

LE MANCATE RISPOSTE TOTALI NELL'INDAGINE SUI CONSUMI DELLE FAMIGLIE

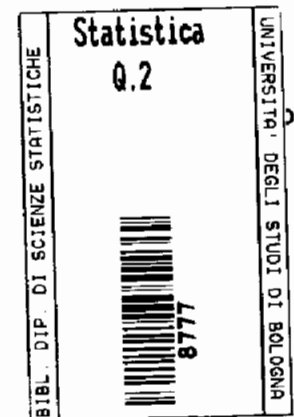
Donato Lucev*
Rapporto di ricerca n.14

CON PRI - La misura dei consumi privati

I lavori raccolti in questa collana hanno avuto origine nell'ambito del progetto di ricerca dell'ISTAT «Le statistiche dei consumi privati nel sistema statistico nazionale» e del progetto di ricerca MURST 40% «La misura dei consumi privati: uno studio sull'accuratezza, coerenza e qualità dei dati». Al progetto di ricerca hanno partecipato i ricercatori dell'ISTAT e dei seguenti Dipartimenti e Istituti universitari:

- Dipartimento di Scienze Statistiche, Bologna
- Dipartimento Statistico, Firenze
- Dipartimento di Scienze Statistiche, Perugia
- Dipartimento di Contabilità Nazionale, Roma
- Istituto di Statistica e Matematica, Istituto Universitario Navale, Napoli
- Istituto di Statistica, Messina.

* Istituto di Statistica e Matematica - Istituto Universitario Navale, Napoli



Dipartimento di Scienze Statistiche "Paolo Fortunati"
dell' Università degli Studi di Bologna
Novembre 1992



INDICE

1. Obiettivi	p.	5
2. I riferimenti alle tematiche d'analisi	"	6
2.1 Gli errori non campionari e le mancate risposte totali	"	6
2.2 Gli studi di identificazione	"	7
2.3 L'indagine sui "Consumi delle famiglie". La formazione del campione	"	9
2.4 Caratteristiche del campione	"	9
3. I dati di base	"	11
3.1 La rilevazione delle mancate risposte totali	"	11
3.2 Il calcolo delle mancate risposte totali in base alla tipologia	"	13
4. L'analisi delle mancate risposte totali	"	14
4.1 Le mancate risposte totali a livello nazionale	"	14
4.2 Le mancate risposte totali a livello regionale	"	17
5. Il trattamento delle mancate risposte totali effettuato dall'Istat	"	20
6. Considerazioni conclusive	"	23
<i>Note</i>	"	26
<i>Riferimenti bibliografici</i>	"	29

1. Obiettivi

Questo lavoro si inserisce nell'ambito della ricerca sui "Consumi privati" che analizza criticamente gli obiettivi e la metodologia dell'indagine Istat sui "Consumi delle famiglie" allo scopo di dare un fondamento adeguato sul piano concettuale ed empirico, ad una sua revisione e, piu' in generale, di suggerire le linee per una nuova valutazione dei consumi privati (Filippucci e Marliani, 1992). Esso si riallaccia ad un primo lavoro riguardante la fase di revisione dell'indagine Istat (Lucev, 1992) e tratta in particolare il problema delle mancate risposte totali.

L'importanza del problema delle mancate risposte totali e degli studi di identificazione ad esse connessi, a cui viene dedicato spazio successivamente, induce ad affrontare il problema delle non risposte con particolare riguardo al momento della revisione in cui esse sono rilevate, che e' essenziale ai fini della identificazione delle caratteristiche dei non rispondenti. Infatti, per i questionari non restituiti all'Istat, o restituiti completamente in bianco, si dispone soltanto delle notizie sulla famiglia contenute nel modello compilato a cura dell'ufficio comunale (W/404) che contiene l'indicazione del capo famiglia, numero dei componenti, indirizzo e ubicazione dell'abitazione. Per i restanti modelli, eliminati nelle diverse fasi della revisione, si dovrebbe disporre di maggiori informazioni essendo compilati totalmente o parzialmente, anche se erroneamente in alcune loro parti.

L'altro aspetto considerato e' quello territoriale in base a diversi livelli di aggregazione: nazionale e regionale, distinguendo nel loro ambito i comuni capoluoghi di provincia, quelli non capoluoghi con piu' di 50 mila abitanti, ed i restanti comuni campionati.

Il riferimento comunale, come e' noto, costituisce il primo stadio di campionamento dell'indagine e la distinzione dei comuni, effettuata, corrisponde al criterio di stratificazione dei comuni stessi ai fini della formazione del campione.

In questo lavoro si e' voluto indagare se tale suddivisione abbia o meno una certa influenza sul manifestarsi delle mancate risposte totali, cioe' se queste ultime siano in qualche modo connesse alla diversa ampiezza demografica dei comuni interessati all'indagine.

Inoltre, si e' considerato l'aspetto temporale delle mancate risposte totali, distinguendole secondo il trimestre di effettuazione dell'indagine. Si e' ritenuto, infatti, che anche l'aspetto temporale possa influenzare le mancate risposte totali. Si pensi per esempio al fatto che in particolari periodi dell'anno si puo' avere un incremento o decremento delle risposte per cause stagionali come le vacanze estive o invernali, durante le quali maggiori sono le difficolta' di reperire le famiglie.

Nel corso dell'analisi si e' proceduto anche ad incrociare fra di loro le singole modalita' territoriali e temporali al fine di evidenziare il loro effetto interattivo nelle varie fasi di produzione di mancate risposte totali.

A livello regionale, invece, si e' fatto ricorso ad analisi statistiche multivariate (componenti principali e corrispondenze), piu' adatte a descrivere in maniera sintetica le relazioni intercorrenti tra le caratteristiche territoriali e la tipologia delle mancate risposte (Lebart e altri, 1982).

L'analisi esposta e' preceduta da una rassegna di alcuni problemi generali connessi alla mancata risposta (errori non campionari, studi di identificazione, ecc.), ed e' completata da una rassegna dei metodi di trattamento delle mancate risposte totali utilizzati dall'Istat, fcosi' come risultano dall'analisi dei documenti interni ed ufficiali dell'Istat.

2. I riferimenti alle tematiche d'analisi

2.1 - Gli errori non campionari e le mancate risposte totali

La teoria dei campioni si e' occupata prevalentemente dell'errore campionario ottenendo notevoli risultati. Meno sviluppati sono stati, invece, gli studi sugli errori non campionari¹ che, specie, negli ultimi anni, hanno attirato sempre piu' l'attenzione di studiosi e di enti pubblici e privati produttori di indagini statistiche.

Cio' e' dovuto dalla constatazione che, in genere, gli errori non campionari risultano molto piu' ampi di quelli campionari e che, la riduzione dell'errore campionario ottenuta tramite un ampliamento del campione, spesso comporta addirittura un aumento dell'errore non campionario che neutralizza gli sforzi, in termini di tempi e di costi, attuati per ridurre il primo tipo di errore (Vajani, 1969).

Tra gli errori non campionari un posto di rilievo e' occupato dalle mancate risposte totali, intese in senso lato, ossia, comprensive sia delle unita' che sono irreperibili o si rifiutano di collaborare all'indagine, sia dei questionari eliminati in fase di controllo per incompletezza o incoerenza delle risposte fornite².

Un primo parametro di valutazione della bonta' dell'indagine, il cosiddetto tasso di risposta (rapporto tra il numero di rispondenti ed il numero delle unita' campionate), tiene conto indirettamente dell'entita' delle mancate risposte totali.

Un basso valore di questo parametro puo' indicare due ordini di problemi:

a- si creano gravi distorsioni nelle stime campionarie, dovute al fatto

che il campione effettivamente osservato si riduce ad una parte di quello programmato;

b- le stime ricavate dai soli rispondenti, il piu' delle volte, non sono estendibili a tutta la popolazione perche', spesso, gli atteggiamenti e i comportamenti dei non rispondenti si diversificano rispetto a quelli dei rispondenti.

Di conseguenza emerge l'esigenza di controllare e ridurre l'entita' delle mancate risposte totali e, persistendo la loro presenza, di effettuare degli "studi di identificazione" per individuare comportamenti e atteggiamenti diversificati tra rispondenti e non rispondenti al fine di proporre il metodo piu' idoneo al trattamento delle mancate risposte totali³.

In base a questa duplice problematica le procedure poste in essere si distinguono in preventive e successive.

Quelle "preventive", sono attuate, di solito, prima o durante le operazioni di raccolta dei dati e mirano ad aumentare la propensione delle unita' campionate a partecipare all'indagine.

Naturalmente la propensione alla collaborazione e' funzione degli specifici contenuti dell'indagine oltre agli altri generici fattori quali: il tipo di intervista (diretta, postale, telefonica, etc.), la struttura del questionario, il modo in cui sono formulate le domande, la capacita' del rilevatore di attirare l'interesse degli intervistati sul tema dell'indagine, i tempi ed i luoghi dell'indagine stessa, etc..

Le procedure "successive", attuate nella fase di revisione e di elaborazione dei dati, sono finalizzate alla rilevazione dell'entita' delle mancate risposte ed alla conoscenza delle caratteristiche dei non rispondenti (studi di identificazione), allo scopo di adottare particolari procedure dette di adeguamento⁴ (*accomodation*) che consentano di correggere, almeno in parte, le distorsioni delle stime provocate dalle mancate risposte totali nell'inferire i risultati campionari alla popolazione.

Nonostante l'impegno profuso dagli organizzatori, in ogni indagine le mancate risposte totali sono ineliminabili e il problema principale resta quello di sottoporle ad analisi per poter scegliere il trattamento piu' idoneo, tra i vari possibili. Per questo motivo nel prosieguo si prestera' maggiore attenzione alle tecniche di tipo successivo ed in particolar modo agli studi di identificazione.

2.2- Gli studi di identificazione

Il problema delle mancate risposte totali ed in particolare le conseguenze che esse comportano e' stato affrontato da numerosi autori, fra cui Cochran

(1977), Kish (1965) e Sukhatme e Sukhatme (1970), i quali pur utilizzando una terminologia diversa, pervengono a risultati molto simili nel calcolare l'errore dovuto al *bias* delle stime impiegando solo le informazioni provenienti dai rispondenti. Questi autori, inoltre, hanno evidenziato l'esistenza di uno stretto legame tra la propensione a rispondere e l'oggetto dell'indagine che determina forse la componente piu' importante di tale distorsione.

