

Il campionamento

Campionamento

Si definisce **campionamento** il procedimento attraverso cui:

dall'**insieme di unità** costituenti l'oggetto dello studio si estrae un **numero ridotto di casi** scelti con criteri tali da consentire la **generalizzazione** dei risultati **all'intera popolazione**.

Risulta necessario che il campione rappresenti correttamente ed efficientemente l'universo:

le statistiche devono essere una buona stima dei parametri della "popolazione di riferimento".

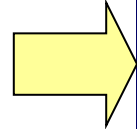


Quanti elementi devono comporre il campione

- Per quanto concerne il **dimensionamento campionario**, a prescindere dal metodo scelto, è possibile affermare che il numero di elementi da includere nel campione varia in funzione del livello desiderato di precisione dei risultati, del grado di accuratezza dell'analisi, dell'estensione ed omogeneità degli elementi della popolazione, delle risorse e del tempo disponibili.
- *“Si tratta di trovare il giusto equilibrio fra necessità della precisione e quella di contenere tempi e costi dell'indagine”.*
- La precisione di una ricerca cresce con l'aumentare della dimensione campionaria, all'aumentare del campione, minori sono i margini di errore. Da evidenziare che la **precisione di un'indagine è funzione solo e soltanto della grandezza del campione e non della grandezza della popolazione di riferimento.**

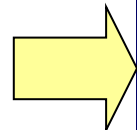
La numerosità campionaria

UNIVERSO O POPOLAZIONE



È l'insieme finito o infinito di unità, **definito nei contenuti, nello spazio e nel tempo**, oggetto dell'indagine statistica.

CAMPIONE



È costituito da un certo numero di unità, estratte con un procedimento da una popolazione, al fine di **rappresentarla** rispetto ai caratteri oggetto di studio.

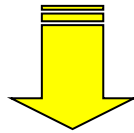
“La numerosità ottima di un campione è quella che consente di ottenere gli obiettivi dell'indagine al minimo costo e sarà il numero minimo in base al quale le stime raggiungeranno il livello di attendibilità atteso. ”

La numerosità campionaria

- E' importante quando si decide l'ampiezza campionaria, stimare quante volte si dovrà suddividere il campione durante l'analisi; ad ogni suddivisione dovrà essere garantita un'adeguata ampiezza del campione.
- Un altro aspetto che facilmente viene trascurato è che l'ampiezza del campione necessaria per l'analisi dei dati varia di fatto da domanda a domanda, a seconda della percentuale di intervistati che non rispondono a ciascuna particolare domanda.

Gli errori di Campionamento

- **Errore casuale:** è determinato dalle variazioni casuali del campione e può essere stimato ed eventualmente ridotto aumentando la numerosità del campione stesso.
- **Errore sistematico:** è più difficile da cogliere poiché direttamente connesso alle problematiche relative al piano di campionamento e quindi al metodo ed alle tecniche di rilevazione utilizzate.



- comportamento degli intervistatori
- comportamento degli intervistati
- errori nella fase di elaborazione dati

Tecniche di campionamento

Vi sono vari tipi di campionamento, che si differenziano sostanzialmente in due categorie:

○ Campionamenti probabilistici

- Si ha un campione probabilistico quando ad ogni unità della popolazione compete una probabilità nota e non nulla di essere estratta.
- Si utilizza la statistica inferenziale (stimatori ed intervalli di confidenza) che forniscono informazioni sulla popolazione completa.

○ Campionamenti non probabilistici

- La scelta degli elementi della popolazione viene effettuata **in base a criteri logici fissati a priori**. In questi casi è consentito esclusivamente descrivere il risultato campionario con gli opportuni strumenti della statistica descrittiva.

Campionamento probabilistico

In questi tipi di campionamento le unità della popolazione hanno prefissate probabilità di essere incluse nel campione:

➤ **casuale semplice:** (**con reinserimento:** ciascuna unità possiede la stessa probabilità di essere selezionata pari ad $1/N$ (N dimensione popolazione); **senza reinserimento:** ogni unità ha probabilità $1/N$ alla prima estrazione, le rimanenti hanno probabilità $1/(N-1)$ e così via)

➤ **stratificato:** (proporzionale, uniforme, ottimale)

Il procedimento di stratificazione consiste:

1. nel raggruppare le unità della popolazione **in strati** il più possibile omogenei al loro interno rispetto al carattere (od ai caratteri) investigati. Gli strati costituiscono una partizione: sono a due a due disgiunti e complessivamente esaustivi; (es. casalinghe, dagli operai, dagli studenti, ecc...)
2. nell'estrarre casualmente un certo numero di unità campione da ciascuno strato.

➤ **sistematico:** Si calcola il "passo di campionamento" $k=N/n$ e si estrae a caso un numero r compreso fra 1 e k . Il numero r identifica la prima unità, dopodiché se ne estraggono sequenzialmente una ogni k . (E' il metodo utilizzato dall'ISTAT per estrarre dalle liste anagrafiche.)

