

COME VALUTARE UNA STIMA

- Problema di stima. Si vuol stimare il valore di una grandezza incognita Θ relativa alla popolazione (ad es. la media)
- Si dispone di un campione di n osservazioni x_1, x_2, \dots, x_n estratte da una popolazione di N unità
- La stima sarà un valore
$$t_n = f(x_1, x_2, \dots, x_n)$$
calcolato sulle n unità campionarie
- Impossibile giudicare della stima sulla base della differenza $t_n - \Theta$ (essendo Θ incognito)
- Si guarda allora alla distribuzione campionaria dello stimatore
$$T_n = f(X_1, X_2, \dots, X_n)$$
ovvero a come la stima si comporta con riferimento all'universo dei campioni (tutte le possibili n -ple estraibili dalle N unità della popolazione)
- Due concetti chiave
 - Correttezza** $E(T_n) = \Theta$
 - Efficienza** (precisione) legata a $\text{Var}(T_n)$
- La $\text{Var}(T_n)$ (e quindi la precisione) è generalmente legata alla variabilità della popolazione (in modo diretto) e alla numerosità campionaria (in modo inverso)

PREGI E DIFETTI DEL CCS

VANTAGGI

facile trattamento teorico

*neutralità della selezione
(buono per qualsiasi
obiettivo d'indagine)*

SVANTAGGI

*trascura eventuali
informazioni a priori*

*può comportare alti costi
(dispersione territoriale)*

*richiede una lista
completa*

Impossibile stimare N

Tecniche alternative di selezione che cercano di sfruttare informazioni a priori con l'intento di aumentare la precisione o diminuire il costo

campione stratificato

campione a stadi e/o a grappoli

campione sistematico

campione con probabilità variabile

La maggior parte delle indagini hanno un piano di campionamento complesso che mette insieme più tecniche (es. campione Forze di lavoro)

CAMPIONE STRATIFICATO

- Si utilizzano informazioni a priori sulla struttura della popolazione
- La popolazione viene suddivisa in classi (strati) rispetto a uno o più caratteri noti
- Da ciascuno strato si estrae in modo indipendente un campione casuale semplice di numerosità n_k in modo tale che la somma $\sum n_k$ sia uguale alla numerosità n desiderata.

OBIETTIVO: ridurre la variabilità **entro** gli strati per aumentare la precisione delle stime
(si dimostra che la varianza dello stimatore di una media dipende dalla varianza entro gli strati e non da quella tra gli strati)

ASPETTATIVA: la riduzione di variabilità dovrebbe essere garantita dalla relazione tra carattere oggetto di indagine e caratteri di stratificazione

un esempio didattico

popolazione di 4 unità
 carattere X (n° sigarette giornaliere)
 carattere Y (sesso)

| | | | |
|----|----|---|----|
| A | B | C | D |
| 10 | 25 | 5 | 40 |
| f | m | f | m |

$$M(X) = 20 \quad \text{Var}(X) = 187,5$$

CAMPIONE di 2 elementi per stimare $M(X)$

Universo
campioni

CCS (\bar{x})

CStrat(\bar{X}_{st})

| | | |
|---------|---------|---------|
| AB 17,5 | AC 7,5 | AD 25 |
| BA 17,5 | BC 15 | BD 32,5 |
| CA 7,5 | CB 15 | CD 22,5 |
| DA 25 | DB 32,5 | DC 22,5 |

| | |
|---------|---------|
| AB 17,5 | AD 25 |
| CB 15 | CD 22,5 |

Distribuzione
campionaria

| \bar{x} | $p(\bar{x})$ |
|-----------|--------------|
| 7,5 | 2/12 |
| 15 | 2/12 |
| 17,5 | 2/12 |
| 22,5 | 2/12 |
| 25 | 2/12 |
| 32,5 | 2/12 |

| \bar{X}_{st} | $p(\bar{X}_{st})$ |
|----------------|-------------------|
| 15 | 1/4 |
| 17,5 | 1/4 |
| 22,5 | 1/4 |
| 25 | 1/4 |

$$E(\bar{x}) = 20$$

$$\text{Var}(\bar{x}) = 62,5$$

$$E(\bar{X}_{st}) = 20$$

$$\text{Var}(\bar{X}_{st}) = 15,75$$

Un esercizio con verifica [STRATI.xls](#)

problemi per la stratificazione

- individuazione caratteri di stratificazioni correlati con quello oggetto di studio
- decisione sul numero di strati
- disponibilità di una lista con le informazioni sui caratteri di stratificazione
- scelta frazione di campionamento in ogni strato (proporzionale vs ottimale)
- criterio di stratificazione buono per un carattere può non esserlo per un altro (problema se i caratteri da stimare sono molti)

CAMPIONE A GRAPPOLI E/O A STADI

- **Tra gli inconvenienti del CCS**
 - *formazione della lista (costo, impossibilità)*
 - *eccessiva dispersione territoriale*
 - *impossibilità di stimare N*
- **Si usano unità di campionamento intermedie (grappoli di unità)**
 - *è sufficiente una lista di grappoli*
 - *si riduce la dispersione territoriale se i grappoli sono unità territoriali*
 - *è possibile stimare N*
- **A STADI:** formazione di una lista di unità nei grappoli estratti e successiva estrazione casuale di unità

Grappolo ideale:

alta varianza **entro** grappoli e bassa **tra** grappoli
(*si dimostra che la varianza dello stimatore di una media dipende dalla varianza tra i grappoli e non da quella entro i grappoli*)

CAMPIONE SISTEMATICO

- **Si dispone di una lista delle N unità della popolazione ordinata da U_1 ad U_N . Fissata la numerosità campionaria n , si sceglie casualmente un numero $1 < i < k = N/n$ e il campione risulta formato dalle unità**

$$U_i, U_{i+k}, U_{i+2k}, \dots, U_{i+(n-1)k},$$

- **Assimilabile al campione stratificato se le liste sono ordinate casualmente (è come se si stratificasse la popolazione in n strati formati da K unità contigue e si estraesse una sola unità da ogni strato in modo non indipendente)**
- **Se la lista è ordinata rispetto ad un qualche criterio (es: territorio) si assicura la presenza nel campione di unità collocate lungo tutta la lista**
- **se la lista è ciclica (es. fatture ordinate per giorno della settimana) si rischia di introdurre una distorsione**

C. CON PROBABILITA' VARIABILE

- **Le unità sono estratte con prob. proporzionale al valore di una variabile nota.**

esempi:

comuni estratti con prob. proporzionale alla loro ampiezza demografica

aziende agrarie estratte con probabilità proporzionale alla loro superficie

- **Nel calcolare una stima, il valore delle singole unità verrà pesato con il reciproco della probabilità di selezione**

π_i = prob. di estrazione della i-esima u.s.



Stima del totale di Y $\rightarrow \sum_i y_i (1/\pi_i)$

Es

x_i =area azienda i; T_x =area totale; $\pi_i=x_i/T_x$;

y_i =produzione grano azienda i.

Stima totale di Y = $\sum_i y_i (1/\pi_i) = \sum_i (y_i/x_i) T_x$

resa per ettaro 
superficie totale 

NB: in questo modo si possono riponderare le famiglie estratte dalle liste elettorali (la probabilità di estrazione dipendendo dal numero di componenti maggiorenni)