Il Campionamento a Grappoli

grappoli a dimensioni costanti

Il campionamento a grappoli:

- Presupposto:
  - popolazione suddivisa, in modo artificiale o naturale, in sottoinsiemi di unità, dette *elementari*, legati da vincoli di contiguità spaziale e/o istituzionale*

    * in relazione ad un criterio logico o naturale che può fondarsi su legami di natura fisica, amministrativa, professionale, relazionale, ecc.

→ I gruppi (grappoli o *cluster*) non hanno necessariamente uguale dimensione

Il campionamento a grappoli (GR):

Esempi di grappoli e unità elementari:
- le famiglie (grappoli) per un'indagine su individui (unità elementari)
- le classi (grappoli) per un'indagine sugli alunni (unità elementari)

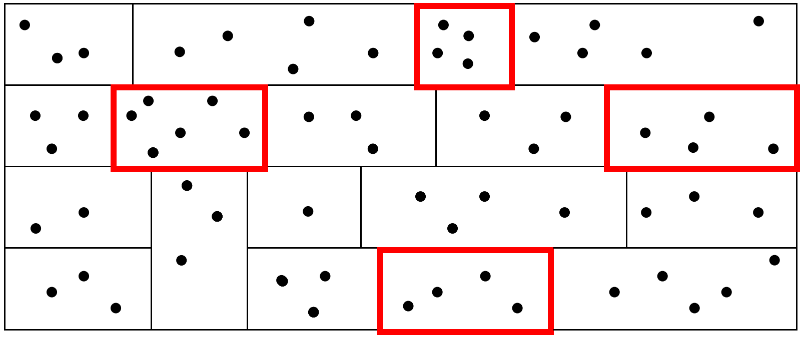
le *unità elementari* (individui-alunni) non vengono scelte in modo diretto, ma vengono estratti alcuni 'grappoli' (famiglie-classi) e tutte le unità del grappolo entrano a far parte del campione

Esempio:
città suddivisa in **aree** definite su una mappa in modo da includere una o più vie e di ampiezza comparabile per numero di abitazioni:

si estrae un campione casuale semplice di aree (grappoli)
si includono nel campione tutte le famiglie residenti nelle aree estratte (unità elementari)

Il campionamento a grappoli (GR):

La figura seguente illustra un esempio di unità della popolazione (i puntini) racchiusi in grappoli (i riquadri). Il GR estrae un numero definito di riquadri (ad esempio quelli in rosso) e il campione sarà formato da tutte le unità contenute nei grappoli estratti.



Il piano di campionamento genera campioni di ampiezza variabile nel caso in cui i grappoli contengano un numero variabile di unità

Quando è utile un GR:

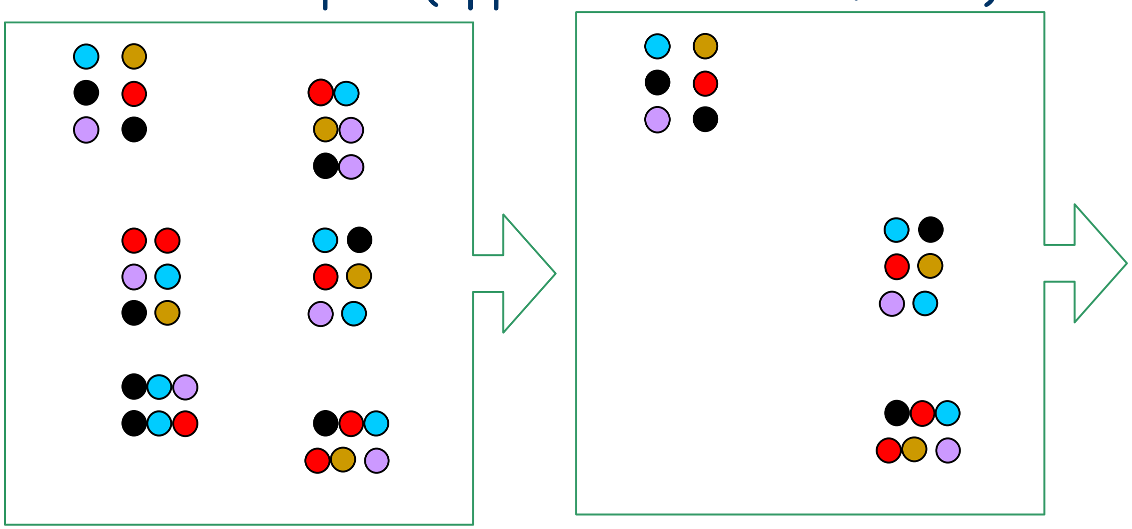
- Non è disponibile una lista delle unità elementari della popolazione, ma si ha quella dei grappoli, usata per l'estrazione;
- i costi di rilevazione aumentano con la distanza fra le unità da intervistare (se i grappoli sono unità geografiche);
- per facilitare l'organizzazione e l'esecuzione della rilevazione;
- nelle analisi dove i cluster costituiscono le unità di interesse.

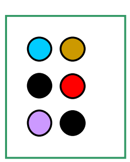
Caratteristiche principali del GR:

- Oggetto di estrazione sono i grappoli e non le unità elementari;
  - Se hanno uguale dimensione i grappoli generalmente sono estratti secondo un campionamento casuale semplice;
  - Se i grappoli hanno numerosità diverse, il valore della dimensione campionaria si configura come una variabile casuale e il campionamento è a probabilità variabili.
- Oggetto della stima può essere la variabile di interesse
  - nei grappoli;
  - nelle unità elementari.

Il campionamento a grappoli

Situazione tipica (opposta allo stratificato)



 grappolo ● Unità elementare

Ipotesi di base:

Affinché il campionamento sia efficace, poiché solo alcuni grappoli fanno parte del campione, essi dovrebbero racchiudere tutta la variabilità della caratteristica da indagare:

- massima eterogeneità interna
- massima omogeneità esterna

→ Non vi è differenza nella scelta dei grappoli

Nota:

Purtroppo, essendo aggregazioni naturali, i grappoli presentano una loro omogeneità interna, in generale inversamente proporzionale alla loro dimensione

- il problema principale riguarda la <u>scelta</u> del grappolo migliore

Confronti
Grappoli - Stratificato

- contiguità fra le unità di un grappolo
- tutte le unità del grappolo fanno parte del campione
- rappresentatività della popolazione non garantita all'interno del grappolo
- generalmente meno efficiente dello stratificato se non è a probabilità variabile

- omogeneità delle unità di uno strato rispetto a una variabile rilevante
- le unità di uno strato vengono campionate
- rappresentanza dello strato nel campione garantita
- generalmente più efficiente del c. a grappoli

Confronti
Grappoli - Casuale semplice

- generalmente meno efficiente per l'intrinseca omogeneità interna al grappolo
- non è garantita l'indipendenza fra le unità interne al gruppo/grappolo
- costi inferiori a parità di numerosità campionaria per la costruzione della lista e la rilevazione delle unità

