

Il Campionamento

La popolazione di riferimento

La popolazione (o universo) di una ricerca è quell'insieme dei casi che teoricamente costituiscono l'oggetto di indagine e che hanno in comune almeno una caratteristica osservabile.

La popolazione di riferimento

La definizione della popolazione dipende dagli obiettivi di ricerca. Ad esempio:

- indagine sui redditi in Italia -> popolazione: individui che percepiscono un reddito in Italia
- indagine sulle prossime elezioni politiche -> popolazione: tutti gli italiani con diritto di voto e con età uguale o superiore a 18 anni
- indagine sull'impatto della riforma universitaria -> popolazione: tutti gli studenti universitari iscritti in uno degli Atenei universitari italiani
- Indagine sul gradimento di un programma televisivo -> popolazione: tutte le persone che hanno visto quel programma

La rilevazione censuaria

In una rilevazione censuaria tutte le unità che compongono la popolazione di riferimento entrano a far parte dell'indagine.

L'esempio più classico è quello del **censimento generale della popolazione e delle abitazioni** svolto dall'Istat ogni 10 anni.

In questa rilevazione tutti i capifamiglia compilano un questionario riguardante tutti i membri della famiglia e i conviventi raggiungendo, così, tutta la popolazione italiana.

La rilevazione campionaria

In quasi tutte le situazioni **reali** di ricerca è impossibile esaminare ogni singola unità della' intera popolazione a causa di:

- limitate risorse – economiche, di personale di tempo – disponibili
- l' intera popolazione da studiare non è fisicamente raggiungibile
- l' intera popolazione da studiare non è del tutto nota

Per superare questi problemi si ricorre ad una rilevazione campionaria ossia una rilevazione che permette di **studiare le caratteristiche di una popolazione attraverso lo studio di una porzione della popolazione stessa (campione) invece dell' intera popolazione.**

Campione e Campionamento

- **Campione:** parte della popolazione selezionata in modo da diminuire, in termini di tempo e costi, l'acquisizione tutte le informazioni utili per la ricerca. Il campione è la rappresentazione in piccolo di tutta la popolazione, sintetizza, cioè tutte le caratteristiche della popolazione originaria.

Scegliere un campione da una popolazione significa effettuare un “campionamento”.

- **Campionamento:** Procedimento attraverso il quale si estrae, da un insieme di unità (**popolazione**), un numero finito di casi (**campione**) che siano rappresentativi di tutta la popolazione e scelti con criteri tali da consentire la generalizzazione (**inferenza**) all'intera popolazione a partire dai risultati ottenuti studiando il campione.

Campioni probabilistici e non probabilistici

Ci sono due famiglie di campionamenti:

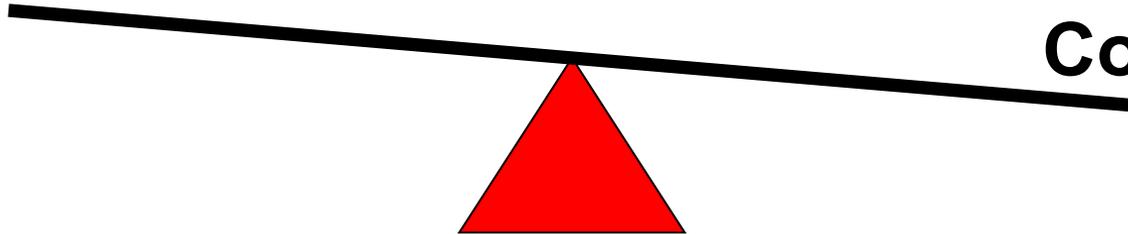
•**campionamento probabilistico**: è la procedura di campionamento in cui i casi vengono scelti in modo che la probabilità che di ciascun caso ha di essere incluso nel campione è nota

•**campionamento non probabilistico**: è la procedura di campionamento in cui i casi vengono scelti partendo da una situazione in cui non si ha accesso alla lista completa della popolazione di riferimento. In questo tipo di campionamento NON si conosce la probabilità che ogni singolo caso ha di essere incluso nel campione

Campionamento

Precisione

Costo



Errore di campionamento

L'EC è un errore casuale insito nelle procedure di formazione del campione.

La “vera” caratteristica della popolazione è (per definizione) ignota, l' errore di campionamento non può mai essere determinato con esattezza.

