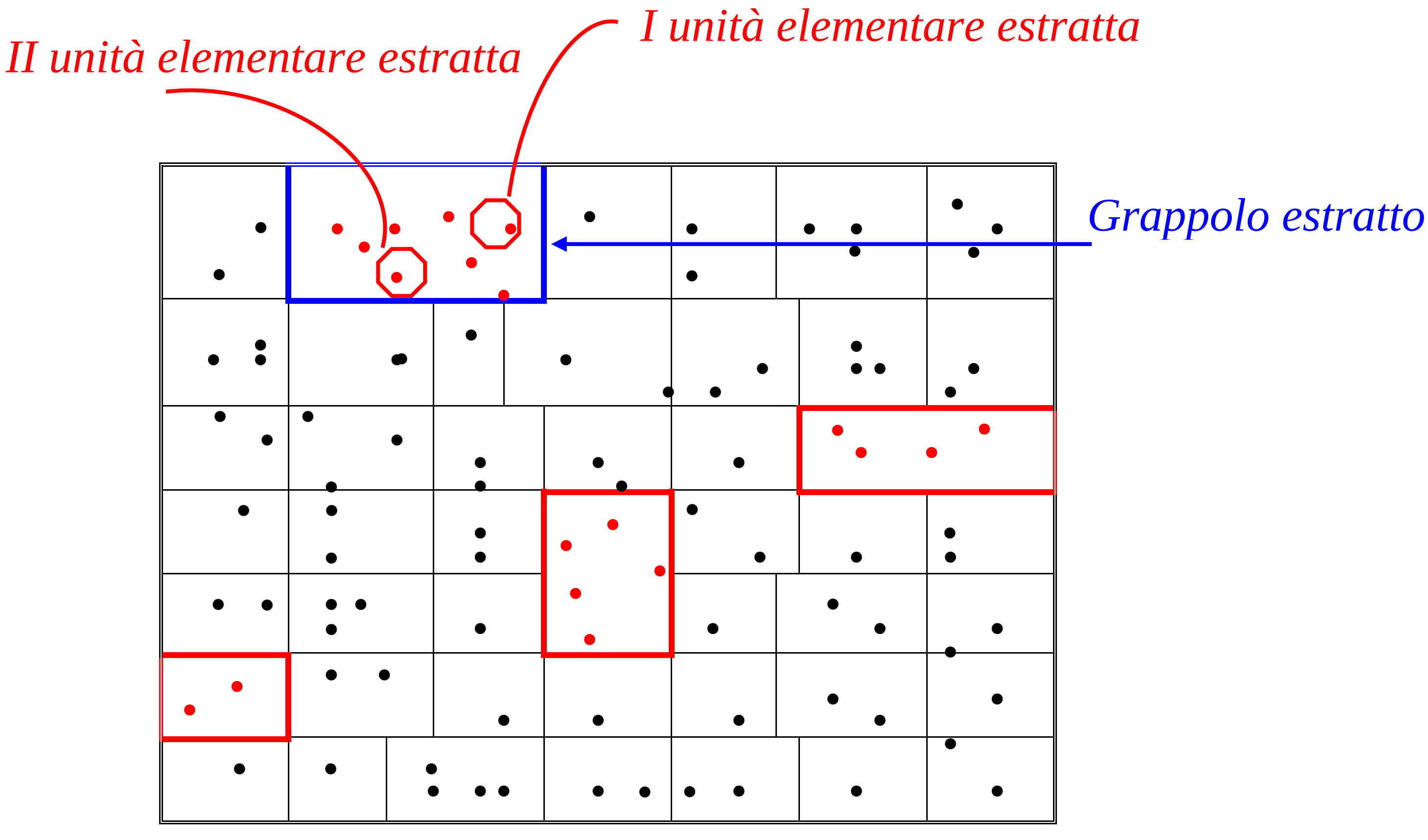
Campionamento a due stadi:

Campionamento senza ripetizione di unità elementari selezionate da gruppi o sottoinsiemi precostituiti detti GRAPPOLI

Presupposto:

-I grappoli sono definiti unità PRIMARIE.

-Le unità in esse contenute sono definite unità ELEMENTARI oppure unità SECONDARIE.



