

Censimento economico e censimento demografico: quali innovazioni possibili ?

di Andrea Mancini

Palermo, 5 ottobre 2007

Industria e Servizi: un Censimento virtuale nel 2011?

- **IL CONTESTO ATTUALE:**

- Nessun paese della Unione Europea fa più un censimento economico con rilevazione esaustiva
- Esso non è richiesto da Unione Europea nè raccomandato a livello internazionale
- L'informazione di dettaglio territoriale sulle imprese è ormai prodotta annualmente da Istat con ASIA-UL
- Dopo il CIS 2001 nessuno ha chiesto di riadottare la Long Form

- **ALLORA QUALE OBIETTIVO PER IL 2011?**

- Garantire un prodotto informativo statistico simile a DWH di CIS 2001

Industria e Servizi: un Censimento virtuale nel 2011?

- **LA STRATEGIA COMPLESSIVA PER IL 2011:**
- Per le imprese usare tutte le informazioni di ASIA – UL
- Eventualmente per le imprese condurre una rilevazione di controllo
- Per le istituzioni pubbliche fare una rilevazione postale con questionario Long Form (anche elettronico) con fogli aggiuntivi (uno per ciascuna UL di Unità Istituzionale)
- Per le istituzioni private non profit fare due rilevazioni (Cfr prossima slide)
- Geocodificare alla sezione di censimento ciascuna Unità Locale mediante uso di archivi comunali di numeri civici

Il Censimento delle istituzioni private non profit

A) TAPPA N°1 : 2° censimento nel 2009, dopo 10 anni dal precedente

- Unità di rilevazione = unità istituzionale
- Metodo = censimento su lista con invio postale
- Organizzazione = simile a quella del 1999
- Questionario = solo Long Form simile a quello del 1999 (ma anche in versione elettronica)
- **Unità di analisi = solo unità istituzionale non profit**

B) TAPPA N°2 : Cis 2011

- Unità di rilevazione = unità istituzionale
- Metodo = censimento su lista con invio postale
- Organizzazione = uguale a quella del 2009
- Questionario = solo Short Form (uno per UL, anche in versione elettronica)
- **Unità di analisi = solo Unità Locale di istituzione non profit**

- **IMPLICAZIONI**

- I risultati dovrebbero garantire parità di informazione rispetto a quella diffusa nel 1999 (non profit) e nel 2001
- Con due approfondimenti: istituzioni pubbliche e istituzioni private non profit
- Per le istituzioni non profit la rilevazione del 2009 può essere occasione per impiantare un registro statistico settoriale da tenere poi aggiornato mediante fonti amministrative
- Per le imprese i risultati della rilevazione 2011 di controllo potrebbero essere occasione di verifica dell'eventuale sottocopertura di ASIA – UL
- Nel 2011 rispetto al 2001 vengono liberate consistenti risorse organizzative degli organi intermedi di censimento a favore del censimento della popolazione

Perchè introdurre innovazioni in metodi e tecniche del Censimento della popolazione?

- Migliorare la coerenza tra dati di struttura demografica dell'anagrafe e del censimento anche in una prospettiva postcensuaria
- Maggiore tempestività (trasmissione ad Eurostat entro marzo 2014 di tutti i dati del Cens. Pop.)
- Riduzione fastidio statistico sulle famiglie
- Rafforzamento degli uffici di statistica
- **ma nel rispetto delle Raccomandazioni UN-ECE:**
 - enumerazione completa (*Universality*)
 - enumerazione individuale (*Individual enumeration*)
 - unicità del tempo di riferimento dei risultati (*Simultaneity*)
 - grande dettaglio territoriale (*Small area-data*)
 - periodicità definita (*Defined periodicity*)

Due scenari di innovazione con moduli comuni ad entrambi

- 1° SCENARIO:
 - “Triennializzazione della rilevazione nei comuni di maggiori dimensioni” per gruppi annuali di sezioni di censimento °
- 2° SCENARIO:
 - “Impiego di liste anagrafiche precensuarie”
- Entrambi gli scenari hanno in comune alcuni moduli di innovazione:
 - A)- diversificazione di metodi e organizzazione tra comuni di diversa classe di ampiezza demografica
 - B)- formazione aree di censimento subcomunali nell’aggiornamento basi territoriali
 - C)- somministrazione di Long Form a campioni di famiglie/individui
 - D)- revisione delle anagrafi coordinata con la determinazione della popolazione legale

