Il campionamento casuale semplice

con ripetizione (ccs)

senza ripetizione (rcs)

CCS - RCS

- Piano di campionamento basilare
  criterio <u>più intuitivo</u> e <u>base</u> per formare un campione
- Selezione delle unità
  - tavole dei numeri casuali
  - opportuni algoritmi computazionali

- **Definizione**
  Si chiama *Campionamento Casuale Semplice senza ripetizione* il campionamento tale che:
  - ciascuna possibile combinazione di n elementi degli N elementi della popolazione abbia la stessa probabilità di essere scelta per il campione (sottoinsieme)

Si ricorda che il numero delle possibili combinazioni (sottoinsiemi) è

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

CCS - RCS

- Il piano di campionamento casuale semplice è quello che assegna una probabilità di selezione di ogni combinazione (sottoinsieme) pari a:

$$P = \frac{1}{\binom{N}{n}}$$

Dimostrazione:

Sia s un campione di dimensione n. *La probabilità che sia estratta l'unità i con j =1,2,…,n* che indica le estrazioni è:

CCS - RCS

$$p\left(i_j \rightarrow \text{prima estrazione}\right) = \frac{n}{N} \; \forall \text{j}$$

$$p\left(i_j \rightarrow \text{seconda estrazione}\right) = \frac{n-1}{N-1} \; \forall \text{j}$$

e così via …. quindi $$P = \frac{n}{N}\frac{n-1}{N-1}\frac{n-2}{N-2}\ldots\ldots\frac{1}{N-n+1} = \frac{1}{\binom{N}{n}}$$

Eventi mutuamente esclusivi
Probabilità composta

P è la probabilità che:

- una combinazione (sottoinsieme) sia il campione estratto
- identifica il piano di campionamento ccs

CCS - RCS

Osservazioni:

Il piano di campionamento casuale semplice senza ripetizione presenta la caratteristica che ogni unità ha la stessa probabilità di selezione pari a

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

Probabilità che una unità ha di far parte del campione

Dimostrazione

la probabilità di scegliere l'unità i alla prima estrazione è $\frac{1}{N}$

la probabilità di scegliere l'unità i alla seconda estrazione?

CCS - RCS

Dimostrazione (segue)

la probabilità di scegliere l'unità i alla **seconda** estrazione è

sia A l'evento $\longrightarrow$ selezione dell'unità alla seconda estrazione
sia B l'evento $\longrightarrow$ non selezione dell'unità alla prima estrazione

per il teorema di Bayes:

$$P(B) = \frac{N-1}{N} \qquad P(A/B) = \frac{1}{N-1}$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \rightarrow P(A \cap B) = P(B)P(A|B) = \frac{N-1}{N}\frac{1}{N-1} = \frac{1}{N}$$

CCS - RCS

Dimostrazione (segue)

teniamo conto anche della terza estrazione
sia A l'evento $\longrightarrow$ selezione dell'unità alla terza estrazione
sia B l'evento $\longrightarrow$ non selezione dell'unità né alla prima né alla seconda estrazione

$$P(A|B)=\frac{1}{N-2} \qquad P(B)=\frac{N-1}{N}\frac{N-2}{N-1}$$

$$P(A\cap B)=\underbrace{\frac{1}{N-2}}_{P(A|B)}\;\underbrace{\frac{N-1}{N}\,\frac{N-2}{N-1}}_{P(B)}=\frac{1}{N}$$

CCS - RCS

Dimostrazione (segue)

ripetuto n volte

La probabilità totale ovvero la probabilità di selezione di una unità i a qualunque estrazione è:

$$\frac{1}{N}+\frac{N-1}{N}\frac{1}{N-1}+\frac{N-1}{N}\frac{N-2}{N-1}\frac{1}{N-2}+....=\underbrace{\frac{1}{N}+\frac{1}{N}+\frac{1}{N}+....\frac{1}{N}}_{n\ volte}=\frac{n}{N}$$

cvd.

CCS - RCS

Probabilità di inclusione

Quanti sono i possibili campioni che contengono l'unità i ?

Si costruiscono tutti i campioni di n-1 unità che non contengono l'unità i

$$\binom{N-1}{n-1}$$

Si inserisce in ogni campione l'unità i in modo da ottenere campioni di dimensione n. Tutti i possibili campioni rimangono:

$$\binom{N-1}{n-1}$$

A: La probabilità di inclusione del primo ordine o *frazione di campionamento* o tasso di sondaggio è:

$$\pi_i = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} = \frac{(N-1)!}{(n-1)!(N-n)!} \times \frac{n!(N-n)!}{N!} = \frac{(N-1)! \times n \times (n-1)!}{(n-1)! \times N \times (N-1)!} = \frac{n}{N}$$

CCS - RCS

B: La probabilità di inclusione del secondo ordine è pari a:

$$\pi_{ij} = \frac{\binom{N-2}{n-2}}{\binom{N}{n}} = \frac{n(n-1)}{N(N-1)}$$

probabilità di inclusione di ordine n

=

probabilità dell'intero campione

Caratteristica solo del ccs-rcs

CCS - RCS

Verifica empirica che ….

campioni di n unità che contengono l'unità i sono $\binom{N-1}{n-1}$

Esempio: N=4 n=3 P=(1,2,3,4)

Tutti i possibili campioni sono 4 e cioè: $\binom{N}{n} = \frac{4!}{3!1!} = 4$

1,2,3 – 1,2,4 – 1,3,4 – 2,3,4

Supponiamo che l'elemento i sia 1. Tutti i possibili campioni di dimensione n-1 che non contengono l'1 sono: $\binom{N-1}{n-1} = \frac{3!}{2!1!} = 3$

E cioè: **2,3 – 2,4 – 3,4**

Se aggiungo l'elemento i (il valore 1) i campioni di dimensione n rimangono sempre 3 e cioè:

1,2,3 – 1,2,4 – 1,3,4

CCS – con ripetizione e ordinato

Due campioni sono considerati distinti anche se formati dalle stesse unità ma con ordine di selezione differente

Definizione
Si chiama *Campionamento Casuale Semplice con ripetizione* il piano di campionamento tale che:

ciascuno degli N^n elementi (campioni/sequenze) possibili ordinati con ripetizione abbia la stessa probabilità di essere estratta

