

TEORIA E TECNICA DELL'INDAGINE STATISTICA E DEL CAMPIONAMENTO (MATR.DISPARI)

CAMPIONAMENTI COMPLESSI

MANUELA SCIONI

Dipartimento Di Scienze Statistiche

manuela.scioni@unipd.it



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



PERCHÉ DISEGNI CAMPIONARI COMPLESSI

- Il CCS è spesso un riferimento teorico poco applicabile nella realtà, ad esempio:
 - Non è realizzabile con interviste faccia a faccia su tutta la popolazione italiana. I tempi e i costi sarebbero altissimi
 - Serve una lista unica di tutta la popolazione. Spesso tale lista non è disponibile, ma sono disponibili liste separate (es. anagrafi), e metterle insieme sarebbe molto dispendioso
- Spesso vogliamo esercitare un maggior controllo sulla selezione, pur mantenendola casuale
 - Es: vogliamo che una popolazione rara sia comunque rappresentata nel campione

LA LISTA

Elenco delle unità che costituiscono la popolazione statistica.

- La lista dovrebbe essere: completa, senza duplicazioni, stabile nel tempo, informatizzata
- Lista semplice:
 - lista di etichette che corrispondono uno a uno alle unità della popolazione

CAMPIONAMENTO CASUALE SEMPLICE

LA LISTA (2)

Problemi:

- Spesso non è disponibile un'unica lista, ma è frammentata (es: le anagrafi comunali)
- La lista può non essere disponibile (es. i clienti di un supermercato)
- Predisporre la lista completa può essere estremamente oneroso, se non impossibile
- Lista complessa, ovvero composta da più liste:
 - Distinte per sottopopolazioni
 - Gerarchiche
 - Dinamiche

Campionamenti complessi

In realtà non è necessario disporre della lista completa (lista semplice): essa va definita univocamente nei contenuti, ma poi, a seconda del disegno di campionamento, è sufficiente disporre delle parti di lista da cui si estrae il campione (lista complessa).

EFFICIENZA

- **Efficienza:** è il livello di precisione dello stimatore in rapporto a uno preso come riferimento

$$\text{Efficienza } (\hat{\mathcal{G}}' | \hat{\mathcal{G}}) = \frac{\text{Var}(\hat{\mathcal{G}}')}{\text{Var}(\hat{\mathcal{G}})}$$

- Il riferimento è generalmente dato dal CCS
 - Alcuni disegni complessi sono più efficienti del CCS
 - Altri disegni sono meno efficienti del CCS

DEFF: EFFETTO DEL DISEGNO DI CAMPIONAMENTO

$Var(\hat{\theta}')$: varianza di stima con un campione “complesso” di numerosità n

$Var(\hat{\theta})$: varianza di stima con un campione casuale semplice di numerosità n

$$DEFF = \frac{Var(\hat{\theta}')}{Var(\hat{\theta})}$$

$DEFF=1$: il campionamento complesso non ha portato nessun vantaggio rispetto al CCS

$DEFF < 1 \implies Var(\hat{\theta}') < Var(\hat{\theta})$ il campionamento complesso è più efficiente del CCS

$DEFF > 1 \implies Var(\hat{\theta}') > Var(\hat{\theta})$ il campionamento complesso è meno efficiente del CCS

UTILIZZI DEL DEFF

- Se il DEFF è noto a priori, può essere utilizzato per stimare la varianza di stima o la numerosità campionaria del campione complesso: una volta calcolato n per CCS, la si moltiplica per il DEFF e si ottiene la numerosità campionaria nel campione complesso necessaria per ottenere la stessa precisione delle stime
- Il DEFF può essere “trasportato” da un’indagine a un’altra con contenuti e struttura simili, per cui è possibile che sia noto a priori
- Ogni variabile di interesse ha un suo DEFF