



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Scienze Economiche, Aziendali e Statistiche		
ANNO ACCADEMICO OFFERTA	2018/2019		
ANNO ACCADEMICO EROGAZIONE	2020/2021		
CORSO DILAUREA	STATISTICA PER L'ANALISI DEI DATI		
INSEGNAMENTO	STATISTICA SOCIALE 2 C.I.		
CODICE INSEGNAMENTO	19840		
MODULI	Si		
NUMERO DI MODULI	2		
SETTORI SCIENTIFICO-DISCIPLINARI	SECS-S/05		
DOCENTE RESPONSABILE	ATTANASIO MASSIMO	Professore Ordinario	Univ. di PALERMO
ALTRI DOCENTI	ATTANASIO MASSIMO	Professore Ordinario	Univ. di PALERMO
	GIAMBALVO ORNELLA	Professore Ordinario	Univ. di PALERMO
CFU	15		
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 GIAMBALVO ORNELLA Martedì 10:00 12:00 Il servizio prenotazione ricevimento è sospeso. Per fissare un appuntamento con la docente si prega di inviare una mail all'indirizzo ornella.giambalvo@unipa.it Mercoledì 12:00 13:00 Il servizio prenotazione ricevimento è sospeso. Per fissare un appuntamento con la docente si prega di inviare una mail all'indirizzo ornella.giambalvo@unipa.it		

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	Conoscenza e capacita' di comprensione Acquisizione: 1. delle tipologie di indagine e del tipo di dato prodotto; 2. del linguaggio proprio dei dati longitudinali; 3. conoscenze di base per l'analisi dei dati di sopravvivenza. Capacita' di applicare conoscenza e comprensione Essere in grado di: 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 (import/export dati; riconoscimento tipologia di dati; utilizzo delle proc di uso comune e relative alle indagini campionarie) Autonomia di giudizio Essere in grado di: 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 provenienti da indagini campionarie Abilita' comunicative Essere in grado di: 1. conoscere gli elementi principali per redigere un rapporto scritto su una semplice tecnica di campionamento o 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 Capacita' d'apprendimento Essere in grado di: 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	La valutazione dell'apprendimento avviene attraverso una prova scritta e un colloquio orale, subordinato al superamento della prova scritta. La Commissione giudicatrice sara' presieduta dai docenti dei due moduli. La prova in itinere sara' stabilita a discrezione del docente e in accordo con gli studenti frequentanti. PROVA SCRITTA 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 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. In entrambe le prove dei due moduli sara' richiesta la conoscenza di SAS, da saper applicare per l'analisi di dati reali PROVA ORALE 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 generale sul testo. I testi in lingua inglese oggetto di esame sono indicati nella sezione "Testi consigliati in inglese". METODO DI VALUTAZIONE FINALE DEL MODULO La valutazione finale dell'esame di ciascuno dei due moduli prendera' in considerazione tre aspetti: i) la padronanza degli argomenti; ii) la capacita' di applicazione delle conoscenze e iii) la proprieta' di linguaggio, valutati nel complesso della prova scritta e orale. Il docente avra' anche la possibilita' 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 disabilita') 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. Il voto risultante per il singolo modulo e' dato dalla media aritmetica semplice delle due prove. VALUTAZIONE FINALE DELL'INSEGNAMENTO La votazione sara' in trentesimi, risultante dalla media aritmetica ponderata dei voti conseguiti nei due moduli (peso 9 per Statistica Sociale 2 e peso 6 per Indagini Campionarie).
ORGANIZZAZIONE DELLA DIDATTICA	lezioni frontali ed esercitazioni. Attivita' di laboratorio anche con il supporto del software SAS per l'analisi di dati reali

