

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

CAMPIONAMENTO STRATIFICATO

Descrizione
Variabili di stratificazione

MANUELA SCIONI

Dipartimento di Scienze Statistiche

manuela.scioni@unipd.it



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



LA STRATIFICAZIONE

Stratificare una popolazione significa ripartirla in sottopopolazioni, dette strati.

PERCHÉ SI STRATIFICA?

1. Perché la popolazione è naturalmente organizzata in sottopopolazioni
2. Per evidenziare insiemi di unità significative per la ricerca (come unità rare, sottoclassi della popolazione, gruppi estremi,...)
3. Per separare sottopopolazioni fisicamente isolate e con caratteristiche speciali (residenti in istituti assistenziali, in aree terremotate o alluvionate,...)
4. Per individuare unità da rilevare con tecniche particolari (es: telefono)
5. Per introdurre il massimo del controllo nella selezione, pur mantenendola casuale
6. Per rendere omogenee le sottopopolazioni rispetto alle variabili da rilevare, cosicché le stime sono più efficienti rispetto a quelle ottenibili col CCS

LE VARIABILI DI STRATIFICAZIONE

- Deve esserci una relazione fra variabili di stratificazione e la variabile di interesse.
- Il campionamento stratificato è efficiente se gli strati sono omogenei al loro interno e molto diversi fra loro (ovvero se è grande la varianza fra gruppi ed è piccola la varianza entro gruppi)
- È più efficiente usare più variabili di stratificazione anziché più modalità della stessa variabile
- Per indagini sul territorio sono sempre utilizzate le ripartizioni amministrative come variabile di stratificazione. Altre variabili utilizzate sono: la densità della popolazione, la distinzione fra zona rurale e urbana, l'altimetria
- Le variabili di stratificazione sono variabili qualitative, o variabili quantitative ridotte in classi

IL CAMPIONAMENTO STRATIFICATO

Il campionamento stratificato consiste nella selezione di un campione probabilistico all'interno di ciascuno degli strati e nell'utilizzo combinato degli stimatori ottenuti in ogni strato. Si può formare un campione con un criterio diverso in ogni strato

Scelte per formare un campione stratificato



NOTAZIONE PER UN CAMPIONAMENTO STRATIFICATO

Strato	h	$1, \dots, h, \dots, H$	Popolazione
Unità componenti	N_h	$N_1, \dots, N_h, \dots, N_H$	N
Unità campionarie	n_h	$n_1, \dots, n_h, \dots, n_H$	n
“Peso”	$W_h = N_h / N$	$W_1, \dots, W_h, \dots, W_H$	1
Varianza interna	S_h^2	$S_1^2, \dots, S_h^2, \dots, S_H^2$	S^2
Frazione di campionamento	$f_h = n_h / N_h$	$f_1, \dots, f_h, \dots, f_H$	f

PARAMETRI DI POPOLAZIONE

$$t_h = \sum_{j=1}^{N_h} Y_{hj}$$

Totale dello strato h

$$t = \sum_{h=1}^H t_h$$

Totale di popolazione

$$\bar{Y}_{hU} = \frac{\sum_{j=1}^{N_h} Y_{hj}}{N_h}$$

Media di strato

$$\bar{Y}_U = \frac{t}{N} = \sum_{j=1}^{N_h} \sum_{h=1}^H \frac{Y_{hj}}{N}$$

Media di
popolazione

$$s_h^2 = \sum_{j=1}^{N_h} \frac{(Y_{hj} - \bar{Y}_{hU})^2}{N_h - 1}$$

Varianza di strato

STIME CAMPIONARIE

$$\bar{y}_h = \frac{\sum_{j \in c_h} Y_{hj}}{n_h}$$

Media campionaria di strato (non distorta)

$$\hat{t}_h = \frac{N_h}{n_h} \sum_{j \in c_h} Y_{hj} = N_h \bar{y}_h$$

Totale campionario di strato (n.d.)

$$s_h^2 = \sum_{j \in c_h} \frac{(Y_{hj} - \bar{y}_h)^2}{n_h - 1}$$

Varianza campionaria di strato

$$\hat{t}_{str} = \sum_{h=1}^H \hat{t}_h = \sum_{h=1}^H N_h \bar{y}_h$$

Stima del totale

$$\bar{y}_{str} = \frac{\hat{t}_{str}}{N} = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_h = \sum_{h=1}^H w_h \bar{y}_h$$

Stima della media

STIME DI
STRATO

STIME
COMPLESSIVE

STIME CAMPIONARIE (2)

La varianza dello stimatore:

$$V(\hat{t}_{str}) = V\left(\sum_{h=1}^H \hat{t}_h\right) = \sum_{h=1}^H V(\hat{t}_h) = \sum_{h=1}^H (1 - f_h) N_h^2 \frac{S_h^2}{n_h}$$

È stimata da:

$$\hat{V}(\hat{t}_{str}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{s_h^2}{n_h}$$

e, di conseguenza:

$$\hat{V}(\bar{y}_{str}) = \frac{1}{N^2} \hat{V}(\hat{t}_{str}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) \left(\frac{N_h}{N}\right)^2 \frac{s_h^2}{n_h}$$

STIMA DI UNA PROPORZIONE

Si noti che in tal caso

$$\bar{y}_h = \hat{p}_h \quad \text{e} \quad s_h^2 = \left(\frac{n_h}{n_h - 1} \right) \hat{p}_h (1 - \hat{p}_h)$$

Quindi:

$$\hat{p}_{str} = \sum_{h=1}^H \frac{N_h}{N} \hat{p}_h$$

$$\hat{V}(\hat{p}_{str}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h} \right) \left(\frac{N_h}{N} \right)^2 \frac{\hat{p}_h (1 - \hat{p}_h)}{n_h - 1}$$