- meno efficiente a parità di dimensione campionaria se la variabilità media della popolazione è minore di quella interna al gruppo
- indipendenza fra le unità del campione garantita
- costi superiori a parità di numerosità campionaria per la costruzione della lista e la rilevazione delle unità

Il campionamento a grappoli

- Notazioni

Descrizione	Popolazione	Campione
Numero dei grappoli	N	n
Numero di unità elementari nel grappolo i	M_i	M_i
Numero complessivo di unità elementari	$M. = \sum_{i=1}^{N} M_i = N \times M$	$m. = \sum_{i=1}^{n} M_i = n \times M$
Valore della variabile oggetto di indagine sull'unità j nel grappolo i	Y_{ij}	y_{ij}
Totale della variabile oggetto di indagine nel grappolo i	$\hat{Y}_{i.} = \sum_{j=1}^{M_i} Y_{ij}$	$\hat{y}_{i.} = \sum_{j=1}^{M_i} y_{ij}$
Totale della variabile oggetto di indagine in tutti i grappoli	$\hat{Y} = \sum_{i=1}^{N} \hat{Y}_{i.} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij}$	$\hat{y} = \sum_{i=1}^{n} \hat{y}_{i.} = \sum_{i=1}^{n} \sum_{j=1}^{M_i} y_{ij}$

Il campionamento a grappoli

- Notazioni

Descrizione	Popolazione	Campione
Media della variabile oggetto di indagine nel grappolo i	$\bar{Y}_{i.} = \frac{\hat{Y}_{i.}}{M_i} = \frac{\sum_{j=1}^{M_i} Y_{ij}}{M_i}$	$\bar{y}_{i.} = \frac{\hat{y}_{i.}}{M_i} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i}$
Media nei grappoli (unità di primo livello)	$\bar{Y} = \frac{\hat{Y}_{i.}}{N} = \frac{\sum_{j=1}^{M_i} Y_{ij}}{N}$	$\bar{y} = \frac{\hat{y}_{i.}}{n} = \frac{\sum_{j=1}^{M_i} y_{ij}}{n}$
Media della variabile oggetto di indagine in tutti i grappoli (unità elementari, di II livello)	$\bar{\bar{Y}} = \frac{\sum_{i=1}^{N} \hat{Y}_{i.}}{M.} = \frac{\sum_{i=1}^{N} \hat{Y}_{i.}}{N \times M} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij}}{N \times M}$	$\bar{\bar{y}} = \frac{\sum_{i=1}^{n} \hat{y}_{i.}}{m.} = \frac{\sum_{i=1}^{n} \hat{y}_{i.}}{n \times M} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{M_i} y_{ij}}{n \times M}$

Il campionamento a grappoli

- Notazioni

Descrizione	Popolazione	Campione
Proporzione della variabile oggetto di indagine nel grappolo i	$P_i = \frac{A_i}{M}$	$p_i = \frac{a_i}{M}$
Proporzione della variabile oggetto di indagine in tutti i grappoli (unità elementari, di II livello)	$P = \frac{\sum_{i=1}^{N} P_i}{N} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} A_{ij}}{NM}$	$p = \frac{\sum_{i=1}^{n} p_i}{n} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{M} a_{ij}}{nM}$

Il campionamento a grappoli

- Notazioni

Descrizione	Popolazione	Campione
Varianza complessiva	$S^2 = \frac{\sum_{i=1}^{N}\left(Y_{ij} - \bar{\bar{Y}}\right)^2}{NM-1}$	$s^2 = \frac{\sum_{i=1}^{n}\left(y_{ij} - \bar{\bar{y}}\right)^2}{nM-1}$
Varianza tra i grappoli	$S_b^2 = \frac{\sum_{i=1}^{N}\left(\bar{Y}_i - \bar{\bar{Y}}\right)^2}{N-1}$	$s_b^2 = \frac{\sum_{i=1}^{n}\left(\bar{y}_i - \bar{\bar{y}}\right)^2}{n-1}$
Varianza entro i grappoli	$S_w^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i}(Y_{ij} - \bar{Y}_i)^2}{N(M-1)}$	$s_w^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{M_i}(y_{ij} - \bar{y}_i)^2}{n(M-1)}$

Il campionamento a grappoli

- Notazioni

Descrizione	Popolazione	Campione
Frazione di campionamento con grappoli a dimensione costante	$f = \dfrac{\#\,\text{unità elementari del campione}}{\#\,\text{unità elementari della popolazione}}$	
Frazione di campionamento con grappoli a dimensione variabile	$f = \dfrac{n}{N} = \dfrac{\#\,\text{grappoli del campione}}{\#\,\text{grappoli nella popolazione}}$	
Frazione di campionamento stimata (usata raramente) per N grande	$\hat{f} = \dfrac{\#\,\text{unità elementari del campione}}{\#\,\text{grappoli}}$	

Il campionamento a grappoli

Se i dati vengono stimati si avrà:

s_b^2 Misura di variabilità fra le medie dei grappoli nel campione

s_w^2 Media ponderata delle misure di variabilità di Y nei singoli gruppi nel campione

$$s^2 = \frac{(n-1)s_b^2 + n(M-1)s_w^2}{nM-1}$$

Variabilità della media nell'intero campione

Se $s_b^2=0$ → massima omogeneità fra grappoli

Se $s_w^2=0$ → massima omogeneità interna, ovvero Yi.=Y.. Per ogni i

Il campionamento a grappoli

La struttura della popolazione

	Grappoli						
Unità	***1***	***2***	***3***	***…***	***i***	***…***	***N***
1	Y_{11}	Y_{21}	Y_{31}	$\cdots$	Y_{i1}	$\cdots$	Y_{N1}
2	Y_{12}	Y_{22}	Y_{32}	$\cdots$	Y_{i2}	$\cdots$	Y_{N2}
$\vdots$	$\vdots$						$\vdots$
j	Y_{1j}	Y_{2j}	Y_{3j}	$\cdots$	Y_{ij}		Y_{Nj}
$\vdots$	$\vdots$						$\vdots$
$M \equiv M_i$	Y_{1M_1}	Y_{2M2}	Y_{3M3}	$\cdots$	Y_{iM_i}	$\cdots$	Y_{NM_N}
Totale	$Y_{1.}$	$Y_{2.}$	$Y_{3.}$	$\cdots$	$Y_{i.}$	$\cdots$	$Y_{N.}$

Il campionamento a grappoli

- *Spazio Campionario*:
  - Insieme di tutti i possibili campioni (grappoli) di dimensione n che si possono formare da una popolazione finita di N unità (Grappoli) in base ad una tecnica predefinita.

$$\binom{N}{n}$$

Numero di tutti i possibili campioni non ordinati senza ripetizione (tipo di campionamento più utilizzato)

Il campionamento a grappoli

- Probabilità di inclusione:
- Probabilità che un'unità, o gruppo di unità, ha di appartenere al campione estratto

↓

Se ogni unità ha la stessa probabilità di inclusione di 1° ordine, il piano di campionamento è detto

AUTOPONDERANTE

Il campionamento a grappoli

Probabilità di inclusione del primo ordine:

$$\pi_{ij} = \frac{n}{N}$$

Dove:

- i = indica l'*i-esimo* grappolo
- j = indica la *j-esima* unità

la probabilità di estrarre l'unità *j* considerando che estrarre l'unità *j* significa dire estrarre il grappolo i-esimo di cui l'unità fa parte.

