

# **Analisi e sintesi degli indicatori di qualità dell'attività di rilevazione nelle indagini campionarie sulle famiglie**

Cosima Mero

Istat, Via Ravà 150, Roma, e-mail: mero@istat.it

Adriano Pareto

Istat, Via Magenta 2, Roma, e-mail: pareto@istat.it

## **1. Introduzione<sup>1</sup>**

Com'è noto, la valutazione e il miglioramento della qualità dei dati delle indagini statistiche è, da sempre, uno degli obiettivi prioritari dell'Istat e, per questo motivo, l'Istituto si è posto il problema di migliorare l'affidabilità della rete di rilevazione attraverso la quale vengono raccolti i dati di base.

Le indagini campionarie sulle famiglie si avvalgono, per la delicata fase di raccolta dei dati, del contributo dei rilevatori comunali, che costituiscono il nodo terminale della "Rete di rilevazione comunale" ovvero l'interfaccia tra l'Istituto e le famiglie.

La fase dell'intervista alla famiglia riveste un ruolo fondamentale, poiché da essa e dal modo in cui viene svolta dipende il buon esito della rilevazione.

In questo contesto ci si propone di analizzare, mediante tecniche di analisi multivariata, le relazioni esistenti tra le principali caratteristiche dei rilevatori della rete comunale e un insieme di indicatori di qualità dell'attività svolta.

La fonte dei dati è costituita dal Sistema Informativo della Rete di rilevazione Comunale (SIRC) che, per ogni indagine, contiene i dati strutturali dei rilevatori coinvolti (età, sesso, titolo di studio, ecc.) e una serie di variabili, riguardanti la raccolta dei dati, che consentono di costruire, per ciascun rilevatore, un set di indicatori del lavoro svolto (tasso di risposta, sostituzione, caduta, ecc.).

Il lavoro proposto ha, quindi, lo scopo di mostrare l'influenza dei rilevatori sulla qualità dei risultati, individuando delle tipologie 'a rischio' sulle quali concentrare i controlli in fase di revisione dei questionari o di monitoraggio dell'attività di rilevazione.

## **2. Il Sistema Informativo della Rete di rilevazione Comunale (SIRC)**

Al fine di gestire, pianificare e programmare la numerosità, la formazione e la qualità dei rilevatori e di monitorarne l'attività, fin dal 1988 è stato realizzato, presso l'Istituto, un archivio informatizzato dei rilevatori e dei comuni (Istat, 2003) che, inizialmente, conteneva solo i dati strutturali degli individui (nome e cognome, data di nascita, sesso, stato civile, residenza, titolo di studio, comune e provincia di lavoro, data di primo ingresso all'attività di rilevatore, rapporto di lavoro con il comune, condizione professionale, indagine a cui partecipa, codice identificativo, ecc.) e informazioni sui comuni appartenenti al campione dell'indagine (indagini a cui partecipa, periodi o

---

<sup>1</sup> I paragrafi 1 e 4 sono stati redatti congiuntamente dagli autori, il paragrafo 2 è a cura di Cosima Mero e il paragrafo 3 è a cura di Adriano Pareto.

cicli di partecipazione, periodo di rilevazione, se comune fisso o ruotante nel campione, numero di famiglie da intervistare per ciascuna indagine, responsabile comunale di ogni indagine, ecc.). Alcune delle informazioni citate scaturiscono dal disegno campionario dell'indagine da effettuare (comuni campione, autorappresentatività e non, famiglie teoriche, ecc.); altre vengono acquisite mediante appositi modelli (modd. ISTAT ARF 1/B e ARF2) opportunamente compilati dai responsabili di indagine, presso i comuni, e dagli stessi rilevatori. L'acquisizione avviene in occasione di ogni indagine nel periodo immediatamente precedente la rilevazione.

Successivamente, considerando che l'attività di rilevazione di ogni indagine campionaria sulle famiglie viene coordinata dall'Istat sia centralmente che a livello periferico tramite gli uffici regionali con compiti di istruzione, organizzazione e controllo, e che i comuni, oltre ad altri compiti, assolvono quelli di scelta, assistenza e controllo dei rilevatori stessi, l'archivio si è andato via via arricchendo di nuove informazioni, per le crescenti esigenze di valutazione e miglioramento della qualità dei dati raccolti. Tali informazioni riguardano anche l'aspetto controllo di qualità dell'attività di rilevazione che viene accertata in due modi: per intervista telefonica alla famiglia e per revisione del questionario compilato.

Di evoluzione in evoluzione, si è arrivati ad implementare un sistema informativo, chiamato SIRC (Sistema Informativo della Rete di rilevazione Comunale), in grado di elaborare non solo i dati strutturali dei rilevatori e dei comuni, ma anche quelle variabili utili ai fini del monitoraggio e del controllo della qualità del lavoro svolto dagli intervistatori.

Attualmente, il SIRC è un sistema realizzato in ambiente oracle e comprende:

1. una banca dati contenente tutti i dati strutturali di cui sopra e una serie di informazioni di interesse, riguardanti lo svolgimento delle rilevazioni, quali le famiglie assegnate, le famiglie intervistate, le famiglie cadute con sostituzione e senza sostituzione, i componenti delle famiglie, le risposte proxy, non proxy e missing per le interviste 'face to face' e per i modelli autocompilati, nonché gli annullamenti in fase di revisione;
2. un sistema di gestione della banca dati che permette sia l'aggiornamento interattivo a livello di singolo comune e rilevatore, sia l'alimentazione in modalità batch per i dati di riepilogo a fine indagine;
3. un sistema per il calcolo di report e indicatori di qualità che permette di disporre di informazioni puntuali sul lavoro svolto dalla rete comunale;
4. un sistema per il monitoraggio della qualità della rilevazione che offre al responsabile dell'indagine la possibilità di valutare e calibrare, opportunamente, eventuali interventi mirati che si dovessero rendere necessari per rimuovere situazioni problematiche che si possono presentare sul territorio nel corso della rilevazione.

