Dimensione campionaria

- 3.1 Scelta della strategia di campionamento;
- 3.2 Numerosità in funzione dei costi di rilevazione;
- 3.3 Numerosità in funzione dell'efficienza degli stimatori;
-

3.1 Scelta della strategia di campionamento

Come si può osservare dai risultati teorici ricavati precedentemente, dimensioni campionarie elevate comportano una riduzione della varianza degli stimatori.
Essa riduce l'ampiezza dell'intervallo di confidenza e quindi l'errore delle stime.

Tuttavia, nella scelta della dimensione del campione bisogna considerare l'incidenza dei costi di realizzazione dell'indagine. Infatti quanto più elevata è la dimensione del campione, tanto più le stime sono attendibili, ma l'aumento della dimensione campionaria determina un aumento dei costi connessi all'indagine.

In generale infatti, la numerosità del campione può essere determinata in relazione a due fattori: costi ed efficienza degli stimatori.

Determinazione della numerosità campionaria:

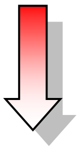
- *In base all'efficienza degli stimatori (precisione di stima);*
- *In base ai costi di rilevazione.*

Aumentare l'*efficienza delle stime*
significa aumentare
la *numerosità del campione*

Aumentare la *numerosità del campione*
significa aumentare
il *costo di rilevazione*

Obiettivo:
trovare il punto di equilibrio tra numerosità campionaria, precisione della stima e costi di rilevazione (nei limiti di spesa prefissati).

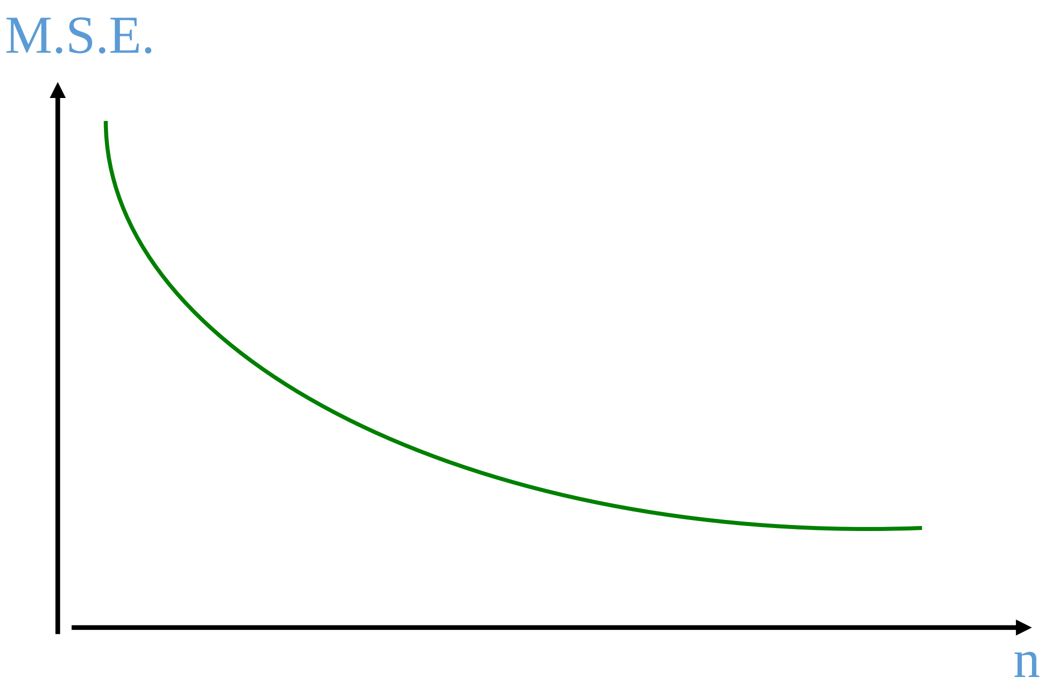
In base alla numerosità campionaria si determina il costo effettivo di indagine



Problema:
il costo è troppo elevato!!!!!

