

\\60\\

**La probabilità individuale di risposta nel
trattamento dei dati mancanti**

di

Andrea Giommi

Ottobre 1990

Dipartimento di Economia Politica
Via Giardini 454
41100 Modena (Italy)

La probabilità individuale di risposta nel trattamento dei dati mancanti

Andrea Giommi - Università di Modena

1. Introduzione

Il fenomeno della mancata risposta (o non risposta) si presenta in termini più o meno marcati in rapporto ai contenuti dell'indagine e ad un insieme di elementi normalmente denominati condizioni di indagine: la metodologia dell'intervista (diretta, postale), il modo in cui sono formulate le domande, la struttura del questionario, la capacità degli intervistatori di attrarre l'interesse dei rispondenti sui temi dell'indagine, i tempi e i luoghi dell'indagine stessa, ecc.

Al variare di contenuti e condizioni varia la quota di non risposta, varia cioè la propensione alla collaborazione. È del tutto naturale tradurre tale propensione in un concetto probabilistico e definire la probabilità individuale di risposta come la probabilità di ottenere l'informazione desiderata da un soggetto chiamato a partecipare all'indagine.

Tale concetto è alla base di un'impostazione emersa negli anni 40 in due lavori di Hartley (1946) e Politz & Simmons (1949) ma sviluppata organicamente soltanto dalla metà degli anni 70. Un'impostazione avente carattere di generalità in quanto nel suo ambito è reinterpretabile la maggior parte delle tecniche utilizzate per fronteggiare la non risposta nella fase di analisi dei dati e che permette una trattazione teorica unificata degli errori campionari e di quelli, non campionari, dovuti alla non risposta.

Con questa comunicazione ci si propone di riassumere ed evidenziare gli aspetti principali di questa impostazione, ricollegando contributi apparsi in numerosi lavori. Tra questi, quelli cui si fa più spesso riferimento sono dovuti a C. E. Särndall & collaboratori (1981, 1983, 1987) e Platek & Gray (1979, 1983, 1988).

2. Probabilità di selezione e di risposta

La definizione di probabilità individuale di risposta data in precedenza può essere riferita sia al complesso delle domande del questionario sia ad ogni singola domanda. In altri termini, l'approccio probabilistico può essere utilizzato indipendentemente dalla tipologia di non risposta che si considera. Normalmente si distinguono le situazioni di non risposta totale (unit non response) da quelle di non risposta ad alcune domande di un questionario (item nonresponse). In questa sede, per motivi di spazio, si fa riferimento essenzialmente alla prima situazione riservando all'altra solo brevi cenni. Inoltre, si considera soltanto la non risposta in ambito campionario anche se buona parte delle considerazioni attinenti all'impostazione descritta possono essere immediatamente estese alle indagini complete.

La teoria del campionamento da popolazioni finite, nella sua impostazione tradizionale o classica, è stata sviluppata sotto l'ipotesi di completa risposta delle unità selezionate, e ruota intorno alla determinazione e alla stima dell'errore campionario. I problemi di non risposta e più in generale gli errori non campionari, pur ricevendo un'attenzione crescente nel tempo, vengono trattati a parte, quasi fosse pacificamente impossibile affiancarli coerentemente a quelli campionari in un unico apparato teorico. E questa impossibilità perdura finché non si ammette che alla distribuzione di probabilità indotta dal metodo di selezione delle unità, di natura oggettiva, possa essere associata una distribuzione di natura prettamente soggettiva legata alle conoscenze ed alle scelte del ricercatore. Nel momento in cui si analizzano i dati in presenza di non risposta, l'adozione di un qualsiasi procedimento inferenziale implica una qualche assunzione soggettiva sul processo che ha originato la non risposta. Non si tratta quindi di ammettere o meno tali assunzioni, ma di scegliere se renderle o non renderle esplicite. La scelta della prima alternativa appare scontata soprattutto in considerazione dei vantaggi per l'utenza finale dei risultati dell'indagine.

Con l'impostazione basata sulla probabilità individuale di risposta si rendono esplicite tali scelte e si possono interpretare in chiave probabilistica assunzioni implicite nelle tecniche comunemente utilizzate nella pratica delle indagini.

Il disegno di campionamento, ossia la distribuzione di probabilità indotta dal processo di selezione delle unità, può essere descritto associando ad ogni unità della popolazione una variabile casuale indicatore, t_i , che assuma valore 1 in corrispondenza dell'evento selezione dell'unità e valore 0 in corrispondenza dell'evento non selezione. Su tale variabile vengono poi definiti valori attesi, varianze e covarianze. Ad

esempio, un campione di n unità estratto da una popolazione di N unità ($n < N$), può essere rappresentato dal vettore:

$$\underline{t} = (t_1, t_2, \dots, t_i, \dots, t_N);$$

con $n = \sum_{i=1}^N t_i$.

