



UNIVERSITÀ DEGLI STUDI DI PALERMO

DIPARTIMENTO	Scienze Economiche, Aziendali e Statistiche		
ANNO ACCADEMICO OFFERTA	2018/2019		
ANNO ACCADEMICO EROGAZIONE	2019/2020		
CORSO DILAUREA MAGISTRALE	SCIENZE STATISTICHE		
INSEGNAMENTO	DATI CATEGORIALI C.I.		
CODICE INSEGNAMENTO	19633		
MODULI	Si		
NUMERO DI MODULI	2		
SETTORI SCIENTIFICO-DISCIPLINARI	SECS-S/01		
DOCENTE RESPONSABILE	SCIANDRA MARIANGELA	Professore Associato	Univ. di PALERMO
ALTRI DOCENTI	ABBRUZZO ANTONINO	Professore Associato	Univ. di PALERMO
	SCIANDRA MARIANGELA	Professore Associato	Univ. di PALERMO
CFU	9		
PROPEDEUTICITA'			
MUTUAZIONI			
ANNO DI CORSO	2		
PERIODO DELLE LEZIONI	1° semestre		
MODALITA' DI FREQUENZA	Facoltativa		
TIPO DI VALUTAZIONE	Voto in trentesimi		
ORARIO DI RICEVIMENTO DEGLI STUDENTI	ABBRUZZO ANTONINO Lunedì 15:00 17:00 DSEAS secondo piano stanza 222 SCIANDRA MARIANGELA Mercoledì 12:00 14:00 DSEAS 2 piano		

DOCENTE: Prof.ssa MARIANGELA SCIANDRA

PREREQUISITI	Il corso richiede una buona conoscenza dei concetti fondamentali della statistica inferenziale e del calcolo delle probabilità.
RISULTATI DI APPRENDIMENTO ATTESI	Conoscenza e capacita' di comprensione. Conoscenza delle metodologie dei network probabilistici e dei modelli per dati categoriali. Acquisizione del linguaggio e della terminologia propri della disciplina. Capacita' di comprensione delle derivazioni, delle proprieta' teoriche e dei nessi interni dei metodi presentati. Capacita' di applicare conoscenza e comprensione Capacita' di affrontare problemi concreti con i metodi acquisiti durante le lezioni frontali. Capacita' di utilizzare l'ambiente statistico R per applicare i metodi acquisiti durante le lezioni frontali e per verificare per via simulativi i risultati teorici. Autonomia di giudizio Essere in grado di comprendere criticamente caratteristiche, potenzialita' e limiti dei metodi trattati. Abilita' comunicative Essere in grado di discutere le caratteristiche di un dato problema. Essere in grado di usare la terminologia statistica e la formalizzazione dei problemi in un'esposizione scritta. Capacita' di apprendimento Essere in grado di consultare la letteratura scientifica sull'argomento; capacita' di apprendere le estensioni dei modelli studiati a lezione; capacita' di apprendimento di software statistico specialistico anche diverso da quello impiegato in aula.
VALUTAZIONE DELL'APPRENDIMENTO	La verifica dell'apprendimento avviene attraverso un colloquio orale, subordinato al superamento della prova scritta. Sebbene il corso sia distinto in due moduli, questi si svolgeranno parallelamente e le prove finali (scritte e orali) si svolgeranno contestualmente in un'unica prova scritta e un'unica prova orale. Prova scritta La prova scritta mira a rilevare le conoscenze e le abilita' possedute dallo studente e la sua capacita' di riportarle in un elaborato scritto che tenga conto anche della proprieta' di linguaggio statistico posseduto. La prova scritta riguardera' la verifica dei concetti di network probabilistici e modelli per l'analisi dei dati categoriali acquisiti durante il corso e sara' svolta anche con l'ausilio del software statistico R, attraverso lo studio di un dataset reale. La prova avra' una durata massima di 3 ore. La soglia di sufficienza (pari a un punteggio di 18 su una scala 18-30), propedeutica al superamento della prova, e' raggiunta se lo studente mostra un utilizzo adeguato dei termini relativi ai concetti in oggetto d'esame. Prova orale La prova orale mira ad approfondire l'elaborato scritto e alla valutazione dell'apprendimento dello studente. Questa consistera' in almeno due domande finalizzate a graduare meglio la valutazione delle conoscenze, competenze e abilita' possedute dallo studente, e la sua capacita' di trasmetterle con idoneo linguaggio statistico. La soglia della sufficienza, della prova orale, sara' raggiunta quando lo studente abbia mostrato conoscenza e comprensione degli argomenti almeno nelle linee generali (definizione dei concetti). Quanto piu', invece, l'esaminando abbia brillantemente superato la prova scritta e abbia dato evidenza, nella prova orale, delle sue capacita' argomentative ed espositive, nonche' di proprieta' di linguaggio statistico e di collegamento con le altre discipline, tanto piu' la valutazione sara' positiva. METODO DI VALUTAZIONE FINALE La valutazione finale dell'esame, espressa di concerto tra i due docenti titolari dei due moduli dell'insegnamento, prendera' in considerazione due aspetti: i) la padronanza degli argomenti; ii) 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 finale e' dato dalla media aritmetica semplice delle due prove.
ORGANIZZAZIONE DELLA DIDATTICA	Il corso sara' suddiviso in lezioni frontali ed esercitazioni. Tutti gli argomenti teorici sviluppati nelle lezioni frontali verranno affrontati in termini applicativi mediante attivita' di laboratorio informatico-statistico con l'utilizzo dell'ambiente di programmazione R.

