

Statistica I

Unità A.2: Concetti preliminari

Tommaso Rigon

Università Milano-Bicocca

Anno Accademico 2020-2021

Argomenti affrontati

- Terminologia e nozioni di base

Riferimenti al libro di testo

- §2.1 — §2.4
- §2.6

Popolazione. Un **insieme di elementi** (persone fisiche, animali, squadre di calcio, visite ad un sito web), che è oggetto di studio statistico. Gli elementi della popolazione sono chiamati genericamente **unità statistiche**.

Campione. Un qualsiasi sotto-insieme della popolazione di riferimento. Il numero di unità statistiche presenti nel campione si chiama **numerosità campionaria**.

- Una popolazione è detta **reale** quando è effettivamente esistente e visibile, mentre viene detta **virtuale** se è astratta o riferita al futuro.
- **Esempio 1.** Le linee degli autobus che hanno circolato a Milano durante il mese scorso sono una popolazione reale.
- **Esempio 2.** I pazienti affetti da una certa patologia (passati e futuri) sono una popolazione virtuale.

Ulteriore terminologia

Variabile. Una **caratteristica** dell'unità statistica che viene rilevata.

Modalità (di una variabile). I **valori distinti** assunti da una variabile.

- **Esempio 1 (continuazione)**. La variabile “numero di passeggeri” saliti sulla linea 61 nell'ultimo mese assume le modalità $0, 1, 2, 3, \dots$
- **Esempio 2 (continuazione)**. La variabile binaria “trattamento” indica se un paziente è stato curato o meno tramite una certa terapia. Può assumere le modalità “Si” oppure “No”.

Tipologie di variabili

- Una variabile è detta **quantitativa** se le modalità sono numeri, mentre viene detta **qualitativa** in tutti gli altri casi.

Quantitative

- Una variabile quantitativa è detta **discreta** se le modalità che assume sono poste in corrispondenza con l'insieme dei numeri naturali (es: anni di età).
- Una variabile qualitativa è detta **continua** o **reali** se le modalità che assume appartengono ad un intervallo reale (es: PIL di una nazione, temperatura).

Qualitative

- Una variabile qualitativa è detta **nominale** o **sconnessa** se le modalità che assume non seguono alcun ordine (es: religione, professione).
- Una variabile qualitativa è detta **ordinale** se le modalità che assume seguono un ordine pre-costituito (es: valutazioni scolastiche, gradi militari).

La raccolta dei dati condiziona la tipologia

- Si consideri una macchina che deve forare delle lastre di metallo. Il **diametro nominale** dei fori è accettabile se è compreso tra 0.94mm e 1.06mm.
- Allora, dati sulla *qualità* della produzione della macchina, potrebbero essere disponibili nella forma:
 - “buono” oppure “difettoso” (dati **dicotomici** o **binari**);
 - “troppo piccolo”, “buono”, “troppo grande” (dati qualitativi ordinali);
 - lunghezza del diametro (dati numerici continui).
- Le differenze non sono semplicemente dovute a come i dati vengono registrati ma possono essere dovute a come i **diametri** vengono effettivamente **misurati**.
- Ad esempio, raccogliere dati sui diametri nella forma "2" è più rapido e richiede strumenti meno costosi di quanto richiesto dalla forma "3".

Dati sperimentali e dati osservazionali

- Nell'analizzare dei dati è bene poi tenere presente il tipo di studio in cui sono stati rilevati. In particolare, è importante la seguente distinzione.
- **Studi sperimentali.** I dati sono stati raccolti in situazioni replicabili e controllate (es: i gli esperimenti di laboratorio).
- **Studi osservazionali.** Lo statistico semplicemente rileva dei dati già esistenti (es: il numero di presenze alberghiere in una stagione, il prezzo di un'azione, etc.).
- Il problema principale degli studi osservazionali è che non sempre è possibile tenere sotto controllo i fattori che influenzano il fenomeno sotto indagine.

Esempio di matrice dei dati

Paziente	Trattamento	Efficacia	Età	Sesso
Antonio	A	Si	47	Maschio
Francesca	B	Si	56	Femmina
Marco	A	No	33	Maschio
Giulia	B	No	60	Femmina
Marta	A	Si	42	Femmina
...
Matteo	B	No	49	Maschio

- Un **campione** di pazienti affetti da una certa patologia per cui sono state registrate le **variabili** trattamento, efficacia, età e sesso.
- La variabile “efficacia” indica se i sintomi della patologia sono scomparsi ad una settimana dal trattamento.