In assenza degli studi di identificazione, lo studioso dispone soltanto delle informazioni relative al piano di campionamento e al numero di mancate risposte totali. In tale caso, la maggior parte dei metodi di adeguamento ha per obiettivo la ricostituzione della struttura della popolazione all'interno del campione.

Specie per le indagini che si ripetono nel tempo si avverte in particolare l'esigenza di indagare se esistano o meno significative divergenze di comportamento tra rispondenti e non rispondenti, tramite opportuni studi di identificazione che permettano di evidenziare se i gruppi dei rispondenti e dei non rispondenti rappresentano entrambi campioni casuali della popolazione di riferimento, oppure se i due gruppi presentano caratteristiche strutturali particolari correlate con la variabile oggetto di studio. Gli studi di identificazione possono utilizzare ulteriori informazioni disponibili a seconda anche del tipo di intervista:

a- informazioni ausiliarie, anche esterne all'indagine, disponibili per entrambi i gruppi, assumendo che non esista tra loro differenza, qualora non sussista differenza nei relativi dati ausiliari;

b- informazioni derivanti da diversi solleciti, tipici delle indagini postali o telefoniche, in cui si assume che i rispondenti dopo il t-mo sollecito diventino sempre piu' simili ai non rispondenti al crescere di t (Hawkins, 1975).

Negli studi che utilizzano informazioni ausiliarie, si distingue il metodo diretto e quello indiretto. Con il metodo indiretto si cerca di provare l'esistenza di differenze rilevanti tra i rispondenti e la popolazione, da cui il campione e' estratto, in base a variabili note per gli elementi della popolazione. I risultati del confronto fra rispondenti e popolazione, positivi o negativi, sono estesi al confronto tra rispondenti e non rispondenti. I difetti di questo metodo sono relativi al fatto che non sempre i dati della popolazione sono aggiornati o comunque attendibili.

Il metodo diretto, invece, utilizza un campione casuale di non rispondenti allo scopo di ottenere informazioni da confrontare direttamente con quelle dei rispondenti. Anche questo metodo presenta alcuni inconvenienti che riguardano i tempi lunghi, i costi elevati, il diverso sistema di raccolta dei dati ed i cambiamenti che possono intervenire tra la prima e la seconda indagine.

2.3- L'indagine sui "Consumi delle famiglie". La formazione del campione

L'indagine sui "Consumi delle famiglie" e' di tipo campionario. Lo schema di campionamento riflette quasi completamente quello della indagine sulle forze lavoro (Fola) dell'Istat, e si concretizza in un campionamento a doppio stadio: nel primo avviene la selezione dei comuni, nel secondo quella delle famiglie.

La selezione dei comuni viene effettuata tramite un duplice procedimento: si inseriscono nel campione tutti i comuni capoluoghi di provincia e quelli con popolazione superiore a 50 mila abitanti, identificati come comuni di tipo AR (autorappresentativi), mentre i restanti comuni sono campionati nell'ambito di un disegno stratificato⁵ all'interno di ciascuna regione. Da ogni strato risultante vengono, poi, selezionati tre comuni dopo averli ordinati in base alla loro popolazione; i comuni di questo secondo gruppo sono identificati di tipo NAR (non autorappresentativi).

Il campione di primo stadio cosi' formato viene ad essere costituito da circa 550 comuni, di cui 150 sono di tipo AR e circa 400 sono di tipo NAR. Le famiglie (circa 39.000) sono estratte dalle anagrafi dei comuni selezionati con estrazione sistematica e rappresentano il secondo stadio del campione.

L'indagine viene effettuata trimestralmente. Il riferimento annuale della stessa comprende quattro indagini a partire da quella effettuata nel primo trimestre (luglio, agosto e settembre) fino all'indagine del quarto trimestre (aprile, maggio e giugno) dell'anno successivo (Istat 1990).

Nell'ambito di ciascun trimestre la periodicitá di rilevazione si differenzia secondo il tipo di comune. Nei comuni di tipo AR la rilevazione e' effettuata ogni mese, mentre in quelli di tipo NAR si tiene conto della posizione di ciascun comune nell'ambito dello strato di appartenenza: i comuni che occupano la prima posizione effettuano la rilevazione nel primo mese del trimestre, quelli in seconda posizione la effettuano nel secondo mese ed infine quelli in terza posizione la effettuano nel terzo mese⁶.

Una descrizione piu' dettagliata riguardo la metodologia dell'indagine e' in Innocenzi (1992) e Filippucci e Marliani (1992).

2.4- Caratteristiche del campione

Nel seguito si fara' riferimento all'indagine sui "Consumi delle famiglie" relativa all'anno 1989-90. Essa ha interessato circa quarantamila famiglie (39.238) distribuite in 549 comuni. Onde avere un riferimento quantitativo a cui rapportare le caratteristiche delle mancate risposte totali, si e' reso necessario classificare i comuni e le famiglie⁷ (il cui numero sottintende quello

dei questionari) per regione e per tipo di comune (AR e NAR, distinguendo i comuni AR in capoluoghi e non capoluoghi e quelli NAR in base al mese di rilevazione nell'ambito del trimestre). Quest'ultimo livello riflette la periodicità di rilevazione nell'ambito del trimestre: ogni mese per i comuni AR e in un solo mese per i comuni NAR.

Le tavv. 1 e 2 riportano le distribuzioni dei comuni e delle famiglie (questionari) per regione e per tipo di comune.

Tav. 1- Distribuzione dei comuni campionati per regione, per tipo e per periodicità di rilevazione, nel periodo d'indagine 1989-90. Totale regionale = 100.

Regioni	Comuni AR		Comuni NAR			Totale
	Capoluoghi	Non capoluoghi	1 mese	2 mese	3 mese	
Piemonte	14,29	7,14	26,19	26,19	26,19	42
Valle d'A.	7,69	0,00	30,77	30,77	30,77	13
Lombardia	18,18	10,91	23,64	23,64	23,64	55
Trentino	6,90	0,00	31,03	31,03	31,03	29
Veneto	21,88	3,13	25,00	25,00	25,00	32
Friuli	30,77	0,00	23,08	23,08	23,08	13
Liguria	20,00	5,00	25,00	25,00	25,00	20
Emilia	21,62	13,51	21,62	21,62	21,62	37
Toscana	26,47	11,76	20,59	20,59	20,59	34
Umbria	13,33	6,67	26,67	26,67	26,67	15
Marche	20,00	5,00	25,00	25,00	25,00	20
Lazio	19,23	11,54	23,08	23,08	23,08	26
Abruzzi	21,05	0,00	26,32	26,32	26,32	19
Molise	10,00	0,00	30,00	30,00	30,00	20
Campania	11,36	27,27	20,45	20,45	20,45	44
Puglia	14,71	23,53	20,59	20,59	20,59	34
Basilicata	14,29	0,00	28,57	28,57	28,57	14
Calabria	11,54	7,69	26,92	26,92	26,92	26
Sicilia	25,00	8,33	22,22	22,22	22,22	36
Sardegna	20,00	5,00	25,00	25,00	25,00	20
Totale	17,49	9,28	24,41	24,41	24,41	549

La maggiore o minore concentrazione della popolazione nei numerosi comuni italiani si rispecchia nel differente profilo di comuni, AR e NAR, inclusi nel campione di primo stadio, sia a livello nazionale che regionale.

Pur in presenza di una disparità tra numero di comuni AR e NAR, le famiglie selezionate nel secondo stadio di campionamento, e quindi, i relativi questionari, si equiripartiscono tra comuni AR (48,5%) e comuni NAR (51,5%).

A livello regionale, nel Lazio si rilevano le incidenze più elevate di famiglie selezionate nei comuni AR e nei capoluoghi di provincia, mentre solo in Campania e in Puglia le incidenze delle famiglie dei comuni non capoluoghi superano quelle dei capoluoghi.

Tav. 2- Distribuzione dei questionari per regione, per tipo di comune e per periodicità di rilevazione, nel periodo d'indagine 1989-90. Totale regionale = 100.

Regioni	Comuni AR		Comuni NAR			Totale
	Capoluoghi	Non capoluoghi	1 mese	2 mese	3 mese	
Piemonte	35,47	9,73	18,24	18,24	18,31	2960
Valle d'Aosta	32,25	0,00	22,42	22,42	22,91	1017
Lombardia	37,06	17,32	15,19	15,19	15,24	3950
Trentino	24,23	0,00	25,23	25,18	25,37	2097
Veneto	36,67	3,69	19,95	19,69	19,69	1950
Friuli	41,99	0,00	19,22	19,22	19,57	1124
Liguria	51,96	11,91	11,91	11,97	12,24	1813
Emilia	42,86	17,46	13,23	13,23	13,23	2268
Toscana	42,46	16,45	13,67	13,67	13,75	2633
Umbria	32,13	9,53	19,06	19,06	20,21	1133
Marche	34,99	6,96	19,33	19,33	19,39	1552
Lazio	67,08	7,57	8,45	8,42	8,49	2852
Abruzzi	37,55	0,00	20,82	20,82	20,82	1153
Molise	21,95	0,00	25,93	26,02	26,10	1157
Campania	25,89	32,16	13,97	14,01	13,97	2920
Puglia	23,42	24,98	17,17	17,22	17,22	2306
Basilicata	20,55	0,00	26,42	26,51	26,51	1090
Calabria	21,97	14,60	21,10	21,10	21,23	1479
Sicilia	41,62	9,75	16,18	16,22	16,22	2595
Sardegna	36,33	6,06	19,18	19,18	19,26	1189
Totale	36,80	11,66	17,14	17,14	17,25	39238

3. I dati di base

3.1- La rilevazione delle mancate risposte totali.

