



ERICA: PROGRAMMA INFORMATICO PER CALCOLARE L'ERRORE CAMPIONARIO ED EXTRACAMPIONARIO IN INDAGINI CON REINTERVISTA E/O COMPENETRAZIONE DELLE ASSEGNAZIONI DEGLI INTERVISTATORI

Susi Dulli

Dipartimento di Matematica Pura ed Applicata, Università di Padova.

Luigi Fabbris, Angela Me

Dipartimento di Scienze Statistiche, Università di Padova.

Il programma informatico ERICA è uno strumento flessibile per l'analisi statistica di dati raccolti in indagini dal disegno di rilevazione complesso. Il programma è predisposto, in particolare, per il calcolo degli errori campionari ed extracampionari di stime basate su dati ottenuti applicando le tecniche di rilevazione della compenetrazione delle assegnazioni dei rilevatori e della reintervista, eventualmente inserite in un campionamento stratificato. Utilizzando algoritmi di stima selezionati, il programma calcola le stime di medie o proporzioni e il loro errore globale, qualunque disegno sia stato adottato per la raccolta dei dati e, nei casi in cui è fattibile, scinde l'errore di campionamento dagli errori di rilevazione attribuibili ai rispondenti e ai rilevatori e dalla covarianza tra l'errore di campionamento e l'errore di rilevazione

1. INTRODUZIONE

Il progetto di cui il programma presentato è il risultato nasce dall'esigenza di disporre di un programma applicativo per l'elaborazione degli errori campionari ed extracampionari che sia portabile e utilizzabile da utenti con media cultura informatica.

Poiché in commercio non esiste un pacchetto applicativo adatto allo scopo, si è ritenuta questa un'utile premessa per la realizzazione di un programma interattivo di immediato utilizzo.

Il programma ERICA (Errori di Rilevazione e CAmpionari) è predisposto per il calcolo degli errori campionari ed extracampionari di stime ottenute sulla base di dati raccolti in indagini statistiche con disegno complesso*. A tal fine, interagendo con l'utilizzatore, seleziona i parametri da stimare in funzione del disegno di indagine (par. 2) e del tipo di variabile considerata (par. 3), calcola automaticamente le stime e stampa le informazioni rilevanti per l'analisi statistica dei dati (par. 4). Le stime si possono calcolare escludendo le mancate risposte e i valori non pertinenti.

Nella nota si descrivono sinteticamente anche la struttura del programma (par. 5) e le procedure attivate nello stesso (par. 6).

2. DISEGNI DI RILEVAZIONE

Nel programma si considerano 4 disegni di rilevazione:

1. stratificazione (par. 2.1);
2. compenetrazione delle assegnazioni dei rilevatori (par. 2.2);
3. reintervista (par. 2.3);
4. reintervista con doppia compenetrazione delle assegnazioni dei rilevatori (par.2.4).

L'obiettivo di disegni che prevedono la compenetrazione delle assegnazioni dei rilevatori e/o la reintervista è la quantificazione, mediante opportuni stimatori, dell'incidenza dell'errore extracampionario sulla distribuzione delle stime in una indagine statistica. Per quantificare gli errori ci si è avvalsi dello schema per l'analisi degli errori di rilevazione che Fellegi (1964) ha proposto sulla scia delle riflessioni di Hansen, Hurwitz e Bershad (1959). Lo schema discende da un modello di scomposizione degli errori in diverse componenti semplici e correlate e campionarie ed extracampionarie. A questo modello sono associati stimatori caratterizzati da livelli di accuratezza variabili (Fellegi, 1964).

2.1 STRATIFICAZIONE

La stratificazione è la suddivisione della popolazione in *strati*. Come è noto, la stratificazione può incrementare l'efficienza delle stime rispetto a campioni casuali semplici, ma non riduce gli errori di rilevazione. Sulla base della stratificazione si possono formare domini di studio entro i quali sono calcolabili le stime oggetto di indagine. I domini di studio coincidono con uno strato o con l'unione di più strati.

* Una prima versione del programma è stata predisposta nell'ambito delle attività per la tesi di laurea di A. Me (1988). La tesi contiene, tra l'altro, riferimenti più dettagliati sull'aspetto statistico dell'argomento.