La probabilità di ogni singolo campione (sequenza) è

$$P(s) = \frac{1}{N^n}$$

CCS – con ripetizione e ordinato

Dimostrazione

Data la sequenza

$$s = \{i_1, i_2, \ldots, i_n\}$$

La sua probabilità – in virtù dell'indipendenza delle estrazioni – è data dal prodotto delle probabilità $\frac{1}{N}$ di selezionare l'unità i_j nella j-esima estrazione

$$p(i_j) = \frac{1}{N}$$

Poiché ciascuna i_j è tale che

$$p(s) = \underbrace{\frac{1}{N} \times \frac{1}{N} \ldots\ldots \frac{1}{N}}_{n \text{ volte}} = \frac{1}{N^n}$$

CCS – con ripetizione e ordinato

Probabilità di inclusione

Probabilità di inclusione del primo ordine dell'unità i

$$\pi_i = P(i_j = i) = P(i \in S) = P[(i_1 = i) \cup (i_2 = i) \cup \ldots \cup (i_n = i)] =$$

$$1 - P[(i_1 \neq i) \cap (i_2 \neq i) \cap \ldots \cap (i_n \neq i)] = \text{ eventi indipendenti}$$

$$P(i_j \neq i) = 1 - P(i_j = i) \rightarrow \boxed{\textit{n volte}} \longrightarrow P(i_j \neq i) = (1 - P_i)^n$$

Pi probabilità che da N oggetti venga estratta la i-esima unità

$$\pi_i = 1 - (1 - P_i)^n = 1 - \left(1 - \frac{1}{N}\right)^n$$

Probabilità che una qualsiasi unità ha di entrare a far parte del campione.
Tale probabilità è uguale per ogni unità in questo tipo di campionamento

CCS – con ripetizione e ordinato

Probabilità di inclusione del secondo ordine

$$\pi_{ij} = P[(i \in S) \cap (j \in S)] = 1 - P[(i \notin S) \cup (j \notin S)] =$$
$$= 1 - P(i \notin S) - P(j \notin S) + P[(i \notin S) \cap (j \notin S)]$$

$$P(i \notin S) = P[(i_1 \notin S) \cap (i_2 \notin S) \cap \ldots \cap (i_n \notin S)] = (1 - P_i)^n$$

$$\pi_{ij} = 1 - (1 - P_i)^n - (1 - P_j)^n + (1 - P_i - P_j)^n =$$

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

CCS – con ripetizione e ordinato

Schema bernoulliano – estrazioni indipendenti

Frequenza attesa di inclusione

$\pi_{i_j} \rightarrow$ *una variabile casuale i.i.d.*

- *1 se l'unità i appare nell'estrazione j*
- *0 altrimenti*

$$\gamma_i = \sum_{J=1}^{n} \pi_{i_j}$$

Numero di volte in cui *l'unità* i si presenta nel campione (***Binomiale***)

$$0 \le \gamma_i \le n$$

$$\phi_i = E(\gamma_i) = \sum_{j=1}^{n} E(\pi_{i_j}) \longrightarrow$$

Numero medio di volte in cui l'unità i si presenta nel campione

$$E(\pi_{i_j}) = 1 \times P(\pi_{i_j} = 1) + 0 \times P(\pi_{i_j} = 0) = 1 \times \frac{1}{N} + 0 \times \frac{1}{N} = \frac{1}{N}$$

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

CCS – senza ripetizione

Media

$$\bar{y}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i \longrightarrow$$ **Media campionaria**

Teorema: $\bar{y}_0$ è uno stimatore non distorto di $\bar{Y}$

$$E(\bar{y}_0) = \frac{\sum_{i=1}^{\binom{N}{n}} (y_1 + y_2 + \ldots + y_n)_i}{n\binom{N}{n}}$$

Il valore atteso di $\bar{y}_0$ è la media di tutte le medie dei possibili campioni

Quanti campioni contengono la generica unità i ? $\binom{N-1}{n-1} = \frac{(N-1)!}{(n-1)!(N-n)!}$

CCS – senza ripetizione

$$E(\bar{y}_0) = \frac{\sum_{i=1}^{\binom{N}{n}} (y_1 + y_2 + \ldots + y_n)_i}{n\binom{N}{n}} = \frac{\frac{(N-1)!}{(n-1)!(N-n)!}(Y_1 + Y_2 + \ldots + Y_N)}{n\binom{N}{n}}$$

$$= \frac{(N-1)!}{(n-1)!(N-n)!} \frac{1}{n\binom{N}{n}} (Y_1 + Y_2 + \ldots + Y_N) =$$

$$= \frac{(N-1)!}{(n-1)!(N-n)!} \frac{1}{n\frac{N!}{n!(N-n)!}} (Y_1 + Y_2 + \ldots + Y_N) =$$

CCS – senza ripetizione

$$= \frac{(N-1)!}{(n-1)!(N-n)!} \frac{1}{n\dfrac{N!}{n!(N-n)!}} (Y_1 + Y_2 + \ldots + Y_N) =$$

$$= \frac{(N-1)!}{(n-1)!(N-n)!} \frac{1}{n\dfrac{N(N-1)!}{n(n-1)!(N-n)!}} (Y_1 + Y_2 + \ldots + Y_N) =$$

$$= \frac{1}{1} \frac{1}{N} (Y_1 + Y_2 + \ldots + Y_N) = \bar{Y}$$

cvd

CCS – senza ripetizione

Osservazione:

$$\sum_{i=1}^{\binom{N}{n}} \left(y_1 + y_2 + \ldots + y_n\right)_i \text{ è più grande di } \left(Y_1 + Y_2 + \ldots + Y_N\right)$$

sottoinsiemi

quanti sono i possibili sottoinsiemi ? $\binom{N-1}{n-1}$

somma dei sottoinsiemi $\binom{N-1}{n-1} Y_1 + Y_2 + \ldots + Y_N$

CCS – senza ripetizione

Ancora:

$$\sum_{i=1}^{\binom{N}{n}} \left(y_1 + y_2 + \ldots + y_n\right)_i = \frac{(N-1)!}{(n-1)!(N-n)!}\left(Y_1 + Y_2 + \ldots + Y_N\right)$$

posso considerare una variabile casuale dicotoma

(1 se l'elemento è del sottoinsieme)

(0 se l'elemento è della popolazione)

$1y_1 + 1y_2 + \ldots + 1y_n + 0y_{n+1} + 0y_{n+2} + \ldots + 0y_N$ - primo sottoinsieme

$1y_1' + 1y_2' + \ldots + 1y_n' + 0Y'_{n+1} + 0Y'_{n+2} + \ldots + 0Y'_N$ - secondo sottoinsieme