•L' errore di campionamento è rappresentato dalla differenza tra i risultati ottenuti dal campione e la vera caratteristica della popolazione che vogliamo stimare.

Normalmente si accetta un errore di campionamento inferiore o pari al 5%.

L'intervallo di confidenza, in questo caso, sarà del 95%.

Tipi di campionamento probabilistico

- Campionamento casuale semplice
- Campionamento sistematico
- Campionamento stratificato
- Campionamento multistadio
- Campionamento a grappoli (*cluster*)
- Campionamento per aree

Il Campionamento casuale semplice

Per **campionamento casuale semplice** si intende un campionamento creato da un universo al cui interno ogni singola unità ha la stessa probabilità di un'altra di essere estratta.

La popolazione considerata è necessariamente finita, altrimenti non si potrebbe definire una lista di campionamento.

Il Campionamento casuale semplice

Come si costruisce questo campione?

Dopo aver definito una lista di campionamento in cui a ogni soggetto viene associato un numero d'ordine, ha inizio l'estrazione dei singoli individui (servendosi, ad esempio, di un generatore di numeri casuali) e, nel caso in cui ogni soggetto estratto non venga più reimmesso per le estrazioni successive, avremo un **campionamento casuale semplice**.

(Es. estrazione della lotteria)

Esempio: Campionamento casuale semplice

1	Alberto D.	25	Monica Q.
2	Riccardo D.	26	Régine D.
3	Bella H.	27	Lucille L.
4	Raimondo L.	28	Geremia W.
5	Stéfania B.	29	Gilles D.
6	Alberto T.	30	Renaud S.
7	Giampaolo V.	31	Piero K.
8	André D.	<u>32</u>	<u>Michele R.</u>
<u>9</u>	<u>Daniele C.</u>	33	Marie M.
10	Antonio Q.	34	Gaétano Z.
11	Giacomo B.	35	Fidèle D.
12	Daniele G.	36	Maria P.
13	Amanda L.	37	Anne-Marie G.
14	Giovanna L.	38	Michele K.
15	Filippa K.	39	Gastone C.
16	Eva F.	<u>40</u>	<u>Aldo M.</u>
17	Piera O.	41	Olivier P.
<u>18</u>	<u>Tommaso G.</u>	42	Ginevra M.
19	Brian F.	43	Berll D.
20	Elena H.	44	Gianna P.
21	Isabella R.	45	Giacomo B.
22	Gianna T.	46	François P.
23	Samanta D.	47	Domenica M.
24	Berll L.	48	Antonio C.

Il Campionamento sistematico

Il **campionamento sistematico** si usa quando le unità che compongono l'universo sono *numerabili* progressivamente e consiste nell'estrarre delle unità di campionamento distanziate da un intervallo costante (cosiddetto passo di campionamento).

Supponiamo di campionare n unità da una popolazione di N unità.

R , detta anche *ragione del campionamento* o *intervallo fisso di estrazione*, rappresenta l'intervallo costante che intercorre tra le singole unità enumerate.

Il Campionamento sistematico

Quindi se l'unità di partenza sarà i , le successive unità di campione saranno determinate nel seguente modo:

- $n_1=i$; $n_2=i+R$; $n_3=i+2R$...

e così procedendo fino all'estrazione di tutte le unità di campione.

Esempio

- $N = 1000$; $n = 100$; $k = 1000/100 = 10$ (R)

- Si sceglie per iniziare un nr compreso tra 1 e 10 e poi se ne prende uno ogni 10

Il Campionamento stratificato

Se l'universo della ricerca non è omogeneo si procede alla costruzione di un **campionamento stratificato**, che si caratterizza per la suddivisione della popolazione in strati.

Gli strati sono classi omogenee a seconda della dispersione supposta di una data variabile (o di più variabili contemporaneamente).

Dopo la suddivisione in k strati omogenei, si procede all'estrazione casuale dei casi in ciascuno dei k strati.

Il campione complessivo è costituito dalla somma dei k subcampioni estratti dagli strati.

Il Campionamento stratificato

Il campione si forma estraendo a caso un certo numero di unità da ogni gruppo che è al suo interno omogeneo.

Alla fine si combinano i diversi campioni così ottenuti creando un campione probabilistico.