Soluzioni:

- *Aumentare il budget disponibile;*
- *Tollerare una minore precisione (a discrezione del committente);*
- *Muoversi in entrambe le direzioni.*



3.2 Numerosità in funzione dei costi di rilevazione

Classificazione dei costi:

Costi

Costi fissi

- definizione degli obiettivi;
- definizione della popolazione e delle variabili da studiare;
- costruzione del questionario;
- reperimento degli strumenti di calcolo;
- analisi ed interpretazione dei risultati;
- report finale.

Costi per la strategia

- preparazione delle liste di unità elementari, dei grappoli o dei strati;
- definizione delle variabili ausiliarie per stimatori per quoziente o per regressione;

Costi per la dimensione

- codifica dei dati raccolti;
- trascrizione su supporto magnetico;

Dipendono dal tipo di intervista:

- Telefonica;
- Posta;
- Intervista diretta.

Per qualunque strategia le funzioni di costo sono così strutturate:

$$C = C_f + C_s + C_d \cdot n$$

dove:

C_f – costi fissi;

C_s – costi di strategia;

C_d – costi legati alla dimensione;

L'espressione precedente può essere semplificata nel modo seguente:

$$C = C_0 + C_1 \cdot n$$

dove:

C_0 – costi che non dipendono dalla dimensione;

C_1 – costi dipendenti dalle unità campionarie.

Determinazione del minimo della funzione C!!!

Si consideri la funzione di costo

funzione di costo

$$C = C_0 + C_1 \cdot n$$

A volte questa può essere scritta anche

$$C = C_0 + C_1 \cdot n + C_2 \cdot \sqrt{n}$$

dove:

C_2 – costo medio di trasferimento per il raggiungimento di una sola unità campionaria (è dimostrato che aumentano in funzione della radice quadrata di n).

Dalla funzione di costo:

$$n = \frac{C - C_0}{C_1}$$

la numerosità campionaria è determinata in funzione dei costi.

3.3 Numerosità in funzione dell'efficienza degli stimatori

Varianza dello stimatore del totale:

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

oppure

$$Var(\hat{Y}) = \left(\frac{1}{n} - \frac{1}{N}\right) N^2 S^2$$

funzione di varianza

Dalla funzione di varianza si ricava la numerosità campionaria:

$$n = \left(\frac{Var_0(\theta)}{N^2 S^2} + \frac{1}{N}\right)^{-1}$$

dove

$Var_0(\theta)$ – varianza voluta.

$$Var_0(\hat{\theta}) = \frac{\varepsilon^2}{z^2_{\alpha/2}}$$

Sostituendo la numerosità nella funzione di costo

$$C = C_0 + C_1 \cdot \left(\frac{Var_0(\theta)}{N^2 S^2} + \frac{1}{N}\right)^{-1}$$

COSTO TOTALE DELL'INDAGINE!!!

Varianza dello stimatore della media:

$$Var(\overline{Y}) = \frac{(1-f)}{n} S^2$$

oppure

$$Var(\hat{Y}) = \left(\frac{1}{n} - \frac{1}{N}\right) S^2$$

funzione di varianza

Dalla funzione di varianza si ricava la numerosità campionaria:

$$n = \left(\frac{Var_0(\theta)}{S^2} + \frac{1}{N}\right)^{-1}$$

dove

$Var_0(\theta)$ – varianza voluta.

$$Var_0(\overline{\theta}) = \frac{\varepsilon^2}{z^2_{\alpha/2}}$$

Sostituendo la numerosità nella funzione di costo

$$C = C_0 + C_1 \cdot \left(\frac{Var_0(\theta)}{S^2} + \frac{1}{N}\right)^{-1}$$

COSTO TOTALE DELL'INDAGINE!!!

Varianza dello stimatore della frequenza relativa:

$$Var(p) = \frac{N-n}{N-1}\frac{P(1-P)}{n}$$ [1]

Dalla funzione di varianza si ricava la numerosità campionaria:

$$n = \left(\frac{Var_0(\theta)}{P(1-P)} + \frac{1}{N}\right)^{-1}$$

dove

$Var_0(\theta)$ – varianza voluta.