Le mancate risposte totali, nel caso dell'indagine sui consumi dell'Istat, possono scaturire da distinte situazioni quali:

- a - gli uffici comunali non mettono in atto gli adempimenti relativi all'indagine;
- b - irreperibilità o rifiuto di collaborare da parte delle famiglie selezionate e impossibilità di trovare una sostituzione;
- c - le famiglie consegnano il questionario completamente in bianco, pur avendo manifestato la disponibilità a collaborare;
- d - i questionari inviati dal comune pervengono all'Istat quando già è avviata l'elaborazione trimestrale dei dati, per cui non è possibile procedere alla loro informatizzazione;
- e - i questionari sono eliminati nella fase di revisione per

incompletezza o incoerenza delle risposte fornite.

L'importanza di tale distinzione, ai fini degli studi di identificazione, e' legata alla individuazione delle caratteristiche dei non rispondenti e dei rispondenti e quindi, successivamente, alla scelta del metodo di adeguamento piu' appropriato. Ad ogni situazione corrisponde una diversa disponibilita' di informazioni.

Si puo' andare dall'assenza totale di informazioni, infatti, nel caso di inadempimento operativo da parte dell'ufficio comunale, alla disponibilita' di poche informazioni di fonte anagrafica, nel caso di irreperibilita' o rifiuto di collaborare da parte delle famiglie o di questionari in bianco, fino ad una maggiore disponibilita', che varia da caso a caso, per i questionari spediti ma pervenuti in ritardo oppure eliminati in fase di revisione.

La rilevazione delle mancate risposte totali avviene nella fase di revisione successiva alla raccolta dei questionari. In base alle operazioni compiute in tale fase viene distinta la revisione quantitativa da quella qualitativa (Masselli, 1992).

Le operazioni compiute nella revisione quantitativa fanno riferimento al riscontro effettuato per ogni comune tra il campione programmato e quello effettivamente osservato, e la differenza costituisce il primo nucleo di mancate risposte totali e corrisponde al numero di questionari non pervenuti. A questi vanno aggiunti i questionari pervenuti completamente in bianco, rilevati prioritariamente nella revisione qualitativa.

La revisione qualitativa si concretizza essenzialmente nella revisione manuale e revisione automatica. Nella revisione manuale le operazioni di controllo riguardano principalmente: la corrispondenza delle informazioni con quanto previsto dal piano di codifica, il rispetto delle istruzioni nella compilazione dei modelli sia da parte delle famiglie che del rilevatore; l'esatto riporto delle spese dai libretti al modello riepilogativo effettuato dal rilevatore; l'esatta attribuzione dei codici alle voci di spesa; infine, almeno per i casi di dubbio, il confronto tra la spesa indicata dalla famiglia con la spesa media per voce di consumo predisposta dall'Istat in relazione ad un gruppo di variabili relative alle famiglie (la ripartizione territoriale, il numero di componenti, posizione nella professione del capofamiglia).

Qualora le informazioni siano incomplete o incoerenti in modo tale da inficiare la validita' del questionario si procede alla sua eliminazione generando una mancata risposta totale.

Superata la fase di revisione manuale, i questionari sono pronti per essere informatizzati e trasferiti su nastro magnetico e, successivamente, sottoposti a controllo automatico che avviene tramite un programma di elaborazione alquanto complesso che comprende diverse procedure (Innocenzi, 1989).

Per grandi linee, il programma consta di due procedure, una riguardante i dati mensili e l'altra i dati trimestrali.

La procedura mensile si puo', a sua volta, suddividere in due parti: la procedura attuata tramite le "griglie di compatibilita'" e la procedura di revisione ordinale.

Le informazioni mensili, registrate su nastro magnetico, sono sottoposte a controllo tramite griglie di compatibilita', allo scopo di: correggere l'errore in base a prefissate regole d'intervento; segnalare l'errore affinche' ne sia verificata la compatibilita'⁸.

Qualora non si possa procedere a correzione e l'errore sia tale da compromettere la validita' delle risposte, il questionario viene eliminato dando luogo a mancata risposta totale.

I questionari non eliminati sono, infine, controllati ordinandoli in base alla spesa complessiva della famiglia (revisione ordinale). In questa fase il controllo viene attuato, al fine di verificarne la coerenza, tramite una supervisione manuale dei questionari ricadenti nelle code della distribuzione, prendendo in considerazione prefissati decili,⁹ che fanno riferimento ai valori piu' bassi o piu' elevati di spesa familiare. Anche in questa fase e' possibile che i questionari vengano eliminati aumentando in tal modo il numero delle mancate risposte totali.

Infine, si procede all'aggregazione trimestrale partendo dal file "pulito" dei dati mensili.

La procedura automatica sui dati trimestrali prevede il trattamento per mancata copertura, l'aggiustamento del campione ed il calcolo dei coefficienti di riporto all'universo. Essa e' descritta con maggiore dettaglio nel paragrafo 5.

Le stime trimestrali sono, infine, sottoposte ad un controllo di coerenza con le corrispondenti stime socio-demografiche dell'indagine sulle forze lavoro, mentre i diversi capitoli di spesa sono confrontati con le stime trimestrali della Contabilita' Nazionale.

3.2- Il calcolo delle mancate risposte totali in base alla tipologia

Per ciascun comune campionato, al fine di ottenere le mancate risposte distinte in base alla loro tipologia, si e' rilevato:

- il numero teorico dei questionari attribuito dal piano di campionamento, quale numero teorico provvisorio (TP), in quanto vi sono alcuni comuni che restituiscono all'Istat un numero di questionari (RP) superiore al numero teorico ad essi attribuito. Tale incremento, sommato al numero teorico provvisorio, definisce per ogni comune il numero teorico modificato (TM);

- il numero di questionari pervenuti (QP). La differenza tra il numero teorico modificato e il numero dei questionari pervenuti permette di calcolare, per comune, il numero di questionari non pervenuti (NP);

- il numero dei questionari eliminati nella fase di revisione manuale (RM);

- il numero dei questionari eliminati nella fase di revisione ordinale (RO).

Come si puo' notare, non si dispone del numero di questionari eliminati nella fase di revisione automatica (RA). Fortunatamente, pero', sono reperibili informazioni supplementari che consentono di ricavare indirettamente tale numero, almeno a livello trimestrale. In un primo momento occorre predisporre le informazioni per ricavare il totale delle mancate risposte (MR). Si parte dal numero teorico modificato (TM) a cui e' aggiunto il numero di questionari creati per mancata copertura (CR) e la somma costituisce il numero teorico effettivo (TE) attribuito a ciascun comune. Inoltre, sempre per comune, si e' rilevato il numero di questionari utilizzati (UT) per produrre i risultati campionari finali. La differenza tra il numero teorico effettivo e il numero di questionari utilizzati fornisce l'entita' delle mancate risposte totali (MR).

A questo punto e' semplice ottenere il numero dei questionari eliminati nella fase di revisione automatica per differenza tra le mancate risposte totali (MR) e la somma dei questionari non pervenuti (NP), di quelli eliminati nella revisione manuale (RM) e nella revisione ordinale (RO), completando cosi' il quadro delle mancate risposte per comune, per trimestre e per tipologia.

Formalmente, l'aspetto quantitativo delle mancate risposte totali e' valutato tramite la relazione:

$$MR = TE - NP$$

mentre l'aspetto qualitativo che fa riferimento ai diversi momenti della fase di revisione in cui esse sono rilevate, e' valutato tramite:

$$MR = NP + RA + RM + RO.$$

4. L'analisi delle mancate risposte totali.

4.1- Le mancate risposte totali a livello nazionale.

Prima di entrare nel merito delle analisi effettuate con riferimento alle caratteristiche descritte e' opportuno fornire un quadro d'insieme delle mancate risposte totali, rilevate nell'anno 1989-90 e classificate in base alle cause che le ha prodotte.

L'indagine ha interessato 39.238 famiglie e sono state rilevate 5.818 mancate risposte totali, pari al 15%, che costituisce un'aliquota abbastanza elevata in grado, certamente, di produrre significativi effetti distorsivi sull'entita' delle stime campionarie. Di queste, la parte piu' cospicua e' rappresentata dai questionari non pervenuti ed e' pari ai 3/4 delle mancate risposte totali e a circa l'11% delle famiglie campionate. Il restante 25% e' relativo a questionari eliminati nelle varie fasi di revisione (Tav. 3) e precisamente in quella manuale (4,4%), automatica (5,4%) e ordinale (12,4%).

Quanto al tipo di comuni, l'incidenza delle mancate risposte totali (MR/TE) risulta maggiore in quelli di tipo NAR, con il 16,8% contro il 12,7% degli AR. Questi ultimi, a loro volta, registrano una considerevole differenziazione tra comuni capoluoghi (8,5%) e non capoluoghi (17,0%).

Le mancate risposte totali in base alla tipologia non presentano differenze rilevanti tra comuni AR e comuni NAR, eccetto una maggiore incidenza nei comuni NAR del rapporto NP/MR. All'interno, dei comuni AR, come si e' detto sopra, si riscontra un andamento del tutto opposto in relazione alle cause delle mancate risposte (non pervenuti o eliminati). Nei comuni capoluoghi di provincia le incidenze dei questionari eliminati nelle fasi di revisione (RM/MR, RA/MR, RO/MR) sono nettamente superiori alle stesse dei comuni non capoluoghi, prevalendo in entrambi l'incidenza degli eliminati nella fase di revisione ordinale, mentre i questionari non pervenuti (NP/MR) si registrano maggiormente nei comuni non capoluoghi (Tav. 3).

Tav. 3- Rapporti caratteristici sulla "non risposta" per tipo di comune, nel periodo d'indagine 1989-90.

Tipo comuni	NP/MR	RM/MR	RA/MR	RO/MR	NP/TE	MR/TE
Comuni AR	77,95	4,30	5,63	12,12	9,91	12,71
-capoluoghi	68,51	5,86	9,11	16,52	5,83	8,51
-non capol.	87,71	2,69	2,02	7,58	22,77	16,96
Comuni NAR	77,65	4,47	6,26	12,65	13,06	16,81
Totale	77,76	4,40	5,41	12,43	11,53	14,83

Legenda sigle: NP = questionari non pervenuti, MR = mancate risposte totali, RM = questionari eliminati nella revisione manuale, RA = questionari eliminati nella revisione automatica, RO = questionari eliminati nella revisione ordinale, TE = questionari teorici effettivi.