Occorre però prestare attenzione alla scelta delle unità, alla dimensione del campione, all'omogeneità delle unità dell'universo.

Esempio: Ricerca sul reddito degli italiani

- Si divide la popolazione per professione in 4 strati: operai impiegati, autonomi e professionisti.
- Estraiamo in modo casuale un campione da ciascuno strato e si unifica il campione.

Il Campionamento multistadio

Per ogni stadio il ricercatore deve decidere le caratteristiche di delimitazione, la numerosità delle unità da estrarre, le probabilità di inclusione e le tecniche adeguate di selezione.

Il campionamento a stadi *non richiede la completezza della lista dei casi*, ma *solo una lista delle unità primarie aggregative* che vengono estratte casualmente.

La lista completa delle unità è necessaria solo in un momento successivo, a livello di unità primarie aggregative estratte.

Il Campionamento multistadio

Esempio di tale situazione è dato dall' anagrafe che non esiste come unico archivio nazionale ma è suddivisa negli 8.103 comuni italiani.

In questo caso si procede:

1. estrazione casuale di un campione di comuni (unità di primo stadio)
2. estrazione di un campione casuale di famiglie (unità di secondo stadio) da ciascuna lista anagrafica per ogni comune selezionato

Il Campionamento a grappoli

Adatto per popolazioni molto grandi (ma finite), questo tipo di campionamento risulta utile per ridurre i costi di rilevazione, attraverso campionamenti intermedi che costituiscono campionamenti di primo, secondo ..., r-esimo stadio.

Lo si utilizza quando la popolazione risulta *naturalmente* divisa in gruppi / grappoli

Si scelgono innanzi tutto delle unità primarie, nell'ambito di queste unità primarie si scelgono delle unità secondarie o di secondo stadio e così via fino a giungere all'unità statistica che l'indagine si prefigge di rilevare (ad esempio >> Istituti superiori/Sezioni/Classi/tutti gli studenti delle classi (grappolo) campionate casualmente

Il Campionamento a grappoli

È un tipo di campionamento a stadi.

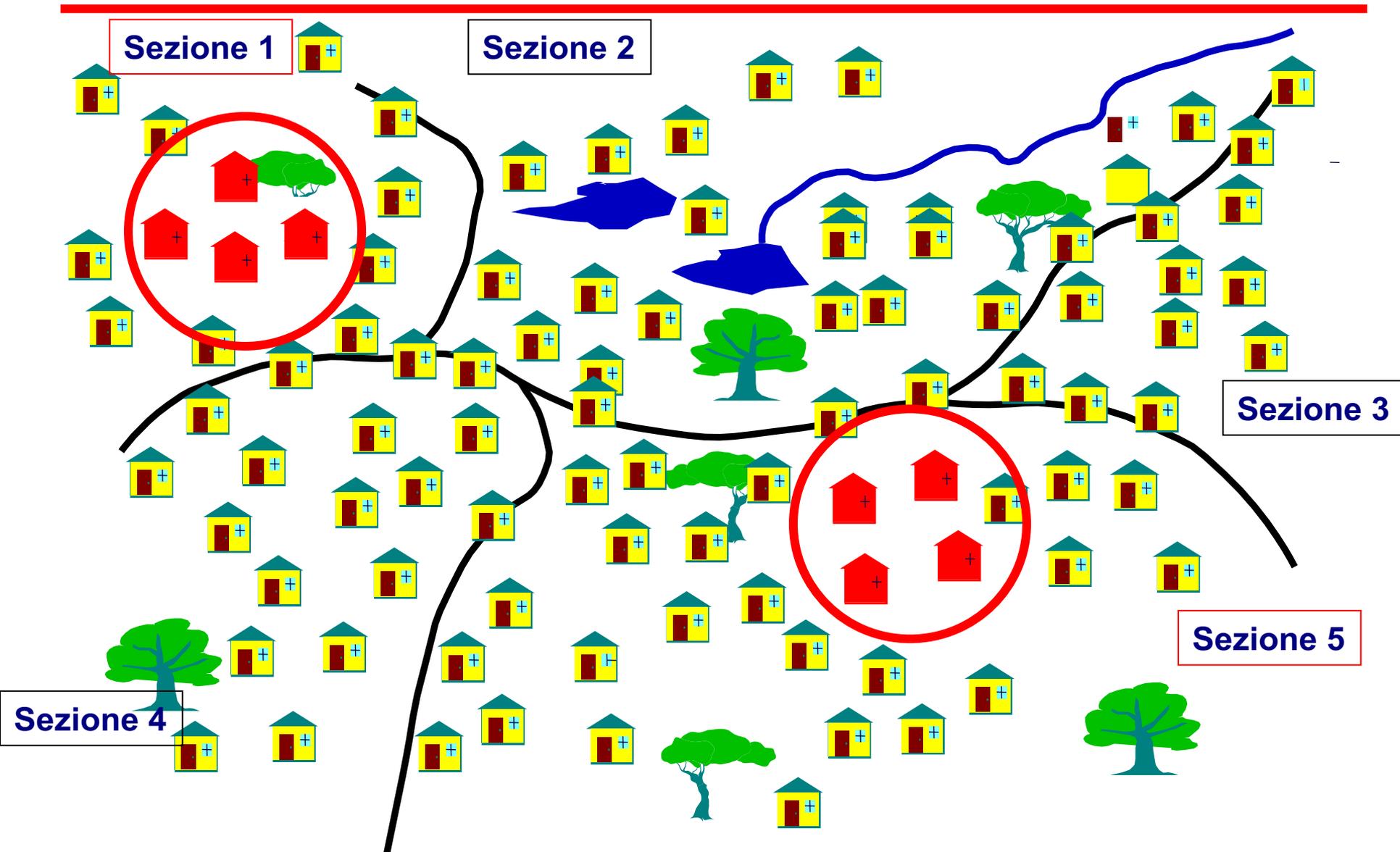
Lo si utilizza quando la popolazione risulta naturalmente divisa in gruppi / grappoli (Esempio: classi scolastiche, reparti aziendali ecc.).

