INDAGINE

Completa (censuaria)

Semplice sul piano teorico ma complessa nella pratica

- Popolazioni non finite
- Osservazione distruttiva

Parziale (campionaria)

Più complessa sul piano teorico ma spesso di più facile attuazione

- Costi limitati
- Tempi ridotti
- Numero elevato di informazioni
- Accuratezza nella rilevazione

POPOLAZIONE

insieme finito o infinito di unità che non interessano prese singolarmente ma per il contributo che danno alle proprietà statistiche dell'insieme di appartenenza

↓

POPOLAZIONE OBIETTIVO

- Elementi componenti
- Estensione spaziale
- Estensione temporale

↓

LISTA

CAMPIONE

Qualsiasi sottoinsieme della popolazione

N → dimensione della popolazione

n → dimensione del campione

Importante distinzione:

Campioni probabilistici:

- è noto l'insieme dei possibili campioni;
- è nota la probabilità di selezione di ciascun campione

Campioni non probabilistici

- Tutti gli altri

SELEZIONE PROBABILISTICA (CASUALE)

- Urna (modello teorico)
- Tavole dei numeri casuali (prima dell'avvento dei calcolatori)
- Programmi informatici (oggi)

Piano (disegno) di campionamento

- selezione del campione

Piano (disegno) di indagine

- definizione della popolazione obiettivo
- caratteri (variabili) da studiare, modo di definirli e di osservarli;
- scelta livelli, spaziali e temporali di indagine;
- definizione dei metodi di raccolta, di codifica e di elaborazione dei dati;
- individuazione dei costi e dei livelli di precisione e accuratezza desiderati;
- stima;
- scelta delle analisi statistiche da affiancare ai metodi di stima;
- metodologia di calcolo degli errori campionari;

- i metodi di controllo rilevazione e correzione degli errori non campionari;
- la presentazione di dati statistici e dei risultati.

STIMA

Procedimento statistico mediante il quale un valore ricavato come funzione (elaborazione) delle osservazioni (stimatore) campionarie viene assunto a rappresentare il valore incognito di una grandezza caratteristica (parametro) della popolazione

Parametri di maggiore interesse:

Totali	(occupati, forza lavoro)
Medie	(reddito pro-capite)
Proporzioni	(tasso di occupazione, attività)
Rapporti (tra totali, medie, ecc.)	
	(tasso di disoccupazione)

Proprietà degli stimatori

Correttezza
Efficienza
Consistenza

Sul piano intuitivo vorremmo che la stima (valore numerico dello stimatore) fosse “vicina” al parametro da stimare

T → Stimatore

t → Stima

ϑ → Parametro

Definiamo:

$D = t - \vartheta$ → Errore di stima

Situazione ideale: D = minimo

D non può essere azzerato nell’indagine campionaria, ma è = 0 nei censimenti.

Come è possibile ridurlo nell’indagine campionaria?

(a) Dimensione campionaria

(b) Piano di campionamento

DIMENSIONE CAMPIONE

↑ n   D ↓

OSSERVAZIONE!
n non può essere aumentato liberamente

PIANI DI CAMPIONAMENTO

(a) Campionamento casuale semplice
(b) Campionamento casuale stratificato
(c) Campionamento sistematico
(d) Campionamento a grappoli e a più stadi

Campionamento casuale semplice

Popolazione:

$U_1 \quad U_2 \quad U_3 \quad U_4$

N = 4 n = 2

Possibili campioni {s}:

$(U_1, U_2)\ (U_1, U_3)\ (U_1, U_4)$
$(U_2, U_3)\ (U_2, U_4)\ (U_3, U_4)$

CCS → P(s) = 1/6

STIMA DELLA MEDIA DA CCS

Y carattere (variabile) di studio
$\overline{Y}$ media da stimare
$\overline{y}$ stimatore

Valori nella popolazione:

$Y_1 = 20,\ Y_2 = 40,\ Y_3 = 36,\ Y_4 = 48,$

Possibili campioni :

$(Y_1\ Y_2)\ (Y_1\ Y_3)\ (Y_1\ Y_4)$
$(Y_2\ Y_3)\ (Y_2\ Y_4)\ (Y_3\ Y_4)$

Valori corrispondenti :