Il riferimento delle mancate risposte al trimestre in cui si verificano, distinguendole anche in base alla tipologia, sicuramente contribuisce ad una migliore comprensione delle cause da cui, esse, possono scaturire. Si rileva, infatti, un andamento decrescente dell'entita' delle mancate risposte

totali nei primi tre trimestri, dal 16% al 12%, aumentando notevolmente (18%) nel quarto trimestre. Tale incremento e' da attribuirsi principalmente all'aumento di questionari eliminati nella fase di revisione ordinale, che abbassa l'incidenza delle altre tipologie di mancate risposte, in quanto l'incidenza dei questionari non pervenuti sul numero programmato di interviste si discosta di poco dalla media degli altri tre trimestri (Tav. 4).

Tav. 4- Rapporti caratteristici sulla "non risposta" per trimestre, nella totalita' dei comuni, nel periodo d'indagine 1989-90.

Trimestre	NP/MR	RM/MR	RA/MR	RO/MR	NP/TE	MR/TE
Lug.- Sett. 89	80,69	2,95	6,35	10,01	12,59	15,61
Ott.- Dic. 89	79,51	4,97	6,18	9,34	10,76	13,54
Gen.- Mar. 90	80,08	5,17	5,08	9,67	9,78	12,21
Apr.- Giu. 90	72,32	4,71	4,25	18,72	12,99	17,96
Totale	72,32	4,40	5,41	12,43	11,53	14,83

Una prima ipotesi sul valore anomalo dei questionari eliminati nella revisione ordinale del quarto trimestre potrebbe riguardare l'approssimarsi delle ferie del periodo estivo, che si riflette in un eccesso o difetto di spesa incoerente nel contesto globale del questionario. Un ulteriore valore anomalo, inferiore alla media degli altri trimestri, e' quello relativo ai questionari eliminati nella fase di revisione manuale del primo trimestre (luglio-settembre). Anche in questo caso una prima spiegazione potrebbe essere collegata alle ferie estive dei revisori.

Tav. 5- Rapporti caratteristici sulla "non risposta" per trimestre, nei comuni AR, nel periodo d'indagine 1989-90.

Trimestre	NP/MR	RM/MR	RA/MR	RO/MR	NP/TE	MR/TE
Lug.- Sett. 89	79,84	2,39	7,03	10,74	12,71	15,91
Ott.- Dic. 89	80,07	4,61	5,91	9,41	9,14	11,42
Gen.- Mar. 90	79,25	6,08	5,87	8,81	7,93	10,01
Apr.- Giu. 90	72,08	4,97	3,57	18,48	9,87	13,52
Totale	77,95	4,40	5,63	12,12	9,91	12,71

Infine, si e' analizzata la tipologia delle mancate risposte per trimestre in relazione sia ai comuni AR che NAR (Tavv. 5 e 6). Nei comuni di tipo AR si abbassa l'incidenza delle mancate risposte relative al quarto trimestre, a causa della diminuzione dei questionari non pervenuti, fermo restando sia

l'incidenza dei questionari eliminati nella revisione ordinale sia l'incidenza dei questionari eliminati nella fase di revisione manuale del primo trimestre.

Nei comuni di tipo NAR le mancate risposte totali aumentano nel quarto trimestre, in corrispondenza di un incremento di questionari non pervenuti, mentre si rileva un leggero incremento dei questionari eliminati nella fase di revisione manuale del primo trimestre.

Tav. 6- Rapporti caratteristici sulla "non risposta" per trimestre, nei comuni NAR, nel periodo d'indagine 1989-90.

Trimestre	NP/MR	RM/MR	RA/MR	RO/MR	NP/TE	MR/TE
Lug.- Sett. 89	81,52	3,49	5,68	9,30	12,49	15,32
Ott.- Dic. 89	79,11	5,22	6,37	9,30	12,28	15,53
Gen.- Mar. 90	80,64	4,56	4,56	10,24	11,52	14,29
Apr.- Giu. 90	71,94	4,56	4,65	18,86	15,93	22,14
Totale	77,65	4,47	5,26	12,65	13,06	16,81

4.2- Le mancate risposte totali a livello regionale

L'analisi delle mancate risposte totali a livello regionale (tav. 7) e' legata essenzialmente da un lato alla stratificazione dei comuni che, come e' noto, avviene in tale ambito e dall'altro, come ovvia conseguenza, alle procedure di adeguamento della struttura del campione alla popolazione che sono espletate facendo riferimento a tale livello territoriale.

Per avere un quadro descrittivo piu' sintetico e maggiormente informativo di quello che potrebbe scaturire dalla osservazione della tav. 7, si e' proceduto ad effettuare un'analisi in componenti principali assumendo come dati di base quelli di detta tavola.

In base allo schema classico della tabella dei dati nelle analisi di tipo fattoriale, le osservazioni sono costituite dalle regioni, mentre le variabili sono relative ai rapporti caratteristici delle mancate risposte totali e precisamente i quozienti che si riferiscono all'aspetto dimensionale:

1- NP/TE, questionari non pervenuti sul totale dei questionari teorici effettivi,

2- MR/TE, mancate risposte totali su totale dei teorici effettivi, ed i quozienti che fanno maggiormente riferimento alla composizione tipologica delle mancate risposte totali:

3- NP/RM, questionari non pervenuti su totale mancate risposte,

4- RM/MR, questionari eliminati nella revisione manuale su totale mancate risposte,

5- RA/MR, questionari eliminati nella revisione ordinale su totale mancate risposte,

6- RO/MR, questionari eliminati nella revisione ordinale su totale mancate risposte.

Tav. 7- Rapporti caratteristici sulla "non risposta" per regione, nel periodo d'indagine 1989-90.

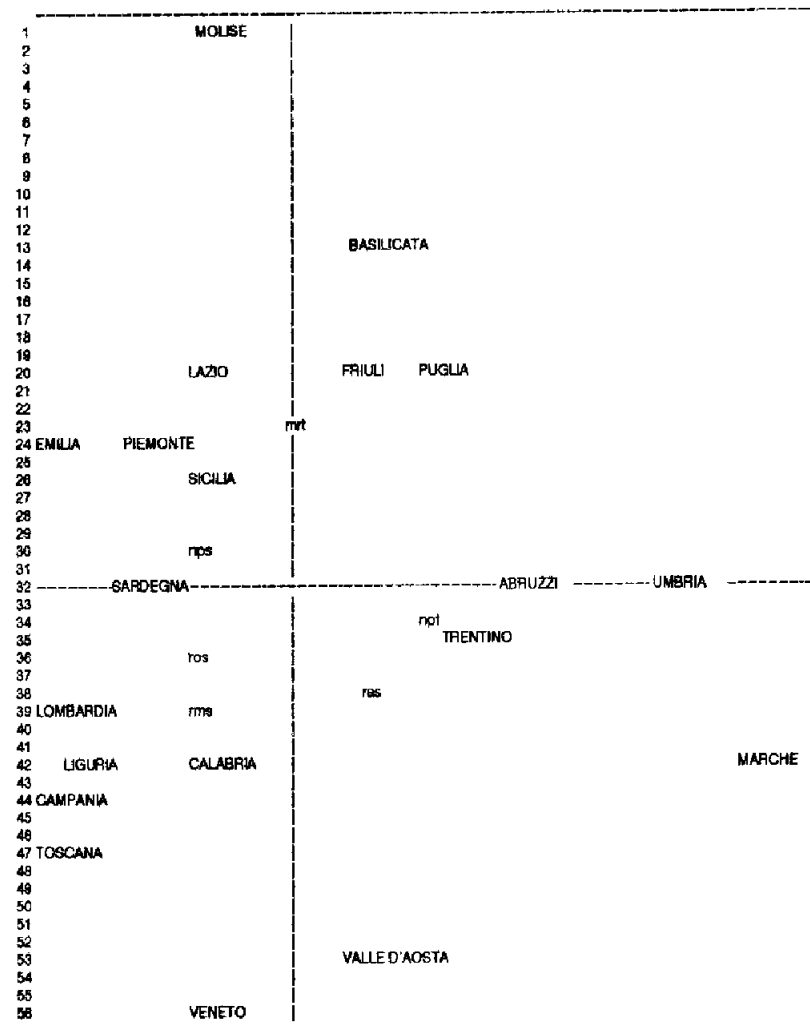
Regioni	NP/MR	RM/MR	RA/MR	RO/MR	NP/TE	MR/TE
Piemonte	92,71	3,10	2,48	1,71	20,20	21,79
Valle d'Aosta	71,76	12,94	15,29	0,00	6,00	8,36
Lombardia	85,62	5,95	2,31	6,12	13,11	15,32
Trentino	45,83	8,33	16,67	29,17	1,57	3,42
Veneto	71,68	13,87	4,05	10,40	6,36	8,87
Friuli	79,41	1,47	17,65	1,47	4,80	6,05
Liguria	75,86	2,96	9,36	11,82	8,49	11,20
Emilia	89,50	1,83	4,57	4,11	8,64	9,66
Toscana	83,97	7,89	3,31	4,83	12,53	14,93
Umbria	41,23	21,05	16,67	21,05	4,15	10,06
Marche	15,00	10,00	25,00	50,00	0,19	2,29
Lazio	77,16	4,63	5,56	12,65	8,77	11,36
Abruzzi	49,04	4,42	15,93	26,55	4,42	9,02
Molise	89,05	2,26	1,89	7,17	21,09	23,68
Campania	82,87	0,58	4,05	12,50	24,52	29,59
Puglia	77,15	0,21	11,32	11,32	15,96	20,69
Basilicata	44,07	3,95	7,91	44,07	7,16	16,24
Calabria	60,39	5,49	3,14	30,98	10,41	17,24
Sicilia	71,72	5,53	2,69	20,06	17,50	24,39
Sardegna	80,53	2,65	5,31	11,50	7,65	9,50
Totale	77,76	4,40	5,41	12,43	11,53	14,83

Considerando il piano formato dai primi due assi fattoriali, a cui resta associato il 75% della variabilita' della tavola (fig. 1), si rileva che:

- lungo il primo asse, si collocano da una parte le regioni: Marche, Umbria, Abruzzi e Trentino caratterizzate da basse incidenze di questionari non pervenuti sul totale previsto e da elevate incidenze di questionari eliminati nella fase di revisione automatica, dall'altra le regioni: Emilia, Lombardia, Campania, Toscana, Liguria, Piemonte e Sicilia in cui ad elevate incidenze di questionari non pervenuti sul totale programmato fa riscontro una bassa incidenza di questionari eliminati nelle singole fasi di revisione;

- lungo il secondo asse, si posizionano da una parte le regioni: Molise, Basilicata, Lazio, Friuli e Puglia caratterizzate da alte incidenze di mancate risposte totali e da ancora basse incidenze di questionari eliminati nelle tre fasi di revisione, dal lato opposto le regioni: Veneto, Valle d'Aosta con basse incidenze di mancate risposte totali e il gruppo formato da Toscana, Campania, Liguria, Calabria e Lombardia che presentano basse incidenze di

FIG. 1. Analisi in componenti principali. Mancate risposte totali per tipologia e p... regioni. Assi 1 E 2.