2.2 COMPENETRAZIONE DELLE ASSEGNAZIONI DEI RILEVATORI

La compenetrazione delle assegnazioni dei rilevatori è un disegno di rilevazione (Mahalanobis, 1946) consistente nel suddividere il campione di n unità in tanti sottoinsiemi casuali di uguale ampiezza quanti sono gli intervistatori e nell'assegnare casualmente un sottoinsieme ad ogni intervistatore. Considerando che ogni assegnazione è un campione casuale della popolazione oggetto di indagine, la variabilità media dei rilevatori si considera proporzionale alla loro propensione a commettere errori. L'applicazione della tecnica della compenetrazione dei subcampioni permette la stima della varianza dell'intervistatore, che è una varianza di rilevazione cosiddetta "correlata" (Fabbris, 1989: cap. 9).

Nel caso che il campione sia stratificato, ogni strato entro il quale si effettua la compenetrazione degli intervistatori, viene considerato come un campione a sé stante.

2.3 REINTERVISTA

Per *reintervista* si intende comunemente l'indagine secondaria, svolta ad una certa distanza dalla principale, talvolta con personale diverso, avente come obiettivo il misurare, per singola unità statistica o in media, la concordanza dei dati ottenuti nelle due fasi. L'eventuale discordanza rilevata nelle due indagini si considera proporzionale alla propensione a dare risposte scorrette da parte delle unità esaminate. La reintervista permette di stimare la cosiddetta "varianza del rispondente", ossia la componente della varianza di rilevazione riferita alle unità rispondenti. Questa forma di varianza si dice "elementare" o "incorrelata".

2.4 REINTERVISTA CON DOPPIA COMPENETRAZIONE DELLE ASSEGNAZIONI DEI RILEVATORI

Se la compenetrazione delle assegnazioni degli intervistatori viene impiegata anche nella reintervista, con gli stessi rilevatori che hanno svolto l'indagine principale, ma facendo in modo che ogni rispondente veda un rilevatore diverso, il disegno che si ottiene è una *reintervista con doppia compenetrazione*. Se si è applicata questa complessa tecnica di rilevazione si possono separare, con buona approssimazione, la varianza dovuta al campionamento da quella imputabile agli errori di rilevazione e identificare, con maggiore precisione, le componenti semplici e correlata dell'errore campionario.

3. TIPI DI VARIABILI

Il programma distingue due tipi di variabili: quelle dicotomiche e quelle quantitative. Nella stima degli errori di rilevazione, per qualche disegno di indagine,

vengono elaborati degli stimatori particolari per le variabili dicotomiche.

4. STIMATORI ADOTTATI

Si consideri una popolazione di N unità dalla quale è stato selezionato un campione di dimensione n . Il campione è suddiviso con una procedura casuale in k sub-campioni (tanti quanti sono gli intervistatori) di uguale dimensione $\bar{n} = n/k$.

Denotiamo con:

$$y_{ij} \quad [i = 1, \dots, k; j = 1, \dots, n_i] \quad (1)$$

il valore osservato presso l'unità campionaria j appartenente all'assegnazione i di numerosità n_i , con

$$Y_j \quad [j = 1, \dots, \bar{n}] \quad (2)$$

il valore medio risultante nei campioni contenenti l'unità statistica j , con

$$\mu_j \quad [j = 1, \dots, N] \quad (3)$$

il "valore vero" dell'unità j , con

$$Y_{..} = \sum_{j=1, N} Y_j / N \quad (4)$$

la media dei valori osservabili e con:

$$\mu = \sum_{j=1, N} \mu_j / N \quad (5)$$

la media della popolazione dei "valori veri".

