



UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

Dipartimento di Comunicazione
ed Economia

Campione e strategie di campionamento



UNIMORE Scienze
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA
della comunicazione

Analisi dei dati per la ricerca sociale
A. A. 2021/22
Elvira Pelle

La popolazione

L'obiettivo ultimo è studiare qualche caratteristica di una **popolazione**.

Si definisce **popolazione** l'insieme degli elementi che costituiscono l'oggetto di studio.

Può essere:

- **finita**: costituita da un numero finito di unità
- **infinita**: costituita da tutte le unità potenzialmente osservabili e non necessariamente già esistenti fisicamente



Può essere:

- *finita*: costituita da un numero finito di unità
- *infinita*: costituita da tutte le unità potenzialmente osservabili e non necessariamente già esistenti fisicamente

Esempi di popolazioni finite:

- la popolazione italiana
- il numero di aziende tessili di Reggio Emilia
- gli studenti immatricolati in un dato a.a. al C.d.L. in SCO

Esempi di popolazioni infinite

- tutte le persone affette da una patologia, oggi o in futuro
- i futuri acquirenti di un certo prodotto
- i pezzi difettosi che può produrre una macchina



In altri termini, una popolazione è una collettività le cui componenti sono dette **unità statistiche**

Si definisce **unità statistica** l'unità elementare su cui vengono osservati i caratteri oggetto di studio

Esempi di unità statistiche:

- I cittadini italiani
- le aziende tessili
- gli studenti immatricolati in un dato A.A.
- i clienti di un negozio.



Censimento

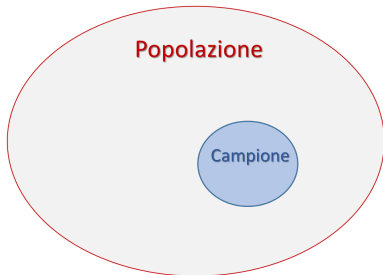
Quando la raccolta dati si attua osservando tutti gli individui di una popolazione parliamo di **censimento**

- Vantaggi
 - Ricchezza delle informazioni raccolte
 - Esaustività
- Limiti:
 - difficoltà nel condurlo: ci sono sempre individui difficili da localizzare. **E questi individui potrebbero avere caratteristiche che li distinguono dal resto della popolazione.**
 - Costi molto elevati.
 - Tempi di elaborazione dei dati molto lunghi



Campionamento

Quando non possiamo (vogliamo) osservare l'intera popolazione facciamo ricorso al campionamento, cioè osserviamo una parte della popolazione



Si definisce **campione statistico** un insieme di unità selezionato, secondo determinati criteri, dalla popolazione



Descrivere e inferire

- ▶ Con **inferenza** si intende il procedimento con cui generalizziamo i risultati parziali osservati sul campione ad una popolazione più generale.
- ▶ Passando dal “piccolo” al “grande”, si introduce naturalmente incertezza. La statistica gestisce questa incertezza utilizzando un linguaggio appropriato, il **calcolo della probabilità**.
- ▶ L'insieme dei metodi che ci permettono di effettuare questa generalizzazione costituiscono il corpo della **statistica inferenziale**.



Come dev'essere il campione

Quando diciamo che il campione sono n individui selezionati nella popolazione, questo non vuol dire che qualunque gruppo di n individui vada bene.

Si definisce **campione “rappresentativo”** un sottoinsieme della popolazione che ne riflette le caratteristiche. (Una versione in miniatura della popolazione.)

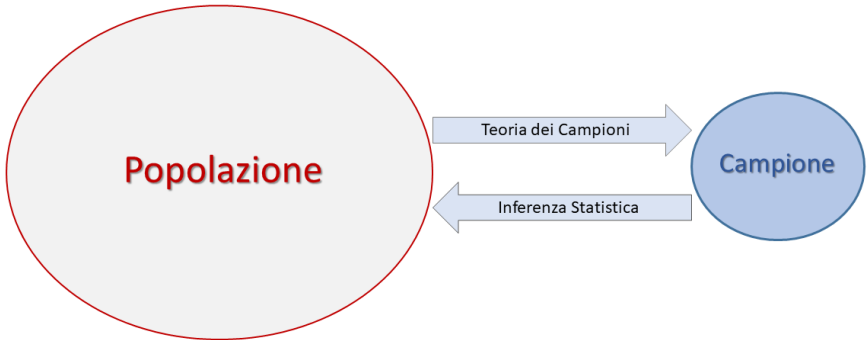
È il fatto che il campione è rappresentativo che consente di estendere alla popolazione i risultati sul campione.