Le classi dei comuni (le soglie sono solo indicative)

Tipo A Comuni con popolazione residente di almeno 50.000 abitanti o comuni capoluogo di provincia;

Tipo B Comuni non capoluogo di provincia con una popolazione residente tra i 10.000 (20.000) e 49.999 abitanti;

Tipo C Comuni con popolazione residente inferiore ai 10.000 (20.000) abitanti.

Basi territoriali e aree di censimento (Cfr. Relazione Fabio Crescenzi)

- **Aggiornamento basi territoriali:** da effettuare a partire dal 2008, mantenendo il più possibile il disegno e i codici delle sezioni di Census 2001; permesse variazioni solo in casi di necessità (ad es: espansione delle località abitate)
- **Aree di censimento:** per ciascun Comune di Tipo A e di Tipo B, in collaborazione con il Comune, saranno individuate aree di censimento **di centro abitato** come insiemi di sezioni limitrofe, contestualmente all'aggiornamento delle basi territoriali
- **Vincoli:** il loro disegno dovrà consentire la ricostruzione, per somma di aree, delle esistenti zone subcomunali a valenza amministrativa (municipi, circoscrizioni, rioni, ecc.)
- **Metodo:** ogni area di censimento deve essere definita come insieme di sezioni di centro abitato limitrofe' che rispetti soglie dimensionali di popolazione e minimizzi le distanze tra vettori di indicatori socio-demografici al 2001 riferiti a ciascuna sezione componente

Short Form – Long Form

Questionari SF e LF nei comuni di Tipo A e B. Nei comuni di Tipo C solo questionari LF

Le variabili SF saranno rilevate esaustivamente, consentendo l'enumerazione di individui e famiglie e l'ottenimento di dati dettagliati per sezione di censimento relativi alle principali caratteristiche demografiche della popolazione (sesso, età, stato civile, cittadinanza, ecc.) e delle famiglie (tipologia di nuclei)

Le informazioni relative alle variabili LF saranno rilevate in ciascun comune di Tipo A e di Tipo B presso un campione di famiglie. Il metodo di campionamento dovrà essere determinato in dipendenza dello scenario che sarà prescelto

1° Scenario : campionamento areale a grappolo

2° Scenario : campionamento casuale semplice o stratificato

1° Scenario: la proposta complessiva

Schema di rilevazione relativo ai comuni di Tipo A in base al tipo di area di censimento

Tipo Area di censimento	Anni/Gruppi di sezioni		
	$G_{t,2}$	$G_{t,1}$	G_t
<i>Di centro abitato</i>	SF	SF	SF LFC
<i>Di altro tipo</i>			LFe

Schema di rilevazione relativo ai comuni di Tipo B in base al tipo di area di censimento

Tipo Area di censimento	Anni/Gruppi di sezioni		
	$G_{t,2}$	$G_{t,1}$	G_t
<i>Di centro abitato</i>			SF LFC
<i>Di altro tipo</i>			LFe

Schema di rilevazione relativo ai comuni di Tipo C in base al tipo di area di censimento

Tipo Area di censimento	Anni/Gruppi di sezioni		
	$G_{t,2}$	$G_{t,1}$	G_t
<i>Di centro abitato</i>			LFe
<i>Di altro tipo</i>			LFe



Short form



Long form campionaria



Long form esaustiva

1° Scenario: le conseguenze sul metodo campionario

- In assenza di liste precensuarie di famiglie o individui il disegno campionario non può che essere **areale a grappolo**
- L'unità campione è la sezione di censimento (=grappolo di famiglie)
- Per i centri abitati dei comuni di Tipo A e di Tipo B ci sono tanti universi di riferimento quante sono le aree di censimento di centro abitato pre-individuate
- Per ciascuna area di censimento di centro abitato l'universo di riferimento è costituito dalle sezioni di censimento ad essa interne
- In ciascuna sezione inserita nel campione la Long Form viene distribuita a tutte le famiglie residenti (= grappolo)
- Nei comuni di Tipo A tutte le sezioni campione LF fanno parte del gruppo Gt, sottoposto a rilevazione nel 2011