## **2.1 I rilevatori comunali**

Il patrimonio informativo della banca dati permette di ottenere un quadro storico e, costantemente, aggiornato delle caratteristiche strutturali della rete di rilevazione comunale a tutti i livelli territoriali. La serie storica disponibile (1994-2003) ha messo in evidenza come la figura del rilevatore tipo sia andata evolvendo negli anni, vuoi in relazione alle diverse politiche di assegnazione del ruolo di rilevatore predisposte dagli enti locali preposti alla selezione, vuoi al variare del mercato dell'offerta, strettamente connesso alle mutazioni sociali del periodo. Non solo, anche l'impegno dei comuni è andato notevolmente aumentando nell'arco di tempo considerato, senza tener conto degli anni in cui, per particolari indagini (es. Multiscopo tematiche), è stato realizzato un ampliamento del campione in specifici contesti territoriali.

Le diverse tipologie di tavole che sono state elaborate riguardano l'impiego dei rilevatori nelle indagini, la frequenza dei rilevatori utilizzati ed il turnover degli stessi nell'ambito di una indagine rispetto ad un periodo di riferimento fissato; il tutto a diversi livelli territoriali.

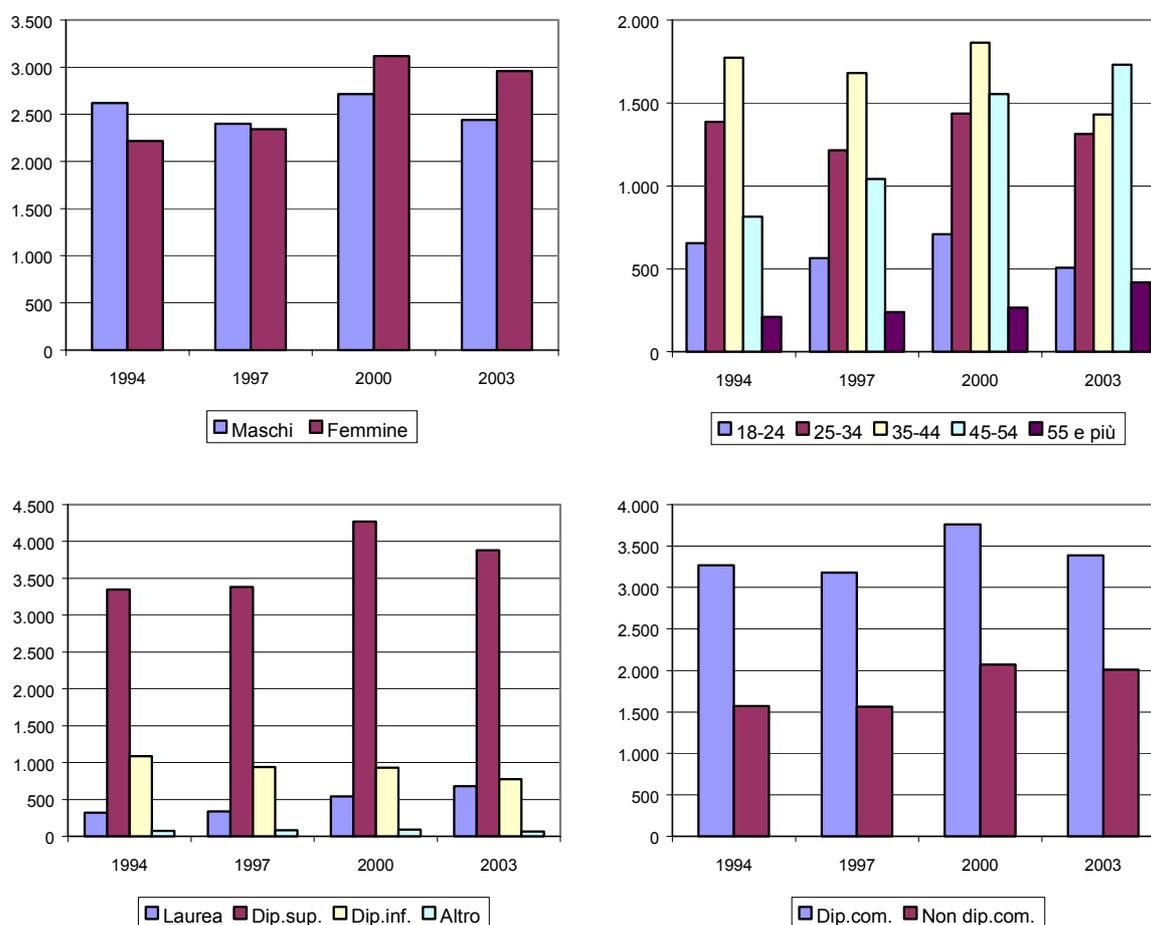
Nelle figure 1 e 2 sono rappresentati i dati relativi a quattro periodi con cadenza fissa di tre anni.

Volendo citare alcuni risultati interessanti, si osserva che la classe di età più rappresentata appartiene ai rilevatori con 35-44 anni, mentre la maggioranza assoluta dei rilevatori maschi registrata fino al 1998, ha poi ceduto in favore delle femmine dal 1999. La caratteristica titolo di studio si concentra nel diploma di scuola media superiore, anche se i laureati acquistano terreno ogni anno e la percentuale dei rilevatori dipendenti comunali risulta sempre nettamente superiore a quella dei rilevatori non dipendenti comunali, sebbene i rapporti di appartenenza a queste due categorie siano andati modificandosi nel tempo (figura 1).