Le probabilità di inclusione del primo ordine risultano definite dai valori attesi:

$$E_1(t_i) = \pi_i$$

dove l'operatore E_1 (il perché dell'indice sarà chiaro tra breve) si estende all'universo dei possibili campioni estraibili dalla popolazione di N unità.

In modo del tutto analogo, scelto un qualunque carattere di indagine \mathcal{Y} , si può definire, per ogni unità inclusa nel campione di n unità, una variabile indicatore d_i ($i=1, \dots, n$) che assume valore 1 se l' i -esima unità del campione fornisce l'informazione richiesta su \mathcal{Y} e valore 0 altrimenti. L'insieme dei rispondenti, di dimensione m ($m \leq n$), è dato dal vettore:

$$\underline{d} = (d_1, d_2, \dots, d_i, \dots, d_n)$$

con $m = \sum_{i=1}^n d_i$

e la probabilità individuale di risposta è immediatamente definibile come:

$$E_2(d_i) = \alpha_i$$

dove E_2 è il valore atteso sull'universo dei possibili insiemi di rispondenti, dato il campione selezionato.

Dalle precedenti espressioni discendono immediatamente quelle relative alle probabilità del secondo ordine di inclusione e di risposta:

$$E_1(t_i, t_j) = \pi_i \pi_j + \text{Cov}_2(t_i, t_j) = \pi_{ij}$$

$$E_2(d_i, d_j) = \alpha_i \alpha_j + \text{Cov}_2(d_i, d_j) = \alpha_{ij}$$

Le covarianze di queste espressioni potranno poi risultare o meno nulle a seconda del disegno di campionamento utilizzato nell'indagine. In un disegno a grappoli, ad esempio, saranno entrambe positive.

3. Stima della media in presenza di non risposta

Come esempio di impiego delle probabilità α_i nella stima della media \bar{Y} , del carattere \mathcal{Y} , si consideri lo stimatore di Horwitz & Thompson applicato ad un campione, ad uno stadio, di n unità; di cui solo m ($m < n$) risultino rispondenti. Si può facilmente verificare la correttezza dello stimatore rettificato mediante le probabilità α_i :

$$\bar{y}_{HT}^* = \sum_{i=1}^m \frac{y_i}{N\pi_i\alpha_i} = \sum_{i=1}^N \frac{t_i d_i y_i}{N\pi_i\alpha_i},$$

$$E(\bar{y}_{HT}^*) = E_1 E_2(\bar{y}_{HT}^*) = E_1 \left(\sum_{i=1}^N \frac{t_i y_i}{N\pi_i} \right) = \bar{Y}.$$

È evidente che nelle espressioni in cui compaiono le variabili t e d , queste rappresentano l'unico elemento aleatorio sul quale applicare in sequenza i due operatori di valore atteso. L'uso delle variabili indicatore nell'espressione dello stimatore è particolarmente utile per ricavare le varianze degli stimatori, ma permette anche di osservare come in assenza di rettifica per probabilità di risposta lo stimatore sia affetto da distorsione. Infatti:

$$E \left(\sum_{i=1}^N \frac{t_i d_i y_i}{N\pi_i} \right) = E_1 \left(\sum_{i=1}^N \frac{t_i y_i \alpha_i}{N\pi_i} \right) = \sum_{i=1}^N \frac{y_i \alpha_i}{N},$$

che differisce da \bar{Y} , risultando sempre una sottostima, per valori positivi della \mathcal{Y} . Il problema del bias sarà ripreso nella successiva sezione.

La varianza dello stimatore può essere ricavata agevolmente tenendo conto del fatto che:

$$V(\bar{y}_{HT}^*) = V_1 E_2(\bar{y}_{HT}^*) + E_1 V_2(\bar{y}_{HT}^*),$$

dove Gli operatori di varianza V_1 e V_2 sono applicati rispettivamente all'insieme dei campioni estratti in base al disegno campionario prescelto e all'insieme dei possibili rispondenti, dato il campione estratto. Ponendo, per semplificare le notazioni (Särnal, 1987):

$$\hat{y}_i = \frac{y_i}{\alpha_i} = \frac{y_i}{\alpha_i \pi_i}$$

si avrà che:

$$\begin{aligned}
V_1 E_2(\bar{y}_{HT}^*) &= \sum^N \hat{y}_i^2 V_1(t_i) + 2 \sum \sum^N \hat{y}_i \hat{y}_j \text{Cov}_1(t_i, t_j) \\
&= \sum^N y_i^2 \frac{1 - \pi_i}{\pi_i} + 2 \sum \sum^N y_i y_j \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j},
\end{aligned}$$