Esempio 1 . Se si vuole effettuare l'indagine sui consumi delle famiglie italiane, per rilevare la caratteristica di studio si può estrarre un campione in cui
_ le unità di primo stadio sono i comuni, dei quali viene estratto un campione
_ all'interno dei comuni estratti viene selezionato un campione di famiglie.

Esempio 2 . Se si vuole effettuare l'indagine sugli studenti degli atenei per verificare l'orientamento alla nuova riforma universitari:
_ stratificazione a priori delle province.
_ le unità di primo [1]stadio sono gli atenei, dei quali viene estratto un campione
_ all'interno degli atenei viene estratto un campione di studenti da intervistare.

Per aumentare l'efficienza del campionamento a due stadi conviene spesso procedere ad una stratificazione delle unità di primo stadio, come avviene in molte indagini su larga scala (indagine ISTAT sui consumi, indagine ISTAT sulle forze di lavoro, ecc.)

L'analogia fra il campionamento a due stadi ed il campionamento a grappoli è che il campionamento a due stadi è un campionamento a grappoli nel quale non vengono osservate tutte le unità all'interno dei grappoli estratti, bensì soltanto un sottoinsieme.

Simbologia:

N – Numero di grappoli nella popolazione;

n – Numero di grappoli nel campione;

M_i – Numero di unità elementari nel grappolo *i*;

m_i – Numero di unità elementari estratte dal grappolo *i* del campione;

$M_.$ – Numero di unità elementari nella popolazione

$$M_{\bullet} = M_1 + M_2 + \ldots.. + M_N = \sum_{i=1}^{N} M_i$$

y_{ij} – Valore della *j-esima* unità estratta nel grappolo *i* del campione;

$\tilde{y}_{i.}$ – Totale della variabile y nel grappolo *i* del campione;

$$\tilde{y}_{i\bullet} = y_{i1} + y_{i2} + \ldots.. + y_{iM_i} = \sum_{j=1}^{M_i} y_{ij}$$

$y_{i.}$ – Totale delle unità secondarie estratte nel grappolo *i* del campione;

$$y_{i\bullet} = y_{i1} + y_{i2} + \ldots .. + y_{im_i} = \sum_{j=1}^{m_i} y_{ij}$$

S^2_{1i} – Varianza della variabile Y nel grappolo *i* della popolazione;

$$S_{1i}^2 = \sum_{i=1}^{M_i} \frac{\left(Y_{ij} - \bar{Y}_{i\bullet}\right)^2}{(M_i - 1)}$$

s^2_{2i} – Varianza del subcampione estratto dall'unità primaria (grappolo) *i*;

$$s_{2i}^2 = \sum_{i=1}^{m_i} \frac{\left(y_{ij} - \bar{y}_{i\bullet}\right)^2}{(m_i - 1)}$$

dove:

$$\bar{y}_{i\bullet} = \frac{y_{i\bullet}}{m_i}$$

Stima del totale nel campionamento a probabilità variabili

Stimatore della media:

-senza ripetizione-

$$\boxed{\hat{Y}_{DSS} = \frac{1}{M_{\bullet}} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \frac{y_{ij}}{\pi_i \cdot \pi_{j/i}} = \frac{1}{M_{\bullet}} \sum_{i=1}^{n} \frac{\hat{Y}_{i\bullet}}{\pi_i}}$$

Stimatore della media.

$\pi_{j/i}$ *probabilità di inclusione dell'unità secondaria j dato che è stato estratto il grappolo i.*

π_i *probabilità di inclusione dell'unità primaria i.*

$\hat{Y}_{i\bullet}$ *stimatore di secondo stadio del totale nell'unità primaria i.*

E' uno stimatore corretto ➔ *stimatore di Horvitz-Thompson*

Varianza dello stimatore:

-senza ripetizione-

$$\boxed{Var(\hat{Y}_{DSS}) = Var_1\left(\sum_{i=1}^{n}\frac{\tilde{y}_{i\bullet}}{\pi_i}\right) + \sum_{i=1}^{N}\frac{Var_2(\hat{Y}_{i\bullet})}{\pi_i}}$$

varianza di primo stadio — *varianza di secondo stadio*

Stimatore della varianza

Var_1 *varianza dello stimatore di Horvitz-Thompson del totale della popolazione nel campionamento ad uno stadio.*

Var_2 *varianza di secondo stadio dello stimatore del totale del grappolo i del campione.*

Dim: varianza dello stimatore!!!