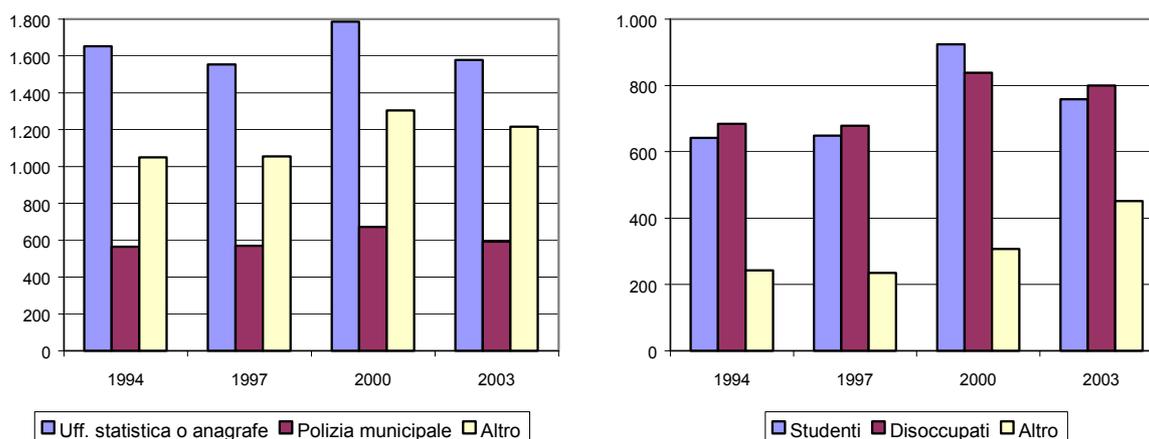
Attualmente, la maggior parte dei rilevatori dipendenti comunali è personale degli uffici di statistica e dell'anagrafe, mentre i rilevatori non dipendenti comunali sono per lo più studenti, disoccupati e individui in cerca di prima occupazione (figura 2). Sul fronte dell'esperienza nell'ambito delle interviste, la percentuale più alta si colloca nella conduzione di una sola indagine, affiancata da un turnover che si attesta mediamente intorno al 44% annuo.

Nel complesso, è importante osservare che i rilevatori più anziani e con più anni di esperienza riguardano i dipendenti comunali con un titolo di studio medio-basso, mentre i giovani, per lo più laureati, svolgono provvisoriamente l'attività di rilevatore, magari in attesa di una occupazione migliore o definitiva.

**Figura 1 – Rilevatori per sesso, età, titolo di studio e rapporto di lavoro. Anni 1994-2003**



**Figura 2 – Rilevatori dipendenti comunali e non dipendenti comunali. Anni 1994-2003**



Dall'esame dei grafici, si rileva un netto incremento dei rilevatori in corrispondenza degli anni 2000 e 2003. Tale aumento è dovuto al fatto che in ciascuno di questi anni si è svolta un'indagine multiscopo tematica con una estensione del campione; ciò ha richiesto un maggior numero di rilevatori, per lo più donne, giovani e dipendenti comunali o studenti.

## 2.2 Gli indicatori di qualità

L'acquisizione delle variabili strettamente connesse all'attività di rilevazione ha consentito di arricchire il patrimonio informativo del sistema mediante una serie di indicatori del lavoro svolto.

L'integrazione del SIRC con indicatori che permettessero di monitorare al meglio, sia dal punto di vista quantitativo che qualitativo, il funzionamento della rete comunale, si è resa necessaria in seguito ai risultati emersi dai lavori di un apposito gruppo di studio Istat che ha visto rappresentate in sé tutte le indagini campionarie sulle famiglie.

Le indagini in oggetto si differenziano in termini di tipologia dell'intervista (se condotta 'face to face', per autocompilazione dei questionari o mista) e di modalità di conduzione (se prevista la sostituzione delle famiglie cadute o no). In ogni caso, considerate le caratteristiche di trasversalità del SIRC rispetto alle indagini, è stato possibile definire un insieme di indicatori standard in grado di soddisfare le esigenze informative espresse nel gruppo.

Le principali variabili considerate sono le seguenti:

- numero di famiglie e di componenti assegnati;
- numero di famiglie e di componenti intervistati;
- numero di famiglie e di componenti caduti dall'elenco base e sostituiti, per motivo della caduta<sup>2</sup>;
- numero di famiglie e di componenti caduti dall'elenco base e non sostituiti, per motivo della caduta;
- numero di famiglie e di componenti caduti dall'elenco suppletivo, per motivo della caduta;
- numero di famiglie e di componenti annullati in fase di revisione (unità per le quali l'informazione rilevata non è conforme alle direttive impartite dall'Istat), per motivo dell'annullamento<sup>3</sup>;

<sup>2</sup> I motivi della caduta sono i seguenti: rifiuto, irreperibilità, morte, momentanea assenza, domicilio in altro comune, emigrazione all'estero, altro motivo.

<sup>3</sup> I motivi dell'annullamento sono: mancato contatto (il rilevatore non si è presentato e ha falsificato l'intervista), intervista telefonica (il rilevatore ha condotto un'intervista telefonica anziché 'face to face').

- numero di interviste e di autocompilati proxy (unità di rilevazione per le quali l'informazione non è stata rilevata direttamente dall'interessato);
- numero di interviste e di autocompilati non proxy (unità di rilevazione per le quali l'informazione è stata rilevata direttamente dall'interessato);
- numero di interviste e di autocompilati con proxy missing (unità di rilevazione per le quali non è possibile determinare se l'intervistato coincide con l'interessato).

