

CORSO DI INFORMATICA

LEZIONE III

Dott. Simone Accordini

Per diventare informazioni fruibili
i dati spesso devono essere CODIFICATI

Esempio:

- dato: uno studente consegna un compito con 8 risposte corrette su 10
- informazione: lo studente ottiene una valutazione pari a “*BUONO*”

⇒ “*buono*” richiama un **codice di valutazione** “*insuff – suff – discreto – buono – ottimo*” e permette una diretta comparazione del risultato

Per poter essere processati da un elaboratore elettronico i dati o le informazioni devono essere codificati

CODIFICA

Operazione mediante la quale i dati e le informazioni raccolte vengono tradotte in codici numerici o alfanumerici.



La codifica è necessaria per

- l'**elaboratore elettronico**: riconosce soltanto sequenze binarie
- il **software utilizzato**: ha delle regole di inserimento dati
- lo **studioso**: deve adottare delle *convenzioni* per poter effettuare calcoli e confronti.

CODIFICA - ESEMPI

Sesso (→ **variabile qualitativa nominale**)

Questionario	Codifica	Software	ASCII	Binario
M	1	0	49	00110001
F	2	1	50	00110010
dato mancante	9	.	46	00101110

Grado di istruzione (→ **variabile qualitativa ordinale**)

Questionario	Codifica
Nessuno	0
Licenza elementare	1
Diploma scuola media inferiore	2
Diploma scuola media superiore	3
Laurea	4
dato mancante	9

Presenza di asma (→ **variabile qualitativa nominale**)

Questionario	Codifica
No	0
Sì	1
dato mancante	9

Altezza in cm (→ **variabile quantitativa**)

Codifica solo per il **dato mancante**: **-99**

*****IMPORTANTE** - La codifica di tutti i possibili valori delle variabili rilevate (e dei dati mancanti) va eseguita **PRIMA** dell'inizio dell'indagine o dell'esperimento.

INDAGINE SULL'INSONNIA TRA GLI STUDENTI DEL CORSO DI INFORMATICA

ID

Data di compilazione

giorno mese anno

Sesso M F

Data di nascita

giorno mese anno

1. Nell'ultimo mese, quante ore ha dormito
mediamente per notte?

n. ore

2. Nell'ultimo mese, ha faticato ad addormentarsi?

Sì No

SE 'NO' VADA ALLA DOMANDA 3, SE 'SI':

2.1 Nell'ultimo mese, mediamente dopo quanto
tempo si è addormentato?

n. ore

3. Nell'ultimo mese, Le è capitato di svegliarsi
durante la notte?

Sì No

SE 'NO' VADA ALLA DOMANDA 4, SE 'SI':

3.1 Nell'ultimo mese, per quanti giorni
si è svegliato durante la notte?

n. giorni

4. Sta prendendo al momento farmaci
contro l'insonnia?

Sì No

5. Ha mai sofferto d'insonnia?

Sì No

SE 'NO' VADA ALLA DOMANDA 6, SE 'SI':

5.1 E' stata confermata da un dottore?

Sì No

6. Nell'ultimo mese, ha dormito sempre all'incirca
negli stessi orari?

Sì No

CODIFICA DELLE DOMANDE DEL QUESTIONARIO SULL'INSONNIA

SESSO: 1 = maschio
2 = femmina
9 = missing

DATA DI NASCITA: - 9 = missing
(3 variabili: giorno, mese e anno)

DOMANDE 1, 2.1, 3.1: - 9 = missing

DOMANDE 2, 3, 4, 5, 5.1, 6: 0 = no
1 = sì
9 = missing

LE TABELLE

Il modo più comodo e leggibile per rappresentare i dati raccolti (*collezione*) è quello di organizzarli in forma tabellare.

Esempio

N° progr.	ID	Cognome	Nome	Sesso	Età	Peso
1	261	Bianchi	Giulio	1	65	68,6
2	8	Rossi	Paolo	1	81	80,1
3	34	Verdi	Giuseppe	1	79	83,1
...			
...			
500	122	Neri	Francesca	2	69	75,9

COLONNA \Leftrightarrow **VARIABILE**

RIGA \Leftrightarrow **SOGGETTO**
(record, osservazione)

VARIABILE = caratteristica della popolazione che si vuole studiare (sesso, età, peso, ...)