Si parta dalla relazione:

$$Var(\hat{Y}_{DSS}) = Var_1\left[E_2(\hat{Y}_{DSS})\right] + E_1\left[Var_2(\hat{Y}_{DSS})\right]$$

E_1 *valore atteso calcolato al variare delle unità primarie.*

E_2 *valore atteso condizionato dalla presenza nel campione delle unità primarie.*

$$1)\ E_2(\hat{Y}_{DSS}) = E_2\left(\sum_{i=1}^{n} \frac{\hat{Y}_{i\bullet}}{\pi_i}\right) = \sum_{i=1}^{n} \frac{1}{\pi_i} E_2\left(\hat{Y}_{i\bullet}\right) = \sum_{i=1}^{n} \frac{\tilde{y}_{i\bullet}}{\pi_i}$$

$$2)\ Var_2(\hat{Y}_{DSS}) = Var_2\left(\sum_{i=1}^{n} \frac{\hat{Y}_{i\bullet}}{\pi_i}\right) = \sum_{i=1}^{n} \frac{1}{\pi_i^2} Var_2\left(\hat{Y}_{i\bullet}\right)$$

in quanto i totali di primo stadio sono a due a due incorrelati tra loro.

Sostituendo i risultati nella formula iniziale:

$$Var(\hat{Y}_{DSS}) = Var_1\left(\sum_{i=1}^{n}\frac{\tilde{y}_{i\bullet}}{\pi_i}\right) + E_1\left[\sum_{i=1}^{n}\frac{1}{\pi_i^2}Var_2\left(\hat{Y}_{i\bullet}\right)\right]$$

Sapendo che:

$$E_1\left[\sum_{i=1}^{n}\frac{1}{\pi_i^2}Var_2\left(\hat{Y}_{i\bullet}\right)\right] = \sum_{i=1}^{N}\frac{1}{\pi_i^2}Var_2\left(\hat{Y}_{i\bullet}\right)\cdot \pi_i$$

Se ne deduce che:

$$Var(\hat{Y}_{DSS}) = Var_1\left(\sum_{i=1}^{n}\frac{\tilde{y}_{i\bullet}}{\pi_i}\right) + \sum_{i=1}^{N}\frac{Var_2(\hat{Y}_{i\bullet})}{\pi_i}$$

c.v.d.!!!

Stima del totale nel campionamento a probabilità costanti

Stimatore corretto del totale:

-senza ripetizione-

Estrazione a probabilità costanti
nel primo e nel secondo stadio

$$\pi_{j/i} = \frac{m_i}{M_i} \qquad \pi_i = \frac{n}{N}$$

$$\boxed{\hat{Y}_{DS} = \sum_{i=1}^{n}\sum_{j=1}^{m_i} \frac{y_{ij}}{\frac{n}{N} \cdot \frac{m_i}{M_i}} = \frac{N}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i} y_{ij}\frac{M_i}{m_i} = \frac{N}{n}\sum_{i=1}^{n}\hat{Y}_{i\bullet}}$$

Stimatore del totale.

Stimatore della varianza:

-senza ripetizione-

Lo stimatore della varianza nel campionamento a due stadi a probabilità variabili è:

$$Var(\hat{Y}_{DSS}) = Var_1\left(\sum_{i=1}^{n}\frac{\tilde{y}_{i\bullet}}{\pi_i}\right) + \sum_{i=1}^{N}\frac{Var_2(\hat{Y}_{i\bullet})}{\pi_i}$$

Varianza di primo stadio a probabilità costanti: stimatore di Horvitz-Thompson.

$$Var_1\left(\sum_{i=1}^{n}\frac{\tilde{y}_{i\bullet}}{\pi_i}\right) = N^2\frac{(1-f_1)}{n}\frac{\sum_{i=1}^{N}(Y_{i\bullet}-Y_{\bullet\bullet}/N)^2}{N-1}$$

Varianza di secondo stadio a probabilità costanti: stimatore di Horvitz-Thompson.

$$Var_2(\hat{Y}_{i\bullet}) = Var_2\left(\sum_{j=1}^{m_i} \frac{M_i}{m_i} y_{ij}\right) = M_i^2 \frac{1-f_2}{m_i} \frac{\sum_{j=1}^{M_i}(Y_{ij} - \overline{Y}_{i\bullet})^2}{M_i - 1}$$