Campionamento probabilistico

- **a grappoli:** Per "grappolo" si intende un sottoinsieme degli elementi della popolazione che viene trattato come un'unica entità ai fini del campionamento. Generalmente i grappoli sono precostituiti:
 - sedi universitarie (che raccolgono gli elenchi degli studenti)
 - aziende industriali (che raggruppano gli operai dell'industria)
 - comuni, province e regioni (che costituiscono dei raggruppamenti delle persone e delle cose comprese nel loro territorio)

- **a più stadi:** Consiste nella selezione delle "unità di primo stadio", all'interno delle quali vengono selezionate alcune unità di uno stadio successivo, e di seguito.

Nel caso in cui gli stadi siano solamente due (ad esempio selezione di comuni nel 1° stadio e di famiglie nel 2°) è possibile individuare le seguenti fasi.

1. Individuazione delle unità di primo stadio, dette unità primarie, che costituiscono dei raggruppamenti delle unità di secondo stadio, dette unità secondarie. Non devono esistere dubbi sull'appartenenza di ogni unità di secondo stadio ad una data unità di primo stadio.
 2. Formazione della lista delle unità di primo stadio.
 3. Selezione casuale di alcune unità di primo stadio.
 4. Selezione casuale di alcune unità di secondo stadio da ciascuna unità di primo stadio estratta. Il campione a due stadi risulta costituito da tutte le unità secondarie prescelte.
- **ripetuti (panel e panel ruotati)** Supponiamo di voler analizzare una popolazione che muta nel tempo e di voler acquisire informazioni su di essa ad intervalli di tempo, non necessariamente regolari. Possiamo procedere in diversi modi per quanto riguarda le modalità di selezione delle unità statistiche da rilevare.
 1. Selezionare CAMPIONI INDIPENDENTI ad ogni occasione di indagine
 2. Mantenere fisso il campione selezionato alla prima occasione ed osservare sempre le stesse unità statistiche nelle rilevazioni successive (PANEL)
 3. Sostituire, in ciascuna occasione, una parte delle unità statistiche della rilevazione precedente con altre unità estratte dalla popolazione (CAMPIONE RUOTATO)

Campionamento non probabilistico

- Campionamento per **scelta ragionata**: La scelta delle unità campionarie non è di natura casuale, bensì viene effettuata sulla base di informazioni riguardanti la popolazione indagata. È appropriato per piccoli campioni. Si identificano le zone dove si trova il maggior interesse per il fenomeno in studio (es. Se per motivi economici l'indagine deve essere limitata a 2 città, queste saranno scelte con criteri mirati, avendo cura che appartengano alla tipologia più rilevante ai fini della ricerca)

- Campionamento **per quote**: **È la tecnica di campionamento non probabilistico più utilizzata, in particolare nelle indagini di mercato e nei sondaggi di opinione.** Si definisce la percentuale di interviste con persone aventi determinate caratteristiche (fumatori-sesso)
 - Un campione per quote si ottiene come segue.
 1. La popolazione viene suddivisa in classi o sottogruppi omogenei, sulla base di caratteristiche legate a quella indagata
 2. Dai dati censuari o da altre fonti si ricava il peso percentuale di ogni classe
 3. Il totale delle unità da inserire nel campione viene suddiviso tra le classi in modo da rispecchiare le proporzioni esistenti nella popolazione e si perviene quindi alla definizione delle quote, cioè il numero di interviste da effettuare in ciascuna classe.
 4. La scelta delle unità da intervistare viene generalmente demandata all'intervistatore stesso, nell'ambito delle quote assegnate.

- Campionamento **tramite testimoni privilegiati**: si intervistano esclusivamente persone esperte del fenomeno in studio (doppio lavoro).

- Campionamento **a valanga**: composto da più fasi, nella prima fase vengono identificate e intervistate alcune persone dotate delle caratteristiche richieste, queste persone sono utilizzate come informatori per identificare altri soggetti che possono essere intervistati in una fase successiva e che a loro volta producono informazioni per identificare altri soggetti con le caratteristiche per essere inclusi nel campione, creando così un effetto a valanga.

Uso dell'oversampling

- A volte il campione anziché essere composto rispetto all'esatta proporzione della composizione dell'universo è volutamente reso più ampio rispetto ad una certa caratteristica.
- A tale espediente si ricorre ***quando si vuole indagare un fenomeno che ha una limitata penetrazione*** sull'universo e di cui si vuole ingrandire la portata.
- **L'Oversampling**, aumentando la sua portata, **augmenta la precisione del fenomeno** stesso permettendo di coglierlo nei contorni desiderati.
- Naturalmente, **se si vuole conoscere l'universo** depurato dai fattori turbativi dell'Oversampling, si potrà ricorrere ad una **riponderazione**, riportando alla sua dimensione naturale quella parte di campione arbitrariamente ingrandita.