



STATISTICA SOCIALE ED INFORMATIZZAZIONE DEI DATI

9 CFU - 1° semestre

Docenti titolari dell'insegnamento

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

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

OBIETTIVI FORMATIVI

▪ SISTEMI PER LA GESTIONE DI BASI DI DATI

Strumenti per la gestione di basi di dati (INF/01)

▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

Fonti e sistemi di indicatori. Strumenti per l'analisi territoriale dei dati (SECS-S/05). Piani di campionamento. Test d'ipotesi non parametrici SECS-S/05)

MODALITÀ DI SVOLGIMENTO DELL'INSEGNAMENTO

▪ 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.

▪ 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.

PREREQUISITI RICHIESTI

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

Nozioni di base di informatica e di gestione dati.

- **FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO**

Conoscenza del primo modulo e fondamenti del calcolo delle probabilità. Principi di inferenza statistica

FREQUENZA LEZIONI

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

Non obbligatoria ma fortemente consigliata.

- **FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO**

Non obbligatoria ma consigliata

CONTENUTI DEL CORSO

- **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.

- **FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO**

Piani di campionamento; Disegno di campionamento; Dimensioni campionarie; Errori di campionamento Criteri di selezione dei campioni; Test su un campione: binomiale, χ^2 , Kolmogorov-Smirnov Test su campioni a due misure dipendenti: McNemar Test su due campioni indipendenti: esatto di Fisher, χ^2 , della mediana, Wilcoxon-Mann-Whitney

TESTI DI RIFERIMENTO

- **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

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

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.

L. Fabbris (1989), L'indagine campionaria, N.I.S., Roma, capp. I (§ 1.4, 1.5, 1.6), pp. 24-39.

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

ALTRO MATERIALE DIDATTICO

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

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

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

Il materiale didattico verrà fornito nel corso delle lezioni.

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, 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

FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

Argomenti	Riferimenti testi
-----------	-------------------

1	Piani di campionamento Disegno di campionamento Dimensioni campionarie Errori di campionamento	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.
2	Criteri di selezione dei campioni	L. Fabbris (1989), L'indagine campionaria, N.I.S., Roma, capp. I (§ 1.4, 1.5, 1.6), pp. 24-39.
3	Test su un campione: binomiale, χ^2 , Kolmogorov-Smirnov Test su campioni a due misure dipendenti: McNemar, dei segni Test su due campioni indipendenti: esatto di Fisher, χ^2 , della mediana, Wilcoxon-Mann-Whitney	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.

VERIFICA DELL'APPRENDIMENTO

MODALITÀ DI VERIFICA DELL'APPRENDIMENTO

▪ SISTEMI PER LA GESTIONE DI BASI DI DATI

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

▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

L'esame sarà scritto. Per i frequentanti sono previste tre 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 quattro esercizi inerenti al modulo in oggetto. Sarà possibile portare con se: formulario, 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 cinque esercizi in 90 minuti. Sarà possibile portare con se: formulario, calcolatrice e tavole di distribuzione di probabilità.

ESEMPI DI DOMANDE E/O ESERCIZI FREQUENTI

▪ 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?

▪ FONTI, SISTEMI DI INDICATORI E PIANI DI CAMPIONAMENTO

1) Campionamento probabilistico e non probabilistico

2) Campionamento casuale A DUE STADI

3) Le fasi di un test non parametrico

Esercizio 1

Un professore di educazione fisica di una scuola superiore ha misurato le performance ottenute da 50 studenti salto in lungo ottenendo la seguente distribuzione di frequenza:

m	n_i
4.05	10
3.80	25
5.16	9
6.02	6
Totale	50

Verificare l'ipotesi che la distribuzione sia normale. ($\alpha=0.05$)

Esercizio 2

All'ingresso di una discoteca, per attirare la clientela, i proprietari sono soliti distribuire dei gadget: unbracciale, un anello, un apribottiglie o un portachiavi. Ci si chiede se l'attribuzione dei gadget sia casuale o legata a qualche fattore specifico come ad esempio il genere. ($\alpha=0.01$)

Bracciale Anello Apribottiglie Portachiavi Tot

30 42 23 20 115
