2. Campionamento Casuale Semplice e Stimatori

2.1 Campionamento casuale semplice

Sebbene non sia molto diffuso nella pratica delle indagini, il campionamento casuale semplice rappresenta il naturale punto di partenza per lo studio di tutti gli altri disegni campionari.

Il campionamento casuale semplice (CCS) costituisce il metodo di campionamento più elementare; può avere valenza autonoma, ma viene più frequentemente utilizzato in congiunzione ad altre tecniche.

Tale tecnica attribuisce *una probabilità di selezione* ad ogni unità della popolazione che può essere **costante** (unità elementari) o **variabile** (unità composte); ne consegue che ogni singola unità della popolazione ha la stessa probabilità di entrare a far parte del campione.

PREGI

- E' privo di errori di selezione: nessuna unità o gruppo di unità è favorito a priori;
- E' molto semplice ed ha quindi un basso costo;

DIFETTI

- E' necessario disporre di una lista completa delle unità;
- Può non essere "rappresentativo": alcune particolari sezioni della popolazione potrebbero non essere rappresentate.

- La rilevazione sul campo può essere molto costosa se l'intervista viene effettuata attraverso un rilevatore e le unità selezionate sono fra loro lontane (alti costi di spostamento).

Se le unità vengono estratte singolarmente e in maniera del tutto casuale rimuovendo dalla popolazione la singola unità estratta si parla si CCS ***senza ripetizione*** *.*

Lo spazio campionario Ω è costituito dall'insieme dei campioni non ordinati e formati da unità tutte tra loro distinte.
Il piano di campionamento associato ad ogni campione, come già anticipato, è pari a:

$$p(c) = \frac{1}{\binom{N}{n}}$$

a cui corrispondono probabilità di inclusione che cambiano a seconda se la selezione è a probabilità costanti o variabili.
A probabilità *costanti* si dimostra che tutte le unità hanno la stessa probabilità di estrazione e dunque di inclusione nel campione (auto-ponderante):

$$\pi_i = \frac{n}{N}$$

Dim !!

Quella precedente è chiamata anche probabilità di inclusione di primo ordine in quanto associata ad una sola unità. Tale rapporto

si chiama anche *frazione di sondaggio* che coincide con la frazione della popolazione campionata.
Le probabilità di secondo ordine sono invece:

$$\pi_{ij} = \frac{n(n-1)}{N(N-1)}$$

La probabilità costante riscontrata nel primo e nel secondo ordine è presente per tutti gli ordini successivi ed è tipica del campionamento casuale semplice.

A probabilità *variabili* le unità (composte) hanno probabilità distinte di essere selezionate e proporzionali alla dimensione.

In generale, per attribuire una probabilità variabile è necessario disporre di un indicatore di importanza relativa detta *misura di ampiezza* o *misura di dimensione* indicate per mezzo di una variabile ausiliaria X (residenti per le città, studenti per atenei/scuole, posti letto per alberghi e strutture ricettive, ecc).

$\mathcal{X}$: *variabile ausiliaria* ritenuta in relazione di approssimata proporzionalità con la variabile oggetto di studio $\mathcal{Y}$.

- X_i: misura d'ampiezza, valore assunto da $\mathcal{X}$ nell'unità *i* della popolazione.

dove $X = \sum_{j=1}^{N} X_j$ è il totale della variabile nella popolazione.

- $P_j = \frac{X_j}{X}$, $j = 1,2,\cdots,N$: misura d'ampiezza normalizzata, rappresenta la probabilità di estrazione dell'unità j-esima;

Tali misure rappresentano i pesi in base ai quali vengono selezionate le unità della popolazione.

-Proprietà-

1) proporzionalità rispetto alle misure di ampiezza X_i, ossia:

πi proporzionale a P_i, $\forall i$;

[contribuisce a calcolare facilmente la probabilità di inclusione del primo ordine]

2) positività delle probabilità di inclusione del secondo ordine: $\pi_{ij}>0$, $\forall(i,j)$;
[è indispensabile perché sussistano stimatori corretti della varianza]

3) non negatività della differenza $\pi_i\pi_j-\pi_{ij}$, $\forall(i,j)$;
[garantisce la non negatività della varianza dello stimatore]

4) soddisfacimento della relazione $\pi_{ij}/\pi_i\pi_j>A$, per $A>0$ non prossima allo zero.
[consente di definire le stime della varianza dello stimatore medesimo]