per la prima componente e:

$$V_2(\bar{y}_{HT}^*) = \sum^N t_i^2 \hat{y}_i^2 V_2(d_i) + 2 \sum \sum_{i < j}^N t_i t_j \hat{y}_i \hat{y}_j \text{Cov}_2(d_i, d_j),$$

e quindi:

$$E_1 V_2(\bar{y}_{HT}^*) = \sum^N \hat{y}_i^2 \frac{1 - \alpha_i}{\alpha_i} \pi_i + 2 \sum \sum^N \hat{y}_i \hat{y}_j (\alpha_{ij} - \alpha_i \alpha_j) \pi_{ij},$$

per la seconda.

Si osserverà che la prima componente è la varianza dello stimatore di Horwitz & Thompson in presenza di completa risposta, mentre la seconda rappresenta l'incremento della varianza dovuto alla correzione per non risposta. Quest'ultimo è infatti nullo se $\alpha_i = 1$, per ogni i , e comunque si riduce a:

$$\sum^N \hat{y}_i^2 \frac{1 - \alpha_i}{\alpha_i} \pi_i$$

se $\alpha_{ij} = \alpha_i \alpha_j$, se cioè le probabilità individuali di risposta sono tra loro mutualmente indipendenti.

4. Probabilità di risposta e tasso di risposta campionario.

Nella precedente sezione si è visto come in assenza di un aggiustamento per non risposta, nella fattispecie mediante le probabilità α_i , lo stimatore \bar{y}_{HT}^* risulti distorto. Nella pratica, senza fare alcun riferimento alle probabilità di risposta, vengono normalmente apportate delle correzioni agli stimatori. La maggior parte dei ricercatori assume la popolazione suddivisa in due strati: quello dei rispondenti e quello dei non rispondenti (in termini probabilistici ciò equivarrebbe a considerare due soli possibili valori per la probabilità di risposta: 0 e 1). Sotto questa assunzione, l'unica informazione campionaria sullo strato dei non rispondenti è la stima della sua dimensione, data dal tasso di non risposta. La correzione più comunemente adottata è rappresentata dalla ponderazione dello stimatore con il reciproco dello stesso tasso di risposta.

Tale ponderazione viene normalmente eseguita all'interno di classi nelle quali sono raggruppate in via preliminare le unità del campione; sostanzialmente una stratificazione a posteriori avente lo scopo di raggruppare le unità campionarie in classi omogenee rispetto alla caratteristica di cui si vuole stimare la media, il totale, ecc. Tali gruppi, o classi, sono detti di riponderazione per non risposta (reweighting groups).

Nell'esempio della precedente sezione, ipotizzando che n sia la dimensione di un qualsiasi gruppo di riponderazione, nel quale figurano m rispondenti, lo stimatore di Horwitz & Thompson, risulterebbe inflazionato dal prodotto per (n/m) . Nel caso in cui le unità della popolazione abbiano la stessa probabilità di inclusione, ponderare per il reciproco del tasso di risposta equivale a stimare la media della popolazione per mezzo di quella dei soli rispondenti.

In ogni caso, la stima risultante sarà corretta solo se la media dei rispondenti nella popolazione è uguale a quella dei non rispondenti per ogni classe di riponderazione. Supponendo, per semplificare l'argomento senza peraltro perdere in generalità, di lavorare sul campione indiviso, il bias dello stimatore della media, $B(\bar{y})$, nell'impostazione classica, è esprimibile come prodotto di due componenti: la quota di non rispondenti nella popolazione, N_{NR}/N e la differenza tra media dei rispondenti, \bar{Y}_R , e media dei non rispondenti, \bar{Y}_{NR} , nella popolazione:

$$B(\bar{y}) = \frac{N_{NR}}{N} (\bar{Y}_R - \bar{Y}_{NR}).$$

Pertanto, in presenza di una quota non trascurabile di non risposta, l'entità del bias dipende dalla differenza tra le medie in parentesi. È su questa differenza che si cerca di intervenire, raggruppando opportunamente le unità rispondenti, se non è possibile ipotizzare che essa sia minima per l'intero campione.

Nell'impostazione classica non si va oltre questo tipo di analisi, alla quale si può anche pervenire partendo dal caso particolare di probabilità di risposta dicotomica.