Lo stimatore della varianza diventa:

$$Var(\hat{Y}_{DSS}) = Var_1\left(\sum_{i=1}^{n} \frac{\tilde{y}_{i\bullet}}{\pi_i}\right) + \sum_{i=1}^{N} \frac{Var_2(\hat{Y}_{i\bullet})}{\pi_i}$$

$$\boxed{Var(\hat{Y}_{DS}) = N^2 \frac{(1-f_1)}{n} \frac{\sum_{i=1}^{N}(Y_{i\bullet} - Y_{\bullet\bullet}/N)^2}{N-1} + \frac{N}{n}\sum_{i=1}^{N} \underbrace{M_i^2 \frac{1-f_2}{m_i} \frac{\sum_{j=1}^{M_i}(Y_{ij} - \overline{Y}_{i\bullet})^2}{M_i - 1}}}$$

varianza di secondo stadio S^2_{2i}

Stima corretta della varianza:

-senza ripetizione-

La stima corretta della varianza nel campionamento a due stadi a probabilità variabili è:

$$s^2(\hat{Y}_{DSS}) = s_1^2\left(\sum_{i=1}^{n}\frac{\hat{Y}_{i\bullet}}{\pi_i}\right) + \sum_{i=1}^{n}\frac{s_2^2(\hat{Y}_{i\bullet})}{\pi_i}$$

Stima corretta della varianza di primo stadio a probabilità costanti: stimatore di Horvitz-Thompson.

$$s_1^2\left(\sum_{i=1}^{n}\frac{\hat{Y}_{i\bullet}}{\pi_i}\right) = N^2\frac{(1-f_1)}{n}\frac{\sum_{i=1}^{n}(\hat{Y}_{i\bullet} - \sum_{i=1}^{n}\hat{Y}_{i\bullet}/n)^2}{n-1}$$

Stima corretta della varianza di secondo stadio a probabilità costanti: stimatore di Horvitz-Thompson.

$$\sum_{i=1}^{n}\frac{s_2^2(\hat{Y}_{i\bullet})}{\pi_i} = \frac{N}{n}\sum_{i=1}^{n}M_i^2\underbrace{\frac{1-f_2}{m_i}\frac{\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_{i\bullet})^2}{m_i-1}}_{\text{varianza di secondo stadio } s^2_{2i}}$$

Stimatore corretto della proporzione:

-senza ripetizione-

Estrazione a probabilità costanti nel primo e nel secondo stadio

$$\boxed{\hat{P}_{DS} = \frac{1}{n}\sum_{i=1}^{n} p_i}$$

dove:

$$p_i = \frac{a_i}{m_i}$$

La relativa varianza dello stimatore:

$$Var(\hat{P}_{DS}) = \frac{(1-f_1)}{n}\frac{\sum_{i=1}^{n}(p_i - \hat{P}_{DS})^2}{n-1} +$$

$$\frac{1}{n \cdot N}\sum_{i=1}^{n}\frac{1-f_{2i}}{m_i - 1} p_i \cdot q_i$$

Esercizio

Un'azienda che produce abbigliamento da lavoro fornisce tute agli operai di 14 stabilimenti industriali il cui numero di dipendenti è distribuito come riportato nella tabella.

Stab.	Dip
1	25
2	33
3	25
4	102
5	38
6	45
7	60
8	108
9	93
10	83
11	47
12	93
13	124
14	80

L'azienda ha proposto due modelli di tute, indicati con A e B, date in prova per una settimana. Allo scopo di mettere in produzione il modello più gradito, l'azienda intervista gli operai di 3 stabilimenti scelti casualmente.
Vengono estratti gli stabilimenti 2, 8 e 14 e si rileva il numero di operai che hanno dichiarato di preferire il modello di tuta A

Stabilimento	Mi	ai
2	33	28
8	108	75
14	80	53

a) Sulla base dei dati campionari, si stimi la proporzione relativa a tutta l'azienda e si determini l'errore standard.
b) Si supponga che, dopo aver estratto gli stabilimenti 2,8,14, di intervistare il 20% dei dipendenti.

Stabilimento	Mi	ai
2	33	5
8	108	17
14	80	10

Si stimi la proporzione e l'errore standard sulla base dei dati riportati di seguito.