Il campionamento a grappoli

Probabilità di inclusione del secondo ordine:

$$\pi_{(ij)(i'k)} = \begin{cases} \dfrac{n}{N} & \text{Se } i = i' \\[2ex] \dfrac{n}{N}\,\dfrac{n-1}{N-1} & \text{Se } i \neq i' \end{cases}$$

Essa varia a seconda che le unità j e k appartengano allo stesso grappolo (i = i') o meno.

Il campionamento a grappoli

con probabilità costanti

$$M_{.} = \sum_{i=1}^{N} M_i$$

Numero di unità elementari nella popolazione

Numero di unità elementari nell' i-esimo grappolo

$$\sum_{i=1}^{N} M_i$$

Numerosità del campione, costante se e solo se M_i = M per ogni grappolo

Le probabilità di inclusione sono uguali per tutti i grappoli

Il campionamento a grappoli

con probabilità costanti

Generalmente si effettua l'estrazione dei grappoli con probabilità costanti quando

- i medesimi sono composti da un numero pressoché uguale di unità elementari

$$M \equiv M_i$$

-quando le diverse dimensioni dei grappoli sono irrilevanti ai fini dell'indagine.

Le probabilità di inclusione del primo ordine sono uguali per tutti i grappoli
Il campionamento a grappoli si presenta con le stesse caratteristiche del campionamento casuale semplice

Il campionamento a grappoli

Stima dei parametri: *la media campionaria*

Media

Unità di primo livello (a)

Unità elementari (b)

Grappoli (grandi unità)

Elementi entro i grappoli (piccole unità)

Il campionamento a grappoli

Stima dei parametri: *la media campionaria - esempio*

	Grappoli						
Unità	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	3	5	5	3	5	5	2
2	5	4	4	4	4	3	5
3	6	5	3	5	3	6	4
4	5	4	2	3	6	7	3
Totale	19	18	14	**15**	18	21	**14**
Media	4,75	4,5	3,5	**3,75**	4,5	5,25	**3,5**

$$\hat{Y}_{i.} = \sum_{j=1}^{M_i} Y_{ij}$$

$$\bar{Y}_{i.} = \frac{\hat{Y}_{i.}}{M_i} = \frac{\sum_{j=1}^{M_i} Y_{ij}}{M_i}$$

$$\hat{y}_{i.} = \sum_{j=1}^{M_i} y_{ij}$$

$$\bar{y}_{i.} = \frac{\hat{y}_{i.}}{M_i} = \frac{\sum_{j=1}^{M_i} y_{ij}}{M_i}$$

Il campionamento a grappoli

Stima dei parametri: *la media campionaria - esempio*

N=7 n=2 M=4 f=2/7=0,286=8/28

$$\hat{Y} = \sum_{i=1}^{N} \hat{Y}_{i.} = 19+18+14+15+18+21+14 = \sum_{i=1}^{N}\sum_{j=1}^{M_i} Y_{ij} = 3+5+6+5+5+4+\ldots+4+3 = 119$$

Media nei grappoli i=1 a N

$$\bar{Y}_{1.} = \frac{\hat{Y}_{1.}}{M_1} = \frac{19}{4} = \frac{\sum_{j=1}^{M_1} Y_{ij}}{M_1} = \frac{3+5+6+5}{4} = 4{,}75 \qquad \bar{y}_{4.} = \frac{\hat{y}_{4.}}{M_4} = \frac{15}{4} = \frac{\sum_{j=1}^{M_4} y_{ij}}{M_4} = \frac{3+4+5+3}{4} = 3{,}75$$

Media nei grappoli unità primo livello

$$\bar{Y} = \frac{\hat{Y}_{i.}}{N} = \frac{\sum_{j=1}^{M_i} Y_{ij}}{N} = \frac{19+18+14+15+18+21+14}{7} = \frac{119}{7} = 17$$

Il campionamento a grappoli

Stima dei parametri: *la media campionaria – esempio*

Media nei grappoli campione unità primo livello

$$\bar{y} = \frac{\hat{y}_{i.}}{n} = \frac{\sum_{j=1}^{M_i} y_{ij}}{n} = \frac{15+14}{2} = 14{,}5$$

Media unità elementari - popolazione

$$\bar{\bar{Y}} = \frac{\sum_{i=1}^{N} \hat{Y}_{i.}}{M.} = \frac{\sum_{i=1}^{N} \hat{Y}_{i.}}{N \times M} = \frac{19+18+14+15+18+21+14}{7 \times 4} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i} Y_{ij}}{N \times M} = \frac{3+5+6+5+5+4+\ldots+4+3}{28} = \frac{119}{28} = 4{,}25$$

Media unità elementari - campione

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{n} \hat{y}_{i.}}{m.} = \frac{\sum_{i=1}^{n} \hat{y}_{i.}}{n \times M} = \frac{15+14}{2 \times 4} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{M_i} y_{ij}}{n \times M} = \frac{3+5+5+3+2+5+4+3}{8} = 3{,}625$$

Il campionamento a grappoli

Stima della media sulle unità di primo livello

Posto M_i=M

$$\bar{Y} = \frac{\sum_{i=1}^{N} \hat{Y}_{i.}}{N} = \frac{\sum_{j=1}^{M_i} Y_{ij}}{N}$$

Media calcolata sui dati della popolazione

$$\bar{y} = \frac{\sum_{i=1}^{n} \hat{y}_{i.}}{n} = \frac{\sum_{j=1}^{M_i} y_{ij}}{n}$$

Media calcolata sui dati del campione

Proprietà

$$E\left(\bar{y}\right) = \bar{Y}$$

Stimatore non distorto

$$V(\bar{y}) = \frac{1-f}{n} \frac{\sum_{i=1}^{N}\left(\bar{Y}_i - \bar{\bar{Y}}\right)^2}{N-1}$$

Varianza dello stimatore media

Il campionamento a grappoli

Varianza della media sulle unità di primo livello *Posto M_i=M*

La varianza totale della variabile oggetto di interesse è:

$$S^2 = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(Y_{ij} - \overline{\overline{Y}}\right)^2 / NM - 1$$