1° Scenario: le conseguenze del campione areale

- Il disegno campionario **areale a grappolo** comporta errori campionari più elevati di quelli ottenibili, a parità di altre condizioni, da campioni casuali semplici o stratificati a causa dell'**effetto cluster**
- Questo fa aumentare la varianza dello stimatore in misura direttamente proporzionale all'omogeneità interna e alla dimensione demografica delle sezioni campione
- Quindi le stime delle variabili LF per singola area di censimento potrebbero essere affette da **errori campionari di livello non accettabile per un censimento**
- Stiamo conducendo delle sperimentazioni su dati 2001 per verificare l'entità dell'errore campionario indotto dall'effetto cluster

1° Scenario: le conseguenze della triennializzazione nei comuni di Tipo A

- **Il problema:**
- aggiornare al 2011 i dati di struttura demografica (individui e famiglie) rilevati con SF nelle sezioni di centro abitato rientranti nei gruppi Gt-2 e Gt-1
- **Flussi di dati già attivi tra i comuni e l'Istat:**
- **dati aggregati** relativi a nascite, morti, trasferimenti di residenza e acquisizioni di cittadinanza;
- periodicità di invio mensile con riepiloghi annuali
- **dati individuali** relativi a nascite, morti, trasferimenti di residenza e acquisizioni di cittadinanza;
- periodicità di invio mensile e riepiloghi annuali
- **i dati individuali consentono** classificazioni per sesso, età, stato civile e cittadinanza;
- attraverso l'indirizzo di residenza consentirebbero l'aggiornamento di gran parte delle variabili *short form*

1° Scenario: le conseguenze della triennializzazione nei comuni di Tipo A

- **Flussi di dati da attivare tra i comuni e l'Istat:**
- dati individuali sui cambi di abitazione all'interno dello stesso comune;
- dati individuali relativi ai cambiamenti di status all'interno della famiglia, con particolare riferimento alla relazione di parentela degli individui.