Sol.:

Stabilimento	Mi	mi	ai	pi
2	33	7	5	0,71
8	108	22	17	0,77
14	80	16	10	0,63

$$\hat{P}_{DS} = \frac{1}{n}\sum_{i=1}^{n} p_i = \frac{1}{3}(0{,}71 + 0{,}77 + 0{,}63) = 0{,}70$$

$$Var(\hat{P}_{DS}) = \frac{(1-f_1)}{n}\frac{\sum_{i=1}^{n}(p_i - \hat{P}_{DS})^2}{n-1} +$$

$$\frac{1}{n \cdot N}\sum_{i=1}^{n}\frac{1-f_{2i}}{m_i - 1} p_i \cdot q_i$$

$$(1-f_1) = 0{,}79$$

$$(1-f_2) = 0{,}80$$

$$\sum_{i=1}^{n}\frac{p_i \cdot q_i}{m_i - 1} = 0{,}0582$$

$$\sum_{i=1}^{n}(p_i - \hat{P}_{DS})^2 = 0{,}0981$$

Esercizio

In un'indagine campionaria si vuole stimare il totale delle ore trascorse giornalmente dagli impiegati davanti ai videoterminali. Si suppone che il piano di campionamento preveda una prima estrazione con probabilità costanti di 4 delle 20 sedi regionali ed una successiva estrazione, sempre a probabilità costanti, con un tasso di sondaggio pari al 15% degli impiegati in ciascuna sede.

Sedi	*Dimensioni sedi*	*Ampiezze campioni*	*Valori osservati*
1	25	4	*6,8,11,3*
2	43	6	*7,2,9,12,3,6*
3	19	3	*8,9,2*
4	48	7	*1,7,13,4,9,3,2*

a) si definiscano, per il piano di campionamento in esame, le frazioni di campionamento di primo e secondo stadio;

b) si stimi il numero totale delle ore trascorse davanti ai videoterminali dagli impiegati delle 20 sedi e si calcoli la varianza dello stimatore utilizzato.

Sol.:

$$f_1 = \frac{n}{N} = \frac{4}{20} = 0{,}2 \quad ; \quad f_2 = \frac{m_i}{M_i} = \frac{4}{20} = 0{,}15\,(\text{costante})$$

Stimatore del totale:

$$\hat{Y}_{DS} = \frac{N}{n}\sum_{i=1}^{n}\hat{Y}_{i\bullet} \quad ; \quad \hat{Y}_{i\bullet} = \sum_{j=1}^{m_i}\frac{M_i}{m_i}y_{ij}$$

$$\hat{Y}_{DS} = \frac{20}{4}[\frac{25}{4}(6+8+11+3)+\ldots+\frac{48}{7}(1+7+13+4+9+3+2)]$$

$$\hat{Y}_{DS} = \frac{20}{4}\cdot(175+279{,}63+120{,}27+267{,}54) = 5\cdot 842{,}44 = 4212{,}2$$

Stima corretta della varianza:

$$\sum_{i=1}^{n}\frac{s_2^2(\hat{Y}_{i\bullet})}{\pi_i} = \frac{N}{n}\sum_{i=1}^{n}M_i^2\underbrace{\frac{1-f_2}{m_i}\frac{\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_{i\bullet})^2}{m_i-1}}_{\text{varianza di secondo stadio } s^2_{2i}}$$

$$s_{21}^2 = \frac{(6-7)^2 + (8-7)^2 + (11-7)^2 + (3-7)^2}{4-1} = 11{,}33$$

$$s_{22}^2 = \frac{(7-6{,}5)^2 + (2-6{,}5)^2 + \ldots + (6-6{,}5)^2}{6-1} = 13{,}90$$

$$s_{23}^2 = \frac{(8-6{,}3)^2 + (9-6{,}3)^2 + (2-6{,}3)^2}{3-1} = 14{,}34$$

$$s_{24}^2 = \frac{(1-5{,}6)^2 + (7-5{,}6)^2 + \ldots + (2-5{,}6)^2}{7-1} = 18{,}62$$