Lo scarto tra il valore osservato in un'indagine e la vera media μ : $y_{ij} - \mu$ si scompone nell'errore di risposta d_{ij} , nell'errore campionario c_j e nella distorsione Δ , dati rispettivamente da:

$$d_{ij} = y_{ij} - Y_j \quad (6)$$

$$c_j = Y_j - Y_{..} \quad (7)$$

$$\Delta = Y_{..} - \mu. \quad (8)$$

In qualunque indagine nella quale si impieghino rilevatori, lo stimatore della media $y_{..} = \sum_{i=1, k} \sum_{j=1, \bar{n}} y_{ij}$ ha varianza (Fellegi, 1964):

$$\text{Var}(y_{..}) = \frac{\sigma_c^2}{n} \cdot \frac{N-n}{N-1} + \frac{\sigma_r^2}{n} [1 + (\bar{n}-1)\delta + (n-\bar{n})\gamma] + 2\alpha\sigma_c\sigma_r \cdot \frac{(\bar{n}-1)(N-n)}{n(N-\bar{n})} \quad (9)$$

dove

$$\sigma_c^2 = \sum_{j=1, N} c_j^2 / N \quad (10)$$

è la componente campionario della varianza e

$$\sigma_r^2 = \sum_{j=1, N} E(d_{ij}^2 | D) / N \quad (11)$$

è la varianza elementare di risposta (che si osserverebbe se la rilevazione avvenisse senza l'intermediazione di intervistatori). $E(\cdot | D)$ indica il valore atteso della statistica condizionatamente al disegno di rilevazione adottato.

Le altre componenti della (9) sono il coefficiente di correlazione intra-

intervistatore:

$$\delta = \frac{1}{\sigma_r^2} \cdot \frac{2}{N(N-1)} \cdot \sum_{j>i} \sum_j E(d_{ij} d_{ij} | D), \quad (12)$$

il coefficiente di correlazione tra errori di risposta ottenuti da intervistatori diversi:

$$\gamma = \frac{1}{\sigma_r^2} \cdot \frac{1}{N(N-1)} \cdot \sum_{j \neq i} \sum_j E(d_{ij} d_{ij'} | D), \quad (13)$$

e la covarianza tra errori di rilevazione ed errori campionari:

$$\alpha \sigma_s \sigma_r = \frac{1}{N(N-1)} \cdot \sum_{j \neq i} \sum_j c_j E(d_{ij} | D). \quad (14)$$

Lo stimatore proposto per $\text{Var}(y_{..})$ è dato dalla varianza tra medie degli intervistatori:

$$\text{var}(y_{..}) = \frac{1}{k(k-1)} \cdot \sum_{i=1, k} (y_i - y_{..})^2. \quad (15)$$

Se il carattere considerato è dicotomico, uno stimatore corretto della frazione p di unità che nella popolazione lo possiedono è dato da:

$$p = \frac{1}{n} \cdot \sum_{j=1, n} y_j = \frac{n'}{n} \quad (16)$$

dove n' è il numero di unità campionarie che possiedono il carattere y .

Lo stimatore proposto per σ_r^2 è:

$$s_r^2 = \frac{2}{m-1} \cdot \sum_{i=1, k} \sum_{j=1, m} (y_{ij}' - y_{ij}'' - y_{i..}' + y_{i..}'')^2 \quad (17)$$

dove si denotano con y' e y'' i valori registrati, rispettivamente, nella prima e nella seconda indagine, con y_{ij}'' il valore ottenuto dall'intervistatore i' che ha visitato l'unità j nella seconda indagine e con m il numero di unità sottoposte a reintervista ($m = m/k$).

Se il carattere è dicotomico, lo stimatore di σ_r^2 è:

$$s_r^2 = \frac{2}{m-1} \cdot [c + b - (c-b)^2/m] \quad (18)$$

dove b e c denotano il numero di unità campionarie rilevate come portatrici dell'attributo, rispettivamente, nell'indagine principale ma non nella secondaria, e nella reintervista ma non nell'indagine primaria.

Supponendo che il campione sia casuale semplice, lo stimatore proposto per $\sigma_r^2 + \sigma_s^2$ è:

$$s_r^2 + s_c^2 = \frac{1}{n-1} \cdot \sum_{i=1, k} \sum_{j=1, n} (y_{ij} - y_{..})^2 \quad (19)$$

mentre quello per "l'indice di incoerenza" (Pritzker e Hanson, 1962), ossia per la frazione di varianza di risposta incorrelata sul totale della varianza incorrelata, è:

$$\hat{\Gamma} = \frac{s_r^2}{s_r^2 + s_c^2} \quad (20)$$

Per calcolare la componente della varianza dovuta all'effetto degli intervistatori nel caso in cui non si effettui la reintervista ma si compenetrino le assegnazioni dei rilevatori, lo stimatore proposto per $\delta\sigma_r^2$ è:

$$\hat{\delta}s_r^2 = k \cdot \text{var}(y_{..}) \frac{1}{k(\bar{n}-1)} \cdot \sum_{i=1,k} p_i (1-p_i) \quad (21)$$

e per δ :

$$\hat{\delta} \cong \delta\hat{\Gamma} = \hat{\delta}s_r^2 / (s_r^2 + s_c^2), \quad (22)$$

dove il numeratore è dato dalla (21) e il denominatore dalla (19).