Una tabella è una lista organizzata in righe e colonne.

- In ciascuna riga sono registrate le informazioni (caratteristiche) relative ad un dato **oggetto**.
- Le colonne identificano le **caratteristiche** di ciascun oggetto.

Righe: sono chiamate **record** (registrazioni)

Colonne: sono chiamate **campi** della tabella

Esempio: Elenco del telefono

Cognome	Titolo	Nome	Profess.	Indirizzo	Frazione	Telefono
Rossi	Dott.	Ugo	Medico chirurgo	48 vl. Rosa	Milano Marittima	748392
Bianchi		Carlo		34 v. Mimosa		254784
Verdi	Rag.	Piero	Studio commerciale	23 c. Giglio		876657
Neri	Ing.	Romeo		45 v. Margherita		875698

I passi da compiere per organizzare una tabella:

1. Individuare l'oggetto di cui si vogliono registrare le informazioni (unità di rilevazione)
2. Individuare le caratteristiche, gli attributi da rilevare dell'oggetto.

Esempio: Elenco del telefono

Oggetto della tabella = gli abbonati al telefono

Caratteristiche descrittive dell'oggetto = i sette attributi riportati nella prima riga

Esempio: Questionario

Questionario					
Nome	_____				
Cognome	_____				
Nato il	_____	A	_____		
Sesso	__	Malattia	_____		



Nome	Cognome	Nato il	A	Sesso	Malattia
Mario	Tizio	24/04/44	Napoli	M	Alfa
Maria	Caio	23/12/52	Ravenna	F	Beta
Carlo	Sempronio	12/10/65	Bologna	M	Gamma

CODICE INDIVIDUALE

Nella codifica dei dati è fondamentale che ad ogni soggetto corrisponda un **codice identificativo univoco (ID)**.

Di solito come ID vengono usati i **numeri naturali** (tranne lo 0): **1, 2, ... , N**.

L'ID garantisce:

- di poter identificare univocamente ciascun record e quindi di mantenere la **CONSISTENZA DELL'INFORMAZIONE**