Scomposizione del numeratore:

$$\sum_{i=1}^{N} \sum_{j=1}^{M} \left(Y_{ij} - \overline{\overline{Y}}\right)^2 = \sum_{i=1}^{N} \sum_{j=1}^{M} \left(Y_{ij} - \overline{Y}_i\right)^2 + M \sum_i \left(\overline{Y}_i - \overline{\overline{Y}}\right)^2$$

Prende in considerazione tutti gli elementi della popolazione

Variabilità nei grappoli

Variabilità fra i grappoli

Il campionamento a grappoli

Varianza della media sulle unità di primo livello

La varianza generale della Y nella popolazione S^2 si scompone in:

Varianza entro i grappoli $\Rightarrow$

$$S_w^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(Y_{ij} - \overline{Y}_i\right)^2}{N(M-1)}$$

Varianza tra i grappoli $\Rightarrow$

$$S_b^2 = \frac{\sum_{i=1}^{N}\left(\overline{Y}_i - \overline{\overline{Y}}\right)^2}{(N-1)M}$$

Il campionamento a grappoli

Nel campione gli stimatori non distorti di S^2, S_B^2, S_W^2 *sono:*

varianza totale $S^2 = \dfrac{\sum_{i=1}^{N}\left(Y_{ij} - \bar{\bar{Y}}\right)^2}{NM - 1}$ **con** $s^2 = \dfrac{\sum_{i=1}^{n}\left(y_{ij} - \bar{\bar{y}}\right)^2}{nM - 1}$

varianza tra i grappoli $S_b^2 = \dfrac{\sum_{i=1}^{N}\left(\bar{Y}_i - \bar{\bar{Y}}\right)^2}{N - 1}$ **con** $s_b^2 = \dfrac{\sum_{i=1}^{n}\left(\bar{y}_i - \bar{\bar{y}}\right)^2}{n - 1}$

varianza entro i grappoli $S_w^2 = \dfrac{\sum_{i=1}^{N}\sum_{j=1}^{M_i}(Y_{ij} - \bar{Y}_i)^2}{N(M - 1)}$ **con** $s_w^2 = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{M_i}(y_{ij} - \bar{y}_i)^2}{n(M - 1)}$

$$E\left(s^2\right) = S^2 \qquad E\left(s_b^2\right) = S_b^2 \qquad E\left(s_w^2\right) = S_w^2$$

Il campionamento a grappoli

Varianza della media sulle unità di primo livello

Scomposizione della varianza

Fonte	Devianze	gdl	Varianze
Entro i grappoli	$\sum_{ij}(Y_{ij}-\overline{Y}_i)^2$	N(M-1)	S_W^2
Tra i grappoli	$\sum_{i}M(\overline{Y}_i-\overline{\overline{Y}})^2$	N-1	S_B^2
Totale	$\sum_{ij}(Y_{ij}-\overline{\overline{Y}})^2$	NM-1	S^2

Il campionamento a grappoli

Varianza della media sulle unità di primo livello

La varianza dello stimatore media (a) può essere espressa anche in termini di S_b^2 tenendo conto della scomposizione della varianza di Y:

$$(NM-1)S^2 = \sum_{j=1}^{M}\sum_{i}^{N}\left(Y_{ij}-\overline{\overline{Y}}\right)^2 = N(M-1)S_w^2 + (N-1)S_b^2$$

quindi

$$S^2 = \frac{N(M-1)S_w^2 + (N-1)S_b^2}{(NM-1)}$$

Per cui

$$V(\bar{y}) = \frac{1-f}{n}\,\frac{\sum_{i=1}^{N}\left(\overline{Y}_i-\overline{\overline{Y}}\right)^2}{N-1} = \frac{1-f}{n}MS_b^2$$

Il campionamento a grappoli

Varianza della media sulle unità di primo livello e sua stima

$$V(\bar{y}) = \frac{1-f}{n} \frac{\sum_{i=1}^{N} \left(\bar{Y}_i - \bar{\bar{Y}}\right)^2}{N-1} = \frac{1-f}{n} MS_b^2 \qquad \hat{v}(\bar{y}) = \frac{1-f}{n} \frac{\sum_{i=1}^{n} \left(\bar{y}_i - \bar{\bar{y}}\right)^2}{n-1} = \frac{1-f}{n} Ms_b^2$$

Più piccola è S_b^2 **minore sarà la varianza dello stimatore**

↓

se i grappoli sono omogenei all'esterno ed eterogenei all'interno conviene il CGR

$$f = \frac{n}{N} = \frac{\text{\# grappoli del campione}}{\text{\# grappoli nella popolazione}}$$

Il campionamento a grappoli

Stima della media sulle unità elementari

Posto $M_{.}=M$

$$\bar{\bar{Y}} = \frac{\sum_{i=1}^{N}\hat{Y}_{i.}}{M.} = \frac{\sum_{i=1}^{N}\hat{Y}_{i.}}{N\times M} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M_i}Y_{ij}}{N\times M}$$

→ **Media calcolata sui dati della popolazione**

$$\bar{\bar{y}} = \frac{\sum_{i=1}^{n}\hat{y}_{i.}}{M} = \frac{\sum_{i=1}^{n}\hat{y}_{i.}}{n\times M} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{M_i}y_{ij}}{n\times M}$$

→ **Media calcolata sui dati del campione**

Proprietà $E\left(\bar{\bar{y}}\right)=\bar{\bar{Y}}$ → **Stimatore non distorto**

Il campionamento a grappoli

Varianza della media sulle unità elementari

Posto M_i=M

Partiamo dalla varianza della media delle unità di I livello

$$\bar{\bar{y}} = \frac{\bar{y}_i}{M} = \frac{\sum_{i,j} y_{ij}}{nM}$$

Siccome valgono le seguenti relazioni e proprietà:

$$\bar{Y} = M\bar{\bar{Y}} \qquad \bar{y} = M\bar{\bar{y}}$$

$$\bar{\bar{y}} = \frac{\bar{y}}{M}$$

deve essere quella di secondo livello
per forza no? vedi anche formula
precedente per S2b

$$V(\bar{\bar{y}}) = \frac{1}{M^2}V(\bar{Y}) = \frac{1-f}{nM^2}MS_b^2 = \frac{1-f}{nM^2}\frac{\sum_{i=1}^{N}(Y_i - \bar{\bar{Y}})^2}{N-1} = \frac{1-f}{nM}S_b^2$$

Il campionamento a grappoli

Varianza della media sulle unità elementari 2

Posto M_i=M

Partiamo dal numeratore della prima formula.
Sommando ed elevando al quadrato, si avrà:

$$\sum_{i=1}^{N}\left(Y_i-\overline{Y}\right)^2=\sum_{i=1}^{N}\sum_{j=1}^{M}\left(y_{ij}-\overline{\overline{Y}}\right)^2+2\sum_{i=1}^{N}\sum_{j<k}^{M}\left(y_{ij}-\overline{\overline{Y}}\right)\left(y_{ik}-\overline{\overline{Y}}\right)$$