$\binom{N-1}{n-1} \longrightarrow$ *Sequenze così fatte*

sommando tutte le sequenze si avrà: $\dfrac{(N-1)!}{(n-1)!(N-n)!}\left(Y_1 + Y_2 + \ldots + Y_N\right)$

CCS – senza ripetizione

Varianza di y_i in una popolazione finita:

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}{N} \longrightarrow S^2 = \frac{\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}{N-1} \longrightarrow s^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n-1}$$

La varianza della media campionaria?

$$\sigma^2_{\bar{y}_0} = \mathrm{var}\left(\bar{y}_0\right) = V\left(\bar{y}\right) = E\left(\bar{y}_0 - \bar{Y}\right)^2 = \frac{S^2}{n}\frac{N-n}{N} = \frac{S^2}{n}\left(1-f\right)$$

$f = \frac{n}{N}$ è la frazione di campionamento

Dimostrazione:
poiché ogni unità appare nello stesso numero di campioni

$E(y_1+y_2+\ldots+y_n)$ sarà multiplo di $Y_1+Y_2+\ldots+Y_N$

$E(y_1+y_2+\ldots+y_n)=k(Y_1+Y_2+\ldots+Y_N)$ $\qquad$ $k=n/N$ $\qquad$ [1]

… relazioni usate

$$E\left[(y_1)+\ldots+(y_n)\right]=\frac{n}{N}\left[(Y_1)+\ldots+(Y_N)\right]$$

se si pone: $y_1^{'}=\left(y_1-\overline{Y}\right)^2$

$$E\left[y_1^{'}+\ldots+\left(y_n^{'}\right)\right]=\frac{n}{N}\left[\left(Y_1^{'}\right)+\ldots+\left(Y_N^{'}\right)\right]$$

$$E\left[\left(y_1-\overline{Y}\right)^2+\ldots+\left(y_n-\overline{Y}\right)^2\right]=\frac{n}{N}\left[\left(Y_1-\overline{Y}\right)^2+\ldots+\left(Y_N-\overline{Y}\right)^2\right]$$

se si pone: $y_1^{'}=\left(y_1-\overline{Y}\right)\left(y_2-\overline{Y}\right)\ldots$

$$E\left[\left(y_1-\overline{Y}\right)\left(y_2-\overline{Y}\right)+\left(y_1-\overline{Y}\right)\left(y_3-\overline{Y}\right)+\ldots+\left(y_{n-1}-\overline{Y}\right)\left(y_n-\overline{Y}\right)\right]=$$

$$=\frac{n(n-1)}{N(N-1)}\left[\left(Y_1-\overline{Y}\right)\left(Y_2-\overline{Y}\right)+\left(Y_1-\overline{Y}\right)\left(Y_3-\overline{Y}\right)+\ldots+\left(Y_{N-1}-\overline{Y}\right)\left(Y_N-\overline{Y}\right)\right]$$

… relazioni usate

$$E\left[\left(y_1-\bar{Y}\right)\left(y_2-\bar{Y}\right)+\left(y_1-\bar{Y}\right)\left(y_3-\bar{Y}\right)+\ldots+\left(y_{n-1}-\bar{Y}\right)\left(y_n-\bar{Y}\right)\right]=$$

$$=\frac{n(n-1)}{N(N-1)}\left[\left(Y_1-\bar{Y}\right)\left(Y_2-\bar{Y}\right)+\left(Y_1-\bar{Y}\right)\left(Y_3-\bar{Y}\right)+\ldots+\left(Y_{N-1}-\bar{Y}\right)\left(Y_N-\bar{Y}\right)\right]$$

la somma dei prodotti è estesa a tutte le coppie di unità y_i e y_j nel campione e nella popolazione (Y_iY_j), che ammontano a

$\frac{n(n-1)}{2} \longrightarrow$ Nel campione
(non considero il prodotto y_i per y_i)

$\frac{N(N-1)}{2} \longrightarrow$ Nella popolazione
(non considero il prodotto Y_i per Y_i)

CCS – senza ripetizione

$$n\left(\bar{y}_0-\bar{Y}\right)=n\left[\frac{\left(y_1+y_2+\ldots+y_n\right)}{n}\right]-n\bar{Y}=\left(y_1-\bar{Y}\right)+\left(y_2-\bar{Y}\right)+\ldots+\left(y_n-\bar{Y}\right) \qquad [2]$$

Considero ora la quantità

$$E\left[\left(y_1-\bar{Y}\right)^2+\ldots+\left(y_n-\bar{Y}\right)^2\right]=\frac{n}{N}\left[\left(Y_1-\bar{Y}\right)^2+\ldots+\left(Y_N-\bar{Y}\right)^2\right] \qquad [3]$$

inoltre

$$E\left[\left(y_1-\bar{Y}\right)\left(y_2-\bar{Y}\right)+\left(y_1-\bar{Y}\right)\left(y_3-\bar{Y}\right)+\ldots+\left(y_{n-1}-\bar{Y}\right)\left(y_n-\bar{Y}\right)\right]=$$
$$=\frac{n(n-1)}{N(N-1)}\left[\left(Y_1-\bar{Y}\right)\left(Y_2-\bar{Y}\right)+\left(Y_1-\bar{Y}\right)\left(Y_3-\bar{Y}\right)+\ldots+\left(Y_{N-1}-\bar{Y}\right)\left(Y_N-\bar{Y}\right)\right] \qquad [4]$$

eleviamo al quadrato la [2]

$$n^2\left(\bar{y}_0-\bar{Y}\right)^2=\left[\left(y_1-\bar{Y}\right)+\left(y_2-\bar{Y}\right)+\ldots+\left(y_n-\bar{Y}\right)\right]^2$$

CCS – senza ripetizione

usando la [3] e la [4] e facendone il valore atteso

$$\boxed{n^2 E\left(\bar{y}_0 - \bar{Y}\right)^2} = \frac{n}{N}\left[\left(y_1 - \bar{Y}\right) + \left(y_2 - \bar{Y}\right) + \ldots + \left(y_N - \bar{Y}\right)\right]^2 =$$

$$= \frac{n}{N}\left\{\left(y_1 - \bar{Y}\right)^2 + \ldots + \left(y_N - \bar{Y}\right)^2 + 2\frac{(n-1)}{N-1}\left[\left(y_1 - \bar{Y}\right)\left(y_2 - \bar{Y}\right) + \ldots + \left(y_{N-1} - \bar{Y}\right)\left(y_N - \bar{Y}\right)\right]\right\} =$$

$$= \frac{n}{N}\left\{\left(y_1 - \bar{Y}\right)^2 + \ldots + \left(y_N - \bar{Y}\right)^2 \underbrace{- \frac{n-1}{N-1}\left(y_1 - \bar{Y}\right)^2 - \ldots - \frac{n-1}{N-1}\left(y_N - \bar{Y}\right)^2}_{} + \right.$$