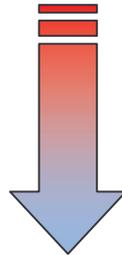
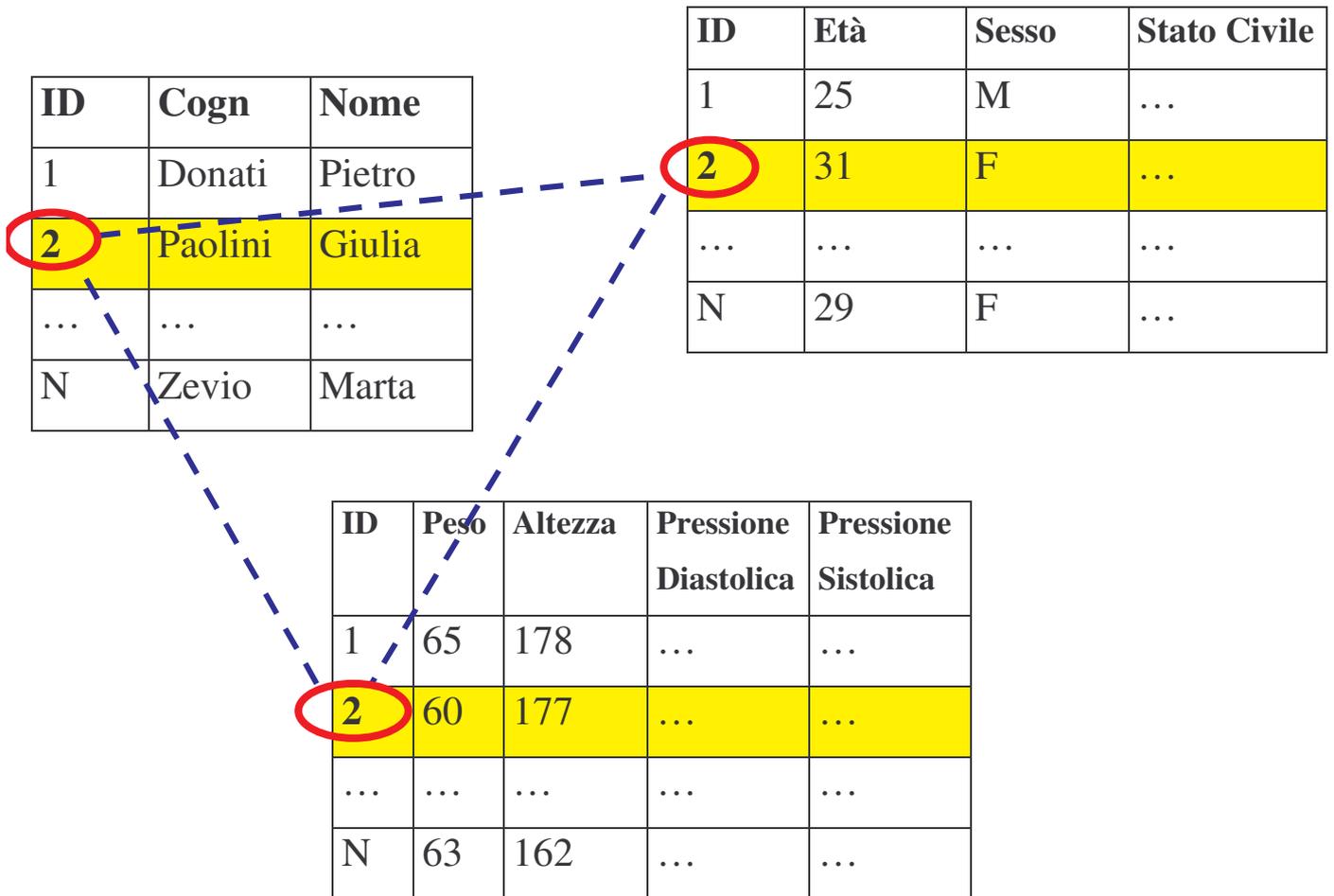
Esempi:

- **Questionario Postale** - applicare l'ID alla lettera che l'intervistato dovrà rispedire consente di: 1) risalire ai dati della persona (cognome, nome, età, sesso); 2) creare un collegamento tra il questionario cartaceo e il record nel file di dati; 3) individuare chi eventualmente contattare una seconda volta per sollecitare la risposta;

Esempi (continua):

- **Questionario somministrato dall'intervistatore e composto da più parti** - applicare l'ID a ciascuna parte del questionario permette di: 1) inserire i dati su fogli e file diversi, in tempi diversi; 2) mantenere collegate tutte le informazioni
- **Analisi di laboratorio:** l'ID su una provetta permette di risalire alla persona che ha fornito quel campione
- la **PRIVACY** del soggetto intervistato o analizzato
 - ⇒ i dati identificativi come il nome, il cognome, ... vengono mantenuti in un **archivio riservato** (nella tabella compare solo l'ID che è anonimo).

Esempio



ID	Et�	Sesso	Stato Civile	Peso	Altezza	Pressione Diastolica	Pressione Sistolica
1	25	M	...	65	178
2	31	F	...	60	177
...
N	29	F	...	63	162

FOGLIO ELETTRONICO

Un **foglio elettronico** è un software dedicato alla gestione di **dati in forma tabellare**, dotato di funzioni per la semplificazione di calcoli algebrici e di routine grafiche per la rappresentazione dei dati.

- **LOTUS 1-2-3** → progenitore di tutti i fogli elettronici
- **Quattro-Pro**
- **Excel** → il più diffuso e usato, basato sulla logica Lotus

FILTRI → selezione di gruppi di record sulla base dei valori delle caratteristiche



INSIEMISTICA E ALGEBRA DELLE PROPOSIZIONI