La struttura della tabella di partenza è simile alla seguente:

Unità popolazione	Superficie *Xi*	*Pi*	Cumulata *Fi*
1	X_1	X_1/X	$F_{1=}\ X_1$
2	X_2	X_2/X	$F_{1=}\ X_1+X_2$
3	X_3	X_3/X	F_3
4	X_4	X_4/X	F_4
5	X_5	X_5/X	F_5
…	…	…	…
N	X_N	X_N/X	$F_N=X$
	X	**1**	

Mentre, a seguito della prima estrazione, eliminando dall'elenco l'unità estratta, diventa come segue:

Unità popolazione	Cumulata *Fi*	*Pi*
1	$F_{1=} X_1$	$X_1/ X\text{-}Xi$
2	$F_{1=} X_1+X_2$	$X_2/ X\text{-}Xi$
3	F_3	$X_3/ X\text{-}Xi$
4	F_4	$X_4/ X\text{-}Xi$
5	F_5	$X_5/ X\text{-}Xi$
…	…	…
N-1	$F_{N-1}=X$	$X_{N-1}/X\text{-}Xi$

Le probabilità di inclusione calcolate sono differenti a seconda delle tecniche di selezione delle unità costituenti l'universo di studio. Quello più utilizzato è il metodo di Yates-Grundy secondo il quale le probabilità di inclusione posso essere così definite:

$$\pi_i = P_i \cdot \left(1 + \sum_{j \neq i}^{N} \frac{P_j}{1 - P_i}\right)$$ *Dim !!*

mentre la corrispondete probabilità di inclusione di secondo ordine è pari a:

$$\pi_{ij} = P_j \frac{P_i}{1 - P_i} + P_i \frac{P_j}{1 - P_j}$$ *Dim !!*

Nella pratica la procedura maggiormente usata è quella senza ripetizione.

L'estrazione è realizzabile se sono soddisfatte le seguenti condizioni:

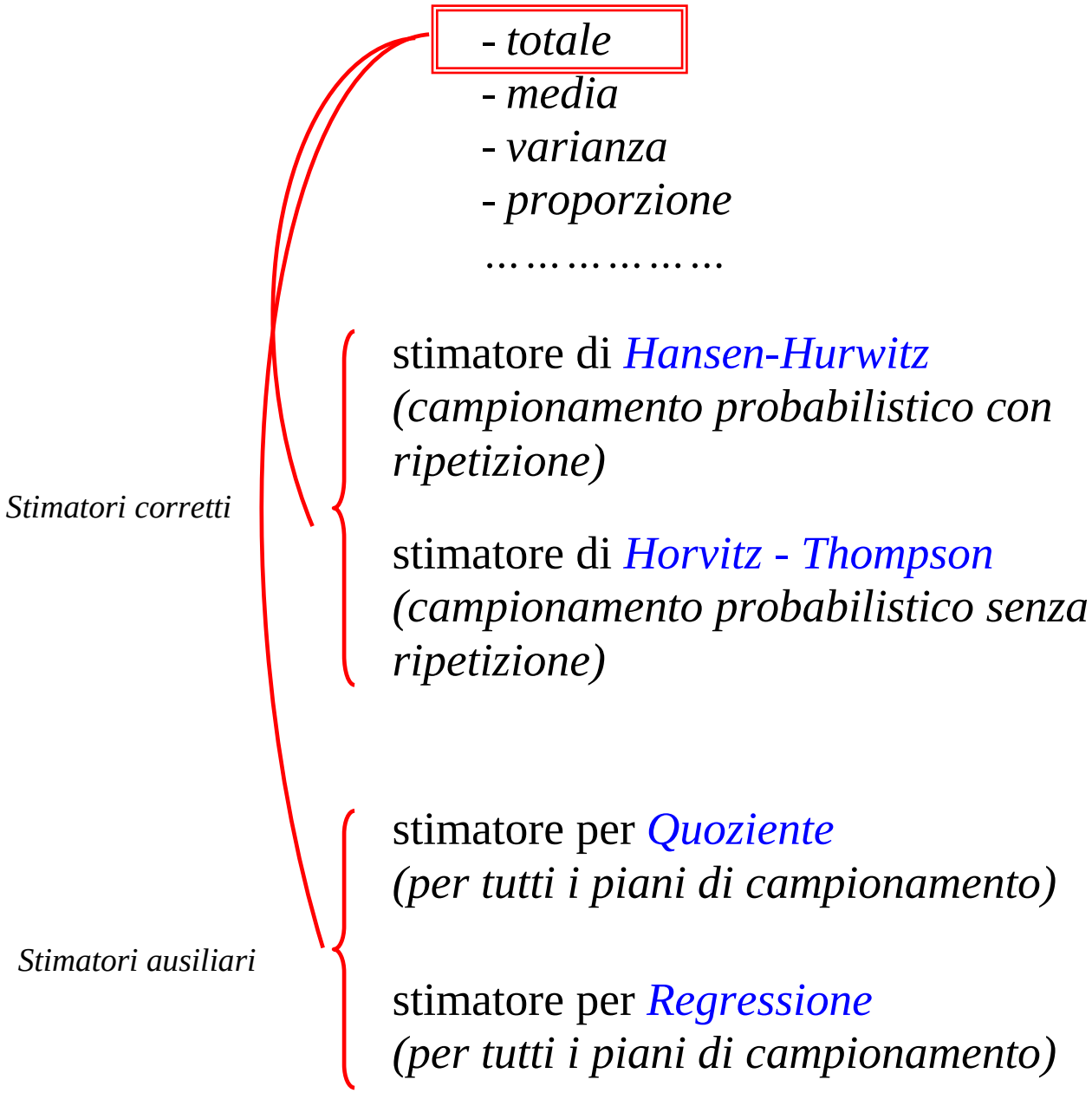
a - la popolazione può essere suddivisa in elementi **distinti** e **identificabili** (unità elementari o composte)