[1] Si consideri la varianza della media in un CCS senza ripetizione:

$$Var\,(\bar{y}) = \frac{1-f}{n}S^2$$

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

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

$$Var_0(p) = \frac{\varepsilon^2 \cdot P^2}{z_{\alpha/2}^2}$$

Sostituendo la numerosità nella funzione di costo

$$C = C_0 + C_1 \cdot \left(\frac{Var_0(\theta)(N-1)}{N \cdot P(1-P)} + \frac{1}{N} \right)^{-1}$$

COSTO TOTALE DELL'INDAGINE!!!

Problema: P è incognito dato che è la variabile oggetto di indagine. Si assume P = 0,5 in modo che è massima la variabilità P(1-P).

ESERCIZI DI RIEPILOGO

Esercizio 1
La Facoltà di Statistica di una Università italiana desidera conoscere il numero medio di studenti attualmente iscritti, che intende scegliere di conseguire la Laurea triennale invece di mantenere l'attuale piano di studi. Allo scopo si decide di effettuare un CCS senza ripetizione tenendo conto che il numero di studenti iscritti è pari a 7831 e che la variabilità prestimata è S=152.
Determinare la numerosità campionaria sapendo che l'errore di campionamento non deve superare i 20 studenti con un livello di probabilità del 98%.

Esercizio 2
Per conoscere la percentuale delle matricole favorevoli alle proposte del Ministro della Pubblica Istruzione relative alla Riforma Universitaria si effettua un sondaggio tra le 1860 matricole. A tale scopo viene estratto un campione semplice senza ripetizione.
Determinare la dimensione campionaria al livello di significatività del 95% con un errore non superiore al 5% nell'ipotesi che, da una ricerca effettuata l'anno precedente in un'altra università, p viene prestimato pari a 0,7.

Esercizio 3
Si vuole determinare la numerosità di un campione di soggetti da intervistare, in una città di 90.000 abitanti, per stimare la proporzione P di possessori di carta di credito sull'intera popolazione. Si desidera che la stima p, al livello di confidenza del 95%, non si discosti da P più del 3% in valore assoluto (errore ammesso).
Si consideri prima l'ipotesi di massima variabilità e poi l'ipotesi in cui in un pre-test la percentuale di possessori della carta di credito sia pari al 31%.
Ipotizzando che il budget disponibile per la conduzione della ricerca sia di 40.000€, che i costi fissi e quelli di strategia in totale ammontano a 4.500€ e che i costi per la singola intervista sono pari a 30€, analizzare i due scenari alternativi.

Esercizio 4:
La popolazione della città X ammonta a 50.000 abitanti. Si vuole condurre un'indagine campionaria al fine di stimare la proporzione di cittadini favorevoli all'allestimento di un nuovo servizio di trasporto. Un'indagine precedente afferma che ben il 70% dei cittadini è favorevole all'allestimento del nuovo servizio.
Si decide di estrarre, con campionamento casuale semplice senza reintroduzione, un campione di 1.000 cittadini. L'amministrazione comunale decide che il servizio sarà allestito solo se i risultati dell'indagine confermeranno che almeno il 70% dei cittadini è favorevole. Si concorda che il livello di fiducia sia del 95%.
a) Calcolare le probabilità di inclusione di primo e secondo ordine degli elementi della popolazione.
b) Calcolare l'errore assoluto associato alla strategia campionaria, senza tenere conto dell'informazione derivante dall'indagine precedente. Che ipotesi è necessario fare sulla varianza di popolazione?
c) Calcolare l'errore assoluto associato alla strategia campionaria, tenendo conto dell'informazione derivante dall'indagine precedente.
d) Calcolare la numerosità campionaria necessaria ad ottenere un errore assoluto pari a 0,02, senza tenere conto dell'informazione derivante dall'indagine precedente.

In seguito all’estrazione del campione (n=1.000), 620 cittadini dichiarano di essere favorevoli all’allestimento del nuovo servizio.
e) Stimare la proporzione di cittadini favorevoli. Stimare il totale di cittadini favorevoli.
f) Mediante la costruzione di un intervallo di confidenza, quale decisione prenderà l’amministrazione sulla base dell’esito campionario, tenendo conto di un 5% di probabilità di errore?