L'impostazione probabilistica permette un'analisi alternativa del bias qualora il carattere di studio sia anch'esso dicotomico. Si assuma, per esemplificare, che nella popolazione sia P la proporzione incognita di unità che possiedono un attributo di interesse e Q il suo complementare; se si stima P dalla proporzione di rispondenti nel campione, \hat{P}_R , che hanno l'attributo, il bias risulta:

$$B(\hat{P}_R) = (N_{NR}/N)(P_R - P_{NR})$$

ed è proporzionale, analogamente a quanto già osservato in precedenza, alla differenza tra proporzione di detentori dell'attributo nello strato dei rispondenti, P_R , e stessa proporzione nello strato dei non rispondenti, P_{NR} . Per analizzare $B(\hat{P}_R)$ in termini di probabilità di risposta occorre, in primo luogo, considerare il legame intercorrente tra probabilità e tasso di risposta. Utilizzando le variabili indicatore introdotte in precedenza è possibile scrivere, per l'intero campione:

$$\frac{m}{n} = \frac{\sum_{i=1}^N t_i d_i}{\sum_{i=1}^N t_i};$$

espressione valida anche nel caso in cui il tasso di risposta sia calcolato all'interno di una classe di riponderazione, nel qual caso n è la dimensione della classe, m il numero di rispondenti nella stessa e N la sua dimensione — generalmente non nota — nella popolazione.

Il valore atteso del tasso di risposta, a meno della distorsione dovuta alla stima rapporto, può essere sviluppato come segue (Platek & Gray, 1988):

$$\begin{aligned} E(m/n) &= E_1 E_2(m/n) = E_1 E_2(\sum t_i d_i / \sum t_i) \\ &= E_1(\sum t_i \alpha_i / \sum t_i) = \sum \pi_i \alpha_i / \sum \pi_i \\ &= \sum w_i \alpha_i = \bar{\alpha}, \end{aligned}$$

dove $\bar{\alpha}$ è la probabilità media di risposta, calcolata come media delle α_i con pesi $w_i = \pi_i / \sum \pi_i$.

La distorsione può allora essere riscritta come segue:

$$B(\hat{P}_R) = \bar{\alpha}^{-1} P(1-P)(\bar{\alpha}_P - \bar{\alpha}_Q)$$

dove $\bar{\alpha}_P$ e $\bar{\alpha}_Q$ sono rispettivamente la probabilità media di risposta di coloro che sono in possesso dell'attributo e di coloro che non ne sono in possesso. Il bias, oltre ad essere inversamente proporzionale alla probabilità media complessiva di risposta, si annulla se la probabilità di risposta di coloro che possiedono l'attributo è uguale a quella di coloro che non lo possiedono.

Passando a considerare il caso più realistico di probabilità variabile tra 0 e 1, estremi inclusi, il bias può essere analizzato attraverso un'ulteriore espressione (Platek & Gray, 1983), valida per \mathcal{Y} quantitativo o dicotomico:

$$B(\bar{y}) = \bar{\alpha}^{-1} \text{Cov}(\alpha_i y_i) = \bar{\alpha}^{-1} \rho(\alpha_i y_i) \sqrt{V(\alpha_i) V(y_i)},$$

dove la covarianza, $\text{Cov}(\cdot)$ e il coefficiente di correlazione, $\rho(\cdot)$, sono calcolati sull'intera popolazione o, alternativamente, sulla popolazione relativa ad una classe di riponderazione.

L'espressione mostra che se le probabilità individuali di risposta non sono uguali a 1 o 0, la distorsione è proporzionale al coefficiente di correlazione tra le stesse probabilità e il carattere di indagine, annullandosi nel caso in cui $\rho=0$. Nel caso particolare in cui $0 < \alpha_i < 1$, ma costante per ogni i , il bias è ancora nullo, indipendentemente dal valore di ρ , essendo $V(\alpha_i)=0$.

L'analisi del bias sotto diverse possibili configurazioni della probabilità di risposta, quando lo stimatore viene rettificato mediante il tasso di risposta, può essere riassunta come segue.

- (a) Probabilità di risposta dicotomica: è equivalente, in termini di disegno campionario, a considerare la popolazione suddivisa nei due strati dei rispondenti e non rispondenti. Le stime risultano distorte se i rispondenti differiscono dai non rispondenti.
- (b) Probabilità di risposta costante: è equivalente a considerare i rispondenti come un campione casuale semplice delle unità del campione originario. Le stime risultano corrette eccettuato per la distorsione dovuta allo stimatore rapporto.
- (c) Probabilità di risposta variabile: equivale a considerare i rispondenti come un campione con probabilità variabile estratto dal campione originario. Le stime sono distorte nella misura in cui la variabile di studio è correlata con la probabilità di risposta.

