Formulario per
Metodi per le indagini campionarie M

Università di Milano-Bicocca, CLAMSES

Le seguenti formule sono date per il totale di un carattere $\mathscr{Y}$, cioè per

$$Y = \sum_{i=1}^{N} Y_i.$$

Si ricordi che si può passare alla media attraverso la relazione $\bar{Y} = Y/N$. Le notazioni sono quelle usate a lezione, in particolare p_i indica la probabilità di estrazione dall'urna, π_i la probabilità di inclusione del primo ordine.

- Stimatore di Horvitz-Thompson

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

- Stimatore di Hansen-Hurwitz

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

CAMPIONAMENTO CASUALE SEMPLICE

Campionamento senza ripetizione.

- Stimatore del totale

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

- Varianza dello stimatore del totale

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

- Stimatore della varianza dello stimatore del totale

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

- Stimatore della media

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

- Varianza dello stimatore della media

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

- Stimatore della varianza dello stimatore della media

$$\hat{\mathrm{Var}}(\hat{\bar{Y}}) = \frac{1-f}{n} s^2$$

- Dimensione campionaria minima per garantire l'accuratezza $P(|Y - \hat{Y}| > d) < \alpha$

$$n = \frac{1}{\frac{1}{n_0} + \frac{1}{N}}, \quad n_0 = \frac{N^2 z_{1-\alpha/2}^2 S^2}{d^2}$$

dove $z_{1-\alpha/2}$ è il quantile della $N(0,1)$ di ordine $1 - \alpha/2$

- Varianza del carattere nella stima delle percentuali

$$S^2 = \frac{N}{N-1} P(1-P), \quad s^2 = \frac{n}{n-1} \hat{p}(1-\hat{p})$$

Campionamento con ripetizione.

- Stimatore del totale

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

- Varianza dello stimatore del totale

$$\mathrm{Var}(\hat{Y}) = \frac{N^2}{n} \sigma^2$$

- Stimatore della varianza dello stimatore del totale

$$\hat{\mathrm{Var}}(\hat{Y}) = \frac{N^2}{n} s^2$$

- Stimatore della media

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

- Varianza dello stimatore della media

$$\mathrm{Var}(\hat{\bar{Y}}) = \frac{1}{n}\sigma^2$$

- Stimatore della varianza dello stimatore della media

$$\hat{\mathrm{Var}}(\hat{\bar{Y}}) = \frac{1}{n}s^2$$

CAMPIONAMENTO DI YATES E GRUNDY

- probabilità di inclusione del primo ordine

$$\pi_i = p_i + \sum_{j \neq i} p_j \cdot \frac{p_i}{(1-p_j)}$$

- probabilità di inclusione del secondo ordine

$$\pi_{i,j} = \frac{p_i p_j}{(1-p_i)} + \frac{p_i p_j}{(1-p_j)}$$

CAMPIONAMENTO STRATIFICATO

- Stimatore del totale

$$\hat{Y} = \sum_{h=1}^{H} \hat{Y}_h = \sum_{h=1}^{H} \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{i,h}$$

- Varianza dello stimatore del totale

$$\mathrm{Var}(\hat{Y}) = \sum_{h=1}^{H} N_h^2 \frac{(1-f_h)}{n_h} S_h^2$$

- Stimatore della media

$$\hat{\bar{Y}} = \sum_{h=1}^{H} W_h \hat{\bar{Y}}_h = \sum_{h=1}^{H} W_h \frac{1}{n_h} \sum_{i=1}^{n_h} y_{i,h}$$

- Varianza dello stimatore della media

$$\mathrm{Var}(\hat{\bar{Y}}) = \sum_{h=1}^{H} W_h^2 \frac{(1-f_h)}{n_h} S_h^2$$

- Ripartizione di Neyman (allocazione ottima)

$$n_h = n\frac{N_h S_h}{\sum_{k=1}^{H} N_k S_k}$$

per ogni $h = 1, \dots, H$

STIMATORI RAPPORTO

Si indichi d'ora in poi con $\mathscr{X}$ il carattere ausiliario.

- Stimatore quoziente del totale

$$\hat{Y}_Q = \frac{\hat{Y}}{\hat{X}} \cdot X$$

- Varianza approssimata dello stimatore del totale

$$\mathrm{Var}(\hat{Y}_Q) \approx \mathrm{Var}(\hat{Y} - R\hat{X})$$

- Stimatore quoziente della media

$$\hat{\bar{Y}}_Q = \frac{\hat{Y}}{\hat{X}} \cdot \bar{X}$$

- Varianza approssimata dello stimatore della media

$$\mathrm{Var}(\hat{\bar{Y}}_Q) \approx \mathrm{Var}(\hat{\bar{Y}} - R\hat{\bar{X}})$$

- lo stimatore quoziente è più efficiente dello stimatore tradizionale corretto se vale la relazione

$$RS_X^2 \leq 2S_{XY}$$

- Stimatore quoziente combinato del totale

$$\hat{Y}_{Qc} = \frac{\hat{Y}_{st}}{\hat{X}_{st}} X$$

- Stimatore quoziente combinato della media

$$\hat{\bar{Y}}_{Qc} = \frac{\hat{Y}_{st}}{\hat{X}_{st}} \bar{X}$$

- Stimatore quoziente separato della media

$$\hat{\bar{Y}}_{Qs} = \sum_{h=1}^{H} W_h \frac{\hat{Y}_h}{\hat{X}_h} \bar{X}_h$$

STIMATORI PER REGRESSIONE

Si indichi d'ora in poi con $\mathscr{X}$ il carattere ausiliario.

- Stimatore per regressione del totale (con β noto)

$$\hat{Y}_R = \hat{Y} + \beta(X - \hat{X})$$

  dove $\beta = S_{XY}/S_X^2$

- Varianza dello stimatore per regressione del totale

$$\mathrm{Var}(\hat{Y}_R) = \mathrm{Var}(\hat{Y})(1-\rho^2)$$

  dove $\rho = S_{XY}/(S_X S_Y)$

- Stimatore per regressione della media (con β noto)

$$\hat{\bar{Y}}_R = \hat{\bar{Y}} + \beta(\bar{X} - \hat{\bar{X}})$$

- Varianza dello stimatore per regressione della media

$$\mathrm{Var}(\hat{\bar{Y}}_R) = \mathrm{Var}(\hat{\bar{Y}})(1-\rho^2)$$

CAMPIONAMENTO CASUALE A GRAPPOLI

Di seguito il simbolo $\bullet$ indica la somma rispetto ad un indice, ad esempio $y_{i\bullet} = \sum_{j=1}^{M_i} y_{ij}$.

- Stimatore del totale

$$\hat{Y}_G = \frac{N}{n} \sum_{i=1}^{n} y_{i\bullet}$$

- Varianza dello stimatore

$$\mathrm{Var}(\hat{Y}_G) = \frac{N^2(1-f)}{n} \cdot \frac{1}{N-1} \sum_{i=1}^{N} (Y_{i\bullet} - Y_{\bullet\bullet}/N)^2$$