Se il disegno di rilevazione dei dati comprende anche la reintervista con doppia compenetrazione delle assegnazioni, lo stimatore proposto per $\delta\sigma_r^2$ è:

$$\hat{\delta}s_r^2 = \frac{m-1}{m} \cdot [s_r^2 - \frac{2k}{m-1} \cdot \sum_{i=1,k} \sum_{j=1,m} (y_{ij}' - y_{ij}'' - y_{i.}' + y_{i.}'')^2] \quad (23)$$

che diventa, se il carattere è dicotomico:

$$\hat{\delta}s_r^2 = \frac{m-1}{m} \cdot [s_r^2 - \frac{2k}{m-1} \cdot (\sum_{i=1,k} (c_i + b_i) - (c_i - b_i)^2/m)]. \quad (24)$$

Lo stimatore di δ è dato da:

$$\frac{\hat{\delta}s_r^2 \cdot 2(m-1)}{\sum_{i=1,k} \sum_{j=1,m} (y_{ij}' - y_{ij}'' - y_{i.}' + y_{i.}'')^2} \quad (25)$$

Lo stimatore proposto per $\alpha\sigma_c\sigma_r$ è:

$$\frac{N-\bar{n}}{2\bar{n}(N-1)} \cdot [\tau + \frac{n-1}{k} \cdot (\frac{\chi}{m(m-1)} s_r^2)] \quad (26)$$

dove τ è dato dalla (21) e χ dalla (23).

Lo stimatore per $(\alpha_1\sigma_{c_1}\sigma_{r_1} + \alpha_2\sigma_{c_2}\sigma_{r_2})/2$ è:

$$\frac{N-\bar{n}}{2\bar{n}(N-1)} \cdot [\frac{\tau_1 + \tau_2}{2} + \frac{cn-1}{k} \cdot (\frac{\chi}{m(m-1)} - s_r^2)] \quad (27)$$

Lo stimatore proposto per $(\delta_1\sigma_{r_1}^2 + \delta_2\sigma_{r_2}^2)/2$ è:

$$\frac{n-1}{n-\bar{n}} \cdot \left[\frac{\chi}{m(m-1)} - s_r^2 \right] \quad (28)$$

Nel seguito, si riporta l'elenco delle grandezze stimate seguendo i diversi percorsi analizzati nella Figura 1.

PERCORSI	STIMATORI
2.2.2.1 1.2.2.1 2.2.2.2 1.2.2.2.	$Y_{..1}, Y_{..2}, \text{Var}(y_{..1}), \text{Var}(y_{..2}), \sigma_{r1}^2 + \sigma_{s1}^2,$ $\sigma_{r2}^2 + \sigma_{s2}^2, (\sigma_{r1}^2 + \sigma_{r2}^2)/2, (\delta\sigma_{r1}^2 + \delta\sigma_{r2}^2)/2, l_1,$ $l_2, \delta\sigma_{r1}^2, \delta\sigma_{r2}^2, \delta_1 + \delta_2, \delta_1, \delta_2, \alpha_1\sigma_{r1}\sigma_{s1}, \alpha_2\sigma_{r2}\sigma_{s2},$ $(\alpha_1\sigma_{r1}\sigma_{s1} + \alpha_2\sigma_{r2}\sigma_{s2})/2.$
2.2.1.1. 2.2.1.2 1.2.1.1. 1.2.1.2	$Y_{..1}, Y_{..2}, \text{Var}(y_{..1}), \text{Var}(y_{..2}), \sigma_{r1} + \sigma_{s1}^2,$ $\sigma_{r2}^2 + \sigma_{s2}^2, (\sigma_{r1}^2 + \sigma_{r2}^2)/2, l_1, l_2.$
2.1.2.1 2.1.2.2 1.1.2.1 1.1.2.2	$Y_{..1}, \text{Var}(y_{..1}), \sigma_{r1}^2 + \sigma_{s1}^2, \delta_1\sigma_{r1}^2, \delta_1.$
2.1.1.1 2.1.1.2 1.1.1.1 1.1.1.2	$Y_{..1}, \text{Var}(y_{..1}), \sigma_{r1}^2 + \sigma_{s1}^2.$