LEGENDA per le sigle delle variabili.

- mrt = Mancate risposte totali sui questionari teorici effettivi (MR/TE)
- npt = Questionari non pervenuti sul totale dei teorici effettivi (NP/TE)
- nps = Questionari non pervenuti sul totale delle mancate risposte (NP/MR)
- rms = Questionari eliminati nella revisione manuale sul totale delle mancate risposte (RM/MR)
- ras = Questionari eliminati nella revisione automatica sul totale delle mancate risposte (RA/MR)
- ros = Questionari eliminati nella revisione ordinale sul totale delle mancate risposte (RO/MR)

questionari eliminati nella fase di revisione.

In sintesi, dall'analisi del piano fattoriale si configurano essenzialmente tre tipologie ben delineate di mancate risposte totali a cui fanno riferimento gruppi di regioni con caratteristiche strutturali di urbanizzazione e demografiche molto simili:

1- elevata incidenza di mancate risposte totali e bassa incidenza di questionari eliminati nelle fasi di revisione. Si identificano in questa configurazione le regioni piu' popolate e con piu' comuni di tipo AR (Emilia, Lombardia, Campani, Toscana, Liguria etc);

2- bassa incidenza di mancate risposte totali ed elevate incidenze di questionari eliminati durante la revisione. Fanno capo a questo gruppo un nucleo di piccole regioni, quali: Marche, Umbria, Abruzzi e Trentino;

3- bassa incidenza di mancate risposte totali ed incidenze elevate in un solo rapporto caratteristico rilevate in: Molise, Basilicata, Veneto e Valle d'Aosta.

Come facilmente si nota, nelle regioni maggiormente urbanizzate si registra una minore collaborazione all'indagine (elevata incidenza di mancate risposte totali), viceversa nelle regioni in cui prevalgono i comuni di tipo NAR, ad una maggiore collaborazione fa riscontro una resa qualitativa peggiore.

Infine, si e' proceduto ad analizzare con maggiore dettaglio il particolare aspetto relativo ai questionari non pervenuti che, a causa della loro incidenza (78%), costituisce la principale fonte di mancate risposte totali.

A tal fine, per poter risalire alla causa della mancata collaborazione, i questionari non pervenuti, nell'ambito comunale, sono stati distinti a seconda che la loro entita':

- coincida con il numero di questionari programmati, avendosi in tal caso "caduta totale" di interviste per quel comune, situazione imputabile, quasi certamente, a mancata operativita' da parte dell'ufficio comunale;

- corrisponda ad una parte del numero di questionari programmati, nel qual caso si parla di "caduta parziale" di interviste, situazione attribuibile principalmente a un mancato reperimento o mancata collaborazione delle famiglie selezionate.

Onde evitare che, specialmente la caduta totale di interviste risultasse un fatto episodico per il comune, si e' evitato il riferimento trimestrale assumendo quello annuale. In questo modo con caduta totale di interviste per un comune viene intesa la mancata rilevazione in tutte e quattro le indagini trimestrali.

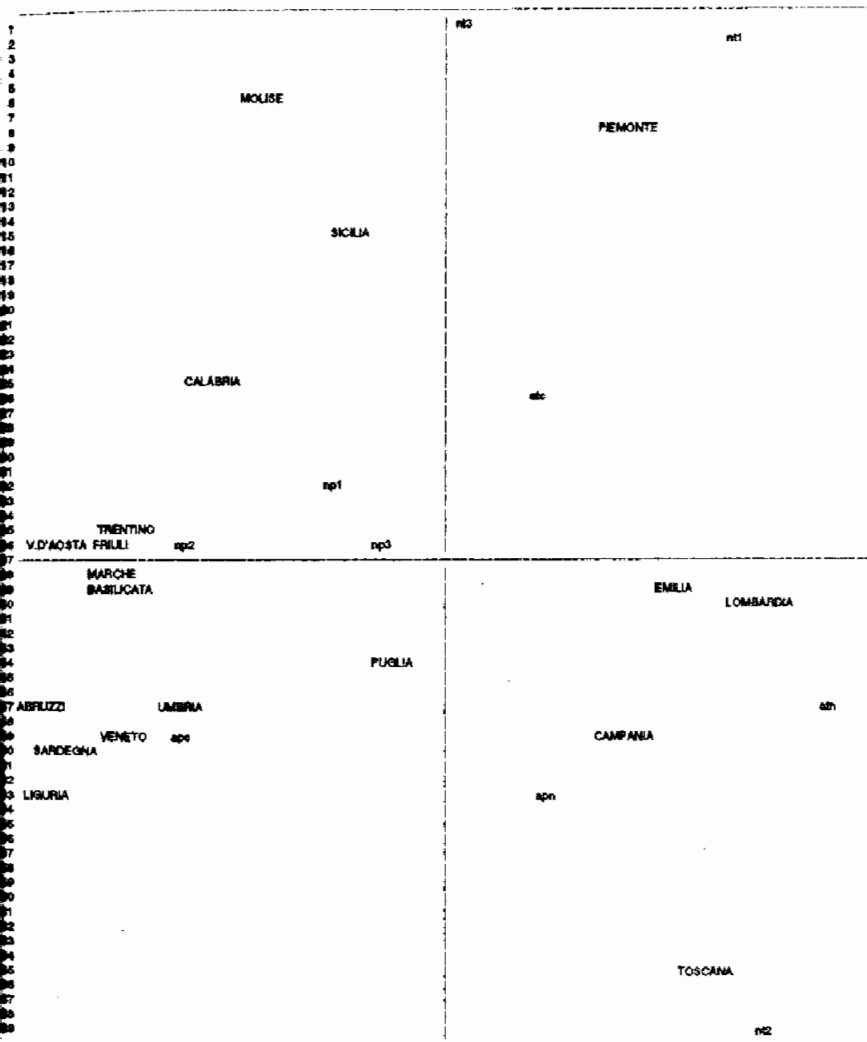
Nella tav. 8 e' fornito il riepilogo per regione di tale classificazione, in cui si procede anche alla distinzione del tipo di comune (AR e NAR), di comuni capoluoghi o non capoluoghi, per quelli di tipo AR, e in base al mese di rilevazione per quelli di tipo NAR.

Anche per l'analisi di detta tavola si e' fatto ricorso ad un metodo

Tav. 8 - Distribuzione delle mancate risposte totali per regione, per tipo comune e per tipo di caduta. Totale regionale = 100. Anno 1989-90

Regioni	Comuni AR				Comuni NAR				Totale
	Totale		Parziale		Totale		Parziale		
	Capol.	Non Cap.	Capol.	Non Cap.	Caduta 1 mese	Caduta 2 mese	Caduta 3 mese	3 mese	
Piemonte	16,74	8,06	3,88	24,19	22,33	4,96	4,34	15,50	645
Valle A.	10,59	10,59		15,87	7,93	34,12	32,94	22,35	85
Lombardia	29,75	12,40	2,64			19,34	3,47	8,60	605
Trentino	6,94	6,94				30,56	27,78	34,72	72
Veneto	33,53	33,53	13,87			2,89	16,76	32,95	173
Fritoli	2,94	2,94				5,88	35,29	55,88	68
Liguria	62,56	62,56	2,46			7,39	15,27	12,32	203
Emilia	23,74	23,74	4,57	16,44		3,65	9,13	9,59	219
Toscana	32,88	32,88	3,31		24,43	10,94	6,87	19,34	393
Umbria	27,48	27,48	23,68			14,04	33,33	13,16	114
Marche	10,00	10,00	5,00			25,00	55,00	5,00	20
Lazio	48,77	48,77	6,17		14,81	8,64	12,65	8,95	324
Abruzzi	48,08	48,08				11,54	33,65	6,73	104
Molise	5,11	5,11			39,42	22,26	16,06	17,15	274
Campania	20,83	9,95	21,99		9,72	4,63	7,06	7,75	864
Puglia	15,09	19,92	22,64		10,06	7,13	19,50	5,66	477
Basilicata						14,12	38,98	30,51	177
Calabria		14,90	4,31		18,82	28,63	17,25	16,08	255
Sicilia	11,37	18,80	2,53	15,17	17,06	11,22	18,33	5,53	633
Sardegna		58,41	1,77			30,09	5,31	4,42	113
Totale	2,48	12,38	8,04	6,60	4,74	11,58	13,51	12,74	5818

FIG. 2. Analisi delle corrispondenze. Mancate risposte totali per regione, per tipo di comune e per tipo di caduta (totale o parziale).



LEGENDA PER LE SIGLE DELLE MODALITA' DELLE VARIABILI.

LE SIGLE SONO COMPOSTE DA TRE CARATTERI ALFANUMERICI:
 IL PRIMO CARATTERE (a o n) DISTINGUE I COMUNI AR E NAR;
 IL SECONDO CARATTERE DISTINGUE L'UTILIZZO (t o p) TOTALE O PARZIALE DEI QUESTIONARI NEI COMUNI;
 IL TERZO PER I COMUNI DI TIPO AR DISTINGUE I COMUNI CAPOLUOGHI DA QUELLI NON CAPOLUOGHI (e o n), MENTRE PER I COMUNI DI TIPO NAR DISTINGUE IL MESE DI RILEVAZIONE DEL TRIMESTRE (1 o 2 o 3).

fattoriale, precisamente quello delle corrispondenze semplici essendo i dati organizzati sotto forma di tabella di contingenza.