L'insieme dei metodi che definiscono come estrarre un campione costituiscono il corpo della **teoria dei campioni**.

Nota Bene: Per grande che sia, un campione non rappresentativo non consente generalizzazioni.



Riassumendo...



Notazione

- *Parametri della popolazione*: costanti che descrivono aspetti peculiari della distribuzione del carattere di interesse nella popolazione (ad esempio, media e varianza)
 - Nel seguito indichiamo con lettere greche i parametri nella popolazione e con lettere latine quelli calcolati sul campione (ad esempio, indichiamo con μ e σ^2 la media e la varianza nella popolazione e con \bar{X} e S^2 la media e la varianza campionarie).
- *Dimensione campionaria*: numero di unità che compongono il campione
 - Nel seguito indichiamo con N la dimensione della popolazione e con n quella del campione.
- *Frazione di campionamento*: rapporto tra dimensione campionaria e dimensione della popolazione ($\frac{n}{N}$)
- *Regola di selezione*: strategia di campionamento (come selezionare le unità da inserire nel campione).



Come ottengo un campione rappresentativo?

L'idea è di selezionare le unità da includere nel campione in modo casuale \Rightarrow *la regola di selezione è di tipo probabilistico (la selezione del campione avviene seguendo una specifica distribuzione di probabilità)*

È necessario individuare:

- l'insieme di tutti i possibili campioni che si possono estrarre dalla popolazione
- la probabilità di essere estratto associata ad ogni campione

Questi due aspetti definiscono la *strategia di campionamento*.



Strategie di campionamento

Ci concentriamo solo su popolazioni finite

- Il più semplice è scegliere n individui in modo **indipendente** e che **ciascun individuo della popolazione abbia la stessa probabilità di essere estratto: campione casuale semplice**
- Altre opzioni sono spesso usate allo scopo di
 - migliorare la rappresentatività;
 - semplificare la procedura (risparmiare quattrini).

tra queste

- campione stratificato;
- campione a grappoli;
- campione a più stadi.

Vale sempre il principio che tutti possono essere estratti, le probabilità possono variare.

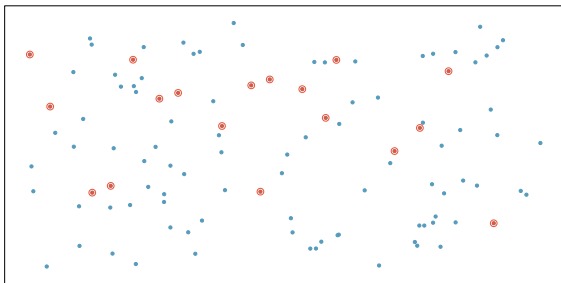


Campionamento casuale semplice (ccs)

I campioni di uguale dimensione hanno tutti la stessa probabilità di essere estratti

Problemi:

- disponibilità di liste di campionamento
- reperibilità di tutte le unità della popolazione
- le unità devono essere estratte in modo casuale



Estrarre le unità casualmente

- Ad ogni unità della popolazione si associa un numero (etichetta)
- si estraggono casualmente i numeri che corrispondono alle etichette

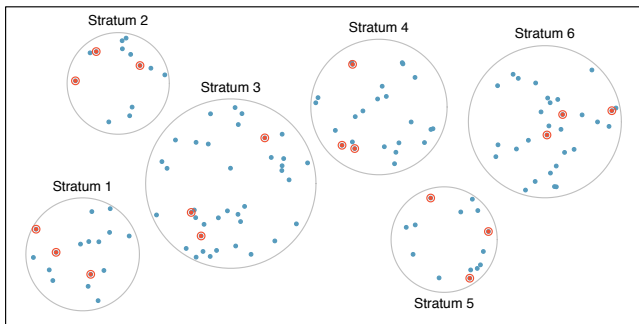
L'estrazione può avvenire

- **Con ripetizione** \Rightarrow la probabilità di includere un'unità nel campione NON cambia da estrazione a estrazione
- **Senza ripetizione** \Rightarrow la probabilità di includere un'unità nel campione CAMBIA da estrazione a estrazione



Campionamento stratificato

La popolazione è divisa in **strati** di unità statistiche simili. Si prende un ccs da ogni strato.