Questa tecnica di campionamento è stata proposta nelle indagini di mercato per evitare di avere una distribuzione dei campioni in un territorio molto vasto. Le procedure da adottare prevedono di dividere la popolazione in *cluster* (solitamente delle zone geografiche). Si estraggono a caso dei *cluster* e si procede per l' estrazione del campione dai *cluster* selezionati.

•Esempio Ricerca sugli studenti

•Istituti superiori/Sezioni/Classi/tutti gli studenti delle classi (grappolo) campionate casualmente

Esempio: Campionamento a grappoli



PROBLEMI LEGATI AL CAMPIONAMENTO PROBABILISTICO

Il primo problema nei campioni di tipo probabilistico è la differenza fra il campione previsto e quello raggiunto >>> **mortalità del campione.**

Un soggetto che non risponde non può essere frettolosamente sostituito con il successivo nella lista di campionamento poiché il suo rifiuto lo rende di per sé differente dal soggetto che lo sostituirà.

In statistica si distingue solitamente tra:

-errori campionari, i quali dipendono dalle modalità di costruzione del campione e riguardano **errori di disegno** e/o stima;

-errori non campionari, che si caratterizzano per tutti quegli altri elementi di distorsione che influiscono sugli stimatori, ma non dipendono dalle operazioni pratiche di definizione del campione.

Il Campionamento NON probabilistico

Oltre ai campioni casuali e probabilistici, esistono altri procedimenti nei quali il criterio di estrazione dei casi si dice “***a scelta ragionata***”.

In tali procedimenti il ricercatore, sulla base delle sue necessità cognitive, determina le regole per stabilire quali casi andranno a far parte del campione.

A questo tipo di campionamento NON è possibile applicare i procedimenti dell' inferenza statistica, quindi i risultati ottenuti con campioni non probabilistici NON sono generalizzabili all' intera popolazione di riferimento.

Tipi di Campionamento NON probabilistico

- Campionamento per quote
- Campionamento a scelta ragionata
- Campionamento bilanciato
- Campionamento a valanga
- Campionamento tipologico-fattoriale

Il Campionamento per quote

La popolazione viene divisa in un certo numero di strati massimamente omogenei al loro interno *rispetto a variabili considerate rilevanti* ai fini della ricerca (come per il campionamento stratificato).

La numerosità interna degli strati è proporzionale alla numerosità dei rispettivi strati nella popolazione.

