioni in campo sociale e medico. Software: SAS
2	Analisi dei dati di sopravvivenza: introduzione e caratteristiche
6	I principali oggetti dell'analisi dei dati di sopravvivenza. La funzione di sopravvivenza, hazard, hazard cumulato (nel continuo e nel discreto) e la speranza di vita.
8	Metodi non parametrici (lo stimatore attuariale e lo stimatore di Kaplan-Meier), Test per il confronto di due funzioni di sopravvivenza. Test log rank e Gehan
9	Metodi semiparametrici (metodo grafico e per la verifica della proporzionalità tra i rischi, il modello di Cox, costruzione della funzione di verosimiglianza parziale, interpretazione dei coefficienti di regressione, della baseline e della funzione di sopravvivenza).

ORE	Esercitazioni
8	calcolo verosimiglianza parziale con 1 variabile dummy con excel. Massimizzazione con newton raphson con excel
2	Tecniche per l'aggregazione di categorie: metodo di Irwin-Lancaster, statistiche test X2 e G2
2	Interazione e aggregazioni di categorie. Analisi grafica per l'individuazione dell'interazione. software: excel + sas
4	Applicazioni al modello di regressione logistica (stima dei parametri, deviance e bonta' adattamento). Software: excel e R
3	Modello di regressione logistica, modelli multinomiale e modelli loglineari. Software: SAS
4	Calcolo funzione di sopravvivenza, hazard, hazard cumulato (nel continuo e nel discreto) e speranza di vita. Software: excel e SAS
7	Verifica grafica e analitica proporzionalita' tra i rischi per l'applicazione del modello di Cox, costruzione della funzione di verosimiglianza parziale, interpretazione dei coefficienti di regressione, della baseline e della funzione di sopravvivenza. Software: Excel, R e SAS
6	Lo stimatore attuariale e lo stimatore di Kaplan-Meier. Calcolo SE. Calcolo Hazard puntuale e intervallare. calcolo hazard cumulato. Test per il confronto di due funzioni di sopravvivenza. Test log rank e Gehan. Software: Excel, SAS e R