5. Probabilità di risposta e imputazione

Si è discusso fino a questo punto di non risposta totale. Le situazioni di non risposta a singoli quesiti del questionario (item nonresponse) vengono nella pratica trattate in modo sostanzialmente diverso. La strategia prevalente è rappresentata dall'imputazione dei valori mancanti, spesso inquadrata nel complesso delle correzioni dei dati raccolti che va sotto il nome di editing. In effetti, in rapporto a singoli item di indagine, la differenza tra non risposta ed errore di risposta è spesso labile ed una metodologia di trattamento comune più che giustificata. Ciò non toglie che i

metodi adottati nella pratica possano essere modificati o reinterpretati in base all'approccio probabilistico qui discusso. Sostanzialmente si tratterà di riferire la probabilità di risposta alle singole variabili di indagine.

L'imputazione di tipo Hot-deck è senz'altro una delle tecniche più diffuse. In senso lato essa, è definibile come attribuzione ad un vuoto di informazione per una determinata variabile di un valore tratto da quelli disponibili nel campione per la stessa variabile. Si parla, utilizzando una terminologia assai diffusa, di unità campionarie riceventi (quelle che hanno dati mancanti) e donatrici (quelle scelte per attribuire dati alle prime). In pratica la scelta delle unità donatrici, sia essa casuale che deterministica, viene effettuata non sull'intero campione ma in classi o gruppi di unità considerate simili, omogenee, vicine, ecc. in base a criteri di valutazione che sfruttano l'insieme dei dati disponibili per tutte le unità campionarie. La ratio sottostante questi metodi è evidente: se un gruppo di unità statistiche sono tra loro simili per una serie di caratteri osservati è probabile che lo siano anche per un altro osservato solo per alcune di esse. Nell'approccio probabilistico ciò si traduce nell'equazione: similarità tra variabili osservate uguale a similarità nelle probabilità di risposta per queste variabili e sperabilmente per quella non completamente osservata. In altri termini, se la classificazione delle unità campionarie dà luogo a gruppi omogenei rispetto alla probabilità di risposta per la variabile non completamente osservata, ci si riconduce ad una situazione del tutto analoga a quella indicata al punto (b) nella precedente sezione e l'imputazione Hot-deck produce stimatori sostanzialmente corretti. In caso contrario — probabilità di risposta variabile — la distorsione dipenderà ancora dalla correlazione tra variabile non completamente osservata e probabilità di risposta.

6. Stima delle probabilità di risposta

Il problema fondamentale dell'approccio probabilistico è rappresentato dal fatto che le probabilità di risposta non sono note e, conseguentemente, devono essere sostituite da stime. Riguardo al come stimare le probabilità di risposta le proposte sono numerose. Il primo esempio di stima viene da Hartley (1946), per la particolare situazione nella quale la non risposta deriva dalla non presenza a casa del soggetto da intervistare. Assimilando la probabilità di risposta a quella di trovarsi a casa al momento dell'intervista Hartley propone di stimarla mediante il rapporto: numero di giorni in cui l'intervistato sarebbe stato trovato a casa, all'ora dell'intervista, sui sei lavorativi precedenti l'intervista stessa. L'informazione in questione può essere ottenuta inserendo nel questionario un'apposita domanda. La probabilità di risposta viene stimata per i soli rispondenti, cioè

i trovati a casa, ma si dimostra con considerazioni analoghe a quelle viste nelle precedenti sezioni che, per le variabili di studio correlate con il tempo passato a casa, gli stimatori rettificati mediante tali probabilità stimate sono meno distorti di quelli non rettificati, comunque basati sui soli rispondenti.

Dopo questo primo esempio di Hartley, c'è un vuoto di circa trenta anni prima che gli autori parlino esplicitamente di probabilità di risposta e della sua stima, senza peraltro concordare sul come effettuarla. Alcuni, tra questi Platek & Gray (1988), ritengono che le probabilità di risposta debbano essere stimate da indagini precedenti di tipo analogo a quella cui ci si riferisce. Come unica alternativa propongono la stima della probabilità di risposta per classi di unità campionarie mediante il tasso di risposta nelle classi stesse. Il loro approccio corrisponde sostanzialmente ad una reinterpretazione probabilistica di quello tradizionale che utilizza il tasso di risposta come elemento di ponderazione dello stimatore. Se nelle classi di riponderazione la probabilità di risposta è effettivamente costante lo stimatore della media, \bar{y} , risulta corretto. Se la probabilità di risposta è solo approssimativamente costante, la distorsione di \bar{y} , dipenderà dall'entità del coefficiente di correlazione $\rho(y_i, \alpha_i)$.