Si noti che il numero di interviste annullate in fase di revisione riguarda il controllo successivo alla rilevazione e deriva dallo svolgimento di telefonate di controllo alle famiglie per verificare l'effettiva realizzazione dell'intervista da parte del rilevatore e la corretta modalità di conduzione della stessa (Camporese *et al.*, 2001).

Generalmente, viene effettuata almeno una telefonata di controllo per ogni rilevatore e, in caso di irregolarità, si procede ad un controllo a tappeto sulle altre.

Pertanto, se il numero di interviste annullate è pari a zero significa che la telefonata di controllo ha avuto esito positivo (intervista valida), altrimenti vuol dire che il controllo ha avuto esito negativo (intervista non valida) e, per quel rilevatore, è possibile determinare la proporzione di annullamenti sul totale delle interviste assegnate.

Gli indicatori selezionati per monitorare il lavoro svolto dai rilevatori e dai comuni riguardano: il tasso di caduta generale e per motivo, il tasso di sostituzione generale e per motivo, il tasso di annullamento generale e per motivo, il tasso di proxy sulle interviste e sugli autocompilati ed altri ancora. Tali indicatori possono essere elaborati a livello territoriale, di singolo rilevatore o per classi di rilevatori.

### 3. L'analisi dei dati

I dati utilizzati riguardano l'insieme dei rilevatori comunali impiegati nell'Indagine Multiscopo annuale nel triennio 2000-2002. Si tratta, nel complesso, di 3.695 unità, il 78,7% delle quali ha partecipato alla rilevazione per un anno, il 13% per due anni e il restante 8,3% per l'intero triennio.

Per ciascun rilevatore, il SIRC contiene due insiemi di variabili: il primo riguarda le caratteristiche strutturali dell'intervistatore (variabili qualitative), mentre il secondo misura la qualità del lavoro svolto (variabili quantitative).

L'analisi dei dati si è svolta in due fasi: nella prima fase è stato condotto uno studio a livello di variabili (studio dell'interdipendenza) e nella seconda fase a livello di unità (ricerca di una tipologia).

#### 3.1 Lo studio dell'interdipendenza

Nella fase di studio dell'interdipendenza è stata effettuata un'analisi in componenti principali (ACP) degli indicatori di qualità della rilevazione. Le variabili utilizzate sono elencate nel prospetto 1 e sono costituite da rapporti percentuali relativi all'operato dei rilevatori.

L'analisi in componenti principali ha fornito tre fattori significativi<sup>4</sup> che descrivono circa il 73% della variabilità complessiva degli indicatori.

La prima componente risulta correlata positivamente con FCAD (0,976), FCADR (0,709), FCADI (0,510) e FCADA (0,713) e negativamente con FINT (-0,974). Pertanto, fa riferimento ad una scala approssimativa che classifica i rilevatori sulla base del tasso di risposta delle famiglie.

---

<sup>4</sup> I fattori con autovalore maggiore di 1.

## Prospetto 1 - Principali indicatori di qualità della rilevazione

Indicatori	Sigla
Tasso di famiglie intervistate	FINT
Tasso di famiglie cadute	FCAD
Tasso di famiglie annullate per intervista non valida	FINV
Tasso di famiglie cadute per rifiuto	FCADR
Tasso di famiglie cadute per irreperibilità	FCADI
Tasso di famiglie cadute per altro motivo	FCADA
Tasso di interviste proxy	I PRO
Tasso di interviste non proxy	IN PRO
Tasso di autocompilati proxy	A PRO
Tasso di autocompilati non proxy	AN PRO

La seconda componente, invece, può essere interpretata come un indicatore del ricorso alle risposte proxy, essendo correlata positivamente con INPRO (0,856) e ANPRO (0,816) e negativamente con I PRO (-0,847) e A PRO (-0,838). La terza componente, infine, è fortemente correlata con FINV (0,996) e contrappone i rilevatori ritenuti ‘affidabili’ (casi in cui le telefonate di controllo hanno avuto esito positivo), ai rilevatori ‘non affidabili’ (casi in cui alcune interviste sono state annullate, in fase di revisione, perché giudicate non valide)<sup>5</sup>.

Sulla base dei risultati ottenuti, si è proceduto alla costruzione di tre variabili categoriali, una per ciascun fattore individuato, in grado di rappresentare, rispettivamente, il fenomeno delle interviste cadute (CAD), proxy (PRO) e annullate (INV)<sup>6</sup>.

Tali variabili sono state messe in relazione con i principali caratteri strutturali dei rilevatori mediante un’analisi delle corrispondenze multiple (ACM) in modo da analizzare le interazioni tra singole modalità. Nella figura 3 è riportata la rappresentazione grafica di tali modalità nel primo piano fattoriale.