b – sia disponibile una **lista** (nominativa o di etichette) delle unità componenti la popolazione

2.2 Stimatore del totale

1. Selezione del piano di campionamento

2. Scelta dello stimatore:



2.2.1 Stimatore di Horvitz-Thompson (SR):

Stimatore per il totale in un piano di campionamento probabilistico - estrazione SENZA ripetizione.

$$\hat{Y} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$

Stimatore corretto del totale di Horvitz-Thompson

π_i : Probabilità di inclusione del primo ordine *(probabilità che l'unità i-esima appartenga al campione).*

dove

$$\pi_i = \sum_{S} \delta_i \, p(s) = E(\delta_i)$$

$$\delta \begin{cases} = 0 & \textit{se l'unità non è presente nel campione} \\ = 1 & \textit{se l'unità è presente nel campione} \end{cases}$$

Lo stimatore H-T è uno stimatore sempre corretto rispetto al totale della popolazione.
DIM.:

$$E[\hat{Y}_{H-T}] = Y$$ *per definizione*

Lo stimatore può essere scritto anche nel modo seguente:

$$\hat{Y}_{H-T} = \sum_{j=1}^{N} \frac{Y_j}{\pi_j} \delta_j$$

δ_j : Variabile dicotomica con valori 0 e 1

$$E[\hat{Y}_{H-T}] = \sum_{j=1}^{N} \frac{Y_j}{\pi_j} E[\delta_j] = Y$$

c.v.d.!!!

Si calcoli l'efficienza dello stimatore corretto del totale.
DIM:

Si consideri una combinazione lineare generica:

$$Y = c_1X_1 + c_2X_2 + \ldots\ldots + c_nX_n$$

con

c_i= costanti
X_i= variabili casuali

$$Var(Y) = E[Y-E(Y)]^2$$

$$= Var(c_1X_1)+Var(c_2X_2)+..+Var(c_nX_n)+...=$$

$$= c_1^2 \cdot Var(X_1) + c_2^2 \cdot Var(X_2) + .. + c_n^2 \cdot Var(X_n) .. =$$

$$= \sum_{i=1}^{n} c_i^2 E(X_i - E[X_i])^2 + \sum_{i=1}^{n}\sum_{j\neq i}^{n} c_i c_j E[(X_i - E[X_i])(X_j - E[X_j])] =$$

da cui:

$$Var(Y) = \sum_{i=1}^{n} c_i^2 Var(X_i) + \sum_{i=1}^{n}\sum_{j\neq i}^{n} c_i c_j Cov(X_i, X_j)$$

... tornando allo stimatore H-T :

$$Var(\hat{Y}_{H-T}) = \sum_{j=1}^{N} \frac{Y_j^2}{\pi_j^2} Var(\delta_j) + \sum_{j=1}^{N} \sum_{i \neq j}^{N} \frac{Y_i}{\pi_i} \frac{Y_j}{\pi_j} Cov(\delta_i \delta_j) =$$

ma:

$Var(\delta_j) = \pi_j (1- \pi_j)$

$Cov(\delta_i;\delta_j) = E(\delta_i;\delta_j) - E(\delta i) E(\delta_j) = \pi_{ij} - \pi_i \pi_j$

$$Var(\hat{Y}) = \sum_{j=1}^{N} \frac{1-\pi_j}{\pi_j} Y_j^2 + \sum_{j=1}^{N} \sum_{i \neq j}^{N} \left(\frac{\pi_{ij}}{\pi_i \cdot \pi_j} - 1\right) Y_i \cdot Y_j$$

c.v.d.!!!

Stima corretta della varianza sul campione:

$$v(\hat{Y}) = \sum_{i=1}^{n} \frac{1-\pi_i}{\pi_i^2} y_i^2 + \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left(\frac{1}{\pi_i \cdot \pi_j} - \frac{1}{\pi_{ij}} \right) y_i \cdot y_j$$

2.2.2 Stimatori a probabilità costanti:

Stima del totale per unità elementari adottando il piano di campionamento casuale semplice SENZA ripetizione.

Le rispettive probabilità di inclusione:

$\pi_j = n / N$ *(costante per tutte le unità della popolazione)*

$\pi_{ij} = n(n-1)/N(N-1)$ *(costante per tutte le unità i,j della popolazione)*

$$1)\hat{Y}_{H-T} = \sum_{i=1}^{n} \frac{y_i}{n/N} = N \cdot \bar{y}$$

Stimatore del totale a probabilità costanti.

Si calcoli l'efficienza dello stimatore corretto del totale a probabilità costanti.

DIM:

$$Var(\hat{Y}_{H-T}) = \sum_{j=1}^{N} Y_j^2 \left(\frac{N}{n} - 1\right) + \sum_{j=1}^{N}\sum_{i \neq j}^{N} \left(\frac{N}{n}\frac{n-1}{N-1} - 1\right) \cdot Y_i \cdot Y_j$$

$$= \frac{N}{n}\sum_{j=1}^{N} Y_j^2 - \sum_{j=1}^{N} Y_j^2 + \frac{N}{n}\frac{n-1}{N-1}\sum_{j=1}^{N}\sum_{i \neq j}^{N} Y_i \cdot Y_j - \sum_{j=1}^{N}\sum_{i \neq j}^{N} Y_i \cdot Y_j$$

sapendo che:

$$Y^2 = \sum_{j=1}^{N} Y_j^2 + \sum_{j=1}^{N}\sum_{i \neq j}^{N} Y_i \cdot Y_j$$

di conseguenza:

$$= \frac{N}{n}\sum_{j=1}^{N} Y_j^2 + \frac{N}{n}\frac{n-1}{N-1}\left(Y^2 - \sum_{j=1}^{N} Y_j^2 \right) - Y^2$$

$$= \frac{N}{n}\sum_{j=1}^{N} Y_j^2 + \frac{N}{n}\frac{n-1}{N-1} Y^2 - \frac{N}{n}\frac{n-1}{N-1}\sum_{j=1}^{N} Y_j^2 - Y^2$$

$$= \left(\frac{N}{n} - \frac{N}{n}\frac{n-1}{N-1} \right)\sum_{j=1}^{N} Y_j^2 - \left(1 - \frac{N}{n}\frac{n-1}{N-1} \right) Y^2$$

considerando poi:

$$\sum_{j=1}^{N} Y_j^2 = NM_2 \quad ; \quad Y^2 = N^2 \overline{Y}^2$$

$$= \left(\frac{N^2}{n} - \frac{N^2}{n}\frac{N-1}{n-1} \right) M_2 + \left(\frac{N^3}{n}\frac{n-1}{N-1} - N^2 \right) M_1^2$$

$$= \frac{N^2}{n}\left(\frac{N-n}{N-1} \right) M_2 - N^2 \left(\frac{N-n}{n(N-1)} \right) M_1^2$$

Considerando inoltre la relazione tra le due varianze:

$$S^2 = \frac{N}{N-1} \cdot \sigma^2$$

e

$$\frac{N-n}{N} = 1 - f$$

Sostituendo:

$$Var(\hat{Y}_{H-T}) = N^2 \frac{1-f}{n} S^2$$

dove:

S^2: è la varianza della popolazione proporzionale a σ^2.
f: tasso di sondaggio (n / N).

