

Campionamento

Un procedimento attraverso cui 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 - idealmente- la generalizzazione all'intera popolazione dei risultati ottenuti studiando il campione.

Se avete bisogno di ulteriori riferimenti, potete leggere il capitolo sul Campionamento in una delle versioni del testo di Corbetta

Il campionamento

Statistica: scienza che descrive lo stato (Gerolamo Ghilini, 1589)

Forme di rilevazione statistica per descrivere lo "stato":
testimonianze nell'Antico Testamento, Egiziani, Sumeri,
Romani.....

I censimenti nascono in Europa nel 1800 (Inghilterra, Francia); in
Italia, il primo nel 1861, poi ogni 10 anni

Censimento: Indagine condotta attraverso questionari, interviste o altri strumenti sugli
elementi di un insieme, per ottenere dati numerici su caratteri dell'insieme stesso;
per antonomasia, quella che si fa sulla popolazione di uno stato, di una regione
ecc.: *l'ultimo censimento della popolazione italiana (Garzanti).*

Oggi, l'Istat:

- Censimenti (tradizionalmente ogni 10 anni; ora ogni anno, ma su campioni rappresentativi)
- indagini campionarie ad hoc (salute, "tempi", violenza...)

Storia del campionamento nelle scienze sociali

1895 Congresso internazionale di statistica a Berna ⇒ campionamento ragionato al posto di indagini censuarie

1934: J.Neyman : basi statistiche del campionamento casuale/probabilistico

Negli Stati Uniti: A partire dagli anni '30 ⇒ sondaggi d'opinione, per inchieste di mercato o sondaggi elettorali

Lo sviluppo del campionamento nelle scienze sociali è legato ai sondaggi elettorali (verifica immediata!)

Sondaggi elezioni S.U: errori clamorosi

1936 : Rivista Literary Digest ⇒ questionario postale, a 10 milioni di nominativi tratti da elenchi telefonici e registri automobilistici (in Usa non esiste un'anagrafe) ⇒ 2 milioni di risposte

Predizioni totalmente errate

?

Problema nelle liste di campionamento: errore di copertura e errore di non risposta

⇒ diffusione del campionamento probabilistico

La logica del campionamento

- Il campione è una selezione della popolazione
- L'idea di campione è legata all'idea di popolazione
- Qualità del campionamento, rappresentatività del campione \Rightarrow possibilità di generalizzare i risultati

In una ricerca si può:

- Studiare tutta la popolazione \Rightarrow rilevazione esaustiva
- Studiare un campione probabilistico/casuale \Rightarrow l'obiettivo è di avere un campione rappresentativo della popolazione
- Studiare un campione non probabilistico \Rightarrow non può essere mai rappresentativo

La logica del campionamento

Studiare tutta la popolazione \Rightarrow valore esatto/certo del parametro (percentuali, medie, correlazioni...) che si vuole conoscere

Studiare il campione \Rightarrow valore esatto del parametro nel campione
MA solo una stima, un valore approssimativo del parametro rispetto alla popolazione \Rightarrow nella popolazione, valore probabilistico

Teoria delle probabilità (necessità di un campione casuale): permette di effettuare tali stime e di calcolare il margine di errore della stima

\Rightarrow Intervallo di Fiducia (Confidence Interval, CI)

\Rightarrow Livello di fiducia: di solito ci accontentiamo di avere il 95% di probabilità di essere "nel giusto"

L'ampiezza dell'Intervallo di Fiducia dipende anche dall'ampiezza del campione e dalla frequenza degli elementi analizzati

Campionamento probabilistico

⇒ evitare le distorsioni (bias)

Ogni elemento della popolazione deve avere la stessa probabilità di essere incluso/escluso dal campione

⇒ ci serve una *lista* completa della popolazione

⇒ il campione non può essere più accurato della *lista di campionamento* da cui è tratto

La popolazione non è un oggetto naturale, è un oggetto teorico ⇒ bisogna definirla con chiarezza

Campionamento probabilistico -> randomizzazione

- Campionamento randomizzato/casuale semplice (sorteggio o tavola dei numeri casuali)
- Campionamento sistematico (selezione sistematica)
- Campionamento casuale stratificato (utile se la popolazione è molto eterogenea rispetto alle variabili di interesse)
 - proporzionale
 - non proporzionale

Campionamento a cluster/a stadi/a grappoli/ per aree

⇒ Quando le persone sono state individuate casualmente, bisogna convincerle a partecipare!!!

Campionamento non probabilistico

- Non probabilistico, per quote
 - ⇒ stratificare il campione in funzione della popolazione di riferimento e degli obiettivi della ricerca ⇒ intervistare le persone nei vari strati fino a raggiungere il numero desiderato
 - ⇒ non è "casuale", ogni persona NON ha la stessa probabilità di essere inclusa
- Di convenienza
- A casaccio ("cheap and dirty")

Nella ricerca qualitativa

La ricerca qualitativa risponde a logiche
e usa strategie diverse:

- Il campione non è rappresentativo, né ha la pretesa di esserlo
- Campionamento intenzionale teorico
- Campionamento "a valanga"

L'ampiezza del campione 1

Grosso modo, più è ampio il campione:

- Più sono precise le stime del fenomeno osservato e più è stretto l'intervallo di fiducia
- Minori sono gli errori nel generalizzare i risultati
- Maggiori sono le possibilità di analisi statistiche

L'ampiezza del campione 2

Nelle inchieste psico-sociali ed epidemiologiche, la scelta dell'ampiezza del campione (oltre che dalle risorse disponibili!) dipende da:

- La frequenza - ipotetica- della variabile oggetto di studio nella popolazione
- L'accuratezza richiesta nelle stime
- L'eterogeneità della popolazione studiata
- Il tipo di analisi statistiche previste

Grosso modo \Rightarrow più è ampio il campione, meglio è ... purchè sia rappresentativo!

\Rightarrow L'ampiezza del campione non sostituisce il fatto che sia casuale e rappresentativo

L'ampiezza del campione 3

Le prime ricerche su comportamento sessuale (USA)

1948: Ricerca Kinsey sulla sessualità ⇒ 18.000 interviste ⇒ in parte volontari auto-selezionati, in parte seguendo la logica - approssimativa - delle quote (100% cell method)

1987: Ricerca Hite sulla sessualità femminile (in seguito sulla sessualità maschile, sull'amore ecc.)

4500 questionari! ... MA ne aveva spediti 100.000

Validità?

⇒ L'ampiezza del campione non sostituisce il fatto che sia casuale e rappresentativo