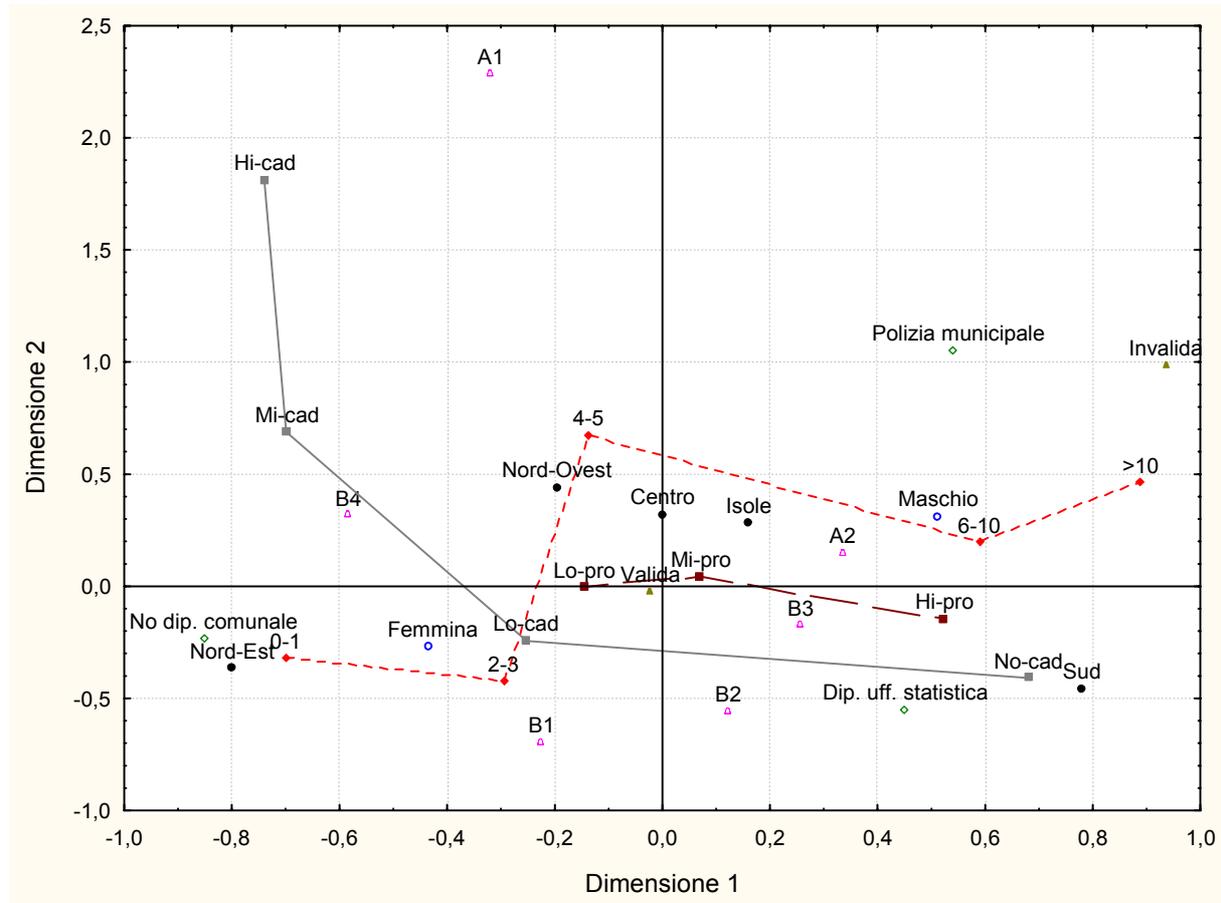
Si osservi che il primo asse risulta correlato con la variabile ‘anni di attività’ che costituisce una ‘spezzata’ attraverso la quale è possibile descrivere alcune variabili strutturali ad essa collegate: il rapporto di lavoro, il sesso e la ripartizione geografica. La similitudine tra le traiettorie di questi caratteri è prova della loro connessione, mentre la vicinanza tra alcune modalità ne riflette il forte grado di interazione. Così, per esempio, l’accoppiamento ‘0-1’-‘No dip. comunale’ significa che la maggior parte dei rilevatori con al massimo un anno di attività non sono dipendenti comunali, mentre la posizione intermedia di ‘6-10’ e ‘>10’ tra i punti modalità ‘Polizia municipale’ e ‘Dip. uff. statistica’ riflette la sostanziale equidistribuzione di queste due categorie tra i rilevatori con più di 6 anni di esperienza.

Inoltre, se la modalità ‘femmina’ risulta collocata pressoché al centro dei punti relativi a ‘0-1’, ‘2-3’ e ‘4-5’, grazie ad una elevata presenza di femmine tra i rilevatori con esperienza inferiore ai 6 anni, la modalità ‘maschio’ si trova quasi all’estremità opposta della spezzata, per la maggiore proporzione di maschi con esperienza superiore ai 6 anni.

<sup>5</sup> Per rilevatore ‘affidabile’ s’intende un rilevatore che ha effettuato soltanto interviste valide; mentre per rilevatore ‘non affidabile’ s’intende un rilevatore che ha effettuato una o più interviste non valide.

<sup>6</sup> La variabile CAD ha le seguenti modalità: ‘No-cad’ (nessuna caduta), ‘Lo-cad’ (fino al 20% di cadute), ‘Mi-cad’ (20% - 40% di cadute), ‘Hi-cad’ (oltre il 40% di cadute). La variabile PRO ha modalità: ‘Lo-pro’ (fino al 20% di proxy), ‘Mi-pro’ (20% - 40% di proxy) e ‘Hi-pro’ (oltre il 40% di proxy). La variabile INV ha modalità: ‘Valida’ (telefonata di controllo con esito positivo: intervista valida) e ‘Invalida’ (telefonata di controllo con esito negativo: intervista non valida).

**Figura 3 - Rappresentazione di alcuni caratteri dei rilevatori nel piano principale dell'ACM**



Infine, per quanto riguarda le ripartizioni geografiche, ‘Nord-Ovest’, ‘Centro’ e ‘Isole’ sono le ripartizioni più equilibrate rispetto alle caratteristiche dei rilevatori, mentre al ‘Nord-Est’ si registra una prevalenza di rilevatori non dipendenti comunali con poca esperienza e al ‘Sud’ una prevalenza di dipendenti comunali con molti anni di esperienza. Tali differenze sono dovute all’eterogeneità nei criteri di arruolamento degli intervistatori da parte dei diversi comuni.