Quando l'ipotesi di probabilità costante all'interno di classi di unità campionarie non è assumibile occorre procedere diversamente. Fra le proposte più interessanti (Särndal, 1981) vi è quella di esprimere la probabilità di risposta in funzione di una (o più) variabile X — nota almeno per tutte le unità del campione — e di uno o più parametri incogniti, θ . Si tratta in altri termini di specificare un modello parametrico:

$$\alpha_i = f(x_i, \underline{\theta}),$$

dal quale ricavare stime di α_i passando per quelle di θ :

$$\hat{\alpha}_i = f(x_i, \hat{\theta}).$$

Si noti che l'utilizzazione del tasso di risposta come stima della probabilità, supposta costante all'interno di gruppi di unità campionarie, equivale all'adozione del semplice modello parametrico:

$$\alpha_{ih} = \theta_h ; \quad (h=1, \dots, H),$$

dove H è il numero di gruppi o classi nelle quali è stato suddiviso il campione e θ_h la probabilità costante per ogni h . Lo stimatore corretto di θ_h è in questo caso:

$$\hat{\theta}_h = m_h/n_h ,$$

cioè il tasso di risposta nel gruppo h .

L'interrogativo principale legato all'uso di modelli di risposta parametrici riguarda l'effetto di una loro errata specificazione. Si deve osservare a questo proposito che: (i) quando non è possibile avanzare alcuna ipotesi su ragionevoli legami tra variabili note e comportamento di risposta/non risposta, qualsiasi metodo si scelga per ridurre la distorsione degli stimatori risulta arbitrario e di difficile valutazione; (ii) modelli non correttamente specificati, ma che approssimino ragionevolmente il meccanismo di risposta, possono offrire risultati notevoli in termini di riduzione della distorsione degli stimatori (Giommi, 1984, 1985).

Un'alternativa alla specificazione di un modello parametrico è rappresentata dalla stima non parametrica delle probabilità di risposta. Sostanzialmente queste sono specificate in funzione delle sole variabili note:

$$\alpha_i = f(x_i)$$

e quindi stimate con procedimenti mutuati dalla teoria degli stimatori Kernel (Giommi, 1986, 1987). Stimatori rettificati con probabilità di risposta così stimate hanno dato buoni risultati in verifiche sperimentali di tipo Monte Carlo.

7. Conclusioni

L'approccio probabilistico consente di inquadrare formalmente la non risposta nella teoria degli errori campionari, rendendo a tempo stesso esplicite le scelte fatte dal ricercatore per minimizzarne gli effetti negativi. L'aspetto cruciale dell'impostazione è indubbiamente rappresentato dalla stima delle probabilità di risposta. È infatti il momento della stima quello in cui le conoscenze e le scelte soggettive si affiancano agli elementi di natura prevalentemente oggettiva legati al disegno campionario. È un pregio dell'impostazione quello di rendere palesi le assunzioni probabilistiche su cui si basano i metodi più comunemente adottati per fronteggiare la non risposta. La debolezza di tali assunzioni dovrebbe spingere spesso i ricercatori verso soluzioni alternative. Purtroppo ciò non si verifica che raramente, forse per la naturale inerzia che vince chi vede nelle nuove metodologie soprattutto complicazioni formali e di calcolo. In realtà le une e le altre sono più presunte che reali dato l'attuale sviluppo e diffusione dei computers e la conseguente facilità con cui possono essere superati i problemi computazionali.

Bibliografia

Giommi A. (1984). A Simple Method for Estimating Individual Response Probabilities in Sampling from Finite Populations, *Metron*, XLII n. 3-4.

Giommi A. (1985). On Estimation in Nonresponse Situations, *Statistica*, XLV, 1, pp. 57-63.

Giommi A. (1986). Procedimenti non parametrici nella stima delle probabilità individuali di risposta, in *Società Italiana di Statistica, Atti della XXXIII Riunione Scientifica, Bari*, vol. II, pp. 231-239.

Giommi A. (1987). Nonparametrics Methods for Estimating Individual Response Probabilities, *Survey Methodology*, 13, pp. 127-134.

Hartley H.O. (1946). Discussion of Paper by F. Yates, *J.R.S.S.*, pp. 109-137.

Nargundkar M. S. & G. B. Joshi (1975). Nonresponse in Sample Surveys, contributed paper, 40th Session of the I.S.I., pp. 626-628, Warsaw.

Platek R. & G. B. Gray (1979). Methodology and Application of Adjustment for Nonresponse, invited paper 42nd Session of the I.S.I., Manila.