5. STRUTTURA DEL PROGRAMMA

Uno dei principi sui quali si basa lo sviluppo del software, è la suddivisione del programma in un numero arbitrario di parti componenti. Ogni singola componente con interfaccia ben definita è detta modulo e il programma basato su tale struttura, modulare.

Per descrivere la struttura del programma, si fa uso di una rappresentazione grafica (Fig.1) dalla quale si deducono i moduli e le loro relazioni.

Si noti che ogni modulo è identificato come SUBROUTINE e che lo scambio di informazioni tra moduli (detto passaggio parametri) è il più ridotto possibile. Per tale motivo non appare nel diagramma qui presentato, ma viene descritto a parte (par. 6).

Con lo schema rappresentato nella Fig. 2 si ottiene una percezione immediata delle relazioni tra i diversi moduli (subordinati e non) e una visione globale delle parti componenti. Le Figg.1 e 2 si integrano a vicenda, essendo la seconda un'espansione e un riferimento al programma applicativo.

Il programma è stato realizzato in FORTRAN sotto MS-DOS; questo ha consentito la stesura di un codice compatto in relazione alla complessità del disegno di rilevazione e del calcolo degli stimatori. Poiché le variabili (numero degli strati, numero delle variabili etc.) sono state definite in modo parametrico, le varie parti componenti il programma sono accessibili ad interventi e modifiche provenienti dall'esterno, pur mantenendo la portabilità.