Il secondo asse rappresenta la tipologia comunale e contrappone i centri dell’area metropolitana ‘A1’, in alto, all’insieme dei comuni con meno di 10.000 abitanti ‘B1’ e ‘B2’, in basso, riservando la zona centrale alle periferie delle aree di grande urbanizzazione ‘A2’ e agli altri comuni con più di 10.000 abitanti ‘B3’ e ‘B4’<sup>7</sup>. Esaminando l’andamento del ‘tasso di caduta’, nel primo piano fattoriale, si rileva una spezzata con andamento obliquo che risulta correlata sia con la tipologia comunale, che con gli anni di attività. Al contrario, il tasso di risposte proxy non risulta particolarmente legato a queste variabili, se non, in minima parte, con quelle connesse al primo asse.

Le categorie ‘Valida’ e ‘Invalida’ suddividono la popolazione in due gruppi: i rilevatori ritenuti ‘affidabili’ (telefonata di controllo con esito positivo) e quelli risultati ‘non affidabili’ (telefonata di controllo con esito negativo). Come previsto, la modalità ‘Valida’, relativa al 97,6% dei rilevatori, risulta collocata vicino all’origine, mentre la modalità ‘Invalida’, essendo una categoria atipica, si trova all’estremità del grafico, nel primo quadrante.

<sup>7</sup> I comuni sono suddivisi nelle seguenti classi: ‘A1’ area urbana centro (comuni di Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari); ‘A2’ area urbana periferia (comuni delle cinture urbane); ‘B1’ altri comuni fino a 2.000 abitanti; ‘B2’ da 2.001 a 10.000 abitanti; ‘B3’ da 10.001 a 50.000 abitanti; ‘B4’ oltre 50.000 abitanti.

### 3.2 Una tipologia di qualità della rilevazione

La presenza di due fattori linearmente indipendenti legati, rispettivamente, alle interviste cadute e alle risposte proxy suggerisce l'esistenza di una partizione delle unità in quattro gruppi, corrispondenti ai diversi quadranti del primo piano dell'ACP.

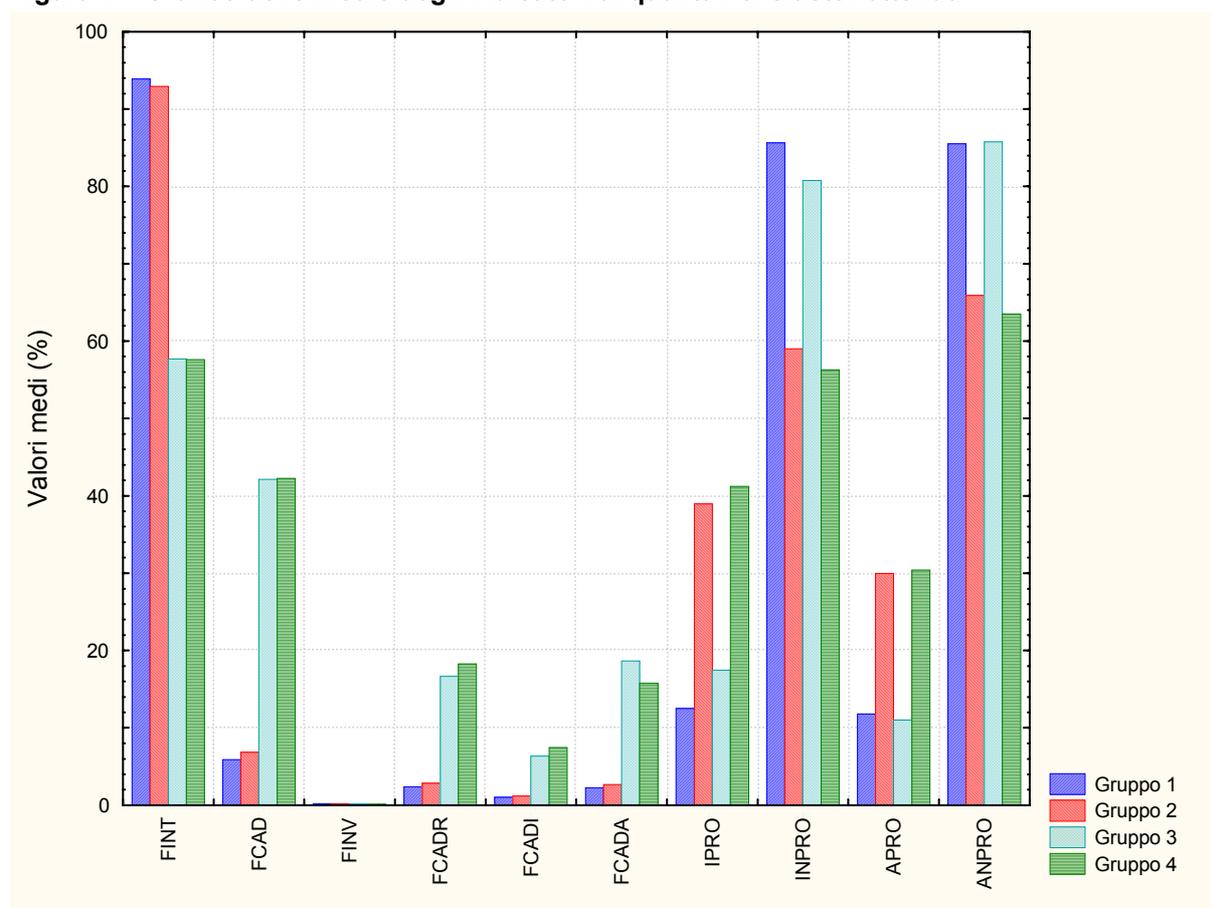
Tali classi tipologiche di rilevatori sono state ottenute mediante un'analisi dei gruppi<sup>8</sup> nello spazio degli indicatori di qualità dell'attività di rilevazione.