Platek R. & G. B. Gray (1983). Imputation methodology: Total Survey Error, in *Incomplete Data in Sample Surveys*, vol. 2 chapter 16-19, Eds. Olkin I. & Madow W. G., Academic Press.

Platek R. & G. B. Gray (1988). Response Probability Approach to Missing Survey Data, *Mimeo*.

Politz A. & W. Simmons (1949). An attempt to get the "Not at home" into the samples without callbacks, *Journ. of the Amer. Stat. Ass.* 44, pp. 9-31.

Särndal C. E. (1981). Frame Work for Inference in Survey Sampling with Applications to Small Area Estimation and Adjustment for nonresponse, invited paper, 43rd Session of the I.S.I., Buenos Aires

Särndal C. E. & T. K. Hui (1981). Estimation for Nonresponse Situations: to what Extent Must We Rely on Models?, in *Current Topics in Survey Sampling*, Eds. Krewsky D. Platek R. Rao J. N. K., Academic Press.

Särndal C. E., C. M. Cassel, J. H. Wretman (1983). Some Uses of Statistical Models in Connection with the Nonresponse Problem, in *Incomplete Data in Sample Surveys*, vol. 3, Eds. Olkin I. & Madow W. G., Academic Press.

Särndal C. E. (1987). *Inférence statistique et analyse des données sous des plans d'échantillonnage complexes*, Presses de l'Université de Montréal

Materiali di discussione

1. Maria Cristina Marcuzzo [1985] "Joan Violet Robinson (1903-1983)", pp.134.
2. Sergio Lugaresi [1986] "Le imposte nelle teorie del sovrappiù", pp.26.
3. Massimo D'Angelillo e Leonardo Paggi [1986] "PCI e socialdemocrazie europee. Quale riformismo?", pp.158.
4. Gian Paolo Caselli e Gabriele Pastrello [1986] "Un suggerimento hobsoniano su terziario e occupazione: il caso degli Stati Uniti 1960/1983", pp.52.
5. Paolo Bosi e Paolo Silvestri [1986] "La distribuzione per aree disciplinari dei fondi destinati ai Dipartimenti, Istituti e Centri dell'Università di Modena: una proposta di riforma", pp.25.
6. Marco Lippi [1986] "Aggregation and Dynamics in One-Equation Econometric Models", pp.64.
7. Paolo Silvestri [1986] "Le tasse scolastiche e universitarie nella Legge Finanziaria 1986", pp.41.
8. Mario Forni [1986] "Storie familiari e storie di proprietà. Itinerari sociali nell'agricoltura italiana del dopoguerra", pp.165.
9. Sergio Paba [1986] "Gruppi strategici e concentrazione nell'industria europea degli elettrodomestici bianchi", pp.56.
10. Nerio Naldi [1986] "L'efficienza marginale del capitale nel breve periodo", pp.54.
11. Fernando Vianello [1986] "Labour Theory of Value", pp.31.
12. Piero Ganugi [1986] "Risparmio forzato e politica monetaria negli economisti italiani tra le due guerre", pp.40.
13. Maria Cristina Marcuzzo e Annalisa Rosselli [1986] "The Theory of the Gold Standard and Ricardo's Standard Commodity", pp.30.
14. Giovanni Solinas [1986] "Mercati del lavoro locali e carriere di lavoro giovanili", pp.66.
15. Giovanni Bonifati [1986] "Saggio dell'interesse e domanda effettiva. Osservazioni sul capitolo 17 della General Theory", pp.42.
16. Marina Murat [1986] "Between old and new classical macroeconomics: notes on Leijonhufvud's notion of full information equilibrium", pp.20.
17. Sebastiano Brusco e Giovanni Solinas [1986] "Mobilità occupazionale e disoccupazione in Emilia Romagna", pp.48.
18. Mario Forni [1986] "Aggregazione ed esogeneità", pp.13.
19. Sergio Lugaresi [1987] "Redistribuzione del reddito, consumi e occupazione", pp. 17.
20. Fiorenzo Sperotto [1987] "L'immagine neopopulista di *mercato debole* nel primo dibattito sovietico sulla pianificazione", pp. 34.