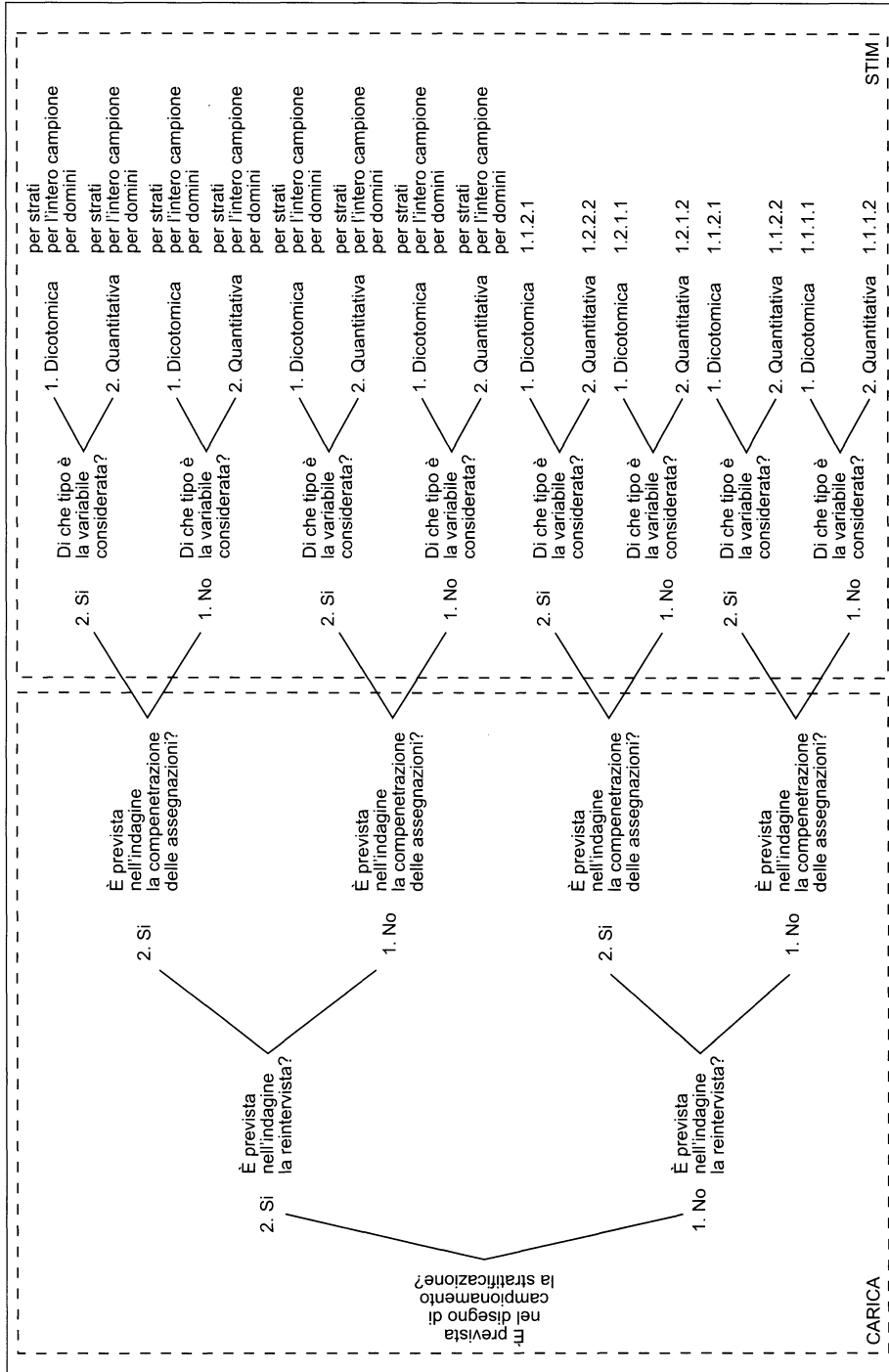


Fig 1: Schema ad albero delle possibili scelte che l'utente può effettuare avvalendosi di ERICA.

6. PROCEDURE ATTIVATE

In questo paragrafo viene riportata la lista delle "procedure" (o moduli), con i relativi parametri formali, che compongono il pacchetto applicativo ERICA. Di ogni procedura, viene fornita una descrizione sintetica della funzionalità ed eventuali commenti sulle variabili che costituiscono i parametri formali.

I nomi degli identificatori del programma possono talvolta non riuscire autoesplicativi, in quanto la sintassi del linguaggio FORTRAN usato ne limita la lunghezza a 6 caratteri.

ACCURA: specifica il significato del termine "accuratezza".

ACOPIA: crea il file di record accoppiati.

AIUTO: aiuto tecnico al programma.

BIBLIO: bibliografia essenziale.

CA: subroutine grafica.

CAMPIO (N, KP, NCP, NPP, CAP, NC, NP, K, PINT, NSTR): calcola, qualora la popolazione sia suddivisa in strati, le numerosità inerenti all'intero campione.

COMPE: specifica il significato del termine "compenetrazione".

CREA1: subroutine grafica.

CREA2: subroutine grafica.

CREA3: subroutine grafica.

DEF: visualizza le espressioni analitiche degli stimatori.

DISTO: specifica il significato del termine "distorsione".

DOMINI: specifica il significato del termine "dominio".

DUEUN (P, CP, X, F, XINT, PINT, PSTR, VAR, FILE1, NUSSC, K, CARICO, A, PJ, STIME, NOMEST, NSTIME, IFORM, TI, VAU, N): calcola le stime inerenti ad una variabile quantitativa qualora il disegno di indagine non preveda né reintervista, né compenetrazione delle assegnazioni.

DUEZE (P, CP, X, F, XINT, PINT, PSTR, VAR, FILE1, NUSSC, K, CARICO, A, PJ, STIME, NOMEST, NSTIME, IFORM, TI, VAU, N): calcola le stime inerenti ad una variabile dicotomica qualora il disegno di indagine non preveda né reintervista, né compenetrazione delle assegnazioni.

ERRCA: specifica il significato del termine "errore casuale".

FINES (MU, IV): conclusa l'analisi di una variabile e prima di passare alla successiva, presenta i percorsi che si possono seguire.

FUNZI (ER, COMA): appare sul margine inferiore dello schermo una linea di diverso colore per specificare le diverse possibilità per l'operazione successiva.

GRA: subroutine grafica.

GRA2 subroutine grafica.

GRA3 subroutine grafica.