Dall'esame del piano (fig. 2) delimitato dai primi due assi fattoriali, a cui resta associata la massima variabilita' della tavola, si evidenzia:

- lungo il primo asse, si posizionano da una parte le regioni: Lombardia, Toscana, Piemonte, Emilia e Campania caratterizzate da elevate incidenze in relazione alla caduta totale dei questionari nei comuni non capoluoghi e nei comuni NAR che effettuano la rilevazione nel primo e secondo mese del trimestre; dall'altra parte le regioni: Abruzzi, Liguria, Marche, Basilicata, Sardegna, Trentino e Friuli caratterizzate da incidenze elevate per caduta parziale sia nei comuni capoluoghi sia nei comuni NAR;

- lungo il secondo asse, da una parte si collocano le regioni: Molise, Piemonte e Sicilia caratterizzate da una caduta totale elevata nei comuni NAR che effettuano la rilevazione nel primo e nel terzo mese del trimestre, dalla parte opposta si posizionano le regioni: Toscana, Sardegna, Liguria, Campania e Veneto caratterizzate da elevata caduta parziale nei comuni che effettuano la rilevazione nel secondo mese e caduta parziale in entrambi i tipi di comuni AR.

In sintesi, e' nelle regioni piu' urbanizzate che si rileva maggiormente la caduta totale dei questionari, interessando entrambi i tipi di comuni, AR e NAR, e nell'ambito AR sia capoluoghi che non capoluoghi; viceversa nelle regioni con scarsa popolazione prevale la caduta parziale nei comuni di tipo NAR.

Piu' preoccupante, invece, e' la situazione rilevata in alcune regioni, quali: Campania, Liguria, Sardegna e Toscana in cui sono elevate le incidenze sia della caduta parziale che di quella totale, sintomo certamente di una maggiore disorganizzazione ed inefficienza nella conduzione dell'indagine.

5. Il trattamento delle mancate risposte totali effettuato dall'Istat

L'elaborazione dei dati trimestrali, effettuata nella seconda fase della revisione automatica, prevede il calcolo del coefficiente di riporto dei dati campionari all'universo comprensivo anche del trattamento delle mancate risposte totali.

Il procedimento seguito dall'Istat, in assenza di identificazione dei non rispondenti, ha lo scopo di ricostruire all'interno del campione la stima della struttura dell'universo sulla base di una o piu' variabili di controllo di cui si conosce la distribuzione nell'universo di riferimento¹⁰. Tale ricostruzione avviene tramite l'utilizzo di un peso attribuito a gruppi omogenei di questionari per attribuire l'importanza che loro compete in base ai dati statistici di struttura.

Questa esigenza si manifesta in presenza di mancate risposte totali o in altri casi in cui il campione risulta diverso da quello desiderato¹¹.

La ricostituzione della struttura della popolazione all'interno del campione avviene, come si è detto, in base ad una o più variabili di controllo, che possono individuarsi tra quelle dello schema di campionamento o ad esso esterne oppure ricorrendo ad entrambi i tipi.

Le variabili considerate dall'Istat sono lo strato territoriale, che è una variabile rientrante tra quelle di campionamento, e le famiglie classificate per numero di componenti che non rientra in tale schema¹².

La considerazione delle due variabili è motivata dal fatto che utilizzando solo lo strato territoriale i risultati potrebbero essere distorti in ordine alla distribuzione delle famiglie per numero di componenti (che costituisce uno dei caratteri in base al quale sono pubblicati i risultati), per cui il peso dovrà scaturire da una combinazione di entrambe le variabili di controllo.

Inoltre, l'applicabilità del metodo, noto come metodo di riponderazione con pesi provenienti dalla popolazione¹³, richiede la conoscenza della distribuzione congiunta di dette variabili nella popolazione di riferimento. Ciò comporta la conoscenza, per trimestre, della distribuzione delle famiglie per strato e per numero di componenti.

Tale distribuzione, come è noto, è disponibile soltanto all'epoca del censimento della popolazione, per cui il riferimento a tale struttura diventa sempre meno attendibile man mano che ci si allontana da tale epoca, in quanto si assiste allo spostamento nel tempo delle famiglie dalle ampiezze maggiori verso quelle minori.

Per ovviare a questo problema l'Istat, dal 1986, fa ricorso alla struttura familiare stimata dall'indagine sulle forze lavoro (occorre ricordare che il concetto di famiglia è al netto delle convivenze) disponibile, però, solo a livello regionale ritenendola come struttura effettiva della popolazione, mentre a livello di strato dispone solo della popolazione complessiva.

La disponibilità solo di queste informazioni comporta due conseguenze metodologiche essenziali e fra di loro interdipendenti: la prima riguarda la tecnica di ponderazione adottata che fa riferimento ad informazioni campionarie in cui il calcolo dei pesi o coefficienti di riponderazione o di riporto avviene tramite una combinazione del reciproco del tasso di risposta a livello di strato e del reciproco del tasso di risposta delle famiglie per numero di componenti a livello regionale; la seconda, facendo riferimento proprio a tale tecnica, implica che il metodo non riduce anche la distorsione dovuta a mancata copertura assicurata, invece, dal metodo di riponderazione con pesi provenienti dalla popolazione.

Questo ultimo motivo induce ad ovviare alle mancate coperture che possono risultare nelle due variabili di controllo prima di procedere al calcolo

dei coefficienti di riponderazione. In ordine alla soluzione del problema si procede, innanzitutto, alla rilevazione di informazioni mancanti per strato¹⁴ poi per dimensione familiare:

a - mancata copertura di strati. Le informazioni mancanti per un intero strato sono ricavate in modo differente a seconda che lo strato comprenda comuni di tipo AR o di tipo NAR. Per i comuni di tipo AR, si procede alla duplicazione di risultati ottenuti da un altro comune di tipo AR della stessa regione. La scelta del comune di riferimento può avvenire o tramite una scelta casuale tra tutti i comuni AR o facendo riferimento ad un comune con caratteristiche simili. Per gli strati NAR si possono verificare due situazioni: almeno un comune dello strato ha proceduto alla rilevazione nell'ambito del trimestre, in tal caso i risultati ottenuti sono estesi anche al comune o ai comuni inadempienti e quindi allo strato; oppure nessun comune ha proceduto a rilevazioni, nel qual caso sono duplicati i risultati di uno strato simile della stessa regione;

b - mancata copertura delle dimensioni familiari. Spesso si verifica che in un comune siano assenti famiglie con un certo numero di componenti, specie quelle con dimensioni più ampie. Il modo di procedere in tal caso è duplice: o viene duplicato il questionario di una famiglia con un numero di componenti ricadente in una delle dimensioni adiacenti a quella mancante, aggiungendo e sottraendo la spesa media per componente a seconda dei casi, oppure duplicando il questionario di una famiglia con lo stesso numero di componenti scelta nello stesso strato, se il comune è di tipo NAR, o della stessa regione, se di tipo AR.

Si dispone, adesso, di tutti gli elementi per il calcolo del coefficiente di riporto, che avviene in due fasi. Anzitutto è calcolato il coefficiente di strato ottenuto dal rapporto tra la popolazione intesa come numero di componenti dell'intero strato (C_i) e del campione (c_i) osservato nello stesso strato i-mo. Poi, dalla distribuzione delle famiglie del campione per dimensione e per strato, tramite il coefficiente di strato, è ottenuta la stima delle famiglie dell'universo-strato, che sommata fornisce la stima delle famiglie di dimensione h a livello regionale (F'_h). Il rapporto, a livello regionale, tra le famiglie effettive (F_h) e quelle stimate dal campione (F'_h) determina il secondo coefficiente di riporto che tiene conto della dimensione familiare. Infine, il prodotto dei due coefficienti, quello di strato e di dimensione familiare, determina il relativo coefficiente di riporto C_{ih} , cioè:

$$C_{ih} = c_i * (F_h / F'_h)$$

che viene associato ad ogni questionario dello strato "i" in relazione alla dimensione familiare "h" (Istat, 1990).

6. Considerazioni conclusive

L'analisi delle mancate risposte totali dell'indagine sui "Consumi delle famiglie", dal momento della rilevazione fino al loro trattamento, porta ad enucleare importanti aspetti tesi a consentire il miglioramento della qualità dell'indagine stessa.

La constatata rilevanza del fenomeno comporta la necessità di sviluppare tutti quegli accorgimenti, individuati in precedenza come metodi preventivi, diretti verso le principali fonti da cui provengono le mancate risposte totali: gli uffici comunali, le famiglie, i rilevatori ed i revisori.

L'analisi svolta ha evidenziato gli aspetti quantitativi e qualitativi delle mancate risposte totali e la considerevole disparità territoriale. Vi sono regioni in cui la mancata risposta si manifesta in forma, per così dire, patologica e si è constatato che si tratta, ad eccezione del Molise, di quelle più popolate, mentre in altre assume dimensioni certamente accettabili, in particolare vedi Marche, Trentino, Umbria e Friuli.

In maniera più specifica, si è guardato alla composizione delle mancate risposte totali (questionari non pervenuti (77%) e questionari eliminati nella fase di revisione (23%)). A causa del rilevante peso dei questionari non pervenuti e con l'obiettivo di perseguire una riduzione della loro entità, si è proceduto ad analizzare tale fenomeno con maggiore dettaglio. A tale fine è stato introdotto nell'analisi il concetto di caduta totale e pur adottando una definizione alquanto restrittiva ad esso resta associata una quota abbastanza elevata (34%) di mancate risposte totali. Inoltre, si è osservato che la caduta totale si concentra principalmente nei comuni non capoluoghi di provincia e nei comuni che effettuano la rilevazione nel terzo mese del trimestre.

Occorre prendere atto che nei comuni interessati da tale fenomeno, nonostante i ripetuti solleciti da parte dell'Istat, permane una netta volontà di non collaborare all'indagine che rende inutile ed infruttuoso qualsiasi tentativo di coinvolgimento.