Dato che:

$$\rho=\frac{2\sum_{i}\sum_{j<k}\left(y_{ij}-\overline{\overline{Y}}\right)\left(y_{ik}-\overline{\overline{Y}}\right)}{(M-1)(NM-1)S^2}$$

ρ = coefficiente di correlazione lineare entro i grappoli

Possiamo scrivere:

$$=(NM-1)S^2+(M-1)(NM-1)\rho S^2=(NM-1)S^2\left[1+(M-1)\rho\right]$$

Il campionamento a grappoli

Coefficiente di correlazione lineare entro i grappoli

$$\rho = \frac{E\left(y_{ij} - \bar{\bar{Y}}\right)\left(y_{ik} - \bar{\bar{Y}}\right)}{E\left(y_{ij} - \bar{\bar{Y}}\right)^2}$$

Poiché tutti i possibili prodotti al numeratore $\left(y_{ij} - \bar{\bar{Y}}\right)$ *sono* $\frac{NM(M-1)}{2}$

e il denominatore è pari a: $E\left(y_{ij} - \bar{\bar{Y}}\right)^2 = \frac{(NM-1)}{NM} S^2$

$$\rho = \frac{E\left(y_{ij} - \bar{\bar{Y}}\right)\left(y_{ik} - \bar{\bar{Y}}\right)}{E\left(y_{ij} - \bar{\bar{Y}}\right)^2} = \frac{2\sum_i \sum_{j<k} \left(y_{ij} - \bar{\bar{Y}}\right)\left(y_{ik} - \bar{\bar{Y}}\right)}{(M-1)(NM-1)S^2}$$

tutti i doppi prodotti all'interno dei grappoli

Il campionamento a grappoli

Varianza della media sulle unità elementari 2

Posto M_i=M

se:

$$\sum_{i=1}^{N}\left(Y_i-\overline{Y}\right)^2=(NM-1)S^2\left[1+(M-1)\rho\right]$$

$$\Downarrow$$

$$V(\bar{\bar{y}})=\frac{1-f}{n}\frac{NM-1}{M^2(N-1)}S^2\left[1+(M-1)\rho\right]$$

quando 1/N è trascurabile

$$\frac{(NM-1)}{M(N-1)}\cong 1$$

però in teoria se voglio far coincidere le varianze allora devo mettere anche questo termine (ricorda!)

Per cui:

$$V(\bar{\bar{y}})=\frac{1-f}{nM}S^2\left[1+(M-1)\rho\right]$$

Il campionamento a grappoli

La media campionaria $\bar{\bar{y}}$ è una stima non distorta di $\bar{\bar{Y}}$ con varianza:

$$V(\bar{\bar{y}}) = \frac{1-f}{nM} S^2 [1 + (M-1)\rho]$$

Mostra di quanto cambia la varianza nell'utilizzare come unità campionaria le unità elementari o le unità di I livello (i cluster):

- $\rho > 0$ $\Longrightarrow$ la varianza delle unità elementari è più piccola
- $\rho < 0$ $\Longrightarrow$ la varianza delle unità di primo livello è più piccola

Il campionamento a grappoli

Varianza della media sulle unità elementari 2

Posto $M_i = M$

Permette di fare un confronto anche con il campionamento casuale semplice!

- ***Per un campionamento casuale semplice di nM elementi la*** $V\left(\bar{\bar{y}}\right)$ ***è uguale***

Tranne che per il fattore $1+(M-1)\rho$

$\downarrow$

Tale fattore esprime come viene modificata la varianza allorché si assume come unità di campionamento il grappolo e non l'unità.

Il campionamento a grappoli

Varianza della media sulle unità elementari

- *$1+(M-1)\rho$ = fattore di Kish. Per i grappoli di dimensione M è il Deff*

  L' effetto del disegno esprime la perdita (o il guadagno) in precisione rispetto al campionamento casuale semplice

- *se $\rho > 0$* ⇒ *$Vccs < V_{CGR}$* *l'aumento dipende da ρ e da M*
- *se $\rho < 0$* ⇒ *$V_{ccs} > V_{CGR}$* quindi grappoli omogenei = rho grande = css migliore di grappoli
- *se $\rho = 0$* ⇒ *$Vccs = V_{CGR}$*

Quando $\rho > 0$?

- *Quando tutti gli scarti sono o tutti positivi o tutti negativi*

  quando le grandezze sono omogenee

Il campionamento a grappoli

Varianza della media delle unità elementari

$$Deff = \frac{V(\overline{\overline{Y}}_{CGR})}{V(\overline{\overline{Y}}_{CCS})} \qquad Deff = [1 + (M-1)\rho]$$

- **È generalmente maggiore dell'unità perché nei grappoli è presente un certo grado di omogeneità** quindi il ccs è generalmente più preciso

il coefficiente di omogeneità nei grappoli tende a diminuire con l'aumentare delle dimensioni dei grappoli quindi per grappoli sufficientemente numerosi (come possono essere, ad esempio, i comuni, le sezioni di censimento, le sezioni elettorali, le ASL, gli ospedali, i distretti scolastici, ecc.)

l'efficienza dei due disegni tende a coincidere

Il campionamento a grappoli

Varianza della media sulle unità elementari 3 …
Coefficiente di correlazione 2

$$\sum_{i=1}^{N}\left(y_i-\bar{\bar{Y}}\right)^2=(N-1)MS_b^2$$

vale anche che

$$\sum_{i=1}^{N}\left(y_i-\bar{\bar{Y}}\right)^2=(NM-1)S^2\left[1+(M-1)\rho\right]$$

vale l'uguaglianza:

$$(NM-1)S^2\left[1+(M-1)\rho\right]=(N-1)MS_b^2$$

per cui

$$\rho=\frac{(N-1)MS_b^2}{(NM-1)S^2\left[1+(M-1)\right]}=\frac{(N-1)MS_b^2-(NM-1)S^2}{(NM-1)(M-1)S^2}=\frac{S_b^2-S^2}{(M-1)S^2}$$

Si ricorda che quando 1/N è trascurabile $\frac{(NM-1)}{M(N-1)}\cong 1$

Il campionamento a grappoli

Coefficiente di correlazione lineare entro i grappoli 2

$$\rho = \frac{E\left(y_{ij} - \overline{\overline{Y}}\right)\left(y_{ik} - \overline{\overline{Y}}\right)}{E\left(y_{ij} - \overline{\overline{Y}}\right)^2} = \frac{\dfrac{\sum_i \sum_{j<k} \left(y_{ij} - \overline{\overline{Y}}\right)\left(y_{ik} - \overline{\overline{Y}}\right)}{NM(M-1)/2}}{\dfrac{(NM-1)S^2}{NM}} = \frac{2\sum_i \sum_{j<k} \left(y_{ij} - \overline{\overline{Y}}\right)\left(y_{ik} - \overline{\overline{Y}}\right)}{(M-1)(NM-1)S^2}$$

$$\rho = \frac{2\sum_i \sum_{j<k} \left(y_{ij} - \overline{\overline{Y}}\right)\left(y_{ik} - \overline{\overline{Y}}\right)}{(M-1)(NM-1)S^2} \Rightarrow \rho = \frac{S_b^2 - S^2}{(M-1)S^2}$$