(20, 40) (20, 36) (20, 48)
(40, 36) (40, 48) (36, 48)

Medie campionarie	Media popolaz.
30 28 34	36
38 44 42	

PRECISIONE DELLO STIMATORE

Reciproco della varianza (o della sua radice quadrata)

Varianza → $V(\overline{y})$

Errore standard → $\sqrt{V(\overline{y})} = ES(\overline{y})$

$V(\overline{y})$ = ((30-36)^2 + (28-36)^2 + (34-36)^2
+ (38-36)^2 + (44-36)^2 + (42-36)^2)/6
= 34,67

$ES(\overline{y})$ = 5,89 → ?

Poco indicativo in termini assoluti
Più utile in termini relativi

n = 3

↓

Possibili campioni

$(U_1\ U_2\ U_3)\ (U_1\ U_2\ U_4)$
$(U_1\ U_3\ U_4)\ (U_2\ U_3\ U_4)$

Valori corrispondenti

(20 40 36) (20 40 48)
(20 36 48) (40 36 48)

Medie campionarie

32 36 34,67 41,33

La media delle quattro medie è ancora uguale a 36 (stimatore corretto); ma la varianza è inferiore:

$V(\overline{y})$ = 11,55 $ES(\overline{y})$ = 3,4

Calcolo alternativo della varianza dello stimatore

Varianza elementare:

$$S^2 = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}{N-1}$$

Varianza dello stimatore:

$$V(\bar{y}) = \frac{S^2}{n}\left(1 - \frac{n}{N}\right)$$

In termini numerici, per n = 3

S^2 = ((20-36)^2 + (40-36)^2 +(36-36)^2
+ (48-36)2)/3
= 138,67

$$V(\bar{y}) = \frac{138{,}67}{3}\left(1 - \frac{3}{4}\right) = 11{,}55$$

Campionamento casuale stratificato

1. Suddivisione popolazione in sub-popolazioni (strati);
2. Selezione di campioni indipendenti da ciascuno strato.

Obiettivi:
- stimatori più precisi
- dominii di studio

$$N_h \Rightarrow \Sigma_h N_h = N$$

$$n_h \Rightarrow \Sigma_h n_h = n \qquad (h = 1, \dots, H)$$

$$W_h \Rightarrow \Sigma_h W_h = 1$$

Parametro da stimare:

$$\bar{Y} = \Sigma_h W_h \bar{Y}_h$$

Stimatore:

$$\bar{y}_{st} = \Sigma_h W_h \bar{y}_h$$

Varianza:

$$V(\bar{y}_{st}) = \Sigma_h W_h^2 V(\bar{y}_{st})$$

Stratificazione proporzionale

Frazione di campionamento costante:

$$f_h = \frac{n_h}{N_h} = \frac{n}{N} = f$$

vantaggi:

(i) per la popolazione generale stime più precise rispetto al ccs;
(ii) facilità di applicazione, con un numero limitato di strati;

svantaggi:

(i) precisione diversa strato per strato;
(ii) difficoltà di applicazione con molti strati.

Stratificazione non proporzionale

Ottimale:

$$f_h \propto \frac{S_h}{\sqrt{c_h}}$$

(c_h = costo di rilevazione unitario)

Uguale precisione in ogni strato:

se possiamo ipotizzare che

$$S_1 \cong S_2 \cong \ldots \cong S_H$$

allora:

$$n_1 \cong n_2 \cong \ldots \cong n_h$$

CAMPIONAMENTO SISTEMATICO

Intervallo di selezione

$$k = \frac{N}{n}$$

Es:

$N = 1500$ $n = 100$

$k = 15$

Numero casuale tra 1 e 15;
Supponiamo 6

6 6+15 6+2x15 . . . . 6+99x15

cioè

6 21 36 1491

CAMPIONAMENTO A GRAPPOLI E A PIÙ STADI

Grappoli: campionamento di aggregati di unità di studio;

Stadi: campionamento di aggregati (o unità) da aggregati di livello gerarchico superiore

Motivazioni:

- Non disponibilità lista unità
- Costi

Attenzione!

Strati → omogenei al loro Interno

Stadi → eterogenei al loro interno