A fronte di tale situazione una possibile azione, volta alla riduzione della caduta totale dei questionari, fermo restando l'attuale struttura dell'indagine, potrebbe essere esperita soltanto con riferimento ai comuni di tipo NAR, tramite la sostituzione del comune inadempiente con un altro dello stesso strato, ipotizzando la sostituibilità dei comuni nell'ambito di uno stesso strato. Migliori risultati potrebbero conseguirsi adottando per i comuni un diverso piano di campionamento, come proposto per altri motivi da Filippucci e Marliani (1992) i quali pervengono ad una ridefinizione dei comuni AR e NAR tenendo conto di nuove soglie di popolazione e della stratificazione per omogeneità territoriale contigua in base al comportamento di consumo, che sicuramente rende meno problematica la sostituzione del comune inadempiente

nell'ambito dello stesso strato.

La restante aliquota di questionari non pervenuti (42%) costituisce la parte più cospicua di mancate risposte totali. In essa confluiscono le mancate risposte dovute ad una serie di cause: interviste non effettuate in uno, due o tre trimestri; questionari arrivati in ritardo all'Istat e considerati come non pervenuti; questionari consegnati dalle famiglie completamente in bianco, che alcuni rilevatori non inviano all'Istat; interviste non effettuate per impossibilità di sostituzione delle famiglie selezionate.

La posizione degli uffici comunali, in questo caso, è nettamente diversa dalla precedente. Non si rileva un rifiuto alla collaborazione e le mancate rilevazioni in qualche trimestre rivestono un carattere episodico o contingente. Una migliore organizzazione operativa della rilevazione ottenuta tramite il potenziamento del sistema ispettivo dell'Istat, la visita alle famiglie nei primi giorni della decade di rilevazione da parte dei rilevatori, l'utilizzo di mezzi più rapidi di consegna dei questionari all'Istat (posta celere, corrieri etc.) in aggiunta ad una revisione del piano di campionamento comportante l'inclusione nel campione di un minor numero di comuni, con conseguente vantaggio in termini di tempo e di costi, dovrebbe ridurre di molto il numero della suddetta tipologia di mancate risposte.

Un altro aspetto da sottolineare, riguardante prevalentemente la qualità dell'indagine, è relativo alle mancate risposte totali connesse alle eliminazioni dei questionari nelle fasi di revisione. Ciò che risalta con maggiore evidenza è l'elevata incidenza di questionari eliminati nella fase di revisione ordinale. L'analisi di tale fenomeno meriterebbe certamente un maggiore approfondimento, in quanto, molto probabilmente, chiama in causa i problemi dell'accuratezza del lavoro dei revisori e l'efficacia degli strumenti di rilevazione. Ma oltre a questi aspetti generali non va trascurato il diverso modo di operare dei revisori nelle varie fasi di controllo che, allo stato attuale, non è possibile identificare ed isolare ma soltanto ipotizzare. Tale incertezza permane anche a livello regionale dove il fenomeno presenta un'ampia variabilità.

Nonostante questi problemi, è però possibile effettuare alcune considerazioni, attinenti al modo con cui si procede all'eliminazione dei questionari in questa particolare fase di revisione. Alla base del procedimento di revisione ordinale seguito dall'Istat vi è l'ipotesi che nelle due fasi precedenti, in quella manuale e in quella automatica, il questionario sia sottoposto e abbia superato anche un efficace controllo delle voci di spesa esaminate nel contesto globale della sua struttura. In tal caso, il riesame dei questionari la cui spesa complessiva ricade negli estremi della distribuzione, dove è più probabile che si accumulino questionari incoerenti, risponde ad un criterio di super-revisione che nel contempo evidenzia sia l'operato dei singoli revisori sia la possibile diversità operativa degli stessi. Tuttavia, l'elevata incidenza dei modelli

eliminati nella fase di revisione ordinale (circa il 56% degli eliminati) fa supporre che il controllo globale nelle due fasi precedenti non sia stato effettuato o venga realizzato in modo parziale e non sistematico.

In tal caso, il controllo tramite la spesa complessiva perde in parte la sua validità, in quanto vengono sottoposti a riesame soltanto i questionari la cui spesa rientra tra gli *outliers*, mentre sfuggono al controllo tutti quelli la cui incoerenza è occultata dall'operazione di totalizzazione.

Si può concludere auspicando che l'organizzazione della fase di revisione venga modificata, senza disperdere le esperienze acquisite e tenendo conto della particolarità dell'indagine, procedendo in particolar modo ad una più efficace identificazione, definizione e ripartizione dei compiti dell'uomo e della macchina e della loro interazione.

Infine, qualche considerazione sul trattamento delle mancate risposte totali. Indubbiamente, il metodo di riponderazione utilizzato dall'Istat dovrebbe ridurre la distorsione degli stimatori anche se comporta un incremento della varianza degli stessi, che discende dalla varianza dei pesi. Quest'ultima è tanto più elevata quante più classi di ponderazione si utilizzano per ripartire la popolazione. Aumentando le classi, diminuisce il numero di rispondenti in esse contenuto e ciò causa una maggiore instabilità dei coefficienti di ponderazione che si riflette poi sulla varianza degli stimatori (Kish, 1965).

Si è usato il condizionale, in quanto la riduzione della distorsione di uno stimatore all'interno di ogni classe è tanto maggiore quanto minore è la differenza tra la media dei rispondenti e quella dei non rispondenti, fino ad eliminarsi se le due medie coincidono dando così luogo ad uno stimatore corretto.

Il punto cruciale è la definizione delle classi in modo tale che all'interno di esse mediamente, quanto più è possibile, non vi sia differenza tra rispondenti e non rispondenti (Cicchitelli e altri, 1992).

Rispetto a tale problematica, con riferimento all'indagine sui consumi delle famiglie, si può osservare:

a)- i questionari relativi a caduta totale o parziale nei comuni che non effettuano l'indagine, che peraltro costituiscono la maggior parte dei questionari non pervenuti, sono da considerare come errori campionari mancando la selezione delle unità secondarie di campionamento, cioè le famiglie, e non come mancate risposte totali;

b)- attribuire ad ogni unità della classe la stessa probabilità di risposta, cioè la stessa probabilità di essere rispondente o non rispondente, ci sembra una forzatura molto discutibile tenendo anche conto delle variabili che intervengono nella costruzione delle classi (strati).

Si sottolinea, di nuovo, la difficoltà nel reperire soluzioni ad entrambe le situazioni fermo restando l'attuale piano di campionamento. Rimandando alla

proposta di un nuovo disegno campionario, richiamata in precedenza, in cui la formazione delle classi territoriali avvenga in base a comportamenti di consumo implica l'ottenimento di classi certamente più omogenee rispetto al fenomeno oggetto dell'indagine.

Nel contempo una più approfondita e sistematica rilevazione delle caratteristiche dei non rispondenti, da parte dell'Istat, renderà possibile formulare una o più alternative nella scelta del metodo più idoneo al trattamento delle mancate risposte totali.

Note

1. In letteratura non esiste una classificazione univoca di errore non campionario. Si parla indifferentemente di errori di osservazione, di misura etc., facendo, spesso, confusione tra le cause da cui essi originano e gli effetti, cioè, le modalità con cui si manifestano (Desabie, 1965), (Kalsbeek, 1980), (Gourieroux, 1981), (Grosbas, 1987), (Dalenius, 1988), (Fabbris, 1991).

Indubbiamente, una classificazione esaustiva delle cause che producono gli errori non campionari comporta notevoli difficoltà per gli aspetti particolari connessi a ciascuna indagine. È preferibile, perciò, fissare le modalità con cui si possono manifestare: mancate risposte totali, parziali e informazioni errate per risalire di volta in volta alle cause che le hanno originate tenendo conto delle caratteristiche dell'indagine in esame.

2. È efficace la distinzione degli errori non campionari (Groves, 1989) fra quelli legati all'assenza di osservazione (mancata copertura, mancata risposta totale, mancata risposta parziale) e quelli legati alla presenza di osservazione costituite dalle informazioni errate.

3. In effetti, gli studi di identificazione fanno riferimento alle ipotesi che possono essere formulate sulla base delle informazioni disponibili circa i motivi che hanno generato la mancata risposta totale. Noti i motivi, si possono formulare due classi di modelli comportamentali: la prima, comprende i modelli di non risposta "non ignorabile" in cui si ipotizza che i non rispondenti costituiscono un particolare sub-campione della popolazione diverso da quello dei rispondenti (tale situazione è tipica delle indagini sui redditi, sul lavoro, di opinioni politiche etc.); la seconda, fa riferimento ai modelli di non risposta "ignorabile" che considerano i non rispondenti come un sub-campione casuale della popolazione.

4. Si parla di imputazione nel caso di trattamento di mancate risposte parziali. La differenza tra i metodi di adeguamento e di imputazione non è nella maggior parte dei casi così netta come potrebbe apparire. Anzi spesso i metodi sono gli stessi e varia soltanto il campo di applicazione, che è l'intero questionario in caso di mancata risposta totale e il singolo *item* nel caso di mancata risposta parziale.

5. La stratificazione dei comuni non capoluoghi e con popolazione inferiore ai 50 mila abitanti avviene in ambito regionale in base alla zona altimetrica, all'attività economica prevalente etc., formando in totale circa 133 strati.

6. Fissate le famiglie da intervistare in un mese, esse vengono equiripartite per decade di rilevazione.

7. I dati di base, a livello comunale, riguardanti sia il campione che le mancate risposte totali sono stati messi a disposizione dall'Istat. Si ringrazia il dott. Innocenzi ed i collaboratori della sua sezione per la disponibilità e per i preziosi suggerimenti.

8. Il controllo automatico non è di tipo interattivo per cui la correzione degli errori è effettuata o tramite forzatura in base al contesto delle risposte fornite o tramite riesame del questionario da parte del revisore.

9. La scelta dei decili varia in base alla distribuzione della spesa, per cui a volte sono considerati il primo decile e l'ultimo ventile, in altri casi il primo decile e l'ultimo quintile e così via.

