



# UNIVERSITÀ DEGLI STUDI DI PALERMO

<b>DIPARTIMENTO</b>	Scienze Economiche, Aziendali e Statistiche		
<b>ANNO ACCADEMICO OFFERTA</b>	2023/2024		
<b>ANNO ACCADEMICO EROGAZIONE</b>	2024/2025		
<b>CORSO DILAUREA MAGISTRALE</b>	STATISTICA E DATA SCIENCE		
<b>INSEGNAMENTO</b>	CATEGORICAL DATA C.I.		
<b>CODICE INSEGNAMENTO</b>	20668		
<b>MODULI</b>	Si		
<b>NUMERO DI MODULI</b>	2		
<b>SETTORI SCIENTIFICO-DISCIPLINARI</b>	SECS-S/01		
<b>DOCENTE RESPONSABILE</b>	SCIANDRA MARIANGELA	Professore Associato	Univ. di PALERMO
<b>ALTRI DOCENTI</b>	ABBRUZZO ANTONINO SCIANDRA MARIANGELA	Professore Associato Professore Associato	Univ. di PALERMO Univ. di PALERMO
<b>CFU</b>	9		
<b>PROPEDEUTICITA'</b>			
<b>MUTUAZIONI</b>			
<b>ANNO DI CORSO</b>	2		
<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>	<b>ABBRUZZO ANTONINO</b> Lunedì 15:00 17:00 DSEAS secondo piano stanza 222 <b>SCIANDRA MARIANGELA</b> Mercoledì 12:00 14:00 DSEAS 2 piano		

DOCENTE: Prof.ssa MARIANGELA SCIANDRA

<b>PREREQUISITI</b>	Il corso richiede una buona conoscenza dei concetti fondamentali della statistica inferenziale e del calcolo delle probabilità.
<b>RISULTATI DI APPRENDIMENTO ATTESI</b>	<p>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.</p> <p>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.</p> <p>Autonomia di giudizio Essere in grado di comprendere criticamente caratteristiche, potenzialita' e limiti dei metodi trattati.</p> <p>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.</p> <p>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.</p>
<b>VALUTAZIONE DELL'APPRENDIMENTO</b>	<p>La verifica dell'apprendimento avviene attraverso un colloquio orale che parte dal commento di un report finale consegnato dagli studenti su un tema stabilito insieme durante il corso.</p> <p>Sebbene il corso sia distinto in due moduli, questi si svolgeranno parallelamente e le prove finali si svolgeranno contestualmente in un'unica prova scritta e un'unica prova orale. La prova scritta si svolgerà in lingua inglese. La prova orale si svolgerà di norma in lingua inglese. Il docente potrà, se lo riterrà opportuno, fare esporre al candidato uno o più argomenti in italiano.</p> <p><b>Prova scritta</b> 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.</p> <p><b>Prova orale</b> 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.</p> <p><b>METODO DI VALUTAZIONE FINALE</b> 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.</p> <p>Entrambe le prove (scritta e orale) sono valutate in trentesimi e si ritengono</p>

	superate con una votazione minima di 18/30. Il voto finale e' dato dalla media aritmetica ponderata con pesi 6 per Categorical Data e 3 per Stochastic Networks.
<b>ORGANIZZAZIONE DELLA DIDATTICA</b>	<p>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. Entrambi i moduli sono svolti in lingua inglese.</p> <p>Per entrambi i moduli saranno organizzati lavori di gruppo e report di analisi con presentazioni autonome degli studenti e svolgimento di attivita' in modalita' homework da discutere in aula. Si utilizzerà il software R per l'analisi dei dataset.</p>

## MODULO STOCHASTIC NETWORKS

*Prof. ANTONINO ABBRUZZO*

### TESTI CONSIGLIATI

Introduction to graphical modelling, Dempster. Capitoli 1, 2, 3, 5, 6, 7 - Springer-Verlag 1995.

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

Dispense del docente.

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

### OBIETTIVI FORMATIVI DEL MODULO

Il corso offre allo studente la possibilità di acquisire una conoscenza sui modelli grafici per l'analisi dei dati categoriali e applicare tali metodologie a casi pratici reali. Alla fine del corso lo studente dovrà essere in grado di riconoscere i pregi e difetti dei modelli grafici per dati categoriali e di descrivere insiemi di dati reali complessi mediante l'uso delle tecniche apprese.

## PROGRAMMA

ORE	Lezioni
4	Modelli grafici log-lineari
4	Network Bayesiani per l'analisi dei dati categoriali
4	Modelli grafici per dati continui
ORE	Esercitazioni
3	Analisi dei dati con i modelli grafici log-lineari
3	Analisi dei dati con i network Bayesiani
3	Analisi dei dati con i modelli grafici Gaussiani

**MODULO  
MODELS FOR CATEGORICAL DATA**

*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) Disponibile presso la biblioteca del DSEAS

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

**OBIETTIVI FORMATIVI DEL MODULO**

Il corso ha come obiettivo fondamentale l'insegnamento dei metodi statistici piu' utili 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 inoltre 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 della loro scala di misura, del modo in cui sono stati generati alla luce delle finalita' conoscitive che ci si pone.

**PROGRAMMA**

<b>ORE</b>	<b>Lezioni</b>
4	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.
6	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
4	Modelli per dati sovradispersi o con eccessi di zer: modelli ZIP and Hurdle
2	Quasi e completa separazione in modi per dati categoriali: la regressione logistica di Firth
<b>ORE</b>	<b>Esercitazioni</b>
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