HELP: parte del programma riservata al glossario di termini strettamente statistici, alla visualizzazione delle possibilità di stima legate al disegno di indagine e alle espressioni analitiche degli stimatori.

IN: subroutine grafica.

IN2: subroutine grafica.

INF (A, AD, IV, NVAR, NSTR, R, ERT, GA): memorizza le informazioni riguardanti le variabili per le quali si vogliono le stime.

INTRO (NSTR, NUSSC, NUSSP, K, VAR, PINT, PSTR, ERX, ERX2, ERX3, CARICO, NC2, NP2, CA2, K2, NSTR2): guida l'utente all'introduzione di alcune informazioni (numerosità strati, variabili considerate, formato record,...).

MENU (P, CP, F, X, XINT, L, R) : presentazione delle tre scelte del programma (inserimento informazioni; elaborazione stime; help).

MOD (GA, NRIC, ERTT, EY, J): memorizza il tipo di ricodifica che si intende far eseguire).

NU2 (NUSSC, NUSSP, K, NSTR, ERR): memorizza la numerosità campionaria e della popolazione e il numero di intervistatori, per strato (per più di 9 strati).

NUMERO (NUSSC, NUSSP, K, NSTR, ERR): memorizza la numerosità campionaria e della popolazione e il numero di intervistatori, per strato (fino a 9 strati).

PRES: subroutine grafica.

QUAUN (P, CP, X, F, XINT, PINT, PSTR, VAR, FILE1, NUSSC, K, CARICO, A, PJ, J, STIME, NOMESE, NSTIME, IFORM, TI, VAU, N): calcola le stime inerenti ad una variabile quantitativa qualora il disegno di indagine preveda solo la penetrazione delle assegnazioni dei rilevatori.

QUAZE (P, CP, X, F, XINT, PINT, PSTR, VAR, FILE1, NUSSC, K, CARICO, A, PJ, J, STIME, NOMESE, NSTIME, IFORM, TI, VAU, N): calcola le stime inerenti ad una variabile dicotomica qualora il disegno di indagine preveda solo la penetrazione delle assegnazioni dei rilevatori.

REINT: specifica il significato del termine "reintervista".

RIC (GA, J, ERTT, EW): memorizza le modalità che si intendono ricodificare.

RICES (GA, N, KKK): procede alla ricodifica dei dati secondo le modalità espresse.

RICES2 (GA, N, KKK): procede alla ricodifica dei dati rilevati nella reintervista secondo modalità da indicare.

SCR: subroutine grafica.

SCRITA: subroutine grafica.

ST: subroutine grafica.

ST1: subroutine grafica.

STAMPA (NOMEST, CIFRE, NSTR, NSTIME, VAU, TI, IVA, ER): visualizza i risultati.

STIM (A,PSTR, VAR, PINT, NUSSC, NUSSP, K, CARICO, T, CP, F, X, XINT, L, R, F1, F2, NSTR, IFORM, NC2, NP2, CA2, K2, NSTR2): calcola gli stimatori per le variabili desiderate, li memorizza in un file, li stampa oppure li fa apparire a video. Le mancate risposte e i valori non pertinenti si possono escludere indicando il loro codice di identificazione.

STIMA (NS, T, A, NSTR, P, NUSSP, NUSSC, K, NP2, NC2, K2, PJ, J, CP, CARICO, CA2, P, F, R, L, STIME, PSTR, F1, F2, VAR, XINT, X, ID, TI, VAU, N3, N2, NSTR2): analizza il tipo di indagine e seleziona la procedura da attivare per il calcolo degli opportuni stimatori.

STIMAT: subroutine grafica.

STRAT: specifica il significato del termine "stratificazione".

TAB: visualizza le possibilità di stima legate al disegno di indagine.

TETRA (PVAR, STR, PSTR, VAR, A1, B, C, D, A, TI, J, IDA): calcola i parametri a, b, c, d per lo studio delle variabili dicotomiche.

TOMINI (N, KP, NCP, NPP, CAP, NC, NP, K, PINT, NSTR, VAU, A): calcola le numerosità all'interno di ogni dominio selezionato.