Nella figura 4 è rappresentato il grafico delle medie degli indicatori di qualità per i cluster ottenuti. Come si può notare, l'insieme dei rilevatori si divide in due gruppi: nel primo gruppo il valor medio di FINT è pari al 93% (FCAD = 6%), mentre nel secondo è inferiore al 58% (FCAD = 42%); all'interno di tali gruppi si distinguono due ulteriori sottoinsiemi aventi, rispettivamente, dei valori medi di IPRO pari, all'incirca, al 15% e al 40% (analoghe considerazioni valgono per APRO).

Pertanto, è possibile definire sinteticamente le quattro classi nel seguente modo:

- gruppo 1: poche interviste cadute e poche risposte proxy (1.496 unità);
- gruppo 2: poche interviste cadute e molte risposte proxy (1.535 unità);
- gruppo 3: molte interviste cadute e poche risposte proxy (355 unità);
- gruppo 4: molte interviste cadute e molte risposte proxy (298 unità).

Figura 4 – Grafico delle medie degli indicatori di qualità nei cluster ottenuti



<sup>8</sup> L'analisi è stata condotta mediante il metodo aggregativo non gerarchico di MacQueen, detto delle *K-medie*, particolarmente utile quando si è in presenza di un elevato numero di osservazioni. Si noti che i metodi non gerarchici, avendo la possibilità di spostare un elemento da un gruppo all'altro, consentono di ottenere dei raggruppamenti più omogenei di quelli prodotti dai metodi gerarchici. Essi, pertanto, risultano più efficaci di quelli gerarchici nell'individuare gruppi con forte omogeneità interna (Delvecchio, 1992).

Una volta individuati i gruppi, si è proceduto alla lettura del profilo dei rilevatori inclusi nei diversi cluster secondo le variabili strutturali più significative, allo scopo di caratterizzare le classi tipologiche ottenute. I risultati sono illustrati nel prospetto 2.

**Prospetto 2 - Composizione percentuale delle principali variabili strutturali nei cluster ottenuti**

Variabili	Gruppo 1	Gruppo 2	Gruppo 3	Gruppo 4	Totale
<b>SESSO</b>					
Maschio	47,4	45,6	43,9	43,3	46,0
Femmina	52,6	54,4	56,1	56,7	54,0
<b>ETA'</b>					
fino a 34 anni	26,1	29,2	34,1	40,9	29,3
35 - 44 anni	25,1	23,3	25,4	24,8	24,3
45 anni e oltre	48,8	47,6	40,6	34,2	46,3
<b>STATO CIVILE</b>					
Single	40,3	42,1	53,5	60,7	44,0
Coniugato	59,7	57,9	46,5	39,3	56,0
<b>TITOLO DI STUDIO</b>					
Diploma inferiore	15,4	14,5	13,8	11,7	14,6
Diploma superiore	72,7	74,0	69,9	69,5	72,7
Laurea	10,6	10,4	14,4	16,4	11,3
Altro	1,3	1,1	2,0	2,3	1,4
<b>RAPPORTO DI LAVORO</b>					
Dipendente comunale					
- Ufficio di statistica o anagrafe	57,3	54,7	47,9	42,6	54,1
- Polizia municipale	11,2	10,6	7,9	10,1	10,6
Non dipendente comunale					
- Studente o disoccupato	24,5	28,9	34,9	44,0	28,9
- Altro	7,0	5,8	9,3	3,4	6,4
<b>RESPONSABILE COMUNALE</b>					
No	87,6	89,8	97,2	97,0	90,2
Si	12,4	10,2	2,8	3,0	9,8
<b>ANNI DI ATTIVITA'</b>					
0 - 1 anni	24,6	7,8	23,1	10,7	16,3
2 - 3 anni	31,7	41,6	31,8	32,9	35,9
4 - 5 anni	12,8	14,5	14,6	25,2	14,7
6 anni e oltre	30,9	36,2	30,4	31,2	33,1
<b>TIPO DI COMUNE</b>					
Area urbana centro	3,7	4,2	32,1	36,2	9,3
Area urbana periferia	10,5	11,3	8,7	6,7	10,4
Fino a 2.000 abitanti	11,2	8,9	3,9	2,3	8,8
Da 2.001 a 10.000 abitanti	36,3	32,7	13,5	11,7	30,6
Da 10.001 a 50.000 abitanti	26,7	28,4	16,6	17,4	25,7
Oltre 50.000 abitanti	11,6	14,5	25,1	25,5	15,3
<b>RIPARTIZIONE GEOGRAFICA</b>					
Nord	46,1	39,3	63,4	55,0	45,7
Centro	16,8	17,1	19,4	17,1	17,2
Sud	27,7	30,4	9,0	11,4	25,7
Isole	9,4	13,2	8,2	16,4	11,4

Dall'esame della tabella risulta evidente che i cluster con poche interviste cadute (gruppi 1 e 2) sono costituiti, prevalentemente, da dipendenti comunali, di oltre 45 anni di età, coniugati, con livello di istruzione medio-basso, operanti nei piccoli centri (l'11% di essi sono anche responsabili di indagine)<sup>9</sup>. Di questi, l'insieme con il minor numero di risposte proxy (gruppo 1) non ha, in buona parte, più di un anno di attività e si trova al Nord, mentre coloro che registrano il maggior numero di risposte proxy (gruppo 2) hanno almeno 2 anni di attività e lavorano al Sud e nelle Isole.