Totali di secondo stadio:

$$\hat{Y}_{i\bullet} = \sum_{j=1}^{m_i} \frac{M_i}{m_i} y_{ij}$$

$$\hat{Y}_{1\bullet} = \frac{25}{4}(6+8+11+3) = 175$$

$$\hat{Y}_{2\bullet} = \frac{43}{6}(7+2+9+12+3+6) = 279{,}24$$

$$\hat{Y}_{3\bullet} = \frac{19}{3}(8+2+9) = 120{,}33$$

$$\hat{Y}_{4\bullet} = \frac{48}{7}(1+7+13+4+9+3+2) = 267{,}54$$

Varianza di primo stadio:

$$s_1^2\left(\sum_{i=1}^{n}\frac{\hat{Y}_{i\bullet}}{\pi_i}\right)=N^2\,\frac{(1-f_1)}{n}\,\frac{\sum_{i=1}^{n}(\hat{Y}_{i\bullet}-\sum_{i=1}^{n}\hat{Y}_{i\bullet}/n)^2}{n-1}$$

$$\sum_{i=1}^{n}\frac{\hat{Y}_{i\bullet}}{n}=\frac{842{,}11}{4}=210{,}53$$

$$\frac{\sum_{i=1}^{n}(\hat{Y}_{i\bullet}-\sum_{i=1}^{n}\hat{Y}_{i\bullet}/n)^2}{n-1}=\frac{(175-210{,}53)^2+..+(267{,}54-210{,}53)^2}{4-1}$$

$$=\frac{17369{,}62}{3}=5789{,}87$$

$$s_1^2\left(\sum_{i=1}^{n}\frac{\hat{Y}_{i\bullet}}{\pi_i}\right)=N^2\,\frac{(1-f_1)}{n}\,\frac{\sum_{i=1}^{n}(\hat{Y}_{i\bullet}-\sum_{i=1}^{n}\hat{Y}_{i\bullet}/n)^2}{n-1}=\frac{20^2(1-\frac{4}{20})}{4}\cdot 5789{,}87$$

$$=\frac{400\cdot 0{,}8}{4}\cdot 5789{,}87=463189{,}6$$

Varianza di secondo stadio:

$$\sum_{i=1}^{n} \frac{s_2^2(\hat{Y}_{i\bullet})}{\pi_i} = \frac{N}{n} \sum_{ì=1}^{n} M_i^2 \frac{1-f_2}{m_i} \frac{\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_{i\bullet})^2}{m_i-1}$$

$$= \frac{20}{4}[25^2 \frac{1-0{,}15}{4} 11{,}33 + \ldots + 48^2 \frac{1-0{,}15}{7} 18{,}62] = 11821{,}84$$

Stima corretta della varianza:

$$s^2(\hat{Y}_{DSS}) = s_1^2 \left(\sum_{i=1}^{n} \frac{\hat{Y}_{i\bullet}}{\pi_i} \right) + \frac{N}{n} \sum_{i=1}^{n} s_2^2(\hat{Y}_{i\bullet})$$

$$s^2(\hat{Y}_{DSS}) = 463189{,}6 + 11821{,}84 = 475011{,}44$$

Stima per quoziente del totale nel campionamento a probabilità costanti

Riduzione della variabilità della stima del totale per n => +∞

Stimatore per quoziente del totale:

-senza ripetizione-

$$\boxed{\hat{Y}_{DSq} = \frac{\sum_{i=1}^{n} \hat{Y}_{i\bullet}}{\sum_{i=1}^{n} M_i} M_{\bullet}}$$

Stimatore per quoziente del totale.

Varianza dello stimatore per quoziente:

Si consideri la trasformata Z = Y - R · X

dove

$$R = \frac{\sum_{i=1}^{N}\sum_{i=1}^{M_i} Y_{ij}}{\sum_{i=1}^{N}\sum_{i=1}^{M_i} X_{ij}} = \bar{Y}_{\bullet\bullet}$$

Varianza dello stimatore del totale a probabilità costanti:

$$Var(\hat{Y}_{DS}) = N^2 \frac{(1-f_1)}{n} \frac{\sum_{i=1}^{N}(Y_{i\bullet} - Y_{\bullet\bullet}/N)^2}{N-1} +$$

$$\frac{N}{n}\sum_{i=1}^{N} M_i^2 \underbrace{\frac{1-f_2}{m_i} s_{2i}^2}$$

$$\boxed{Z_{i\bullet} = Y_{i\bullet} - M_i \cdot \bar{Y}_{\bullet\bullet} \quad ; \quad Z_{\bullet\bullet} = Y_{\bullet\bullet} - M_{\bullet} \cdot \bar{Y}_{\bullet\bullet}}$$

