



---

# 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

---

## OBIETTIVI FORMATIVI

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

Il corso è finalizzato alla conoscenza di fonti, sistemi di indicatori e piani di campionamento.

### ▪ SISTEMI PER LA GESTIONE DI BASI DI DATI

Il corso mira a offrire la informazioni di base per la gestione di basi di dati relazionali e fornire inoltre nozioni sui Big Data, sulla loro gestione e sull'impatto di questi nella società contemporanea.

---

## MODALITÀ DI SVOLGIMENTO DELL'INSEGNAMENTO

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

Il corso affronta i principali argomenti relativi alla gestione di un'indagine statistica. Gli argomenti sono trattati seguendo il percorso logico delle fasi di un'attività di ricerca: la specificazione del problema, la raccolta dei dati, l'elaborazione, l'analisi dei dati e la valutazione dei risultati conseguiti secondo specifiche dimensioni d'analisi spazio-temporale.

### ▪ SISTEMI PER LA GESTIONE DI BASI DI DATI

L'insegnamento è organizzato in una serie di lezioni frontali. Esercitazioni sono svolte in classe all'interno delle lezioni stesse. Un paio di speakers esterni sono normalmente invitati per esporre su

temi avanzati di Computational Social Science.

## PREREQUISITI RICHIESTI

- **FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO**  
Principi di matematica e statistica
- **SISTEMI PER LA GESTIONE DI BASI DI DATI**  
Nozioni di base di informatica e di gestione dati.

---

## FREQUENZA LEZIONI

- **FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO**  
Non obbligatoria ma consigliata
- **SISTEMI PER LA GESTIONE DI BASI DI DATI**  
Non obbligatoria ma fortemente consigliata.

---

## CONTENUTI DEL CORSO

- **FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO**  
Le fonti delle informazioni statistiche; modelli di rilevazione delle informazioni da fonti istituzionali; i sistemi statistici informativi, procedure d'integrazione fra data-bases; le trasformazioni dei dati; classificazione tipologica di tassi, rapporti, numeri indici; gli indicatori; le tecniche di costruzione e di indicatori composti.
- **SISTEMI PER LA GESTIONE DI BASI DI DATI**  
Il corso intende dare una panoramica dei sistemi di gestione delle basi di dati sia da un punto di vista teorico che pratico. L'enorme quantità di dati di tipo "sociale" oggi disponibile rende necessaria una corretta gestione informatica di tali dati al fine di poter comprendere al meglio ed analizzare in maniera più proficua i fenomeni sociali che possono essere estratti da questa enorme quantità di dati. Strumenti quali Facebook, la lettura di quotidiani online, lo scambio di email, ecc., forniscono oggi allo scienziato sociale una nuova dimensione di analisi difficilmente gestibile con tecniche convenzionali per via della numerosità dei dati stessi.

Si intende formare lo studente con le conoscenze teoriche necessarie alla comprensione delle tecniche di archiviazione delle basi di dati. Inoltre si vuole dare una conoscenza pratica e metodologica di un sistema di gestione delle basi di dati di larga diffusione. Esploreremo inoltre i principi dei Big Data e della loro interazione con le scienze sociali.

---

## TESTI DI RIFERIMENTO

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

STAT (2011), Navigando tra le fonti demografiche e sociali, ISTAT, Roma,  
[http://www3.istat.it/dati/catalogo/20100325\\_01/Navigando\\_tra\\_le\\_fonti\\_demografiche\\_sociali.pdf](http://www3.istat.it/dati/catalogo/20100325_01/Navigando_tra_le_fonti_demografiche_sociali.pdf)  
Bonarini F. (2006), Guida alle fonti statistiche socio-demografiche, CLEUP, Padova, pp. 1-142;  
231-308.  
Cavaleri P. e Venturini F. (a cura di) (2004), Documenti e dati pubblici sul web. Guida  
all'informazione di fonte pubblica in rete, Il Mulino, Bologna.  
D. F. Iezzi (2009), Statistica per le Scienze Sociali, Carocci, Roma (Cap. 10 e 11).  
OECD (2008), Handbook on Constructing Composite Indicators. Methodology and user guide.,  
[www.oecd.org/publishing](http://www.oecd.org/publishing).