Come formare gli strati? Informazioni aggiuntive sulla popolazione



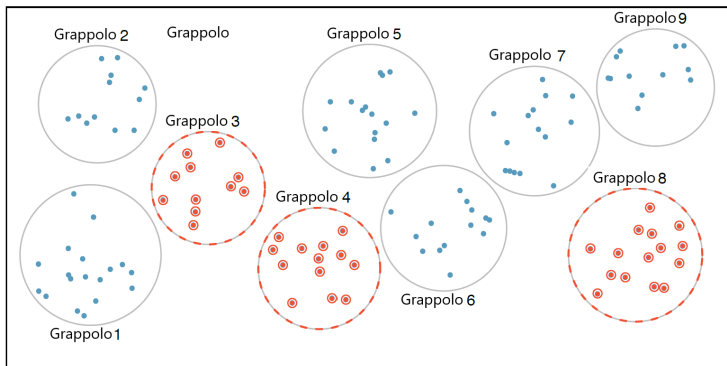
Campionamento stratificato vs ccs

- migliori stime (se gli strati sono ben fatti)
- possibilità di ottenere stima per i singoli strati (oltre che per l'intera popolazione)



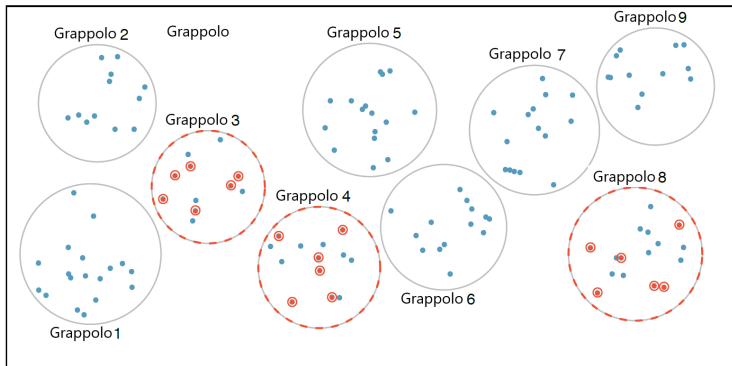
Campionamento a grappolo

La popolazione è divisa in **grappoli**, che usualmente sono formati da osservazioni non omogenee. Si prende un ccs di grappoli.



Campionamento a due stadi

Si estrae un ccs di grappoli, poi un ccs da ciascun grappolo.



Popolazione e parametro

- ▶ Consideriamo una *popolazione*,
- ▶ Di questa popolazione interessa una certa caratteristica, questa è detta *parametro*,



Campione

- ▶ Non possiamo/vogliamo osservare tutta la popolazione (troppo grande, non c'è tempo),
- ▶ Scegliamo **in modo casuale** alcuni individui: il *campione*.

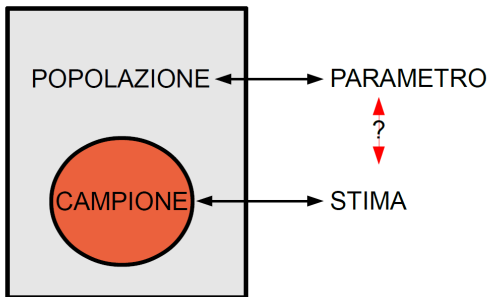


Campione e stima

- ▶ Osserviamo le caratteristiche rilevanti su questi individui.
- ▶ Ad esempio:
- ▶ Sulla base di questi otteniamo una **stima** del parametro,



Schematicamente: popolazione-parametro, campione-stima



- Non è ovvio, ancora, cosa sia una stima e come sia collegata al parametro.
- Per intanto una stima è una caratteristica quantitativa del campione che si ritiene 'simile' al parametro.