



UNIVERSITÀ DEGLI STUDI DI PALERMO

DEPARTMENT	Scienze Economiche, Aziendali e Statistiche		
ACADEMIC YEAR	2018/2019		
BACHELOR'S DEGREE (BSC)	STATISTICS FOR DATA ANALYSIS		
INTEGRATED COURSE	SOCIAL STATISTICS 2 - INTEGRATED COURSE		
CODE	19840		
MODULES	Yes		
NUMBER OF MODULES	2		
SCIENTIFIC SECTOR(S)	SECS-S/05		
HEAD PROFESSOR(S)	ATTANASIO MASSIMO	Professore Ordinario	Univ. di PALERMO
OTHER PROFESSOR(S)	ATTANASIO MASSIMO	Professore Ordinario	Univ. di PALERMO
	GIAMBALVO ORNELLA	Professore Ordinario	Univ. di PALERMO
CREDITS	15		
PROPAEDEUTICAL SUBJECTS			
MUTUALIZATION			
YEAR	3		
TERM (SEMESTER)	1° semester		
ATTENDANCE	Not mandatory		
EVALUATION	Out of 30		
TEACHER OFFICE HOURS	ATTANASIO MASSIMO Wednesday 10:00 - 12:00 dseas ed 13 studio 104 GIAMBALVO ORNELLA Tuesday 10:00 - 12:00 Il servizio prenotazione ricevimento e sospeso. Per fissare un appuntamento con la docente si prega di inviare una mail all'indirizzo ornella.giambalvo@unipa.it Wednesday 12:00 - 13:00 Il servizio prenotazione ricevimento e sospeso. Per fissare un appuntamento con la docente si prega di inviare una mail all'indirizzo ornella.giambalvo@unipa.it		

PREREQUISITES	The course requires the knowledge of descriptive and inferential statistics, the foundations of mathematics and probability theory and SAS software. The course of Social statistics 1 is a mandatory pre-requisite.
LEARNING OUTCOMES	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
ASSESSMENT METHODS	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).
TEACHING METHODS	frontal lessons and practice exercises. Laboratory of SAS for real data analysis

MODULE SOCIAL STATISTICS 2

Prof. MASSIMO ATTANASIO

SUGGESTED BIBLIOGRAPHY

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)

AMBIT	50250-Statistico, statistico applicato, demografico
--------------	---

INDIVIDUAL STUDY (Hrs)	147
-------------------------------	-----

COURSE ACTIVITY (Hrs)	78
------------------------------	----

EDUCATIONAL OBJECTIVES OF THE MODULE

At the end of the course the student should be able to: 1. recognize and describe the essential features of the logistic regression with dichotomous and polytomous response; 2. recognize and describe the essential features of survival data with the more 'common statistical survival tools'; 3. recognize and interpret mathematical terms, probabilistic and statistical the most common objects (hazard function, survival function and cumulative hazard);4.analyze and interpret in an elementary way a Cox regression model. It must also be able to use spreadsheets and statistical software SAS.

SYLLABUS

Hrs	Frontal teaching
2	Outline and aims.
2	Introduction to types of data (prospective vs retrospective, aggregate vs individual, experimental vs survey, longitudinal vs single spell), Main measures of association used in contingency tables. Data analysis.
4	The representation of multidimensional data. Recalls logistic regression model.
2	Data building through a survey and use of data collected by others (fundamental research in the medical, social, demographic)
2	Data analysis and categorization techniques
2	Interaction: definition, functional forms and graphic analysis.
2	Goodness of fit of a logistic regression model: calculation of deviance. applications. Software: SAS
7	Multinomial models. Proportional odds.Partial proportional Odds. Real data sets. Software: SAS
2	Analysis of survival data: introduction and features
6	The main objects of the analysis of survival data. The survival function, hazard, cumulative hazard (in continuous and discrete) and life expectancy
8	Nonparametric methods (the actuarial estimator and the Kaplan-Meier estimator), test for comparison of two survival functions. Test rank and Gehan log.
9	semi-parametric methods (graphical method and for the verification of proportionality 'between risks, the Cox model, the construction of the partial likelihood function, interpretation of the regression coefficients, and the baseline of the survival function).
Hrs	Practice
4	Computation of partial likelihood with 1 dummy covariate. software: excel. Maximization via newton raphson. Software: excel e sas
2	Irwin-Lancaster method, X2 e G2 tests
2	Interaction and categories aggregation. graphical analysis for the detection of the interaction.software: excel + sas
2	Applications to the logistic regression model (parameter estimation, deviance and goodness' adaptation). Software: excel and R
3	Logistic regression models, multinomial models and log-linear models. Software: SAS
4	survival function calculation, hazard, cumulative hazard (in continuous and discrete) and life expectancy. Software: excel and SAS

7	Check graphic and analytical proportionality 'between the risks for the application of the Cox model, the construction of the partial likelihood function, interpretation of the regression coefficients, and the baseline of the survival function. Software: Excel, R e SAS
6	The actuarial estimator and the Kaplan-Meier estimator. SE calculation. Calculation Hazard punctual and interval. calculating cumulative hazard. Test to compare two survival functions. Test rank and Gehan log. Software: Excel, SAS, and R

MODULE SAMPLE SURVEYS AND OPINION POLLS

Prof.ssa ORNELLA GIAMBALVO

SUGGESTED BIBLIOGRAPHY

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

Dispense del corso

Boscaino G., Giambalvo O. (2014) Sampling, in Probability and Statistics: A Didactic Introduction, pagg. 176-209. Ed. Jose' I. Barragues; 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

AMBIT	50250-Statistico, statistico applicato, demografico
INDIVIDUAL STUDY (Hrs)	98
COURSE ACTIVITY (Hrs)	52

EDUCATIONAL OBJECTIVES OF THE MODULE

The main aim of this module regards basic statistical methods for survey sampling (questionnaire, dimension, and schemes), estimation of an unknown parameter of a population (unknown) and its standard error. At the end of the module, students will be able to perform a simple survey and collect data. Student will be more confident with Excel worksheet and SAS.

SYLLABUS

Hrs	Frontal teaching
1	Introduction: aims, exams procedures.
1	Introduction to survey sampling: history and sampling from finite population.
6	Simple survey sampling: estimation of the mean, proportion, and total. Setting optimum sampling dimension. The random sampling using SAS.
15	Stratified survey sampling: estimation of the mean, proportion, and total. Allocation: proportional, equal, and optimum. The stratified sampling using SAS.
4	Comparisons between simple and stratified sampling (for the three different allocations).
5	Collecting data: questionnaire building and administration.
Hrs	Practice
6	Simple survey sampling: estimation of the mean, proportion, and total. Setting optimum sampling dimension. Estimators and their variances using SAS.
14	Stratified survey sampling: estimation of the mean, proportion, and total. Allocation: proportional, equal, and optimum. Estimators and their variances using SAS.