Capitolo 2

Un po' di terminologia: I dati e le variabili

- Prima di fare cose "divertenti" con i dati, è necessario conoscere un po' di gergo per chiamare le cose con il nome giusto!

> ESEMPIO 1: Nell'esempio sul numero di casi di spina bifida nel gruppo trattato e nel gruppo di controllo, sulle donne in gravidanza sono state anche osservate età, statura, peso, e gruppo sanguigno.

- I **DATI** sono semplicemente una raccolta di informazioni.
- L'insieme (di individui, o animali, o oggetti, o …) a cui si fa riferimento costituisce l'insieme delle **UNITA' STATISTICHE** (casi).
- L'insieme di tutte le unità statistiche è detto convenzionalmente **POPOLAZIONE** di riferimento. Invece, un aggregato di unità statistiche selezionate (tramite un esperimento di campionamento) da una popolazione è detto **CAMPIONE**.
- La dimensione del campione può variare da poche unità a molte migliaia di osservazioni, perché a ogni osservazione è legato un "costo di rilevazione".

> ESEMPIO 1: La popolazione oggetto di studio è l'insieme virtuale di tutte le donne in gravidanza, comparabili per abitudini di vita. Il campione è costituito dalle n = 4000 donne che sono entrate nell'esperimento.

- Le caratteristiche di interesse (caratteri) rilevate sulle unità statistiche vengono chiamate **VARIABILI** (ad esempio, età, peso, trattamento, …).
- Una variabile è quindi una caratteristica di diretto interesse che si studia sul campione (o popolazione), che può assumere una pluralità di valori (almeno 2), che devono essere esaustivi e non sovrapposti.
- I valori distinti assunti da una variabile sono dette **MODALITA'** della variabile. Le modalità in genere sono note preliminarmente.
- La scala delle modalità delle variabili può produrre i seguenti tipi di dati:
  - Dati **QUALITATIVI** o **CATEGORIALI** (ordinali, sconnessi), e **DICOTOMICI**;
  - Dati **QUANTITATIVI** o **NUMERICI** (discreti, continui);
  - Dati **TRASFERIBILI**.

ESEMPIO 2: Un piccolo insieme di osservazioni

Si vuole investigare sull'interesse verso terapie mediche non appartenenti alla medicina ufficiale come l'omeopatia, l'agopuntura, etc.

NOME	ETA'	SESSO	TITOLO.STUDIO	PESO	ATTIVITA'
Lucia	30	F	laurea	72	dipendente
Filippo	28	M	laurea	55	disoccupato
Carla	26	F	diploma	79	casalinga
Franco	29	M	laurea	63	dipendente
Antonio	22	M	diploma	24	studente

Individui (unità elementari)

Caratteri (variabili) di interesse che assumono differenti valori

- In particolare, in statistica si parla di dati:
  - **QUALITATIVI** o **CATEGORIALI** quando le modalità della variabile sono espresse in forma verbale (sesso, istruzione, religione,…). A loro volta questi dati possono essere:
    - **SCONNESSI** o **NOMINALI** se non esiste nessun ordinamento tra le modalità (religione, colore occhi, modalità di somministrazione …);
    - **ORDINALI** se è possibile individuare un ordinamento naturale delle modalità (istruzione, giudizio, …).

    Se le modalità sono solo due si parla di dati **DICOTOMICI** o **BINARI** (sesso, presenza/assenza, …). A volte vengono assegnati dei valori numerici, ma il valore del numero non vuol dire assolutamente nulla!
  - **NUMERICI** o **QUANTITATIVI** quando le modalità sono espresse da numeri (età, pressione, peso, …). A loro volta questi dati possono essere:
    - **INTERI** o **DISCRETI** quando le modalità sono numeri interi (numero di figli, numero di visite mediche in un anno, ….);
    - **CONTINUI** o **REALI** quando le modalità sono numeri reali (volume di una massa tumorale, altezza, dose,…); eventuale suddivisione in classi;
    - **TRASFERIBILI** se si può cedere tutta o una parte del carattere posseduto a un'altra unità (reddito, numero di operai dipendenti,…).

Analisi statistiche

- DATI **QUALITATIVI** vs **NUMERICI**
  - Le diversità dei dati permettono diversi tipi di analisi statistiche.
  - Ci sono degli "strumenti statistici" appositi per studiare tipi diversi di dati.
  - Tra le varie tipologie di dati è implicita una gerarchia (le variabili quantitative possono essere discretizzate, le variabili quantitative discrete possono essere tradotte in variabili qualitative ordinali, quelle ordinali posso essere considerate nominali). Le analisi statistiche sono più ricche, per così dire, ascendendo la gerarchia.

- DATI **UNIVARIATI** vs **MULTIVARIATI**
  - Le analisi univariate considerano una sola variabile rilevata sulle unità.
  - Nello studio congiunto di due variabili si parla di analisi bivariata.
  - Lo studio congiunto di piu' di due variabili è detto analisi multivariata (ovviamente il multivariato include il bivariato).

ESEMPIO 3: Un piccolo esempio sui trattamenti (per fissare la terminologia)

Vogliamo sapere quale tra due trattamenti, detti A e B, per una certa patologia, è migliore.

La popolazione di riferimento è l'insieme di tutti i pazienti che hanno quella particolare patologia (oggi, ma anche domani,…). Le unità statistiche sono i pazienti. In questo caso la popolazione è virtuale.

25 pazienti sono trattati con A e 25 pazienti con B. Alla fine della prima settimana si valuta, per ogni paziente, se i sintomi sono scomparsi. Il campione è costituito dai n = 50 pazienti trattati e per cui è nota la risposta (dopo una settimana) al trattamento.

DATI

paziente	trattamento	risposta
1	A	SI
2	A	NO
…..	…..	……
25	A	NO
…..	……	……
50	B	NO

VARIABILI

trattamento = con modalità A e B
risposta (dopo una settimana) = con modalità SI e NO

Esercizi

1. La media della pressione sistolica sanguigna in un maschio adulto sano è ritenuta pari a circa 129. Si è misurata la pressione di 100 maschi adulti sani appartenenti ad una comunità le cui abitudini dietetiche potrebbero essere causa dell'aumento dei valori della pressione.
   a) La pressione sistolica sanguigna è un esempio di variabile:
   ∇ continua ∇ discreta ∇ dicotomica
   b) L'unità statistica è:
   ∇ la comunità ∇ il singolo maschio ∇ il singolo valore della pressione

2. 100 appezzamenti di terreno di uguale dimensione e coltivati con un certo ortaggio sono stati divisi in 4 gruppi di 25 appezzamenti ciascuno. Ciascun gruppo è stato poi fertilizzato usando 4 diverse dosi di una certa sostanza (dose 1 = 1hg, dose 2 = 2hg, dose 3 = 3hg, dose 4 = 4hg).
   La variabile "dose di fertilizzante" è un esempio di variabile:
   ∇ continua ∇ discreta ∇ categoriale

3. Una azienda in cerca di personale ha effettuato una selezione tra tutti i candidati che si sono presentati. Qual è l'unità statistica?

4. Indicare la natura e le modalità delle seguenti variabili: Sesso, Numero di figli, Reddito familiare, Prezzo all'ingrosso di una sostanza, Corso di laurea frequentato, Altezza, Religione.