Stima della varianza sul campione:

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

2.3 Stimatore della media

Lo stimatore della media campionaria ponderata è deducibile da quello del totale:

$$\hat{\bar{Y}}_{H-T} = \frac{1}{N}\hat{Y}_{H-T} = \frac{1}{N}\sum_{i=1}^{n}\frac{y_i}{\pi_i}$$

Mentre a probabilità di inclusione costanti è:

$$\hat{\bar{Y}}_{H-T} = \frac{1}{N}\hat{Y}_{H-T} = \frac{1}{N}\cdot N\cdot \bar{y} = \bar{y}$$

Stimatore della media a probabilità costanti.

La varianza dello stimatore risulta:

$$Var\,(\hat{\bar{Y}}_{H-T}) = Var\left(\frac{\hat{Y}_{H-T}}{N}\right) = \frac{1}{N^2}\cdot Var\left(\hat{Y}_{H-T}\right) = \frac{1-f}{n}S^2$$

Stima della varianza sul campione:

$$v(\hat{\bar{Y}}_{H-T}) = \frac{1-f}{n}s_y^2$$

2.4 Stimatore della proporzione:

Se il carattere qualitativo Y è dicotomico, ala i-esima unità della popolazione viene assegnato il valore $Y_i=1$ o $Y_i=0$ a seconda se il carattere è presente o meno nell'unità osservata.

$$\hat{P}_{H-T} = \frac{1}{n}\sum_{i=1}^{n} yi = \frac{A}{n}$$

Per misurare il livello di efficienza si parte dalla varianza dello stimatore della media:

$$Var\ (\hat{\bar{Y}}_{H-T}) = \frac{1-f}{n} S^2$$

Si dimostra che, se il carattere è dicotomico (si ipotizza la distribuzione bernoulliana), per la proporzione vale la relazione:

$$S^2 = \frac{N}{N-1} P(1-P)$$

DIM:

Essendo, per definizione:

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i^2 + \bar{Y}^2 - 2Y_i\ \bar{Y})$$

per un carattere dicotomico, essendo P la media:

$$\sum_{i=1}^{N} Y_i^2 = \sum_{i=1}^{N} Y_i\ = N \cdot P$$

per cui

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 + \sum_{i=1}^{N} \bar{Y}^2 - \sum_{i=1}^{N} 2Y_i\,\bar{Y}\right] =$$

$$= \frac{1}{N-1}\left[NP + NP^2 - 2NP^2\right] == \frac{N}{N-1}P(1-P)$$

$$= \frac{n}{n-1}P(1-P)$$

se calcolato in un campione

$$Var(\hat{P}_{H-T}) = \frac{1-f}{n-1} \cdot p(1-p)$$

2.5 Stimatore per quoziente

In precedenza si è visto come, nell'ambito del campionamento con probabilità variabili, gli stimatori corretti si giovino dell'impiego di una variabile ausiliaria X nota per tutte le unità della popolazione.

Con questo approccio la variabile ausiliaria viene utilizzata a livello di campionamento anche se poi si riflette sullo stimatore dato che compaiono nelle probabilità di inclusione.

Con gli stimatori ausiliari si vuole utilizzare la variabile X a livello di stimatore. Ad esempio si immagini di voler stimare la produzione totale delle imprese di un certo settore. Si ammetta di conoscere il totale degli addetti nel settore e che vi sia, a livello di singola impresa, una relazione di proporzionalità tra numero di addetti e produzione. In tal caso, estratto un campione casuale semplice senza ripetizione, il seguente stimatore può essere ragionevolmente considerato come stimatore del totale:

$$\hat{Y}_q = \frac{\hat{Y}}{\hat{X}} X = \hat{Y} \frac{X}{\hat{X}}$$

Stimatore non corretto del totale: *stimatore per quoziente.*

con X noto.

VANTAGGIO - Stimatore applicabile per qualunque piano di campionamento: con e senza ripetizione, a probabilità uguali o variabili.

$\frac{X}{\hat{X}}$ } *fattore di correzione per ridurre la variabilità campionaria dello stimatore corretto intorno al totale Y.*

Stimatore "non corretto" del totale in funzione della variabile ausiliaria.

Si calcoli la distorsione dello stimatore per quoziente per una bassa numerosità campionaria.

DIM:

$$E[\hat{Y}_q] = X \cdot E\left[\frac{\hat{Y}}{\hat{X}}\right] = X \cdot E\left[\hat{R}\right] \neq Y$$

Stimatore distorto.