(aggiungendo e sottraendo la stessa quantità)

$$\overbrace{+ \frac{n-1}{N-1}\left(y_1 - \bar{Y}\right)^2 + \ldots + \frac{n-1}{N-1}\left(y_N - \bar{Y}\right)^2}^{} +$$

$$\left. + 2\frac{n-1}{N-1}\left(y_1 - \bar{Y}\right)\left(y_2 - \bar{Y}\right) + \ldots + 2\frac{n-1}{N-1}\left(y_{N-1} - \bar{Y}\right)\left(y_N - \bar{Y}\right)\right\} =$$

CCS – senza ripetizione

$$= \frac{n}{N}\left(1-\frac{n-1}{N-1}\right)\left[\left(y_1-\bar{Y}\right)^2+\ldots+\left(y_N-\bar{Y}\right)^2\right]+\frac{n-1}{N-1}\Big[\underbrace{\left(y_1-\bar{Y}\right)+\ldots+\left(y_N-\bar{Y}\right)}_{\text{= 0 per la prima proprietà della media}}\Big]^2 =$$

$$= \frac{n}{N}\left(\frac{N-1-n+1}{N-1}\right)\left[\left(y_1-\bar{Y}\right)^2+\ldots+\left(y_N-\bar{Y}\right)^2\right]=\frac{n}{N}(N-n)\underbrace{\frac{\left(y_1-\bar{Y}\right)^2+\ldots+\left(y_N-\bar{Y}\right)^2}{N-1}}_{= S^2} =$$

$$= \frac{n}{N}(N-n)S^2 \longrightarrow E\left(\bar{y}_0-\bar{Y}\right)^2=\frac{n(N-n)}{Nn^2}S^2=\frac{S^2}{n}\frac{N-n}{N} \qquad \text{cvd}$$

CCS

Verifica empirica che:

$$E\left[\left(y_1+y_2+y_3+\ldots+y_n\right)\right]=\frac{n}{N}\left[\left(Y_1+Y_2+Y_3+\ldots.+Y_N\right)\right]$$

$$E\left[\left(y_i+y_j+y_k\right)\right]=\frac{\left(y_1+y_2+y_3\right)+\left(y_1+y_2+y_4\right)+\left(y_1+y_3+y_4\right)+\left(y_2+y_3+y_4\right)}{4}$$

$$=\frac{\left(3y_1+3y_2+3y_3+3y_4\right)}{4}=\frac{3}{4}\left(y_1+y_2+y_3+y_4\right)$$

CCS – senza ripetizione

Se non conosciamo S^2 della popolazione, occorre stimarlo:

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y}_0)^2}{n-1}$$

s^2 è uno stimatore non distorto di S^2

$$E\left(s^2\right) = S^2 = \frac{\sum_{i=1}^{N} \left(Y_i - \bar{Y}\right)^2}{N-1}$$

Sostituendo la stima di S^2 nella formula della varianza della media, otteniamo la stima della varianza della media pari a:

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

la stima della varianza della media è leggermente distorta

CCS – senza ripetizione

Errore standard della stima

$$\sigma_{\bar{y}_0} = \frac{S}{\sqrt{n}}\sqrt{\frac{N-n}{N}} = \frac{S}{\sqrt{n}}\sqrt{(1-f)}$$

Stima dell'Errore standard della stima

$$\hat{s}_{\bar{y}_0} = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N}} = \frac{s}{\sqrt{n}}\sqrt{(1-f)}$$

1-f: fattore di correzione per popolazioni finite
può essere ignorato se

$$\frac{n}{N} < 0{,}05$$

CCS – senza ripetizione

Totale

$$\hat{y} = \frac{N}{n}\sum_{i=1}^{n} y_i = N\bar{y} \longrightarrow$$ *Totale campionario*

anche chiamato stimatore per espansione

$$\mathrm{var}(\hat{y}) = E(\hat{y}-Y)^2 = \mathrm{var}(N\bar{y}) = N^2\,\mathrm{var}(\bar{y}) = \frac{N^2S^2}{n}\frac{N-n}{N} = \frac{N^2S^2}{n}(1-f)$$

var(kx)=k2var(x)

N.B.
Come per la media anche per il totale, se si usa s^2 al posto di S^2, le stime saranno leggermente distorte:

$$\hat{v}(\bar{y}_0) = s^2_{\bar{y}_0} = \frac{s^2}{n}\frac{N-n}{N}$$

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

CCS – senza ripetizione

La varianza del totale nel campionamento casuale semplice, **nel caso di estrazioni con ripetizione**, è pari a

$$\mathrm{var}(\hat{y}) = E(\hat{y} - Y)^2 = \frac{N^2 S^2}{n} = \frac{N^2 S^2}{n}$$

CCS – senza ripetizione

Formula più semplice per s^2

$$s^2 = \frac{1}{n-1}\sum\left(y_i - \bar{y}\right)^2 = \frac{1}{n-1}\left[\sum y_i^2 + n\bar{y}^2 - 2\bar{y}\sum y_i\right] =$$

$$= \frac{1}{n-1}\left[\sum y_i^2 + n\frac{\sum y_i^2}{n^2} - 2\frac{\sum y_i}{n}\sum y_i\right] = \frac{1}{n-1}\left[\sum y_i^2 - \frac{\left(\sum y_i\right)^2}{n}\right]$$

CCS – senza ripetizione

Supponiamo di avere una popolazione che può essere suddivisa in due categorie: C e $\overline{C}$