TREUN (P, CP, X, F, XINT, PINT, PSTR, VAR, FILE1, NUSSC, K, CARICO, A, PJ, J, STIME, NOMEST, NSTIME, IFORM, TI, VAU, N, FILE2, NC2, K2, NP2, CA2): calcola le stime inerenti ad una variabile quantitativa qualora il disegno di indagine preveda solo la reintervista.

TREZE (P, CP, X, F, XINT, PINT, PSTR, VAR, FILE1, NUSSC, K, CARICO, A, PJ, J, STIME, NOMEST, NSTIME, IFORM, TI, VAU, N, FILE2, NC2, K2, NP2, CA2): calcola le stime relative ad una variabile dicotomica qualora il disegno di indagine preveda solo la reintervista.

UNOUN (P, CP, X, F, XINT, PINT, PSTR, VAR, FILE1, NUSSC, K, CARICO, A, PJ, J, STIME, NOMEST, NSTIME, IFORM, TI, VAU, N, FILE2, NC2, K2, NP2, CA2): calcola le stime inerenti ad una variabile quantitativa qualora il disegno di indagine preveda sia la reintervista che la compenetrazione delle assegnazioni dei rilevatori.

UNOZE (P, CP, X, F, XINT, PINT, PSTR, VAR, FILE1, NUSSC, K, CARICO, A, PJ, J, STIME, NOMEST, NSTIME, IFORM, TI, VAU, N, FILE2, NC2, K2, NP2, CA2): calcola le stime inerenti ad una variabile dicotomica qualora il disegno di indagine preveda sia la reintervista che la compenetrazione delle assegnazioni dei rilevatori.

VALMAN: specifica il significato del termine "dato mancante".

VERIF (T, NSTR): scorrono sul video le informazioni memorizzate precedentemente, affinché l'utente ne possa verificare l'attendibilità.

VIDEO (IN, IFIN, NOMEST, CIFRE, NSTIME, TI, IVA, ER): disegna sul video la tabella dei risultati.

VIDEO2 (NOMEST, CIFRE, NSTR, NSTIME, VAU, TI, SR1, SR2, IVA, ER): carica i risultati in un file.

VIDEO3 (NOMEST, CIFRE, NSTR, NSTIME, VAU, TI, SR1, SR2, IVA, ER): stampa i risultati.

VISUA (NOMEST, CIFRE, NSTR, NSTIME, VAU, TI, SR1, SR2, IVA, ER): visualizza il menù nel quale le scelte si riferiscono alle diverse possibilità di presentazione dei risultati.

N. B. Il programma può essere richiesto a SUMMA, Via Udine 6, 35100 Padova, tel. 049/ 8803685.

BIBLIOGRAFIA

- Fabbris L.: *L'indagine campionaria. Metodi, disegni e tecniche di campionamento*, La Nuova Italia Scientifica, Roma, 1989.
- Fellegi I.P.: *Response variance and its estimation*, Journal of the American Statistical Association, 59, 1016–1041; 1964.
- Hansen M., Hurwitz W. e Bershadt: *Measurement errors in censuses and surveys*, Bulletin of the International Statistical Institute, 38, 359–378; 1959.
- Me A.: *ERICA, programma informatico per Personal Computer per il calcolo dell'errore campionario e degli errori dei rispondenti e del rilevatore*. Tesi di laurea in Scienze Statistiche e Demografiche, Facoltà di Scienze Statistiche, Demografiche ed Attuariali, Università di Padova, 1988.
- Mahalanobis P.C.: *Recent experiments in statistical sampling in the Indian Statistical Institute*, Journal of the Royal Statistical Society, 109, 325–370; 1946.
- Pritzker L. e Hanson R.H.: *Measurement errors in the 1960 census of population*, American Statistical Association, Proceedings of the Social Statistics Section: 80–90, 1962.

SUMMARY

The computer package ERICA is a flexible tool for the statistical analysis of data collected in complex surveys. The package is suitable for working out the sampling and response errors of estimates based on data obtained by applying the interpenetrated sub-sampling and/or reinterview techniques, eventually nested in a stratified sample. The estimators have been selected by evaluating the suggestions of various authors, of which reference is given. The package computes the estimate of means and proportions and their total errors, whatever the data collection design. Whenever applicable, the sampling error, the response error attributable to respondents and to enumerators, and the covariance between sampling and response errors are given apart.