La distorsione può essere quantificata come segue:

$$d(\hat{Y}_q) = E[\hat{Y}_q] - Y$$

$$= X \cdot E[\hat{R}] - E\left[\hat{Y}\right]$$

$$= E\left[\hat{X}\right] \cdot E[\hat{R}] - E\left[\hat{X} \cdot \hat{R}\right] = -Cov(\hat{X}, \hat{R})$$

Un altro modo per esprimere la distorsione, più semplice per le procedure pratiche, è la seguente:

$$d[\hat{Y}_q] = -Cov(\hat{R}, \hat{X})$$

$$Cov(\hat{R}, \hat{X}) = \rho(\hat{R}, \hat{X}) \sqrt{Var(\hat{R}) Var(\hat{X})}$$

$$d[\hat{Y}_q] = -\rho(\hat{R}, \hat{X}) \sqrt{Var(\hat{R}) Var(\hat{X})}$$

$$\left| d[\hat{Y}_q] \right| \leq \sqrt{Var(\hat{R}) Var(\hat{X})}$$

sapendo che

$$\sqrt{Var(\hat{X})} = CV(\hat{X}) \cdot X$$

si ha:

$$\left| d[\hat{Y}_q] \right| \leq \sqrt{Var(\hat{R})} \cdot CV(\hat{X}) \cdot X$$

$$Var(\hat{Y}_q) = X^2 \cdot Var(\hat{R})$$

$$\left| d[\hat{Y}_q] \right| \leq CV(\hat{X}) \cdot \sqrt{Var(\hat{Y}_q)}$$

c.v.d.!!!

Teorema:
Lo stimatore del totale per quoziente è uno stimatore approssimativamente corretto asintoticamente (per n>30).

DIM:

Si ponga:

$$\delta_y = \frac{\hat{Y} - Y}{Y} \quad ; \quad \delta_x = \frac{\hat{X} - X}{X}$$

allora si può scrivere:

$$\hat{Y}_q = Y(1+\delta_y)(1+\delta_x)^{-1} .$$

Se si assume che:

$$|\delta_x| < 1 \quad \Rightarrow \quad 0 < \hat{X} < 2X$$

allora la funzione $(1+\delta_x)^{-1}$ può essere sviluppata in serie:

$$(1+\delta_x)^{-1} = 1 - \delta_x + \delta_x^2 - .....$$

Sostituendo:

$$\hat{Y}_q = Y(1+\delta_y)(1-\delta_x+\delta_x^2-.....)$$
$$= Y(1+\delta_y-\delta_x-\delta_y\delta_x-\delta_x^2+.....)$$

(…il limite per n che tende all'infinito rende trascurabili i fattori con l'esponente uguale o maggiore di due…)

$$\hat{Y}_q = Y(1+\delta_y-\delta_x) = Y + Y(\delta_y-\delta_x)$$

$$E[\hat{Y}_q] = Y + Y \cdot \underbrace{E(\delta_y-\delta_x)}_{=0} = Y$$

c.v.d.!!!

Teorema:
Nelle ipotesi di grandi campioni e di R=Y/X>0, lo stimatore per quoziente del totale è più efficiente dello stimatore corretto se:

$$\rho(\hat{X},\hat{Y}) > \frac{CV(\hat{X})}{2CV(\hat{Y})}$$

DIM:

I fase: determinazione della varianza dello stimatore

$$E.Q.M.(\hat{Y}_q) = Var(\hat{Y}_q) + d(\hat{Y}_q)^2$$

Per grandi campioni:

$$MSE(\hat{Y}_q) = Var(\hat{Y}_q) \quad ; \quad d(\hat{Y}_q) = 0$$

$$Var(\hat{Y}_q) = E.Q.M.(\hat{Y}_q) = E(\hat{Y}_q - Y)^2$$

dato che

$$\hat{Y}_q = Y + Y(\delta_y - \delta_x)$$

si ha:

$$Var(\hat{Y}_q) = Y^2 E(\delta_y - \delta_x)^2 =$$

$$Var(\hat{Y}_q) = Y^2 E\left(\frac{\hat{Y} - Y}{Y} - \frac{\hat{X} - X}{X}\right)^2 =$$

$$Var(\hat{Y}_q) = Y^2 E\left(\frac{X\hat{Y} - XY - \hat{X}Y + XY}{XY}\right)^2 =$$