Indichiamo con A il numero di soggetti appartenenti a C
Nella popolazione avrò:

$$P = \frac{A}{N}$$

Proporzione

Nel campione avrò:

$$p = \frac{a}{n}$$

Proporzione campionaria

se $y_i \begin{cases} 1 \text{ se } i \in C \\ 0 \text{ se } i \notin C \end{cases}$

$$\overline{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{A}{N} = P$$

$$\overline{y}_o = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{a}{n} = p$$

$$\sum_{i=1}^{N} Y_i^2 = A; \sum_{i=1}^{n} y_i^2 = a$$

$$A = NP; a = np$$

CCS – senza ripetizione

S^2 e s^2 in funzione di P e p

$$S^2 = \frac{1}{N-1}\sum\left(y_i - \overline{Y}\right)^2 = \frac{1}{N-1}\left[\sum y_i^2 + N\overline{Y}^2 - 2\overline{Y}\sum y_i\right] = \frac{1}{N-1}\left[\sum y_i^2 + N\overline{Y}^2 - 2\frac{\sum y_i}{N}\frac{\sum y_i}{N}N\right] =$$

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

1-P = Q

$$s^2 = \frac{1}{n-1}\sum\left(y_i - \overline{y_o}\right)^2 = \frac{1}{n-1}\left[\sum y_i^2 - n\overline{y}_0^2\right] = \frac{n}{n-1}\times pq$$

- quando la popolazione è dicotoma … la distribuzione di campionamento è una Binomiale
- possiamo stimare la proporzione di successi perché ci riferiamo a variabili non metriche
- concettualmente (p o y_0 o il totale) non sono gli stessi stimatori

$\overline{y}_0$ o il totale riguardano variabili metriche

p o A riguardano variabili categoriali/categorializzabili

CCS – senza ripetizione

$$p = \frac{a}{n}$$

Stimatore corretto di P E(p)=P

la varianza di p v(p) è:

$$v(p) = \frac{N}{(N-1)}\frac{PQ}{n}\left(\frac{N-n}{N}\right) = \left(\frac{N-n}{N-1}\right)\frac{PQ}{n}$$

$$v(\bar{y}_0) = \frac{S^2(N-n)}{nN}$$

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

CCS – senza ripetizione

$$p = \frac{a}{n}$$ Stimatore corretto di P E(p)=P

una stima non distorta di v(p) sui dati campionari è:

$$\hat{v}(p) = \frac{N-n}{N(n-1)} pq$$

Dimostrazione:

Consideriamo la stima della varianza della media

$$\hat{v}(\bar{y}_0) = \frac{s^2(N-n)}{nN} \text{ come stima non distorta di } v(\bar{y}_0)$$

se

$$s^2 = \frac{n}{n-1} pq$$

CCS – senza ripetizione

avremo che:

$$\hat{v}(p) = s_p^2 = \frac{s^2}{n}\frac{N-n}{N} = \frac{\mathrm{n}}{\mathrm{n}\text{ - }1}\frac{pq}{n}\frac{N-n}{N} = \frac{N-n}{N(n-1)}pq$$

se N è molto grande rispetto a n cioè
n/N abbastanza piccolo

$$\boxed{\hat{v}(p) \approx s_p^2 = \frac{pq}{(n-1)}}$$

CCS – senza ripetizione

A=NP $\hat{A} = Np$

$$\hat{y} = N\overline{y}_0$$

$$V\left(\hat{A}\right) = \frac{N^2 PQ}{n}\frac{N-n}{N-1} = N^2 V(P) \longrightarrow$$

Se è nota la varianza della popolazione

$$v\left(\hat{A}\right) = \frac{N\left(N-n\right)}{n-1} pq = N^2 v(p) \longrightarrow$$

Se non è nota la varianza della popolazione

Stima non distorta di V(A)

CCS – con reimmissione

La varianza della media campionaria, nel caso di **campionamento con ripetizione**, è pari a:

$$\sigma_{\bar{y}}^{2} = \mathrm{var}(\bar{y}) = V_{ccr}(\bar{y}) = E(\bar{y} - \bar{y})^{2} = \frac{S^{2}}{n}$$

Quindi la varianza della media del campionamento casuale semplice con reinserimento è maggiore di quella senza reinserimento.

$$\sigma_{\bar{y}ccs}^{2} > \sigma_{\bar{y}rcs}^{2}$$

lo stesso tipo di relazione vale anche per
- totale
- proporzione

CRS – stima della dimensione campionaria

Stima della dimensione campionaria n

a) limite massimo di errore che si vuole commettere
b) precisione della stima campionaria da ottenere
c) costi da sostenere
d) tempo a disposizione, informazioni disponibili, funzioni di varianza….

Precisione desiderata legata all'aumentare massimo di errore tollerato nelle stime campionarie

Qual è il livello di errore tollerabile ? 5% ? 1%?
Errore:

a) assoluto
b) relativo (r*apporto fra errore assoluto e valore del parametro da stimare).*

CCS – stima della dimensione campionaria

Errore assoluto d $P\left|\hat{\theta}-\theta\right| \leq d$

Errore relativo r $P\left|\frac{\hat{\theta}-\theta}{\theta}\right| \leq r = P\left\{\left|\hat{\theta}-\theta\right| \leq r\theta\right\}$

$$\hat{\theta} = \begin{cases} \bar{y} \rightarrow v(\bar{y}) = \dfrac{S^2}{n}(1-f) \\ p \rightarrow v(p) = \dfrac{N-n}{N-1}\dfrac{PQ}{n} \\ \hat{y} \rightarrow v(N\bar{y}) = N^2\dfrac{S^2}{n}(1-f) \end{cases}$$

$$\frac{\hat{\theta}-\theta}{\sqrt{v(\hat{\theta})}} \sim N(0,1)$$

l'errore theta hat è la varianza dello stimatore osservata nella popolazione

CCS – stima della dimensione campionaria

In genere viene chiesto un errore assoluto pari a d

Nella pratica si può determinare un'ampiezza campionaria tale che sia minima la probabilità di commettere l'errore di superare tale valore "d"

$$P\left|\hat{\theta}-\theta\right|\leq d$$

$$P\left\{\left|\frac{\hat{\theta}-\theta}{\sqrt{V\left(\hat{\theta}\right)}}\right|\leq\frac{d}{\sqrt{V\left(\hat{\theta}\right)}}\right\}=1-\alpha$$

$$=P\left\{\left|z_{\alpha/2}\right|\leq\frac{d}{\sqrt{V\left(\hat{\theta}\right)}}\right\}=1-\alpha$$