## ▪ SISTEMI PER LA GESTIONE DI BASI DI DATI

1. Slides fornite dal docente
2. Albano-Ghelli-Orsini, Basi di Dati Relazionali e a Oggetti, Zanichelli, 1997
3. Ullman, Basi di Dati e Basi di Conoscenza
4. Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà. Viktor Mayer-Schönberger, Kenneth N. Cukier e R. Merlino
5. Machine Learning: The Art and Science of Algorithms That Make Sense of Data, Peter Flach, Cambridge University Press
6. Basi di Dati, Modelli e linguaggi di interrogazione, Atzeni, Ceri, Paraboschi, Torlone, terza edizione, McGraw-Hill 2002

## ALTRO MATERIALE DIDATTICO

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

Il materiale didattico verrà fornito nel corso delle lezioni.

### ▪ SISTEMI PER LA GESTIONE DI BASI DI DATI

Slides fornite dal docente disponibili sul sito: <http://www.dmi.unict.it/~ggiuffrida/>

---

## PROGRAMMAZIONE DEL CORSO

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

Argomenti	Riferimenti testi
1 Le fonti delle informazioni statistiche. Modelli di rilevazione delle informazioni da fonti istituzionali	ISTAT (2011), Navigando tra le fonti demografiche e sociali, ISTAT, Roma, <a href="http://www3.istat.it/dati/catalogo/20100325_01/Navigando_tra_le_fonti_demografiche_sociali.pdf">http://www3.istat.it/dati/catalogo/20100325_01/Navigando_tra_le_fonti_demografiche_sociali.pdf</a>

---

- 2 I sistemi statistici informativi. Procedure d'integrazione fra data-bases. Bonarini F. (2006), Guida alle fonti statistiche socio-demografiche, CLEUP, Padova, pp. 1-142; 231-308. Cavaleri P. e Venturini F. (a cura di) (2004), Documenti e dati pubblici sul web. Guida all'informazione di fonte pubblica in rete, Il Mulino, Bologna.
- 
- 3 Le trasformazioni dei dati. Classificazione tipologica di tassi, rapporti, numeri indici. La questione degli indicatori. Le tecniche di costruzione e di standardizzazione degli indicatori. D. F. Iezzi (2009), Statistica per le Scienze Sociali, Carocci, Roma (Cap. 10 e 11). OECD (2008), Handbook on Constructing Composite Indicators. Methodology and user guide., [www.oecd.org/publishing](http://www.oecd.org/publishing).

## 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, Basi di Dati, Modelli e linguaggi di interrogazione, terza edizione, McGraw-Hill 2002. • Albano-Ghelli-Orsini, Basi di Dati Relazionali e a Oggetti, Zanichelli, 1997 Ullman, Basi di Dati e Basi di Conoscenza
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**

L'esame sarà scritto. Per i frequentanti sono previste due prove intermedie. Le date verranno concordate di volta in volta con gli studenti per venire incontro alle esigenze degli stessi. Si provvederà, tuttavia, a darne informazione tramite avviso con notevole anticipo. Per svolgere la verifica di apprendimento del modulo successivo è necessario aver superato la verifica di apprendimento del modulo precedente. La prova consisterà nella risoluzione di due esercizi e la risposta a tre domande inerenti il modulo in oggetto. Sarà possibile portare con sé: calcolatrice e tavole di distribuzione di probabilità.

Per chi decidesse di sostenere l'esame per intero, la prova si svolgerà negli appelli ufficiali, consisterà nella risoluzione di due esercizi e la risposta a quattro domande in 90 minuti. Sarà possibile portare con sé: calcolatrice e tavole di distribuzione di probabilità.

#### ▪ **SISTEMI PER LA GESTIONE DI BASI DI DATI**

L'esame è in formato scritto con una serie di domande a risposta multipla V/F.

#### **ESEMPI DI DOMANDE E/O ESERCIZI FREQUENTI**

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

- 1) L'indagine multiscopo
- 2) Le fonti delle informazioni statistiche.
- 3) Le fonti delle informazioni statistiche.

##### **Esercizio 1**

Il risultato ottenuto rilevando il reddito al primo impiego negli ultimi 7 anni dei laureati di una prestigiosa università romana è il seguente:

Anno	2002	2003	2004	2005	2006	2007	2008
Reddito	780	640	610	570	640	620	610

Avvalendosi degli strumenti statistici appresi, analizzare e commentare l'evoluzione del fenomeno considerato.

##### **Esercizio 2**

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

▪ **SISTEMI PER LA GESTIONE DI BASI DI DATI**

Il modello relazionale dei dati è quello tipicamente utilizzato nei DBMS: V/F?

Un ipertesto è un testo distribuito su vari computer: V/F?

Un DBMS aiuta nella soluzione al problema della ridondanza dei dati: V/F?

---