10. I metodi utilizzati per il trattamento delle mancate risposte totali si possono distinguere in due gruppi in cui la discriminante fa riferimento all'esistenza o meno di diversità di comportamento dei rispondenti e dei non rispondenti rispetto all'indagine in studio. Rientrano nel primo gruppo i metodi che ipotizzano la casualità delle mancate risposte totali e quindi nessuna differenza tra le unità campionate e si pongono in genere l'obiettivo di ricostruire all'interno del campione la struttura della popolazione al fine di migliorare le stime campionarie. I principali metodi sono quelli di: duplicazione, riponderazione, sostituzione, stratificazione a posteriori, del quoziente e di regressione, etc., e le loro varianti poste in essere per contenere l'incremento della varianza degli stimatori pur in presenza di una riduzione della loro distorsione.

Nel secondo gruppo l'ipotesi di base presuppone, anche a seguito degli studi di identificazione, un diverso comportamento tra rispondenti e non rispondenti. La varietà dei metodi di questo gruppo è sostanzialmente legata alle caratteristiche particolari dei non rispondenti, che si possono supporre o che sono manifeste, le quali si dovranno, poi, tradurre in specifici modelli. Uno dei primi metodi è quello noto come sottocampionamento o doppio campionamento proposto da (Hansen, Hurwitz 1946) in cui si risale dalle mancate risposte totali di un'indagine postale alle caratteristiche dei non rispondenti tramite un sottocampione di questi intervistato direttamente.

Attuali sono, invece, i metodi che fanno ricorso ai metodi probabilistici bayesiani. Il metodo base presuppone la possibilità di categorizzare la variabile d'interesse per l'indagine e di disporre di una serie di variabili ausiliarie le cui determinazioni siano note sia per i rispondenti che per i non rispondenti. Ciò permette di calcolare la probabilità condizionata che un elemento del campione appartenga ad una specifica categoria della variabile oggetto di studio, noti i relativi valori delle variabili ausiliarie. Per le numerose varianti a tale metodo cfr. Lodwick (1966), Lichten (1972), Lusted (1968), Cornfield (1964), Daniel e altri (1982), Ericson (1967), Rubin (1977), Albert (1982).

11. Ciò si può verificare quando la selezione del campione avviene da liste della popolazione improprie, come ad es. nel caso del ricorso a liste elettorali al posto di liste anagrafiche non disponibili.

12. Si prescinde da altri caratteri di classificazione in base ai quali sono pubblicati i risultati campionari (come la condizione socio-economica del capofamiglia, etc.) ritenendo che il campione sia sufficientemente ampio in modo da garantire la significatività dei risultati campionari nella rappresentazione della popolazione di riferimento.

13. Nell'ambito del metodo di riponderazione si distinguono vari metodi che variano tra di loro per il modo con cui viene ad essere calcolato il peso, oltre a numerose varianti, nell'ambito di ciascuno di essi, poste in essere per ovviare a particolari aspetti negativi introdotti dalle singole procedure. Si ricorda che l'obiettivo principale del metodo è quello di ridurre l'errore totale della stima campionaria agendo contemporaneamente sulle due sue componenti: la distorsione (bias) e la varianza degli stimatori. Il metodo di riponderazione richiede la conoscenza nella popolazione della distribuzione delle variabili di controllo prescelte per effettuare la ricostituzione della loro struttura all'interno del campione, dopo aver suddiviso entrambi (popolazione e campione) in opportune classi omogenee. È questa la situazione ottimale in quanto i pesi provengono dagli elementi della popolazione. Non disponendo della distribuzione delle variabili nella popolazione, si può far ricorso per il calcolo dei pesi agli stessi elementi del campione. In tale caso, dopo aver suddiviso le variabili in classi, i pesi sono ottenuti dal reciproco del tasso di risposta relativo ad ogni classe. Entrambe le tecniche di riponderazione, con pesi provenienti dalla popolazione o dal campione, mentre riducono la distorsione aumentano la varianza degli stimatori che cresce in funzione del numero di classi considerate e quindi della varianza dei pesi. Per ovviare almeno in parte a questo problema si utilizza il metodo raking, introducendo delle variabili ausiliarie. In effetti, con questa tecnica i pesi sono costruiti in modo da conformare la distribuzione marginale di dette variabili a quella che le stesse hanno nella popolazione o nel campione se si dispone soltanto di informazioni campionarie.

Una ulteriore tecnica di riponderazione basa il calcolo dei pesi sulle probabilità individuali di risposta, ipotizzando che la propensione a collaborare all'indagine dipenda dai contenuti e dall'insieme delle condizioni in cui essa si svolge e che tale propensione sia possibile tradurla in termini probabilistici. Le prime applicazioni dovute a Politz e Simmons (1949) collegavano il problema delle mancate risposte totali, traducendolo in probabilità, con il fattore "non a casa". In tempi più recenti si è registrato il ricorso ad altri fattori; Platek e Grey (1979,1983) ad esempio hanno proposto stime della probabilità di risposta basata su informazioni esterne all'indagine e Sarndal (1981) stime di probabilità ottenute tramite modelli parametrici.

14. Il concetto di strato nella fase di trattamento delle mancate risposte totali è esteso anche ai comuni di tipo AR considerando ogni comune uno strato a se' stante.

Riferimenti bibliografici

- ALBERT J.H. (1982), *A Bayesian Treatment of Nonresponse when Sampling from a Dichotomous Population*, Proceedings of ASA SSRM, pp. 314-317.
- COCHRAN W.G. (1977), *Sample Techniques*, J. Wiley, New York.
- CORNFIELD J. (1964), *Bayes's theorem*, in Proceedings of 6th IBM Medical Symposium, pp. 163-196.
- DALENIUS T. (1988), *Handbook of Statistics: Sampling*, vol. 6, Krishnaiah Rao, Pittsburgh.
- DANIEL W. W., SCHOTT B., ATKINS F.C., DAVIS A. (1982), *An adjustment for Nonresponse in Sample Survey*, Educational and Psychological Measurement, 42, pp. 57-65.
- DESABIE J. (1965), *Theorie et Pratique des Sondages*, Dunod, Paris.
- ERICSON W.A. (1967), *Optimal Sample Design With Nonresponse*, Jasa, n. 62.
- FABBRIS L. (1991), *Abbinamento tra fonti d'errore nella formazione dei dati e misure dell'effetto degli errori sulle stime*, in Bollettino SIS n. 22, Roma.
- FILIPPUCI C., MARLIANI G. (1992), *La misura dei consumi delle famiglie: una riflessione a partire dall'esperienza italiana*, Relazione alle "Giornate di Statistica Economica 1992", Bressanone, 16-18 Settembre.
- GOURIEROUX C. (1981), *Theories des Sondages*, Esa, Paris.
- GROSBAS J.M. (1987), *Methodes Statistiques des Sondages*, Esa, Paris.
- GROVES R. M. (1989), *Survey errors and survey costs*, J. Wiley, New York.
- HANSEN M.H., HURWITZ W.N. (1946), *The problem of Nonresponse in Sample Survey*, Jasa, n. 41, pp. 516-529.
- HAWKINS D. F. (1975), *Estimation of non response bias*, Sociological Methods & Research, vol.3, n. 4, pp. 404-407.
- INNOCENZI G. (1989), *Qualita' dei dati della rilevazione sui consumi delle famiglie*, nota interna, Istat, Roma.
- INNOCENZI G. (1992), *Procedure di rilevazione ed elaborazione dei dati dell'indagine Istat sui bilanci delle famiglie e cenni ad alcune indagini eseguite in altri paesi*, CON.PRI. Rapporto di ricerca n. 3, Bologna, dipartimento di Scienze Statistiche.
- ISTAT (1989), *Manuale di tecniche d'indagine, 6- il sistema di controllo della qualita' dei dati*, Note e Relazioni, n. 1, Roma.
- ISTAT (1990), *I consumi delle famiglie - anno 1988*, Collana d'informazione, n. 30, Roma, pp. 7-39
- KALSBECK W.D. (1980), *A conceptual review of survey error due to non response*, Proceedings of ASA SSRM, pp. 131-136.
- KISH L. (1965), *Survey Sampling*, J. Wiley, New York.
- LEBART L., MORINEAU A., FENELON J.P. (1982), *Traitement des donnees statistiques*, Dunod, Paris.
- LICHTEN S. (1972), *Conditional non-independence of data in practical Bayesian decision task*, Organization Behavior and Human Performance, n. 8, pp. 21-25.
- LODWICK G.S. (1966), *Computer-aided diagnosis in radiology*, Investigative Radiology, n 1, pp. 72-80.
- LUSTED L.B. (1968), *Introduction to medical decision making*, Thomas, Springfield Illinois.
- LUCEV D. (1992), *La revisione manuale dei dati dell'indagine sui consumi delle famiglie dell'Istat*, CON. PRI. Rapporto di ricerca n. 18, Bologna, Dipartimentio di Scienze Statistiche.
- MASSELLI M. (1989), *La revisione*, in Istat, *Manuale di tecniche d'indagine, 6- il sistema di controllo della qualita' dei dati*, Note e Relazioni, n. 1, Roma.
- PLATEK R., GRAY G.B. (1979), *Methodology and application of adjustment for nonresponse*, ISI, 42 Session, Manila.
- PLATEK R., GRAY G.B. (1983), *Imputation methodology: total survey error, in Incomplete Data in Sampling Surveys*, vol. 2, Academic Press, Washington D.C.
- POLITZ A. N. , SIMMONS W.R. (1949), *An attempt to get the "not at home" into the sample without call-backs*, Jasa, 44.
- RUBIN D. B. (1977), *Formalizing Subjective Notions about the Effect of Nonrespondents in Sample Surveys*, n. 72, JASA, pp. 538-543.
- SARNDAL C.E. (1981), *Frameworks for inference in survey sampling with application to small area estimation and adjustment for non response*, ISI, 43 Session, Buenos Aires.
- SUKHATME P.V., SUKHATME B. V. (1970), *Sampling Theory of Survey with Applications*, Iowa State University Press.
- VAJANI L. (1969), *Metodi statistici nelle ricerche di mercato*, Etas Kompass, Milano.