CCS – stima della dimensione campionaria

Lo stesso vale per r

$$P\left\{\left|\frac{\hat{\theta}-\theta}{\theta}\right| \le r\right\} = 1-\alpha$$

$$= P\left\{\left|\hat{\theta}-\theta\right| \le r\theta\right\} = 1-\alpha$$

$$= P\left\{\left|\frac{\hat{\theta}-\theta}{\mathrm{var}\left(\hat{\theta}\right)}\right| \le \frac{r\theta}{\mathrm{var}\left(\hat{\theta}\right)}\right\} = 1-\alpha$$

$$z_{\alpha/2} = \frac{r(\theta)}{\sqrt{V\left(\hat{\theta}\right)}} \Rightarrow r(\theta) = z_{\alpha/2}\sqrt{V\left(\hat{\theta}\right)}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Errore assoluto d

$$P\left\{\left|p-P\right|\leq d\right\}=1-\alpha$$

$$P\left\{\left|\frac{p-P}{\sqrt{V(p)}}\right|\leq\frac{d}{\sqrt{V(p)}}\right\}=1-\alpha$$

$$P\left\{\left|z_{\alpha/2}\right|\leq\frac{d}{\sqrt{V(p)}}\right\}=1-\alpha \qquad V(p)=\frac{PQ}{n}\frac{N-n}{N-1}$$

$$z_{\alpha/2}=\frac{d}{\sqrt{V(p)}}\Rightarrow d=z_{\alpha/2}\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{PQ}{n}}$$

$$z^2\frac{PQ}{n}\frac{N-n}{N-1}=d^2\Rightarrow z^2PQN-z^2PQn=d^2n(N-1)$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Errore assoluto d

$$d^2 n(N-1) + z^2 PQn = z^2 PQN$$

$$z^2 PQN = d^2 n(N-1) + z^2 PQn = n\left[(N-1)d^2 + z^2 PQ\right]$$

$$n = \frac{z^2 PQN}{\left[(N-1)d^2 + z^2 PQ\right]}$$

$$= \frac{z^2 PQN}{d^2 N - d^2 + z^2 PQ} = \frac{\dfrac{z^2 PQN}{d^2 N}}{\left[\dfrac{Nd^2}{Nd^2} - \dfrac{d^2}{Nd^2} + \dfrac{z^2 PQ}{Nd^2}\right]} = \frac{\dfrac{z^2 PQN}{d^2 N}}{\left[1 - \dfrac{1}{N} + \dfrac{z^2 PQ}{Nd^2}\right]} =$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Errore assoluto d

$$n = \frac{\dfrac{z^2_{\alpha/2} PQ}{d^2}}{1 + \dfrac{1}{N}\left(\dfrac{z^2_{\alpha/2} PQ}{d^2} - 1\right)} = \frac{n_0}{1 + \dfrac{1}{N}(n_0 - 1)}$$

$$\boxed{\frac{z^2 PQ}{d^2} = n_0}$$

se $N \rightarrow \infty$

$\dfrac{n_0}{N} \rightarrow 0 \quad n_0 \cong \mathrm{n}$ altrimenti $\boxed{n = \dfrac{n_0}{1 + \dfrac{(n_0 - 1)}{N}} \cong \dfrac{\mathrm{n}_0}{1 + \dfrac{\mathrm{n}_0}{\mathrm{N}}}}$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Errore relativo r

basta sostituire rP a d

$$n = \frac{\dfrac{z_{\alpha/2}^2 PQ}{r^2 P^2}}{1 + \dfrac{1}{N}\left(\dfrac{z_{\alpha/2}^2 PQ}{r^2 P^2} - 1\right)} = \frac{n_0}{1 + \dfrac{1}{N}\left(n_0 - 1\right)}$$

dove $$n_0 = \frac{z_{\alpha/2}^2 PQ}{r^2 P^2} = \frac{z_{\alpha/2}^2}{r^2}\frac{Q}{P}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Errore relativo r

basta sostituire rP a d

$$P\left\{\left|\frac{p-P}{P}\right| \le r\right\} = P\{|p-P| \le rP\} = 1-\alpha$$

$$P\left\{\left|\frac{p-P}{\sqrt{V(p)}}\right| \le \frac{rP}{\sqrt{V(p)}}\right\} = 1-\alpha \qquad P\left\{\left|z_{\alpha/2}\right| \le \frac{rP}{\sqrt{V(p)}}\right\} = 1-\alpha$$

$$z_{\alpha/2} = \frac{rP}{\sqrt{V(p)}} \Rightarrow rP = z_{\alpha/2}\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{PQ}{n}}$$

$$V(p) = \frac{PQ}{n}\,\frac{N-n}{N-1}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Errore relativo r

$$z_{\alpha/2} = \frac{rP}{\sqrt{V(p)}} \Rightarrow rP = z_{\alpha/2}\sqrt{\frac{N-n}{N-1}}\sqrt{\frac{PQ}{n}}$$

$$z^2 \frac{PQ}{n}\frac{N-n}{N-1} = r^2P^2 \Rightarrow z^2PQN - z^2PQn = r^2P^2n(N-1)$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Errore relativo r

$$r^2P^2n(N-1)+z^2PQn=z^2PQN$$

$$z^2PQN=r^2P^2n(N-1)+z^2PQn=n\left[(N-1)r^2P^2+z^2PQ\right]$$

$$n=\frac{z^2PQN}{\left[(N-1)r^2P^2+z^2PQ\right]}$$

$$=\frac{z^2PQN}{r^2P^2N-r^2P^2+z^2PQ}=\frac{\dfrac{z^2PQN}{r^2P^2N}}{\left[\dfrac{Nr^2P^2}{Nr^2P^2}-\dfrac{r^2P^2}{Nr^2P^2}+\dfrac{z^2PQ}{Nr^2P^2}\right]}=\frac{\dfrac{z^2PQN}{r^2P^2N}}{\left[1-\dfrac{1}{N}+\dfrac{z^2PQ}{Nr^2P^2}\right]}=$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Errore relativo r

$$n = \frac{\dfrac{z^2_{\alpha/2}PQ}{r^2P^2}}{1+\dfrac{1}{N}\left(\dfrac{z^2_{\alpha/2}PQ}{r^2P^2}-1\right)} = \frac{n_0}{1+\dfrac{1}{N}(n_0-1)}$$

$$\frac{z^2PQ}{r^2P^2} = n_0$$

se $N \to \infty$

$\dfrac{n_0}{N} \to 0 \quad n_0 \cong \mathrm{n}$ altrimenti $n = \dfrac{n_0}{1+\dfrac{(n_0-1)}{N}} \cong \dfrac{\mathrm{n}_0}{1+\dfrac{\mathrm{n}_0}{\mathrm{N}}}$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della media