Il campionamento a grappoli

Coefficiente di correlazione lineare entro i grappoli 3

considerato che:

$$S_w^2 = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(Y_{ij} - \overline{Y}_i\right)^2}{N\left(M-1\right)}$$

e

$$\sum_{i}^{N}\left(Y_{i.} - \overline{Y}\right)^2 = M(N-1)S_B^2$$

è possibile riscrivere l'equazione della scomposizione della varianza

$$(NM-1)S^2 = N(M-1)S_W^2 + (N-1)S_B^2$$

in questo modo:

$$(NM-1)S^2 = N(M-1)S_W^2 + \frac{\sum_{i=1}^{N}\left(Y_{i.} - \overline{Y}\right)^2}{M}$$

Il campionamento a grappoli

Dalla formula della devianza complessiva isoliamo S_b^2 e sostituiamo S_w^2

$$S^2\rho = S^2 - S_W^2 = S^2(1-\rho)$$

$$S_b^2 = \frac{(NM-1)S^2 - N(M-1)S_w^2}{N-1} = \frac{(NM-1)S^2 - N(M-1)(S^2 - \rho S^2)}{N-1} =$$

$$= \frac{NMS^2 - S^2 - NMS^2 + NS^2 + N(M-1)S^2\rho}{N-1} = \frac{(N-1)S^2 + N(M-1)S^2\rho}{N-1} =$$

$$= S^2\left[1 + \frac{N(M-1)\rho}{N-1}\right]$$

Il campionamento a grappoli

Per N grande avremo:

Posto M_i=M

$$\tilde{S}^2 = \frac{S_b^2 + (M-1)S_w^2}{(M-1)}$$

Se ρ=1

$$S_w^2 \approx 0 \qquad S_b^2 \approx MS^2$$

In questo caso non conviene il campionamento a grappoli

Il campionamento a grappoli

Coefficiente di correlazione – coefficiente di omogeneità

$$S^2\rho = S^2 - S_W^2 \qquad \rho = 1 - \frac{S_w^2}{S^2} = 1 - \delta$$

δ **noto come coefficiente di omogeneità**

Determiniamo il campo di variazione di ρ

Poiché

$$(N-1)S_B^2 \geq 0$$

$$(NM-1)S^2 \geq N(M-1)S_W^2$$

quindi

$$\frac{S_W^2}{S^2} \leq \frac{NM-1}{N(M-1)}$$

Il campionamento a grappoli

Coefficiente di correlazione – coefficiente di omogeneità

$$\rho = 1 - \frac{S_w^2}{S^2} = 1 - \delta$$

$$\rho = 1 - \frac{S_w^2}{S^2} \geq 1 - \frac{NM - 1}{N(M-1)} = -\frac{N-1}{N(M-1)}$$

per cui: $$-\frac{NM-1}{N(M-1)} \leq \rho \leq 1$$

e in forma approssimata per N grande $$-\frac{1}{M-1} \leq \rho \leq 1$$

Il campionamento a grappoli

Considerazioni sul valore di ρ

Se i grappoli sono omogenei (formati da unità elementari molto simili) il valore di

$$\rho$$

tende a 1. Questa è la situazione peggiore per il campionamento a grappoli.

Se i grappoli sono eterogenei (formati da unità elementari molto diversi fra loro) il valore di

$$\rho$$

tende a

$$-\frac{NM-1}{N(M-1)}$$

Questa è la situazione ideale per il campionamento a grappoli.

Il campionamento a grappoli

Stima del totale nelle unità elementari

$$\hat{Y} = \sum_{i=1}^{N} \hat{Y}_{i.} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij}$$ → *Totale della variabile Y nella popolazione*

$$\hat{y} = \sum_{i=1}^{n} \hat{y}_{i.} = \sum_{i=1}^{n} \sum_{j=1}^{M_i} y_{ij}$$ → *Totale della variabile Y nel campione*

Proprietà

$$E(\hat{y}_{CCG}) = \hat{Y}$$ → *Stimatore non distorto*

$$\hat{y}_{CGR} = \frac{N}{n} \sum_{i=1}^{n} \sum_{j=1}^{M_i} y_{ij}$$

Il campionamento a grappoli

Varianza del totale nelle unità elementari

$$V(\hat{y}) = N^2 \frac{1-f}{n} \frac{\sum_{i=1}^{N} \left(\hat{Y}_i - \bar{Y}\right)^2}{N-1} = N^2 \frac{1-f}{n} \frac{\sum_{i=1}^{N} \sum_{j=1}^{M} \left(Y_{ij} - \bar{Y}\right)^2}{N-1}$$

$$V(\hat{y}) = N^2 \frac{1-f}{n} S_b^2$$

$\bar{Y} = Y / N$ → *Media della variabile nei grappoli della popolazione*

totale di Y / N (numero di grappoli, non di unità)

Il campionamento a grappoli

Varianza del totale (in termini di Sb)

Se M_i=M

$$S_b^2 = \frac{\sum_{i=1}^{N}\left(\overline{Y}_i - \overline{\overline{Y}}\right)^2}{(N-1)M}$$

È la varianza tra i grappoli

Stima corretta di $V(\hat{y})$

$$\hat{v}(\hat{y}) = N^2 \frac{1-f}{n} s_b^2$$

dove

$$s_b^2 = \frac{\sum_{i=1}^{n}\left(\bar{y}_i - \bar{\bar{y}}\right)^2}{n-1}$$

$$f = \frac{n}{N} = \frac{\#\,\text{grappoli del campione}}{\#\,\text{grappoli nella popolazione}}$$

Il campionamento a grappoli

Varianza del totale

Si esprima la varianza del totale in funzione di S^2 :

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N} M^2(\bar{y}_i - \bar{Y})^2$$ segue che sostituendo $MS_b^2 = S^2$

si avrà:
$$V(\hat{y}_{CGR}) = N^2\frac{1-f}{n}S^2 = N^2\frac{1-f}{n}MS_b^2$$

Quindi maggiore sarà S_B^2 maggiore sarà la varianza della stima del totale

Se si esprime la varianza del totale in funzione di ρ ovvero del

Coefficiente di correlazione lineare:

$$\rho = \frac{E\left[\left(y_{ij} - \bar{\bar{Y}}\right)\left(y_{ik} - \bar{\bar{Y}}\right)\right]}{E\left(y_{ij} - \bar{\bar{Y}}\right)}$$