Gli individui non vengono, però, estratti casualmente da ciascuno strato, bensì scelti dagli intervistatori che sono solo obbligati a rispettare le quote.

Esempio: indagini telefoniche con interviste stabilite per quote (sesso, classe di età, istruzione, ecc.).

Il Campionamento a scelta ragionata

Questo tipo di campionamento è usato nel caso in cui il campione sia di ampiezza limitata per evitare oscillazioni casuali rispetto a determinate caratteristiche della popolazione che si vogliono tenere sotto controllo.

Le unità campionarie vengono scelte sulla base di alcune loro caratteristiche.

Il Campionamento a scelta ragionata

Esempio: Campione dell' Istituto Cattaneo IARD

- I Comuni italiani sono stati classificati in 20 strati derivanti dall'incrocio fra la dimensione del comune (meno di 5000 abitanti, 10-50 mila; 50-100 mila; oltre 100 mila) e zona geografica (Nord, Centro, Sud e Isole).
- Per ogni strato si è scelto un numero di comuni sulla base del peso demografico della popolazione residente (Es. la popolazione dei comuni del nord con 5000 abitanti è il 6% della popolazione totale sono stati scelti 6 comuni)

Il Campionamento bilanciato

È un tipo di campione a scelta ragionata nel quale si selezionano i casi tenendo sotto controllo il rispetto delle distribuzioni di queste *proprietà*: la selezione dei casi si effettua in maniera tale che le medie o le proporzioni del campione risultino prossime a quelle note della popolazione per determinate variabili (es reddito, stato di occupazione, patologie croniche etc.)

Il Campionamento a valanga

Molto utilizzato per lo studio di fenomeni non istituzionalizzati o di natura clandestina, questo tipo di campionamento si caratterizza per la concatenazione dei contatti.

Si basa, infatti, sul legame di rete che intercorre tra i soggetti coinvolti nella ricerca: si parte dall'intervista di un gruppo ristretto di soggetti noti, ai quali si chiede di indicare altri soggetti appartenenti alla stessa categoria o che svolgono lo stesso ruolo.

Il Campionamento a valanga -2

Questo tipo di campionamento è particolarmente utile nei casi in cui non si conosca a priori né l'entità del fenomeno né la lista dei soggetti su cui effettuare la rilevazione.

La concatenazione dei contatti, però, comporta il *pericolo* di ottenere una certa omogeneità delle informazioni dal momento che gli intervistati tenderanno ad indicare soggetti simili con esperienze simili oppure i soggetti più attivi all'interno di una determinata categoria sociale.

Esempio: Indagine sugli immigrati senza permesso di soggiorno

Si contatta un immigrato, lo si sottopone ad intervista e poi gli si chiede di indicare un altro immigrato di sua conoscenza disposto a rilasciare l'intervista.

Parametri

- L'**intervallo di confidenza** esprime il *marginale statistico d'errore*. Ad esempio, in un campione il 47% ha risposto "sì" a una certa domanda. Con un intervallo di confidenza 4 (cioè del 4%) la percentuale di persone che risponderebbero "sì", nell'eventualità di un'intervista a tappeto a tutta la popolazione, sarebbe compresa fra il 43% ($47\% - 4\%$) e il 51% ($47\% + 4\%$).
- Il **livello di confidenza** esprime il *grado di certezza* del risultato. Continuando con l'esempio precedente, porre il livello di confidenza al 95% significa che col 95% di probabilità la percentuale di persone che risponderebbero "sì", nell'eventualità di un'intervista a tappeto a tutta la popolazione, sarebbe compresa fra il 43% ($47\% - 4\%$) e il 51% ($47\% + 4\%$).

Calcolo delle ampiezze

- <https://www.demetra.com/calcolo-numerosita-campione/>

Campionamento probabilistico

Siti:

<https://www.random.org>

<https://www.blia.it/utali/casuali/>

Fattori di *distorsione* nei dati

(cfr. lezione questionario)

- Cattiva qualità dello strumento
- Contesto situazionale inadeguato
- Effetti distorcenti delle risposte (desiderabilità sociale...)
- **Insufficiente professionalità degli intervistatori**
- **Alto tasso di non risposta totale (al questionario)**
- **Alto tasso di non risposta parziale (ad alcune parti)**
- Inadeguatezza di codifica e analisi dei dati