Varianza dello stimatore per quoziente del totale

$$Var(\hat{Y}_{DS}) = N^2 \frac{(1-f_1)}{n} \frac{\sum_{i=1}^{N}(Y_{i\bullet} - M_i \cdot \bar{Y}_{\bullet\bullet})^2}{N-1} + \frac{N}{n}\sum_{i=1}^{N} M_i^2 \frac{1-f_2}{m_i} s_{2i}^2$$

stima corretta per quoziente della varianza

$$s^2\left(\hat{Y}_{DSq}\right) = N^2 \frac{(1-f_1)}{n} \frac{\sum_{i=1}^{n}(\hat{Y}_{i\bullet} - M_i \cdot \hat{\bar{Y}}_{\bullet\bullet})^2}{n-1} + \frac{N}{n}\sum_{i=1}^{n} M_i^2 \frac{1-f_2}{m_i} s_{2i}^2$$

ottenuta sostituendoli valore della trasformata Z!!

Esercizio

Si vuole stimare il numero di pazienti ricoverati per incidenti automobilistici, nel corso di un anno, negli ospedali non specializzati di un'area composta da 10 comuni. I dati relativi alla popolazione sono riportati in tabella:

Comuni	1	2	3	4	5	6	7	8	9	10
N.Ospedali	2	3	1	1	4	2	4	2	3	1
N.Posti letto	677	198	300	32	410	275	570	311	408	65

Si decide di effettuare un campionamento a due stadi, con i comuni come unità primarie e gli ospedali come unità secondarie. Al primo stadio vengono scelti, con probabilità costante, tre comuni; al secondo viene scelto un ospedale per comune. I risultati ottenuti dall'indagine sono i seguenti:

Comuni	*Casi di ricovero per incidente*	*N. posti letto*
1	14	310
7	10	223
8	6	150

a) proporre uno stimatore che non tenga conto di variabili ausiliarie e stimare il numero di ricoveri per incidenti automobilistici nel corso di un anno;
b) Proporre uno stimatore che tenga conto delle variabili ausiliarie e stimare il numero di ricoveri per incidenti automobilistici;
c) Commentare la differenza dei risultati ottenuti nei punti precedenti.

Sol.:

a)

$$\hat{Y}_{DS} = \frac{N}{n}\sum_{i=1}^{n}\hat{Y}_{i\bullet} \quad ; \quad \hat{Y}_{i\bullet} = \sum_{j=1}^{m_i}\frac{M_i}{m_i}y_{ij}$$

$$\hat{Y}_{DS} = \frac{10}{3}[(2\cdot 14 + 4\cdot 10 + 2\cdot 6)] = 266,4$$

b)

$$\hat{Y}_{DSq} = \frac{\sum_{i=1}^{n}\hat{Y}_{i\bullet}}{\sum_{i=1}^{n}M_i}M_{\bullet}$$

$$\sum_{i=1}^{n}\hat{Y}_{i\bullet} = \sum_{i=1}^{n}\frac{M_i}{m_i}(y_{1\bullet} + y_{7\bullet} + y_{8\bullet}) = \frac{677}{310}14 + \frac{570}{223}10 + \frac{311}{150}6 = 68,57$$

$$\hat{Y}_{DSq} = \frac{68,57}{M_1 + M_7 + M_8}M_{\bullet} = \frac{68,57}{677+570+311}3246 = 142,86$$

- - ESERCIZI - -

Esercizio 1

In un'indagine campionaria si vuole stimare il totale delle ore trascorse giornalmente dagli impiegati davanti ai videoterminali. Si suppone che il piano di campionamento preveda una prima estrazione con probabilità costanti di 4 delle 20 sedi regionali ed una successiva estrazione, sempre a probabilità costanti, con un tasso di sondaggio pari al 15% degli impiegati in ciascuna sede.

Sedi	Dimensioni sedi	Ampiezze campioni	Valori osservati
1	25	4	6,8,11,3
2	43	6	7,2,9,12,3,6
3	19	3	8,9,2
4	48	7	1,7,13,4,9,3,2

a) si definiscano, per il piano di campionamento in esame, le frazioni di campionamento di primo e secondo stadio;
b) si stimi il numero totale delle ore trascorse davanti ai videoterminali dagli impiegati delle 20 sedi e si calcoli la varianza dello stimatore utilizzato.

