

# IL CAMPIONAMENTO

## Popolazione e campione

- **Campionamento:** procedimento attraverso il quale si estrae, da un insieme di unità (popolazione) costituenti l'oggetto dello studio, un numero ridotto di casi (campione) scelti con criteri tali da consentire la generalizzazione all'intera popolazione dei risultati ottenuti studiando il campione.

- La rilevazione campionaria presenta:
  - a) vantaggi nei costi di rilevazione;
  - b) vantaggi nei tempi di raccolta dati e di elaborazione;
  - c) vantaggi organizzativi,
  - d) vantaggi di approfondimento e di accuratezza
- Problema del campionamento nella ricerca sociale → problema della selezione dei casi

- **Popolazione:** un insieme  $N$  (ampiezza della popolazione) di unità (dette anche «unità statistiche» o «unità di analisi») che costituiscono l'oggetto del nostro studio
- **Campione** l'insieme delle  $n$  (ampiezza del campione) unità campionarie (che chiamiamo «casi») selezionate tra le  $N$  unità che compongono la popolazione, allo scopo di rappresentarla (da cui l'uso dell'espressione «campione rappresentativo») ai fini del nostro studio.
- La popolazione è l'oggetto da conoscere; il campione è lo strumento della conoscenza.
- **Campionamento** la procedura che seguiamo per scegliere le  $n$  unità campionarie dal complesso delle  $N$  unità della popolazione

- La stima del campione sarà affetta da un errore, che chiamiamo **errore di campionamento**.
- Se indichiamo con **V** il valore (sconosciuto) del parametro nella popolazione (una media, una proporzione, un coefficiente di correlazione, ecc.), con **v** il valore trovato nel campione (cioè la sua stima) e con **e** l'errore di campionamento, possiamo scrivere:

$$V = v \pm e$$

$V$	=	$v$	$\pm$	$e$
parametro della popolazione (incognito)		stima del campione		errore di campionamento

## Campioni probabilistici: il campione casuale semplice

- Un campione si dice probabilistico quando ogni unità che lo compone è estratta con una probabilità nota (e diversa da zero)
- Abbiamo un campionamento casuale semplice quando tutte le unità della popolazione hanno la stessa probabilità di essere incluse nel campione
- Errore di campionamento:

[1]

$$e = z \frac{s}{\sqrt{n}} \sqrt{1-f}$$

dove:

$z$  = coefficiente dipendente dal livello di fiducia della stima;

$s$  = deviazione standard campionaria della variabile studiata;

$n$  = ampiezza del campione;

$1-f$  = fattore di correzione per popolazioni finite, dove  $f$  = frazione di campionamento =  $n/N$ .

- **L'errore di campionamento** è direttamente proporzionale al livello di fiducia che vogliamo avere nella stima e alla variabilità del fenomeno studiato, mentre è inversamente proporzionale all'ampiezza del campione.

## Ampiezza del campione

- Quanto grande deve essere il campione?

$$[3] \quad n = \left( \frac{zs}{e} \right)^2 \quad n \equiv n - 1 = \frac{z^2 pq}{e^2}$$

dove:

$z$  = coefficiente dipendente dal livello di fiducia che vogliamo avere nella stima;

$s$  oppure  $\sqrt{pq}$  = variabilità del fenomeno studiato;

$e$  = errore della stima.

## Altri campioni probabilistici:

- Campionamento sistematico
- Campionamento stratificato
- Campionamento a stadi
- Campionamento a grappoli
- Campionamento per aree
- Campioni complessi



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

**I 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.
2. nell'estrarre casualmente un certo numero di unità campione da ciascuno strato.

➤ C. 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.)

## ➤ C. a più stadi:

Consiste nella selezione delle “unità di primo stadio”, all’interno delle quali vengono selezionate alcune unità di uno stadio successivo, e via 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.
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.

➤ C. 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)

# Il campionamento nella ricerca sociale

- la procedura di campionamento produce un **errore di selezione**
- l'errore di selezione può essere distinto in ulteriori tre componenti:
  - **errore di copertura,**
  - **errore di campionamento**
  - **errore di non risposta**

## Errore di copertura. Lista della popolazione

- In tutti i casi in cui non esiste una lista della popolazione occorre rinunciare a tecniche di campionamento probabilistico.
- In questi casi non si può assegnare a tutte le unità della popolazione una certa probabilità di estrazione.
- In questi casi il ricercatore sociale ha di fronte a sé tre possibilità:
  - ridefinire la popolazione
  - trascurare gli esclusi
  - procedere a un'integrazione del campione

## **Errore di campionamento. Ampiezza del campione**

- La dimensione del campione non può più essere determinata a priori dal ricercatore sulla base delle distribuzioni delle singole variabili studiate, ma dovrà far riferimento al tipo di analisi che egli intende utilizzare
- la dimensione del campione viene a dipendere:
  - a) dalla distribuzione delle variabili studiate;
  - b) dal tipo di analisi che si intende fare

## Errore di non risposta

- può avere due cause ben diverse
  - a) il mancato contatto con i soggetti estratti, oppure
  - b) il rifiuto a farsi intervistare di una parte di essi



## Campioni non probabilistici

- Quando il disegno probabilistico non può essere impostato, oppure si sa a priori che non potrà essere attuato nella fase di rilevazione, si ricorre fin dall'inizio ai cosiddetti campioni non probabilistici
  - Campionamento per quote
  - Disegno fattoriale
  - Campionamento a scelta ragionata
  - Campionamento bilanciato
  - Campionamento a valanga
  - Campionamento telefonico
  - Campionamento di convenienza

### ➤ **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.

E' 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:**

E' 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 (es: 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.

# Ponderazione

- Procedura con la quale modifichiamo artificialmente (in sede di elaborazione dei dati, tramite operazioni matematiche) la composizione del campione onde renderla più prossima alla distribuzione della popolazione
- Si realizza attribuendo un peso alle unità campionarie, che varia a seconda delle loro caratteristiche

## Bontà di un campione

- *L'errore campionario* dipende in gran parte dall'ampiezza del campione, mentre *l'errore di copertura* e *l'errore di non risposta* dipendono dall'accuratezza della procedura di campionamento (un campione è tanto più accurato quanto più si avvicina al modello del campione probabilistico). **Ampiezza e accuratezza** rappresentano le due caratteristiche che definiscono la bontà di un campione
- Un campione è **rappresentativo** quando fornisce un'immagine in piccolo ma senza distorsioni della popolazione

# Esercizio

Provate a costruire un *campionamento per quote*:

- Popolazione di riferimento : **Castel di Sangro**
- Intervallo di confidenza (MoE) del 5%
- Livello di confidenza del 95%
- Variabili (dicotomizz.) da considerare:  
Genere+Laurea+Marimonio