Il campionamento a grappoli

Varianza del totale

Dalla formula della devianza complessiva isoliamo S_b^2 e sostituiamo S_w^2

$$S_b^2 = \frac{(NM-1)S^2 - N(M-1)S_w^2}{N-1} = \frac{(NM-1)S^2 - N(M-1)(S^2 - \rho S^2)}{N-1} =$$

$$= \frac{NMS^2 - S^2 - NMS^2 + NS^2 + N(M-1)S^2\rho}{N-1} = \frac{(N-1)S^2 + N(M-1)S^2\rho}{N-1} =$$

$$= S^2\left[1 + \frac{N(M-1)\rho}{N-1}\right]$$

$$V(\hat{y}_{GR}) = N^2 \frac{1-f}{n} MS^2 \left[1 + \frac{N(M-1)\rho}{N-1}\right]$$

o, in forma approssimata, per N grande:

questo rho è come il primo visto con doppi prodotti per grappolo

$$V(\hat{y}_{GR}) = \underbrace{N^2 \frac{1-f}{n} MS^2}_{\text{coincide con ccs}} \underbrace{[1 + (M-1)\rho]}_{\text{aggiunta al grappolo}}$$

Il campionamento a grappoli

Varianza del totale

- Il primo addendo al secondo membro coincide con la varianza dello stimatore nel campione CCS di *nM* unità elementari
-il secondo addendo esprime l'incremento della varianza *se ρ > 0,* (o il decremento se ρ <0) dovuto al campionamento a grappoli.

Valgono, pertanto, le stesse considerazioni fatte per la varianza della media

Il campionamento a grappoli

Varianza del totale

$$\rho = \frac{E\left[\left(y_{ij} - \bar{\bar{Y}}\right)\left(y_{ik} - \bar{\bar{Y}}\right)\right]}{E\left(y_{ij} - \bar{\bar{Y}}\right)}$$

Può anche essere espresso nei seguenti termini: $\rho = 1 - \dfrac{S_W^2}{S^2}$

da cui: $S^2\rho = S^2 - S_W^2$ **ovvero** $S_W^2 = S^2 - \rho S^2$

Poiché $(N-1)S_B^2 \geq 0$ **si ottiene:** $\dfrac{S_W^2}{S^2} \leq \dfrac{NM-1}{N(M-1)}$

da cui: $\rho = 1 - \dfrac{S_W^2}{S^2} \geq 1 - \dfrac{NM-1}{N(M-1)} = -\dfrac{N-1}{N(M-1)}$ $\qquad -\dfrac{N-1}{N(M-1)} \leq \rho \leq 1$

Il campionamento a grappoli

$$-\frac{1}{M-1} \leq \rho \leq 1$$

Per N grande:

Se ρ =1, i grappoli sono omogenei (la situazione peggiore per effettuare un campionamento a grappoli).

$$\boldsymbol{\rho = -\frac{1}{M-1}}$$ **= elevata eterogeneità tra le unità elementari nei grappoli**

CGR è più efficiente di CCS

Il campionamento a grappoli

Varianza del totale

il coefficiente di omogeneità nei grappoli tende a diminuire con l'aumentare delle dimensioni dei grappoli quindi per grappoli sufficientemente numerosi l'efficienza dei due disegni tende a coincidere

è generalmente maggiore dell'unità perché nei grappoli è sempre presente una certa omogeneità

Il campionamento a grappoli

Stima della proporzione

Y è qualitativa
oggetto di stima è la frequenza assoluta o relativa delle unità della popolazione che possiedono l'attributo:
la frequenza relativa prende il nome di **proporzione**

Caso più semplice → **Carattere dicotomo** → **Possiede l'attributo** / **Non possiede l'attributo**

$$P = \frac{A}{N}$$ → *Proporzione di soggetti che ∈ alla classe C nella popolazione*

$$p = \frac{a}{n}$$ → *Proporzione di soggetti che ∈ alla classe C nel campione*

Il campionamento a grappoli

Stima della proporzione

Posto che Mi~M

$$P_i = \frac{A_i}{M}$$ ⇒ *Proporzione di elementi che nell' i-esimo grappolo Є alla classe C*

$$p_i = \frac{a_i}{M}$$ ⇒ *Proporzione campionaria dell' i-esimo grappolo*

Stimatore della proporzione

$$p = \frac{1}{n}\sum_{i=1}^{n}\frac{a_i}{M_i} = \frac{1}{n}\sum_{i=1}^{n} p_i$$

coincide con la media aritmetica delle proporzioni dei grappoli estratti

Proprietà

$$E(p) = P$$ ⇒ *Stimatore non distorto di P*

Il campionamento a grappoli

Varianza della proporzione

Posto che Mi~M

La varianza dello stimatore è simile alla varianza del totale, basta dividerla per N^2 e sostituire S^2 con

$$\frac{\sum_{i=1}^{N}(p_i - P)^2}{N-1}$$

che è la varianza tra le proporzioni dei grappoli. Si ha pertanto:

$$V(p) = \frac{1-f}{n}\frac{\sum_{i=1}^{N}(p_i - P)^2}{N-1} \Rightarrow \hat{v}(p) = \frac{1-f}{n}\frac{\sum_{i=1}(p_i - p)^2}{n-1}$$

E per N grande $$\hat{v}(p_{CGR}) \cong \frac{N-n}{N^2 n}\sum_1^N (p_i - p)^2$$

Nota: $$s_b^2 = \frac{\sum (p_i - p)^2}{n-1}$$

Il campionamento a grappoli

Varianza della proporzione

La stima della varianza di p considerando un CCS di nM elementi sarà:

$$\hat{v}(p_{CCS}) = \frac{NM - nM}{NM - 1}\frac{pq}{nM}$$

E per N grande

$$\hat{v}(p_{CCS}) \cong \frac{N-n}{N}\frac{pq}{nM}$$

Effetto del disegno

$$\frac{\hat{v}(p)}{\hat{v}(p_{CCS})} \approx \frac{M\sum_{i=1}^{n}(p_i - p)^2}{Npq}$$

Il campionamento a grappoli

Dimensione campionaria

Bisogna determinare il valore

n_0 = # grappoli nel campione

che minimizza il costo per una data varianza o viceversa

N.B.: se i grappoli sono pressoché uguali, si può stimare la numerosità ottima m_0, anche se risulta orientativa, vista l'impossibilità di deciderla a priori. Noi abbiamo operato con m costante o m medio.