MODULO STATISTICA SOCIALE 2

Prof. MASSIMO ATTANASIO

TESTI CONSIGLIATI

Agresti A. (1990) An Introduction to Categorical Data Analysis, Wiley (pp.1 e 2.1)
 Collett D. (1999), Modelling binary data. Chapman & Hall. (capp 3 e 7)
 Stokes M., Davis C.S., Koch G.G (2012) Categorical Data Analysis Using SAS®, Third Edition,(cap.9) SAS Institute
 Collett D. (1994), Modelling survival data in medical research, Chapman & Hall. (capp.1,2,3 e 5.1 e 5.2)
 Kleinbaum (1996) Survival Analysis – A Self learning text, Springer (capp 1,2,3 e 4)
 Attanasio M. Dispense sull'aggregazione di categorie e collapsamento di variabili.
 Dispense di Statistica Sanitaria (a cura di F.Aiello, M.Attanasio e L.Librizzi). Universita' di Palermo.
 Testi in inglese consigliati. Collett D. (1994), Modelling survival data in medical research, Chapman & Hall. (capp.1,2)

TIPO DI ATTIVITA'	B
AMBITO	50250-Statistico, statistico applicato, demografico
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	147
NUMERO DI ORE RISERVATE ALLE ATTIVITA' DIDATTICHE ASSISTITE	78

OBIETTIVI FORMATIVI DEL MODULO

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 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 SAS

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
4	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

2	Applicazioni al modello di regressione logistica (stima dei parametri, deviance e bontà d'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 di proporzionalità 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

MODULO INDAGINI CAMPIONARIE E SONDAGGI DEMOSCOPICI

Prof.ssa ORNELLA GIAMBALVO

TESTI CONSIGLIATI

Cochran (1977) Sampling Techniques. Wiley & Sons. Cap da 1 a 5.

Dispense del corso

Boscaïno G., Giambalvo O. (2014) Sampling, in Probability and Statistics: A Didactic Introduction, pagg. 176-209. Ed. Jose' I. Barraganes; Adolfo Morais; Jenaro Guisasola, by CRC press Taylor & Francis Group. ISBN 9781482219777

Testo per la prova in inglese: Cochran (1977) Sampling Techniques. Wiley & Sons. Cap.1 - par. 1 e par. 2

TIPO DI ATTIVITA'	B
AMBITO	50250-Statistico, statistico applicato, demografico
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	98
NUMERO DI ORE RISERVATE ALLE ATTIVITA' DIDATTICHE ASSISTITE	52

OBIETTIVI FORMATIVI DEL MODULO

Il corso ha come principale obiettivo lo studio delle procedure per la raccolta dei dati di un'indagine campionaria, la stima di un parametro incognito della popolazione e il calcolo dell'errore standard. Si punterà l'attenzione anche sulla scelta della migliore dimensione campionaria per ottenere il risultato più preciso. Gli studenti, alla fine del corso, devono essere in grado di costruire gli strumenti base per la raccolta dei dati, condurre un'indagine campionaria semplice in funzione dei vincoli posti. Lo studente acquisirà maggiore familiarità con l'uso del foglio di calcolo Excel e SAS.

PROGRAMMA

ORE	Lezioni
1	Introduzione al modulo. Descrizione degli obiettivi formativi, del programma e delle modalità della prova finale.
1	Fondamenti delle indagini campionarie e dei sondaggi d'opinione. Cenni storici.
6	Il campionamento da popolazione finita. Principali piani di campionamento probabilistico: campionamento casuale semplice. Stima della media, della proporzione e del totale. Determinazione della numerosità campionaria. Il campionamento casuale semplice con il SAS.
15	Il campionamento casuale stratificato. Stima della media, della proporzione e del totale. Determinazione della numerosità campionaria nei casi sopra citati. Il campionamento stratificato con il SAS.
4	Confronto fra il campione casuale semplice e stratificato nelle allocazioni uguale, proporzionale e ottimo.
5	Tecniche di raccolta dei dati nelle indagini campionarie: il questionario e la scheda di intervista. Costruzione e somministrazione.

ORE	Esercitazioni
6	Campionamento casuale semplice. Stima della media, della proporzione e del totale. Determinazione della numerosità campionaria. Calcolo degli stimatori e delle varianze degli stimatori con il SAS.
14	Campionamento casuale stratificato. Stima della media, della proporzione e del totale. Determinazione della numerosità campionaria. Calcolo degli stimatori e delle varianze con il SAS.