



# UNIVERSITÀ DEGLI STUDI DI PALERMO

<b>DIPARTIMENTO</b>	Scienze Economiche, Aziendali e Statistiche
<b>ANNO ACCADEMICO OFFERTA</b>	2017/2018
<b>ANNO ACCADEMICO EROGAZIONE</b>	2019/2020
<b>CORSO DILAUREA</b>	STATISTICA PER L'ANALISI DEI DATI
<b>INSEGNAMENTO</b>	STATISTICA SOCIALE 2
<b>TIPO DI ATTIVITA'</b>	B
<b>AMBITO</b>	50250-Statistico, statistico applicato, demografico
<b>CODICE INSEGNAMENTO</b>	06705
<b>SETTORI SCIENTIFICO-DISCIPLINARI</b>	SECS-S/05
<b>DOCENTE RESPONSABILE</b>	ATTANASIO MASSIMO Professore Ordinario Univ. di PALERMO
<b>ALTRI DOCENTI</b>	
<b>CFU</b>	8
<b>NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE</b>	128
<b>NUMERO DI ORE RISERVATE ALLA DIDATTICA ASSISTITA</b>	72
<b>PROPEDEUTICITA'</b>	06649 - STATISTICA 2 18159 - STATISTICA SOCIALE I C.I.
<b>MUTUAZIONI</b>	
<b>ANNO DI CORSO</b>	3
<b>PERIODO DELLE LEZIONI</b>	1° semestre
<b>MODALITA' DI FREQUENZA</b>	Facoltativa
<b>TIPO DI VALUTAZIONE</b>	Voto in trentesimi
<b>ORARIO DI RICEVIMENTO DEGLI STUDENTI</b>	ATTANASIO MASSIMO Mercoledì 10:00 12:00 dseas ed 13 studio 104

<b>PREREQUISITI</b>	Il corso prevede la conoscenza della statistica descrittiva e inferenziale, di basi della matematica e del calcolo delle probabilità. Il corso di Statistica Sociale 1 è propedeutico.
<b>RISULTATI DI APPRENDIMENTO ATTESI</b>	<p>Conoscenza e capacita' di comprensione</p> <p>Acquisizione:</p> <ol style="list-style-type: none"> <li>1. del linguaggio proprio dei dati longitudinali;</li> <li>2. delle tipologie di indagine e del tipo di dato prodotto;</li> <li>3. conoscenze di base per l'analisi dei dati di sopravvivenza.</li> </ol> <p>Capacita' di applicare conoscenza e comprensione</p> <p>Essere in grado di:</p> <ol style="list-style-type: none"> <li>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;</li> <li>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.</li> </ol> <p>Autonomia di giudizio</p> <p>Essere in grado di:</p> <ol style="list-style-type: none"> <li>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.);</li> <li>2. descrivere un data set di dati di sopravvivenza e di saper commentare in maniera elementare i risultati di un modello di Cox.</li> </ol> <p>Abilita' comunicative</p> <p>Essere in grado di:</p> <ol style="list-style-type: none"> <li>1. conoscere gli elementi principali per redigere un rapporto scritto su una semplice tecnica di campionamento o su uno studio di sopravvivenza;</li> <li>2. esporre oralmente i risultati di un'analisi di sopravvivenza a seconda del destinatario dello studio, anche non esperto in statistica.</li> <li>3. essere in grado di colloquiare con esperti non statistici su temi statistici di sopravvivenza</li> </ol> <p>Capacita' d'apprendimento</p> <p>Essere in grado di:</p> <ol style="list-style-type: none"> <li>1. consultare la letteratura nazionale ed internazionale di base;</li> <li>2. aumentare le conoscenze acquisite nel corso frequentando corsi di livello superiore;</li> <li>3. distinguere i testi a seconda del loro contenuto statistico nei diversi ambiti di applicazione;</li> <li>4. rielaborare quanto appreso attraverso l'adattamento alle condizioni e ai limiti imposti all'eventuale committente e dal tipo di problema da risolvere.</li> </ol>
<b>VALUTAZIONE DELL'APPRENDIMENTO</b>	<p>La valutazione dell'apprendimento avviene attraverso una prova scritta e un colloquio orale, subordinato al superamento della prova scritta. La Commissione giudicatrice sara' presieduta dal docente titolare dell'insegnamento. La prova in itinere sara' stabilita a discrezione del docente e in accordo con gli studenti frequentanti. Essa copre in genere la prima parte del programma sui modelli logistici.</p> <p><b>PROVA SCRITTA</b></p> <p>La prova scritta (anche in itinere) mira a rilevare le conoscenze e le abilita' possedute dallo studente anche informatiche (Excel e altri software utilizzati) e la sua capacita' di commentare dei 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.</p> <p>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 e all'orale il candidato dimostra 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><b>PROVA ORALE</b></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. Infine, in base alla delibera del Consiglio di CdS L-41 del 28/09/2015, la prova orale prevede anche due 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</p>

	<p>generale sul testo.</p> <p>I testi in lingua inglese oggetto di esame sono indicati nella sezione "Testi consigliati in inglese"</p>
<b>OBIETTIVI FORMATIVI</b>	<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 politomica; 2. riconoscere e descrivere le caratteristiche essenziali dei dati di sopravvivenza con gli "strumenti statistici di sopravvivenza piu' comuni"; 3. riconoscere e interpretare in termini matematici, probabilistici e statistici gli oggetti piu' 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 R e SAS.</p>
<b>ORGANIZZAZIONE DELLA DIDATTICA</b>	lezioni ed esercitazioni
<b>TESTI CONSIGLIATI</b>	<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 &amp; 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 &amp; 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). Universita' di Palermo.</p> <p>Testi in inglese consigliati. Collett D. (1994), Modelling survival data in medical research, Chapman &amp; Hall. (capp.1,2)</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	Bonta' 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 proporzionalita' 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
2	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
1	Interazione e aggregazioni di categorie. Analisi grafica per l'individuazione dell'interazione.
2	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
2	Calcolo funzione di sopravvivenza, hazard, hazard cumulato (nel continuo e nel discreto) e speranza di vita. Software: excel e R
6	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 e R

<b>ORE</b>	<b>Esercitazioni</b>
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 e R