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

CAMPIONAMENTO STRATIFICATO: ALLOCAZIONE PROPORZIONALE

MANUELA SCIONI
Dipartimento di Scienze Statistiche
manuela.scioni@unipd.it

Università degli Studi di Padova
Dipartimento di Scienze Statistiche

ALLOCAZIONE PROPORZIONALE

- **Allocazione proporzionale:** la frazione di campionamento è la stessa all'interno di ogni strato

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

- Il campione è **autoponderante.**
- Stima: considerando che $W_h = \frac{N_h}{N} = \frac{n_h}{n}$, le formule sono così semplificate:

$$\bar{y}_{pr} = \sum_{h=1}^{H} W_h\,\bar{y}_h = \sum_{h=1}^{H} \frac{n_h}{n}\,\bar{y}_h = \sum_{i=1}^{n} \frac{y_i}{n}$$

$$V(\bar{y}_{pr}) = \sum_{h=1}^{H}\left(1-\frac{n_h}{N_h}\right)W_h^2\frac{S_h^2}{n_h} = \sum_{h=1}^{H}(1-f)\frac{n_h^2}{n^2}\frac{S_h^2}{n_h} = \frac{1-f}{n}\sum_{h=1}^{H} W_h S_h^2$$

La varianza di stima dipende dalla varianza entro gli strati

CAMPIONE STRATIFICATO PROPORZIONALE VS. CCS (1)

$$V_{ccs}\left(\overline{y}\right) - V_{pr}\left(\overline{y}_{str}\right) = \frac{1-f}{n}\left(S^2 - \frac{1}{N}\sum_{h=1}^{H} N_h\, S_h^2\right)$$

Ma:

$$\begin{aligned} S^2 &= \frac{1}{N-1}\left[\sum_{h=1}^{H}\sum_{j=1}^{N_h}\left(Y_{hj} - \overline{Y}_{hU} + \overline{Y}_{hU} - \overline{Y}_U\right)^2\right] \\ &= \frac{1}{N-1}\left[\sum_{h=1}^{H}(N_h - 1)S_h^2 + \sum_{h=1}^{H} N_h(\overline{Y}_{hU} - \overline{Y}_U)^2\right] \end{aligned}$$

CAMPIONE STRATIFICATO PROPORZIONALE VS. CCS (2)

Posto $N_h-1 \cong N_h$ e $N-1 \cong N$:

$$S^2 \cong \frac{1}{N}\left[\sum_{h=1}^{H} N_h\, S_h^2 + \sum_{h=1}^{H} N_h \left(\overline{Y}_{hU} - \overline{Y}_U\right)^2\right]$$

$$V_{ccs}\left(\overline{y}\right) - V_{pr}\left(\overline{y}_{str}\right) \cong \frac{1-f}{n\, N} \sum_{h=1}^{H} N_h \left(\overline{Y}_{hU} - \overline{Y}_U\right)^2$$

Il campionamento stratificato proporzionale è più efficiente del campionamento casuale semplice a parità di dimensione campionaria. L'efficienza è direttamente proporzionale alla varianza fra strati.

CAMPIONE STRATIFICATO PROPORZIONALE VS. CCS (3)

Se si svolgono gli stessi calcoli usando l'espressione non approssimata di S^2 si ottiene:

$$V_{ccs}(\overline{y}) - V_{pr}(\overline{y}_{str}) = \frac{1-f}{n(N-1)} \cdot \left[\sum_{h=1}^{H} N_h \left(\overline{Y}_{hU} - \overline{Y}_U \right)^2 - \frac{1}{N} \sum_{h=1}^{H} (N - N_h) S_h^2 \right]$$

In realtà tale quantità è positiva se:

$$\sum_{h=1}^{H} N_h \left(\overline{Y}_{hU} - \overline{Y}_U \right)^2 > \frac{1}{N} \sum_{h=1}^{H} (N - N_h) S_h^2$$

Ovvero se la variabilità tra le medie risulta sufficientemente grande a confronto della variabilità interna agli strati.

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

CAMPIONAMENTO STRATIFICATO: ALLOCAZIONE OTTIMALE

MANUELA SCIONI
Dipartimento di Scienze Statistiche
manuela.scioni@unipd.it

ALLOCAZIONE OTTIMALE

- Qualora le varianze e/o i costi all'interno degli strati siano notevolmente differenti, è consigliabile campionare con frazioni più elevate gli strati per cui:
  - la varianza entro lo strato è elevata
  - non è costoso campionare nello strato
  - lo strato è molto numeroso.
- Si dimostra che l'allocazione ottima richiede che n_h sia proporzionale a $\frac{N_h S_h}{\sqrt{c_h}}$

$$n_h = n \frac{W_h S_h / \sqrt{C_h}}{\sum_h W_h S_h / \sqrt{C_h}}; \quad \text{Se i costi sono costanti}: \; n_h = n \frac{W_h S_h}{\sum_h W_h S_h}$$

Allocazione di Neyman

VARIANZA DI STIMA ED EFFICIENZA DELL'ALLOCAZIONE DI NEYMAN

$$V_{NEY}(\bar{y}_{str}) = \frac{1}{n}\left(\sum_{h=1}^{H} w_h S_h\right)^2 - \frac{1}{N}\sum_{h=1}^{H} w_h S_h^2$$

$$V_{prop}(\bar{y}_{str}) - V_{NEY}(\bar{y}_{str}) = \frac{1}{n}\sum_{h=1}^{H} w_h (S_h - \bar{S})^2$$

Il guadagno dell'allocazione di Neyman rispetto all'allocazione proporzionale è tanto maggiore quanto più le varianze di strato sono diverse tra loro.

Attenzione: il campione non è più autoponderante.