



UNIVERSITÀ DEGLI STUDI DI PALERMO

DEPARTMENT	Scienze Economiche, Aziendali e Statistiche		
ACADEMIC YEAR	2021/2022		
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	18159 - SOCIAL STATISTICS I - INTEGRATED COURSE		
MUTUALIZATION			
YEAR	3		
TERM (SEMESTER)	1° semester		
ATTENDANCE	Not mandatory		
EVALUATION	Out of 30		
TEACHER OFFICE HOURS	<p>ATTANASIO MASSIMO Wednesday 10:00 - 12:00 dseas ed 13 studio 104</p> <p>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</p> <p>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</p>		

<p>PREREQUISITES</p>	<p>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.</p>
<p>LEARNING OUTCOMES</p>	<p>Knowledge and understanding Acquisition: 1. the types of survey and the type of data; 2. the language of the longitudinal data; 3. basic knowledge for the analysis of survival data. Ability to apply knowledge and understanding. To be able to: 1. recognize the nature of the data (individual or aggregate, longitudinal or instantaneous, spatial or precise, censored or complete, etc.) and identify the way of getting them; 2. recognize survival (or event history) data and describe health and epidemiological data, through indices and through statistical models used in biostatistics and epidemiology. 3. be able to use SAS at an intermediate level; recognition of type of data; knowing how to use the univariate freq procs; sort; means; logistic; lifestest; phreg (for the Social statistics module); proc surveymeans and various options (for the sample surveys module). Autonomy. To be able to: 1. interpreting results obtained in relation to the type of data (data from surveys or experimental data; longitudinal or instantaneous data; etc.), and to the area (social, medical-health, educational, etc.); 2. to describe a data set of survival data and to be able to interpret the results of a Cox model. 3. to be able to evaluate the potential of SAS in problems of analysis of real data coming from sample surveys Communication skills. To be able to: 1. know the main elements to write a report on a simple sampling technique or on a survival study; 2. present the results of a survival analysis depending on the audience, even if not expert in statistics. 3. be able to communicate with non-statistical experts on simple sample surveys Learning skills To be able to: 1. consult basic national and international literature; 3. distinguish texts according to their statistical contents in the various fields of application; 4. re-work what has been learned by adapting to the conditions and limits imposed on any customer and the type of problem to be solved; 5. evaluate the use of the most common IT tools (excel, R and SAS) in comparative terms.</p>
<p>ASSESSMENT METHODS</p>	<p>La valutazione dell'apprendimento avviene attraverso una prova scritta e un colloquio orale, subordinato al superamento della prova scritta. La Commissione giudicatrice sarà presieduta dai docenti dei due moduli. La prova in itinere sarà 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 abilità possedute dallo studente, anche informatiche (Excel e SAS), e la sua capacità di commentare i dati reali. La prova, della durata totale massima di 3 ore per il modulo da 9 cfu, e di 2 ore per il modulo da 6 cfu, prevede da 4 a 5 quesiti di natura pratica e/o teorica, a risposta aperta. Per la prova in itinere il tempo sarà di 90 minuti e avrà la medesima struttura. La soglia di sufficienza (18-21) è raggiunta se lo studente mostra un utilizzo sufficiente delle conoscenze di base in termini esecutivi. La soglia 21-24 è raggiunta quando i quesiti sono svolti tutti ma presentano imprecisioni di calcolo e interpretazione. La soglia 24-27 viene assegnata quando il compito è svolto in ogni sua parte, dimostrando una conoscenza approfondita (sa dimostrare proprietà con un adeguato formalismo), mentre la votazione da 28 a 30 e lode viene raggiunta quando alle caratteristiche precedenti si aggiunge la capacità di spaziare e generalizzare con altri ambiti della statistica ed è 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 sarà 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 consisterà in almeno due domande finalizzate a valutare la capacità 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 capacità di formulazione di una risposta. La prima verterà sulla traduzione di una frase, la seconda sarà una domanda di carattere generale sul</p>

	<p>testo. I testi in lingua inglese oggetto di esame sono indicati nella sezione "Testi consigliati in inglese".</p> <p>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.</p> <p>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).</p>
TEACHING METHODS	<p>frontal lessons and practice exercises. SAS Lab for data analysis. During the course, the teacher will share with the students a short article, a book chapter, or a part thereof in English of an informative nature, which will be discussed with the students.</p>

**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 collassamento 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)	141
-------------------------------	-----

COURSE ACTIVITY (Hrs)	84
------------------------------	----

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 Techinques. 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 Techinques. Wiley & Sons. Cap.1 - par. 1 e par. 2
 Testi per il campionamento con SAS: Selecting Unrestricted and Simple Random With Replacement Samples Using Base SAS® and PROC SURVEYSELECT David. D. Chapman, Consultant, Alexandria, VA e 1.http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_surveysselect_details04.htm

AMBIT	50250-Statistico, statistico applicato, demografico
INDIVIDUAL STUDY (Hrs)	94
COURSE ACTIVITY (Hrs)	56

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 software SAS.

SYLLABUS

Hrs	Frontal teaching
1	Introduction: aims, exams procedures.
1	Introduction to survey sampling: history and sampling from finite population.
4	Simple survey sampling: estimation of the mean, proportion, and total. Setting optimum sampling dimension.
2	The ranodm sampling with SAS
10	Stratified survey sampling: estimation of the mean, proportion, and total. Allocation: proportional, equal, and optimum.
2	Comparisons between simple and stratified sampling (for the three different allocations).
5	SAS procedures and statmenst for surveys; the proc surveymeans for randoma sampling and stratified sampling.
2	The stratified sampling using SAS.
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.