**MODULO
NETWORK PROBABILISTICI**

Prof. ANTONINO ABBRUZZO

TESTI CONSIGLIATI

Introduction to graphical modelling, Dempster. Capitoli 1, 2, 3, 5, 6, 7

Graphical Models with R, Søren Højsgaard, David Edwards. Steffen Lauritzen, Springer, 2012. Capitoli 1, 2, 3, 4, 6, 7.

TIPO DI ATTIVITA'	C
AMBITO	21031-Attività formative affini o integrative
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	54
NUMERO DI ORE RISERVATE ALLE ATTIVITA' DIDATTICHE ASSISTITE	21

OBIETTIVI FORMATIVI DEL MODULO

Il corso si propone di guidare lo studente alla conoscenza delle metodologie dei network probabilistici e all'acquisizione della capacità di applicare tali metodologie a casi pratici reali. Alla fine del corso lo studente dovrà essere in grado di riconoscere i pregi e difetti dei network probabilistici e di descrivere insiemi di dati reali complessi sfruttando e integrando le tecniche apprese.

PROGRAMMA

ORE	Lezioni
4	Modelli grafici per dati categoriali
4	Modelli grafici per dati continui
2	Selezione dei modelli grafici
2	Modelli grafici per dati ad alta dimensionalità

ORE	Esercitazioni
3	Applicazione a dataset reali dei modelli grafici per dati categoriali
3	Applicazione a dataset reali dei modelli grafici per dati continui
3	Applicazione dei modelli grafici per dati ad alta dimensionalità

MODULO DATI CATEGORIALI

Prof.ssa MARIANGELA SCIANDRA

TESTI CONSIGLIATI

Agresti A. (2002) The analysis of categorical data (2nd ed.), Academic Press, London. (Chs. 1 to 9) Disponibile presso la biblioteca del DSEAS

Johnson, Valen E., Albert, James H. (1999), Ordinal Data Modeling, SpringerVerlag New York (Chs. 3 and 4) Acquistabile online o presso la libreria universitaria.

TIPO DI ATTIVITA'	C
AMBITO	21031-Attività formative affini o integrative
NUMERO DI ORE RISERVATE ALLO STUDIO PERSONALE	108
NUMERO DI ORE RISERVATE ALLE ATTIVITA' DIDATTICHE ASSISTITE	42

OBIETTIVI FORMATIVI DEL MODULO

Il corso ha come obiettivo fondamentale l'insegnamento dei metodi statistici piu' opportuni per l'analisi di dati categoriali. L'attenzione e' rivolta, quindi, al problema del trattamento dei dati non metrici e all'individuazione dei piu' comuni e usati metodi per analizzarli. Inoltre, il corso fornisce strumenti propedeutici al trattamento dei dati che riguardano lo schema che ha generato il set di dati. Cio' perche' si vuole che lo studente sia in grado di trattare i dati tenendo conto del loro livello di misura, del modo in cui sono stati generati e delle finalita' conoscitive.

PROGRAMMA

ORE	Lezioni
6	1.1 Concetti e definizioni di base: variabili categoriali, matrici di variabili categoriali, analisi di relazioni simmetriche e asimmetriche: approccio con e senza formalizzazione probabilistica. 1.2 Richiamo alle tradizionali distribuzioni discrete.
10	2.1 Principali misure di associazione e dipendenza • Modelli logit-lineari e log-lineari 2.2 Estensioni a tabelle IxJ 2.3 Modelli per dati a risposta politomica 2.4 Modelli per variabili categoriali ordinali
8	3.1 Misure e modelli di dipendenza e associazione in tabelle a tre vie 3.2 Procedure di selezione del modello
ORE	Esercitazioni
4	Introduzione ai principali pacchetti R
8	Tabelle di contingenza a due vie : come trattarle in R
6	Tabelle di contingenza a piu' vie: specificazione in ambiente R