$$= \frac{Y^2}{X^2 Y^2} E\left(X^2\hat{Y}^2 + \hat{X}^2 Y^2 - 2XY\hat{X}\hat{Y}\right) =$$

$$= \frac{1}{X^2}\left[E(X^2\hat{Y}^2) + E(\hat{X}^2 Y^2) - E(2XY\hat{X}\hat{Y})\right] =$$

$$= \frac{1}{X^2} X^2 E(\hat{Y}^2) + \frac{1}{X^2} Y^2 E(\hat{X}^2) - \frac{1}{X^2} 2XYE(\hat{X}\hat{Y}) =$$

$$= E(\hat{Y}^2) + \frac{Y^2}{X^2} E(\hat{X}^2) - \frac{2Y}{X} E(\hat{X}\hat{Y}) =$$

$$= E(\underbrace{\hat{Y} - R \cdot \hat{X}}\,)^2$$

per cui

$$= E(\hat{Z})^2.$$

Che rappresenta anche la varianza dello stimatore in quanto:

$$Var(\hat{Z}) = E(\hat{Z} - \underbrace{E\left[\hat{Z}\right]}_{0})^2$$

Infatti:

$$E(\hat{Z}) = E(\hat{Y} - R \cdot \hat{X}) = Y - RX = 0$$

Pertanto

$$Var(\hat{Y}_q) = Var(\hat{Z})$$

II fase: efficienza dello stimatore

Abbiamo dimostrato che:

$$Var(\hat{Y}_q) = Var(\hat{Z})$$

ossia:

$$Var(\hat{Y}_q) = Var(\hat{Y} - R \cdot \hat{X})$$

$$= Var(\hat{Y}) + R^2 \cdot Var(\hat{X}) - 2RCov(\hat{X}, \hat{Y})$$

Condizione di maggiore efficienza dello stimatore per quoziente rispetto a quello corretto:

$$Var(\hat{Y}_q) < Var(\hat{Y})$$

Sostituendo:

$$= Var(\hat{Y}) + R^2 \cdot Var(\hat{X}) - 2RCov(\hat{X}, \hat{Y}) < Var(\hat{Y});$$

$$R^2 \cdot Var(\hat{X}) - 2RCov(\hat{X}, \hat{Y}) < 0$$

$$Cov(\hat{X},\hat{Y}) > \frac{R \cdot Var(\hat{X})}{2}$$

$$\rho(\hat{X},\hat{Y})\sqrt{Var(\hat{X})Var(\hat{Y})} > \frac{1}{2}\frac{Y}{X} \cdot Var(\hat{X})$$

da cui:

$$\rho(\hat{X},\hat{Y}) > \frac{CV(\hat{X})}{2CV(\hat{Y})}$$

c.v.d.!!!

L'uso dello stimatore per quoziente è conveniente se:

- *la correlazione tra gli stimatori è **positiva**;*
- *la variabilità dello stimatore X non è sostanzialmente superiore a quella di Y.*

ESERCIZI

Esercizio 1

Allo scopo di definire un programma di prevenzione dei furti nei supermercati, la direzione del gruppo AUCHAN ha deciso di testare un meccanismo di controllo in due supermercati sui 6 presenti in Emilia Romagna. Il numero di furti annui per ciascun supermercato è di seguito riportato

Negozi	N. Furti annui	Importo del danno (in migliaia di €)
1	12	70
2	15	81
3	11	65
4	7	30
5	10	44
6	11	62

a) Definire lo spazio campionario per un campione casuale SSR per campioni non ordinati di ampiezza 2.
b) Estrarre con il metodo di Yates un campione casuale SSR di ampiezza n=2;
c) Calcolare le probabilità di inclusione del primo e del secondo ordine per le unità estratte;
d) Stimare il valore economico totale del danno arrecato annualmente al gruppo AUCHAN dai furti nel supermercato

Esercizio 2

L'ufficio commerciale di una grande azienda con 7 filiali vuole conoscere quante commesse ricevute annualmente vengono evase con più di due settimane di ritardo. A tal fine viene deciso di effettuare un'indagine campionaria per iniziare a valutare i fatturati delle singole filiali; a tal proposito si estrae, con probabilità variabili, un campione di 2 filiali così distribuite:

Filiali	N° addetti	Fatturato (mln di €)
1	18	2,1
2	15	2,2
3	21	5,1
4	35	6,8
5	30	6,5
6	36	7,1
7	26	4,3

a) Estrarre con il metodo di Yates un campione casuale SSR di ampiezza n=2;
b) Calcolare le probabilità di inclusione del primo ordine per le unità estratte;
c) Stimare il totale del fatturato utilizzando lo stimatore corretto per il campionamento casuale SSR e successivamente costruire lo stimatore corretto ipotizzando un campionamento casuale SCR;

Esercizio 3

Durante la campagna promozionale di un nuovo prodotto per la casa, viene data una confezione in omaggio a un campione di 100 casalinghe. Dopo 15 giorni, 57 donne risultano favorevoli all'acquisto del prodotto. Si vuole definire, al livello di confidenza del 95%, un intervallo di fiducia per P, sapendo che la dimensione della popolazione è N=3.350. Quali sarebbero stati gli estremi dell'intervallo se il campione fosse stato di 200 donne fermo restando la percentuale di donne favorevoli?

Esercizio 4

Per conoscere il valore del consumo annuo pro-capite di generi alimentari in un comune di 5.785 residenti è stato scelto un campione SSR di n=20 famiglie tra le N=1.866 famiglie ottenendo i seguenti risultati:

Famiglie	Consumi totali	Numero componenti
1	16,80	2
2	30,64	4
3	39,05	5
4	20,07	3
5	30,88	4
6	21,51	3
7	19,29	3
8	30,68	4
9	39,85	5
10	29,32	4
11	53,54	6
12	14,00	2
13	22,29	3
14	36,28	4
15	10,96	2
16	21,45	3
17	21,78	3
18	5,68	1
19	29,52	4
20	5,62	1
Tot.	499,01	66

Calcolare il consumo totale sia con lo stimatore per quoziente sia con lo stimatore corretto e verificare quale dei due risulta più efficiente.
[si consideri che $s_x^2=1,69$; $s_y^2=143,14$; $s_{xy}=15,32$]

Esercizio 5

L'impresa Y, costituita da N=150 dipendenti, vuole stimare il tempo totale giornaliero che i suoi dipendenti dei diversi reparti amministrativi dedicano alle telefonate extra-lavorative. Con un piano di campionamento SCR si estraggono 3 reparti i cui tempi (espressi in minuti) sono:

$$y_1=28.1,\ y_2=25.3,\ y_3=31.6$$

Sapendo che : $N_1=13$, $N_2=10$, $N_3=15$, definire lo spazio campionario e determinare il tempo totale e la varianza.

Esercizio 6

Ad una popolazione di quattro unità sono associati i valori:

$$Y_1=50,\ Y_2=12,\ Y_3=84,\ Y_4=42$$

con probabilità di estrazione iniziali:

$$p_1=0{,}21,\ p_2=0{,}25,\ p_3=0{,}37,\ p_4=0{,}17$$

a) Si definisca lo spazio dei campioni SCR di ampiezza n=2;
b) Si scelga un campione e si stimi il totale della variabile Y e la relativa varianza.

Esercizio 7

Un proprietario agricolo vuole prevedere la produzione dell'anno corrente di un terreno di 1000 ettari. Per i primi 10 appezzamenti (scelti a probabilità costanti) ha già ottenuto la produzione nell'anno in corso (Y) misurata in numero di piante, e ha inoltre conservato i dati relativi a tali appezzamenti nell'anno passato (X).

Appezzamenti	1	2	3	4	5	6	7	8	9	10
Anno in corso	25	15	22	24	13	18	35	30	10	29
Anno precedente	23	14	20	25	12	18	30	27	8	31

Sapendo che la produzione totale dell'anno passato è stata di 21000 piante:

1. stimare il valore della produzione totale dell'anno in corso utilizzando lo stimatore corretto e lo stimatore per quoziente;
2. sulla base dei risultati ottenuti si dica quale dei due stimatori è preferibile.

Esercizio 8

Un'azienda commerciale possiede 15 punti vendita la cui superficie di vendita è in metri quadri in media pari a 65 con variabilità pari a 680. Al fine di stimare il fatturato totale vengono estratti casualmente senza ripetizione 4 punti vendita con le seguenti caratteristiche.

PV	Sup (mq)	Fatturato (mln di €)
A	80	1,2
B	70	0,92
C	50	0,5
D	110	1,8

Stimare con due differenti strategie il fatturato totale dell'azienda. Confrontare i risultati ottenuti e commentare i risultati.