Esercizio 2

Si vuole stimare il numero di pazienti ricoverati per incidenti automobilistici, nel corso di un anno, negli ospedali non specializzati di un'area composta da 10 comuni. I dati relativi alla popolazione sono riportati in tabella:

Comuni	1	2	3	4	5	6	7	8	9	10
N.Ospedali	2	3	1	1	4	2	4	2	3	1
N.Posti letto	677	198	300	32	410	275	570	311	408	65

Si decide di effettuare un campionamento a due stadi, con i comuni come unità primarie e gli ospedali come unità secondarie. Al primo stadio vengono scelti, con probabilità costante, tre comuni; al secondo viene scelto un ospedale per comune. I risultati ottenuti dall'indagine sono i seguenti:

Comuni	Casi di ricovero per incidente	N. posti letto
1	14	310
7	10	223
8	6	150

a) Proporre uno stimatore che non tenga conto di variabili ausiliarie e stimare il numero di ricoveri per incidenti automobilistici nel corso di un anno;
b) Proporre lo stesso stimatore e uno alternativo che tengano conto della variabile ausiliaria per stimare il numero di ricoveri per incidenti automobilistici;
c) Commentare la differenza dei risultati ottenuti nei punti precedenti.

Esercizio 3

Un'azienda che produce abbigliamento da lavoro fornisce tute agli operai di 14 stabilimenti industriali il cui numero di dipendenti è distribuito come riportato nella tabella.

Stab.	Dip
1	25
2	33
3	25
4	102
5	38
6	45
7	60
8	108
9	93
10	83
11	47
12	93
13	124
14	80

L'azienda ha proposto due modelli di tute, indicati con A e B, date in prova per una settimana. Allo scopo di mettere in produzione il modello più gradito, l'azienda intervista gli operai di 3 stabilimenti scelti casualmente.
Vengono estratti gli stabilimenti 2, 8 e 14 e si rileva il numero di operai che hanno dichiarato di preferire il modello di tuta A

Stabilimento	Mi	ai
2	33	28
8	108	75
14	80	53

c) Sulla base dei dati campionari, si stimi la proporzione relativa a tutta l'azienda e si determini l'errore standard.
d) Si supponga che, dopo aver estratto gli stabilimenti 2,8,14, di intervistare il 20% dei dipendenti. Si stimi la proporzione e l'errore standard sulla base dei dati riportati di seguito.
e)

Stabilimento	Mi	ai
2	33	5
8	108	17
14	80	10

Esercizio 4

L'amministrazione provinciale di una certa provincia del nord Italia vuole stimare il numero di ore di straordinario realizzato in una settimana dal personale dipendente dei 217 Comuni di cui una provincia si compone corrispondenti a 6510 dipendenti. Allo scopo sceglie casualmente 17 Comuni e successivamente il 25% dei dipendenti delle amministrazioni estratte ottenendo i seguenti risultati.

Comuni.	*Numero Dipendenti*	*Numero ore settimanali di straordinario*	*Varianze s_i^2*
1	27	40	17.64
2	15	23	4.41
3	6	11	2.25
4	32	37	18.49
5	19	31	8.41
6	11	18	2.56
7	64	76	42.25
8	81	103	98.01
9	23	34	16.00
10	29	36	15.21
11	41	54	22.09
12	9	10	1.69
13	18	33	13.69
14	25	35	9.61
15	13	19	4.00
16	35	48	26.01
17	53	63	33.64
TOTALE	**501**		

Stimare il numero totale di ore di straordinario effettuate dai dipendenti con le strategie a disposizione e confrontare i risultati ottenuti.

Esercizio 5

Un istituto bancario a diffusione nazionale desidera valutare il numero di clienti che sarebbero disposti a sottoscrivere un conto corrente on-line, disponendo dei seguenti dati:

	Nord	*Centro*	*Sud*	*Isole*
Clienti	*16552*	*14892*	*10689*	*4586*

Si predispone allo scopo un campione di clienti con una frazione di sondaggio del 5%, che fornisce i seguenti risultati:

	Nord	*Centro*	*Sud*	*Isole*
Totale clienti che aderirebbero al conto on-line	*824*	*658*	*310*	*214*
S_h^2	*1229*	*897*	*1458*	*1664*

a) Fornire una stima ad intervallo del totale dei clienti che aderirebbero ad un conto on-line e la relativa varianza.