Errore assoluto d

$$z_{\alpha/2} S_{\bar{y}_0} = d$$

$$z^2_{\alpha/2} \frac{S^2}{n} \frac{N-n}{N} = d^2 \Rightarrow z^2_{\alpha/2} S^2 (N-n) = d^2 nN$$

$$z^2_{\alpha/2} S^2 N - z^2_{\alpha/2} S^2 n = d^2 nN$$

$$z^2_{\alpha/2} S^2 N = z^2_{\alpha/2} S^2 n + d^2 nN$$

$$z^2_{\alpha/2} S^2 N = n\left(d^2 N + z^2_{\alpha/2} S^2\right)$$

$$n = \frac{z^2_{\alpha/2} S^2 N}{d^2 N + z^2_{\alpha/2} S^2}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della media

Errore assoluto d

dividiamo per d^2N

$$n = \frac{\dfrac{z_{\alpha/2}^2 S^2 N}{d^2 N}}{\dfrac{d^2 N}{d^2 N} + \dfrac{z_{\alpha/2}^2 S^2}{d^2 N}} = \frac{\dfrac{z_{\alpha/2}^2 S^2}{d^2}}{1 + \dfrac{1}{N}\left(\dfrac{z_{\alpha/2}^2 S^2}{d^2}\right)}$$

$$n_0 = \frac{z_{\alpha/2}^2 S^2}{d^2}$$

$$n = \frac{n_0}{1 + \dfrac{n_0}{N}}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della media

Errore relativo r

$$P\left\{\left|\frac{\bar{y}_0-\bar{Y}}{\bar{Y}}\right|\leq r\right\}=P\left\{\left|\bar{y}_0-\bar{Y}\right|\leq r\bar{Y}\right\}=1-\alpha$$

$$P\left\{\frac{\left|\bar{y}_0-\bar{Y}\right|}{\sigma_{\bar{y}_0}}\leq\frac{r\bar{Y}}{\sigma_{\bar{y}_0}}\right\}=P\left\{z_{\alpha/2}\leq\frac{r\bar{Y}}{\sigma_{\bar{y}_0}}\right\}=1-\alpha$$

$$r\bar{Y}=z_{\alpha/2}\sigma_{\bar{y}_0}=z_{\alpha/2}\sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della media

Errore relativo r

$$r\overline{Y} = z_{\alpha/2}\sigma_{\overline{y}_0} = z_{\alpha/2}\sqrt{\frac{N-n}{N}}\frac{S}{\sqrt{n}}$$

$$r^2\overline{Y}^2 = z^2_{\alpha/2}\sigma^2_{\overline{y}_0} = z^2{}_{\alpha/2}\frac{N-n}{N}\frac{S^2}{n}$$

$$r^2\overline{Y}^2 = \frac{z^2{}_{\alpha/2}NS^2 - z^2{}_{\alpha/2}nS^2}{Nn} \qquad r^2\overline{Y}^2 Nn = z^2{}_{\alpha/2}NS^2 - z^2{}_{\alpha/2}nS^2$$

$$n = \frac{z^2{}_{\alpha/2}NS^2}{\left(r^2\overline{Y}^2 N + z^2{}_{\alpha/2}S^2\right)}$$

$$n = \frac{\dfrac{z_{\alpha/2}S^2}{\left(r\overline{y}\right)^2}}{1 + \dfrac{1}{N}\left(\dfrac{z_{\alpha/2}S}{r\overline{y}}\right)^2}$$

$$n_0 = \frac{z^2_{\alpha/2}S^2}{\left(r\overline{y}\right)^2} \Rightarrow n = \frac{n_0}{1+\dfrac{n_0}{N}}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale

Errore assoluto d

$$P\left\{\left|N\bar{y} - N\bar{Y}\right| \leq d\right\} = 1 - \alpha$$

$$z_{\alpha/2} S\hat{y} = d$$

$$z^2_{\alpha/2} N^2 \frac{S^2}{n} \frac{N-n}{N} = d^2 \Rightarrow z^2_{\alpha/2} N^2 S^2 (N-n) = d^2 nN$$

$$z^2_{\alpha/2} N^2 S^2 N - z^2_{\alpha/2} N^2 S^2 n = d^2 nN$$

$$z^2_{\alpha/2} N^2 S^2 N = n\left(d^2 N + z^2_{\alpha/2} N^2 S^2\right)$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale

Errore assoluto d

$$z^2_{\alpha/2} N^2 S^2 N = n\left(d^2 N + z^2_{\alpha/2} N^2 S^2\right)$$

$$n = \frac{z^2_{\alpha/2} N^2 S^2 N}{d^2 N + z^2_{\alpha/2} N^2 S^2}$$

dividiamo per d^2N

$$n = \frac{\dfrac{z^2_{\alpha/2} N^2 S^2 N}{d^2 N}}{\dfrac{d^2 N}{d^2 N} + \dfrac{z^2_{\alpha/2} N^2 S^2}{d^2 N}} = \frac{\dfrac{z^2_{\alpha/2} N^2 S^2}{d^2}}{1 + \left(\dfrac{z^2_{\alpha/2} N^2 S^2}{d^2}\right)}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale

Errore assoluto d

$$n = \frac{\dfrac{z_{\alpha/2}^2 N^2 S^2}{d^2}}{1 + \left(\dfrac{z_{\alpha/2}^2 N^2 S^2}{d^2}\right)}$$

$$n_0 = \frac{z_{\alpha/2}^2 N^2 S^2}{d^2}$$

$$n_0 = \frac{n_0}{1 + n_0}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale

Errore relativo r

$$P\left\{\left|\frac{N\bar{y}-N\bar{Y}}{N\bar{Y}}\right|\leq r\right\}=\left(\left|\bar{y}-\bar{Y}\right|\leq rN\bar{Y}\right)=1-\alpha$$

$$z_{\alpha/2}S\hat{y}=rN\bar{Y}$$

$$r^2\bar{Y}^2N^2=z_{\alpha/2}^2\sigma^2\hat{y}=z^2{}_{\alpha/2}N^2\frac{N-n}{N}\frac{S^2}{n}$$

$$r^2\bar{Y}^2N^2=\frac{z^2{}_{\alpha/2}N^2NS^2-z^2{}_{\alpha/2}nN^2S^2}{Nn}=$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale

Errore relativo r

$$r^2 \overline{Y}^2 N^2 = \frac{z^2{}_{\alpha/2} N^2 N S^2 - z^2{}_{\alpha/2} n N^2 S^2}{Nn} =$$

$$r^2 \overline{Y}^2 N^2 N n = z^2{}_{\alpha/2} N^2 N S^2 - z^2{}_{\alpha/2} n N^2 S^2$$

$$r^2 \overline{Y}^2 N^2 N n + z^2{}_{\alpha/2} n N^2 S^2 = z^2{}_{\alpha/2} N^2 N S^2$$

$$n\left(r^2 \overline{Y}^2 N^2 N + z^2{}_{\alpha/2} N^2 S^2\right) = z^2{}_{\alpha/2} N^2 N S^2$$

$$n = \frac{z^2{}_{\alpha/2} N^2 N S^2}{r^2 \overline{Y}^2 N^2 N + z^2{}_{\alpha/2} N^2 S^2}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale

Errore relativo r

$$n = \frac{z^2{}_{\alpha/2} N^2 N S^2}{r^2 \overline{Y}^2 N^2 N + z^2{}_{\alpha/2} N^2 S^2}$$

dividiamo tutto per $r^2 \overline{Y}^2 N^2$

$$n = \frac{\dfrac{z^2{}_{\alpha/2} N N^2 S^2}{r^2 \overline{Y}^2 N}}{\left(\dfrac{r^2 \overline{Y}^2 N^2 N}{r^2 \overline{Y}^2 N^2} + \dfrac{z^2{}_{\alpha/2} N^2 S^2}{r^2 \overline{Y}^2 N^2} \right)}$$

$$n = \frac{n_0}{N + \dfrac{z_{\alpha/2}{}^2 S^2}{r^2 \overline{y}^2}}$$

$$n_0 = \frac{z^2{}_{\alpha/2} N \; S^2}{r^2 \overline{Y}^2} \Rightarrow n = \frac{n_0}{N(1 + n_0)}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Esempio: *Errore (assoluto) massimo tollerabile 3%*

N=48.000

$$P\{|p-P| \le d\} = 1-\alpha$$

$$P\{|p-P| \le 0{,}03\} = 1-\alpha$$

$$P\left\{\left|\frac{p-P}{\sqrt{V(p)}}\right| \le \frac{0{,}03}{\sqrt{V(p)}}\right\} = 1-\alpha$$

$$P\left\{|z_{\alpha/2}| \le \frac{0{,}03}{\sqrt{V(p)}}\right\} = 1-\alpha$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Esempio: *Errore (assoluto) massimo tollerabile 3%*
N=48.000

con α=0,05

$$1{,}96 = \frac{0{,}03}{\sqrt{V(p)}} \Rightarrow V(p) = \frac{0{,}03^2}{1{,}96^2} = 0{,}000234$$

nell'ipotesi cautelativa in cui P=Q=0,5

$$V(p) = \frac{PQ}{n}\frac{N-n}{N-1} = \frac{0{,}5 \times 0{,}5}{n}\frac{N-n}{N-1} = 0{,}000234$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima della proporzione

Esempio: *Errore (assoluto) massimo tollerabile 3%*
N=48.000

$$V(p) = \frac{0{,}5 \times 0{,}5}{n} \frac{N-n}{N-1} = 0{,}000234$$

$$= 0{,}25N - 0{,}25n = nN0{,}000234$$

$$0{,}25N = nN0{,}000234 + 0{,}25n = n\left(N0{,}000234 + 0{,}25\right)$$

$$n = \frac{0{,}25N}{N0{,}000234 + 0{,}25} = \frac{0{,}25 \times 48000}{48000 \times 0{,}000234 + 0{,}25} = 1045$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale A

Esempio: *Errore (relativo) massimo tollerabile 10%*

con α=0,05

$$P\left\{\left|\frac{Np-NP}{NP}\right|\leq r\right\}=\left(\left|p-P\right|\leq rP\right)=1-\alpha$$

$$P\left\{\left|z_{\alpha/2}\right|\leq\frac{0{,}10Np}{N\sqrt{V(p)}}\right\}=0{,}95$$

$$=0{,}10Np=1{,}96N\sqrt{V(p)}\qquad V(p)=\frac{S^2}{n}\frac{N-n}{N}=\frac{PQ}{n}\frac{N-n}{N-1}$$

N ????? se non lo conosciamo

$$V(p)\approx\frac{PQ}{n}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale A

Esempio: *Errore (relativo) massimo tollerabile 10%*

nell'ipotesi cautelativa in cui P=Q=0,5

$$= 0{,}10 \times 0{,}5 = 1{,}96 \frac{\sqrt{0{,}5 \times 0{,}5}}{\sqrt{n}} \Rightarrow 0{,}05 = \frac{1{,}96 \times 0{,}5}{\sqrt{n}}$$

$$\sqrt{n} = \frac{1{,}96 \times 0{,}5}{0{,}05} \Rightarrow n = \frac{1{,}96^2 0{,}5^2}{0{,}05^2} = 384{,}16 \approx 385$$

caso in cui N=10000

$$0{,}10 \times 0{,}5 = 1{,}96 \frac{\sqrt{0{,}5 \times 0{,}5}}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$0{,}10 \times 0{,}5 = 1{,}96 \frac{\sqrt{0{,}5 \times 0{,}5}}{\sqrt{n}} \sqrt{\frac{10000-n}{9999}}$$

CCS – stima della dimensione campionaria

Dimensione campionaria per la stima del totale A

Esempio: *Errore (relativo) massimo tollerabile 10%*

$$0{,}05 = \frac{0{,}98}{\sqrt{n}} \sqrt{\frac{10000 - n}{9999}} =$$

$$5{,}102^2 = \frac{10000 - n}{n} \Rightarrow n = \frac{10000}{27{,}028} = 369{,}98 \approx 370$$

caso in cui N=1000

$$0{,}10 \times 0{,}5 = 1{,}96 \frac{\sqrt{0{,}5 \times 0{,}5}}{\sqrt{n}} \sqrt{\frac{1000 - n}{999}}$$

$$n = \frac{1000}{3{,}6} = 278$$

CCS – stima della dimensione campionaria

a proposito di n

a parità di altre condizioni, al crescere del coefficiente di confidenza (z), aumenta rapidamente la dimensione campionaria

$$n \rightarrow z_{\alpha/2}^{2}$$

n dipende dalla varianza della popolazione o dalla sua stima (V)

Popolazione più variabile

$$n \rightarrow V$$

n più grande

n dipende dalla precisione (d) (errore massimo che si può commettere)

$$n \rightarrow d^{2}$$

Errore grande → n piccolo