



---

# DATA MINING, STATISTICA SOCIALE ED INFORMATIZZAZIONE DEI DATI

9 CFU - 1° semestre

## Docenti titolari dell'insegnamento

**ROSARIO GIUSEPPE D'AGATA** - Modulo FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO - SECS-S/05 - 6 CFU

**Email:** rodagata@unict.it

**Edificio / Indirizzo:** Palazzo Reburdone, Via Vitt. Emanuele II, 8

**Telefono:** 095 70305273

**Orario ricevimento:** Lunedì 10-12

**GIOVANNI GIUFFRIDA** - Modulo SISTEMI PER LA GESTIONE DI BASI DI DATI - INF/01 - 3 CFU

**Email:** ggiuffrida@dmi.unict.it

**Edificio / Indirizzo:** Palazzo Reburdone, Viale Vittorio Emanuele 8 I Piano

**Telefono:** 095 70305265

**Orario ricevimento:** Mercoledì 10-13

---

## MODALITÀ DI SVOLGIMENTO DELL'INSEGNAMENTO

### ▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

Le lezioni saranno frontali e prevederanno esercitazioni con l'ausilio di excel

## PREREQUISITI RICHIESTI

### ▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

Lo studente deve essere in possesso delle nozioni fondamentali di matematica e di statistica descrittiva ed inferenziale

---

## FREQUENZA LEZIONI

### ▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

La frequenza non è obbligatoria ma fortemente consigliata

---

## CONTENUTI DEL CORSO

### ▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

Il primo modulo si concentrerà sulla progettazione di un'indagine statistica a partire dall'analisi delle fonti statistiche ufficiali, nazionali ed internazionali. In questa fase un'attenzione particolare sarà rivolta alle strategie di campionamento e alla costruzione del questionario compresi alcuni cenni sulle tecniche di scaling.

Il secondo modulo prevede un approfondimento del concetto di indicatore, nello specifico si tratteranno gli argomenti legati alla costruzione di un indicatore composito. La seconda parte del modulo prevede, infine, si concentrerà sui test non parametrici più utilizzati in ambito sociale.

---

## TESTI DI RIFERIMENTO

### ▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

Modulo 1

D. F. Iezzi (2009), *Statistica per le Scienze Sociali*, Carocci, Roma. Parte prima (da pag. 23 a pag. 190)

G. Cicchitelli, A. Herzel e G. E. Montanari (1992), *Il campionamento statistico*, Il Mulino, Bologna, cap. III (§§ 1, 2, 3, 4, 5, 6, 7), pp. 69-84.

Modulo 2

D. F. Iezzi (2009), *Statistica per le Scienze Sociali*, Carocci, Roma. Parte seconda (da pag. 283 a pag. 314)

*Handbook on Constructing Composite Indicators. Methodology and user guide.*, [www.oecd.org/publishing](http://www.oecd.org/publishing).

S. Siegel e N. J. Castellan jr. (1992), *Statistica non parametrica*, McGraw-Hill, Milano, cap. III (§§ 3.4, 3.4.1, 3.4.2) e cap. IV (§§ 4.1, 4.2, 4.3), pp. 67-92; cap. V (§§ 5.1,5.2), pp. 113-130; cap. VI (§§ 6.1, 6.2, 6.3, 6.4), pp. 151-191.

---

## PROGRAMMAZIONE DEL CORSO

### SISTEMI PER LA GESTIONE DI BASI DI DATI

---

Argomenti	Riferimenti testi
1 Dai dati all'informazione: Sistemi informativi e informatici; Dato e informazione; Organizzazione relazionale dei dati; Interrogazione; Sistemi di interrogazione evoluti	Atzeni,Ceri,Paraboschi,Torlone, <i>Basi di Dati,Modelli e linguaggi di interrogazione</i> , terza edizione, McGraw-Hill 2002. • Albano-Ghelli-Orsini, <i>Basi di Dati Relazionali e a Oggetti</i> , Zanichelli, 1997Ullman, <i>Basi di Dati e Basi di Conoscenza</i>

---

- 2 Introduzione alla Computational Social Science; Nozioni di «Big Data» e Artificial Intelligence; Concetti e cenni di algoritmi di «profilazione utente»; Social Networks e Social Network Analysis; slides fornite dal docente; Machine Learning: The Art and Science of Algorithms That Make Sense of Data, Peter Flach, Cambridge University Press; Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà. Viktor M
- 
- 

## VERIFICA DELL'APPRENDIMENTO

### MODALITÀ DI VERIFICA DELL'APPRENDIMENTO

#### ▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

Il corso prevede una verifica intermedia la cui data verrà concordata con gli studenti e resa pubblica nel più breve tempo possibile. Per poter svolgere la prova finale è necessario aver superato la verifica intermedia. In caso contrario l'esame si svolgerà interamente per iscritto nelle date stabilite da calendario.

### ESEMPI DI DOMANDE E/O ESERCIZI FREQUENTI

#### ▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

- 1) Indagini strutturali ed indagini correnti
- 2) I rapporti statistici
- 3) Campionamento probabilistico e non probabilistico
- 4) Le fasi di un test non parametrico
- 5) Le indagini multiscopo
- 6) Campionamento casuale A GRAPPOLI

#### **Esercizio 1**

Si costruisca un indicatore composito di integrazione degli immigrati considerando tre indicatori semplici (attrazione, inserimento sociale ed inserimento occupazionale) calcolato su cinque province. Si commenti la procedura e i risultati cui si perviene.

Prov	Attraz	InsSoc	InsOcc
Torino	40,6	37,3	39,4
Milano	65	46,2	55,8
Trento	46,1	62,2	37,1
Vicenza	57,3	64,7	51,8
Treviso	58,4	58,3	47,6

## Esercizio 2

Sia data la seguente distribuzione di pesi rilevati su 51 studenti. Verificare l'ipotesi che la distribuzione dalla quale è estratto il campione sia normale. ( $\alpha=0.05$ )

Kg	40- 45	46-50	51-55	56-60	Totale
$n_i$	7	12	17	15	51

## Esercizio 3

La tabella mostra l'esito di un'indagine condotta su 113 studenti sull'opinione in merito alla legalizzazione delle droghe leggere. Ad un livello di significatività  $\alpha=0.05$  è possibile stabilire che vi è una differenza statisticamente significativa tra maschie e femmine?

	Favorevoli	Contrari	Totale
Femmina	35	33	68
Maschio	15	30	45
Totale	50	63	113