I cluster con molte interviste cadute (gruppi 3 e 4) sono caratterizzati da studenti o disoccupati, di età inferiore ai 35 anni, single, con livello di istruzione medio-alto, operanti nei grandi centri urbani. Anche in questo caso, il gruppo con minore incidenza di risposte proxy (gruppo 3) si distingue per una maggiore presenza, rispetto alla media, di rilevatori alle prime armi, concentrati soprattutto al Nord e al Centro; mentre i rilevatori con maggiore incidenza di risposte proxy (gruppo 4) tendono ad avere più anni di esperienza e agiscono al Nord e nelle Isole.

Se da una parte, quindi, il tasso di caduta risulta influenzato principalmente dalla dicotomia centri urbani - piccoli centri; il tasso di proxy sembra, invece, legato agli anni di attività e alla ripartizione geografica, come già evidenziato dall'ACM.

Nel prospetto 3 è riportata, per ciascuna delle classi ottenute, la stima della probabilità che un rilevatore sia 'non affidabile' ovvero che abbia una probabilità maggiore di zero di effettuare una intervista non valida<sup>10</sup>.

### Prospetto 3 - Probabilità che un rilevatore sia 'non affidabile'

Descrizione	Gruppo 1	Gruppo 2	Gruppo 3	Gruppo 4	Totale
N. di rilevatori	1.496	1.535	355	298	3.684
- di cui con interviste non valide	34	32	12	9	87
Probabilità di individuare un rilevatore 'non affidabile'	0,023	0,021	0,034	0,030	0,024
Probabilità di effettuare una intervista non valida per un rilevatore 'non affidabile'	0,099	0,112	0,084	0,063	0,098
Probabilità che un rilevatore sia 'non affidabile'	0,23	0,19	0,40	0,48	0,24

<sup>9</sup> Il responsabile di indagine è colui che coordina i rilevatori nell'ambito del proprio Comune. Per la natura della funzione che svolge, il responsabile di indagine non dovrebbe essere un rilevatore, ma ci sono contesti in cui questo requisito è difficile da soddisfare. Pertanto, ci sono rilevatori che sono anche responsabili di indagine e, in tal caso, si tratta di dipendenti comunali.

<sup>10</sup> Definiamo i seguenti eventi:

R = il rilevatore è affidabile /  $\bar{R}$  = il rilevatore non è affidabile;

C = il controllo ha esito positivo (intervista valida) /  $\bar{C}$  = il controllo ha esito negativo (intervista non valida).

Allora, la probabilità di individuare un rilevatore 'non affidabile' è data dalla formula:

$$P(\bar{C}) = P(\bar{C} | R) P(R) + P(\bar{C} | \bar{R}) P(\bar{R}) \quad \text{dove } P(\bar{C} | R) = 0,$$

da cui

$$P(\bar{R}) = P(\bar{C}) / P(\bar{C} | \bar{R}).$$

Nel complesso, si ha:

$$P(\bar{C}) = 87 / 3.684 = 0,024,$$

mentre  $P(\bar{C} | \bar{R})$  può essere stimata, sulla base dei dati relativi agli 87 rilevatori con interviste non valide, mediante il valor medio del rapporto tra il numero di interviste annullate e il numero di interviste effettuate ed è pari a 0,098.

Quindi:

$$P(\bar{R}) = P(\bar{C}) / P(\bar{C} | \bar{R}) = 0,024 / 0,098 = 0,24.$$

Si osservi che i valori della probabilità nei cluster con tassi di caduta più alti (gruppi 3 e 4) sono sensibilmente superiori ai valori corrispondenti ai cluster con tassi di caduta più bassi (gruppi 1 e 2). In particolare, i rilevatori del gruppo 4 hanno una probabilità di effettuare delle interviste non valide pari a 0,48; pertanto si stima che, tra i rilevatori con elevati tassi di caduta e di risposte proxy, uno su due non sia affidabile.

#### **4. Conclusioni**

I risultati ottenuti, seppur limitati ad un'unica indagine in un dato triennio, mostrano un interessante legame tra gli indicatori di qualità dell'attività di rilevazione e le caratteristiche strutturali dei rilevatori. In particolare, lo studio dell'interdipendenza ha evidenziato un forte legame tra il tasso di interviste cadute e la tipologia comunale (area metropolitana – piccoli centri), mentre il tasso di risposte proxy sembra maggiormente connesso agli anni di attività del rilevatore e alla ripartizione geografica del comune di appartenenza.

L'analisi dei dati e le riflessioni sui risultati possono inoltre contribuire al miglioramento e all'affinamento delle tecniche di monitoraggio della fase di rilevazione nelle indagini campionarie sulle famiglie. Per quanto riguarda, infatti, lo studio delle tipologie di rilevatori, appare evidente l'utilità di concentrare le telefonate di controllo sui rilevatori con elevati tassi di caduta e di risposte proxy che operano nei grandi centri urbani, in modo da individuare il maggior numero di interviste non valide, migliorando, così, la qualità dei dati.

L'uso del SIRC, come strumento per monitorare il funzionamento della rete comunale, costituisce quindi un valido ausilio nel controllo degli errori non campionari prodotti nella fase di rilevazione, con particolare riferimento all'attività dell'intervistatore.

#### **Bibliografia**

- BOUROCHE J.M. E G. SAPORTA (1983). *L'analyse des données*. CLU, Napoli.
- CAMPORESE R., QUATTROCIOCHI L. E SABBADINI L.L. (2001). *Strategie di prevenzione delle mancate risposte e degli errori non campionari nelle indagini Multiscopo*. SIEDS.
- DELVECCHIO F. (1992). *Analisi statistica di dati multidimensionali*. Cacucci, Bari.
- ISTAT (2003). *I rilevatori comunali delle indagini campionarie sulle famiglie, Serie storica 1994-2001* (documento interno).