21. M. Cecilia Guerra [1987] "Benefici tributari del regime misto per i dividendi proposto dalla Commissione Sarcinelli: una nota critica", pp. 9.
22. Leonardo Paggi [1987] "Contemporary Europe and Modern America: Theories of Modernity in Comparative Perspective", pp. 38.
23. Fernando Vianello [1987] "A Critique of Professor Goodwin's 'Critique of Sraffa'", pp. 12.
24. Fernando Vianello [1987] "Effective Demand and the Rate of Profits: Some Thoughts on Marx, Kalecki and Sraffa", pp. 41.
25. Anna Maria Sala [1987] "Banche e territorio. Approccio ad un tema geografico-economico", pp. 40.
26. Enzo Mingione e Giovanni Mottura [1987] "Fattori di trasformazione e nuovi profili sociali nell'agricoltura italiana: qualche elemento di discussione", pp. 36.
27. Giovanna Procacci [1988] "The State and Social Control in Italy During the First World War", pp. 18.
28. Massimo Matteuzzi e Annamaria Simonazzi [1988] "Il debito pubblico", pp. 62.
29. Maria Cristina Marcuzzo (a cura di) [1988] "Richard F. Kahn. A disciple of Keynes", pp. 118.
30. Paolo Bosi [1988] "MICROMOD. Un modello dell'economia italiana per la didattica della politica fiscale", pp. 34.
31. Paolo Bosi [1988] "Indicatori della politica fiscale. Una rassegna e un confronto con l'aiuto di MICROMOD", pp. 25.
32. Giovanna Procacci [1988] "Protesta popolare e agitazioni operaie in Italia 1915-1918", pp. 45.
33. Margherita Russo [1988] "Distretto industriale e servizi. Uno studio dei trasporti nella produzione e nella vendita delle piastrelle", pp. 157.
34. Margherita Russo [1988] "The effects of technical change on skill requirements: an empirical analysis", pp. 28.
35. Carlo Grillenzoni [1988] "Identification, estimation of multivariate transfer functions", pp. 33.
36. Nerio Naldi [1988] "Keynes' concept of capital" pp. 40.
37. Andrea Ginzburg [1988] "Locomotiva Italia?" pp. 30.
38. Giovanni Mottura [1988] "La 'persistenza' secolare. Appunti su agricoltura contadina ed agricoltura familiare nelle società industriali" pp. 40.
39. Giovanni Mottura [1988] "L'anticamera dell'esodo. I contadini italiani dalla 'restaurazione contrattuale' fascista alla riforma fondiaria" pp. 40.
40. Leonardo Paggi [1988] "Americanismo e riformismo. La socialdemocrazia europea nell'economia mondiale aperta" pp. 120.
41. Annamaria Simonazzi [1988] "Fenomeni di isteresi nella spiegazione degli alti tassi di interesse reale" pp. 44.
42. Antonietta Bassetti [1989] "Analisi dell'andamento e della casualità della borsa valori" pp. 12.
43. Giovanna Procacci [1989] "State coercion and worker solidarity in Italy (1915-1918): the moral and political content of social unrest" pp. 41.
44. Carlo Alberto Magni [1989] "Reputazione e credibilità di una minaccia in un gioco bargaining"

pp. 56.

45. Giovanni Mottura [1989] "Agricoltura familiare e sistema agroalimentare in Italia" pp. 84.
46. Mario Forni [1989] "Trend, Cycle and 'Fortuitous Cancellations': a Note on a Paper by Nelson and Plosser" pp. 4.
47. Paolo Bosi, Roberto Golinelli, Anna Stagni [1989] "Le origini del debito pubblico e il costo della stabilizzazione" pp. 26.
48. Roberto Golinelli [1989] "Note sulla struttura e sull'impiego dei modelli macroeconomici" pp. 21.
49. Marco Lippi [1989] "A Short Note on Cointegration and Aggregation" pp. 11.
50. Gian Paolo Caselli and Gabriele Pastrello [1989] "The Linkage between Tertiary and Industrial Sector in the Italian Economy: 1951-1988. From an External Dependence to an Internal One" pp. 40
51. Gabriele Pastrello [1989] "François Quesnay: dal Tableau Zig-Zag al Tableau formule: una ricostruzione" pp. 48
52. Paolo Silvestri [1989] "Il bilancio dello stato" pp. 34
53. Tim Mason [1990] "Tre seminari di Storia Sociale Contemporanea" pp. 26
54. Michele Lalla [1990] "The Aggregate Escape Rate Analysed through the Queueing Model" pp. 23
55. Paolo Silvestri [1990] "Sull'autonomia finanziaria delle Università" pp. 11
56. Paola Bertolini, Enrico Giovannetti [1990] "Uno studio di 'filiera' nell'agroindustria. Il caso del Parmigiano Reggiano" pp. 164
57. Paolo Bosi, Roberto Golinelli, Anna Stagni [1990] "Effetti macroeconomici, settoriali e distributivi dell'armonizzazione dell'IVA" pp. 24
58. Michele Lalla [1990] "Modelling Employment Spells from Emilian Labour Force Data" pp. 18
59. Andrea Ginzburg [1990] "Politica nazionale e commercio internazionale" pp. 22