Il campionamento a grappoli

Dimensione campionaria con una funzione di costo

$$C' = C_f + nC_1 + \sum_{i \in C} M_i C_{2i}$$

Costi fissi

Costo di estrazione di un grappolo

Costo delle unità elementari del grappolo i

C' = variabile casuale dipendente da $\sum M_i$, ovvero dalle diverse dimensioni degli N grappoli di cui si compone la popolazione

Il campionamento a grappoli

Dimensione campionaria con una funzione di costo

$$E(C) = nC_1 + n\overline{M}C_2$$

$C'-C_f$ $\quad \sum_1^N M_i / N$ $\quad C_{2i}$

<u>Costo totale</u> dell'indagine inteso come costo atteso di tutti i possibili campioni di n grappoli

Il Campionamento a Grappoli

grappoli a dimensioni variabili

grappoli a probabilità variabili

Il campionamento a grappoli

con dimensioni variabili

$$M_o = \sum_{i=1}^{N} M_i$$

La stima del totale per il grappolo i

$$\hat{y}_i = \sum_{i=1}^{M_i} y_{ij} = M_i \bar{y}_i$$

$$\bar{Y} = M_i \bar{\bar{Y}}$$

Una stima non distorta del totale è $\hat{y} = \dfrac{N}{n} \sum_{i=1}^{n} \hat{y}_i$

$$\bar{y}_i = M_i \bar{\bar{y}}$$

con varianza:

$$V(\hat{y}) = \frac{N^2(1-f)}{n} \frac{\sum_{i=1}^{N} (y_i - \bar{Y})^2}{N-1}$$

La cui stima è:

$$v(\hat{y}) = \frac{N^2(1-f)}{n} \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{n-1}$$

Il campionamento a grappoli

con dimensioni variabili

La stima del totale spesso è distorta. Tale risultato si presenta quando:

a) le medie di grappolo variano poco nei grappoli
b) gli Mi variano molto fra i grappoli

Se gli Mi variano, varia anche

$$\hat{y}_i = \sum_{i=1}^{M_i} y_{ij} = M_i \bar{y}_i$$

e ne risente anche la varianza che sarà molto grande.

In questo caso si preferisce ricorrere al campionamento a probabilità variabili.

Il campionamento a grappoli

con probabilità variabili

viene utilizzato se:

- Gli M_i sono diversi per ogni i;
- Vi è proporzionalità fra Y (= variabile da investigare) e M.

si assegna ad ogni grappolo una probabilità di estrazione $P_i=M_i/M_.$

i grappoli più numerosi hanno un probabilità di estrazione più alta

Si devono conoscere a priori le numerosità di tutti i grappoli

Il campionamento a grappoli

con probabilità variabili

- se in realtà gli *N* grappoli sono costituiti da un numero differente di unità elementari (caso molto più frequente);
- se si ritiene che vi sia una relazione di approssimata proporzionalità tra *Y e M, allora* si dovrà procedere all'estrazione di *n* grappoli con probabilità proporzionale alla dimensione *Mi* dei singoli grappoli, cioè assegnando a ciascuno di essi una probabilità di estrazione pari a

$$P_i = \frac{M_i}{M}$$

Per potere utilizzare questo metodo è necessario conoscere a priori la dimensione esatta dei grappoli.

La probabilità di estrazione di un grappolo (la scelta di ogni unità di I livello, grande unità) è proporzionale alla grandezza dell'universo stesso.

Il campionamento a grappoli

con probabilità variabili

Esempio: Si supponga di dover estrarre i grappoli con probabilità variabili. Nella tabella seguente vengono descritte le dimensioni per ogni grappolo (unità) e le cumulate delle dimensioni.

Tabella 1: Grappoli nella popolazione e relativa numerosità.

Unità	Dimensione	Cumulata
1	7	7
2	5	12
3	8	20
4	4	24
5	3	27

La probabilità che un'unità elementare venga estratta è costante ed è pari a: $\frac{1}{M_i} \times \frac{M_i}{M_0} = \frac{1}{M_0}$

Il campionamento a grappoli

con probabilità variabili

Estraiamo casualmente un numero da 1 a 27 (estremo superiore delle cumulate):

se, ad esempio, il numero è 19 allora abbiamo estratto il 3° grappolo.
se N è grande si può utilizzare la *tecnica di Lahiri*

Si individua il grappolo di dimensioni maggiori (*max dim*) e si estraggono a caso due numeri *i* ed *m* tali che:

$$\begin{cases} 1 \leq \mathrm{i} \leq \mathrm{N} \\ 1 \leq \mathrm{m} \leq \max \dim \end{cases}$$

Se m≤$M_0 = 27$ si estrae proprio il grappolo i;

Se m>M_i si procede all'estrazione di un'altra coppia di numeri.

Il campionamento a grappoli

con probabilità variabili / estrazioni con ripetizione

La stima del totale

STIMATORE di Hansen Hurwitz

$$\hat{Y}_{HHGR} = \frac{M_0}{n} \sum_{i=1}^{N} \bar{y}_i$$

con varianza:

$$V\left(\hat{Y}_{HHGR}\right) = \frac{M_0}{n} \sum_{i=1}^{N} M_i \left(\bar{Y}_i - \bar{\bar{Y}}\right)^2$$

La cui stima è:

totale diviso M0 è la media di secondo livello stimata

$$\hat{v}\left(\hat{Y}_{HHGR}\right) = \frac{M_o^2}{n} \sum_{i=1}^{N} \left(\bar{y}_i - \frac{\hat{Y}_{HHGR}}{M_0} \right)^2$$

Il campionamento a grappoli

con probabilità variabili / estrazioni senza ripetizione

La stima del totale

STIMATORE di Horvitz/Thompson

$$\hat{Y}_{HTGR} = \sum_{i=1}^{N} \frac{y_i}{\pi_i}$$

con varianza:

$$V\left(\hat{Y}_{HTGR}\right) = \sum_{i=1}^{N} \sum_{j>1}^{N} \left(\pi_i \pi_j - \pi_{ij}\right) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2$$

La cui stima è:

$$\hat{v}\left(\hat{Y}_{HTGR}\right) = \sum_{i=1}^{N} \sum_{j=1}^{M_o} \left(\frac{\pi_i \pi_j}{\pi_{ij}} - 1\right) \left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2$$

Il campionamento a grappoli

con probabilità variabili / estrazioni senza ripetizione

La stima della proporzione

$$P_{GR} = \frac{\sum_{i=1}^{N} A_i}{\sum_{i=1}^{N} M_i}$$

è

$$p_{GR} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} M_i}$$

Il campionamento a grappoli

con probabilità variabili / estrazioni senza ripetizione

con varianza:

$$V(p_{GR}) = \frac{1-f}{n\overline{M}^2} \frac{\sum_{i=1}^{N} (a_i - PM_i)^2}{N-1}$$

La cui stima è:

$$\hat{v}(p_{GR}) = \frac{1-f}{n\hat{\overline{M}}^2} \frac{\sum_{i=1}^{N} (a_i - p_{GR}M_i)^2}{n-1}$$

dove:

$$\hat{\overline{M}} = \sum_{i=1}^{n} \frac{M_i}{n}$$

Rappresenta la numerosità media dei grappoli nel campione