1° Scenario: le conseguenze della triennializzazione nei comuni di Tipo A

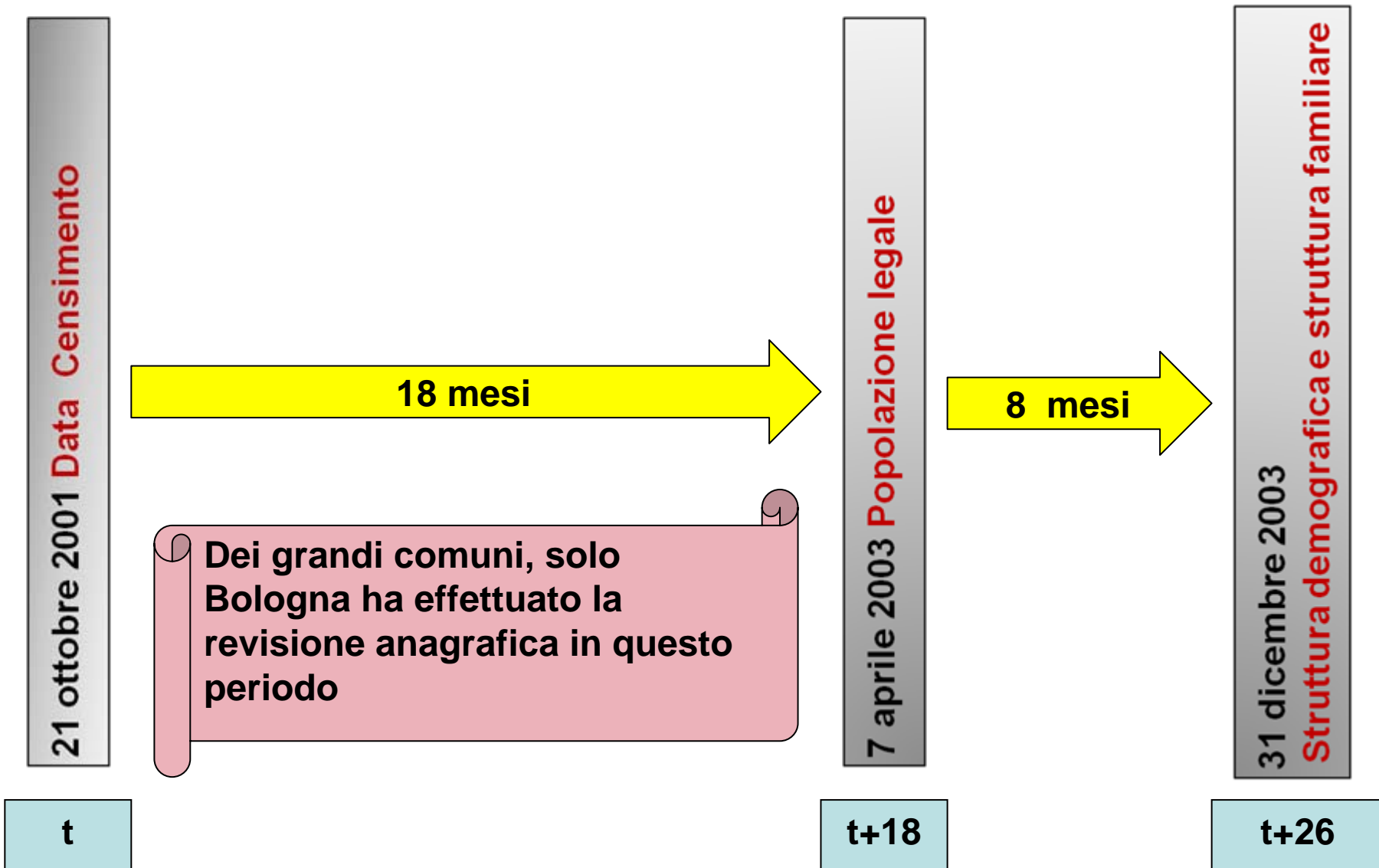
- **Le difficoltà emergenti:**
- Per aggiornare al *tempo t* occorre fare il linkage tra schede anagrafiche individuali e di famiglia e i questionari rilevati al tempo t-2 e t-1;
- i dati individuali spesso risultano incompleti nell'indirizzo di residenza, rendendo impossibile la geocodificazione alla sezione di censimento
- il livello di completezza delle informazioni individuali di fonte anagrafica richiede specifici controlli sui tempi e sulla completezza della registrazione degli eventi in anagrafe
- **Cosa stiamo facendo:**
- studiando l'impiego di modelli probabilistici per operare l'aggiornamento al tempo t

1° Scenario: la triennializzazione come opportunità organizzativa

- A)- dilazionare i tempi della rilevazione nei comuni di più ampie dimensioni può contribuire a migliorare il grado di copertura delle loro operazioni censuarie
- Post-enumeration survey 2001:
- A livello nazionale il tasso di copertura è stato del 98,5 per cento;
- Tuttavia esso varia da un **massimo del 99,30%** per i comuni sotto i 10mila abitanti ad un **minimo del 96%** per i comuni metropolitani;
- Per Italia la stima degli sfuggiti al censimento è di 833.417 unità; di queste il 44,6% riguarda i comuni metropolitani.

- B)- la triennializzazione consente di dare più tempo ai comuni di maggiori dimensioni per effettuare il confronto censimento-anagrafe e anche per concludere la revisione anagrafica prima della determinazione della popolazione legale
- Analisi degli scostamenti tra popolazione legale e anagrafica
- Al 21/10/2001 Roma, Milano, Napoli, Torino e Palermo coprono il 32% della complessiva differenza negativa tra popolazione legale e anagrafica (-**1.588.332** individui);
- questa quota è 3 volte maggiore del loro peso sulla popolazione legale nazionale (11,2%).

Cosa è successo nel 2001

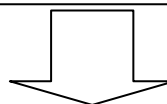


2° Scenario: la proposta complessiva

Lista anagrafica comunale di individui, famiglie e convivenze
(LAC)

Formazione delle liste ausiliarie (LIFA)

Formazione della lista di Numeri Civici (RNC) basata su
Archivio/Rilevazione



Operazioni sul campo

Distribuzione e Raccolta
dei questionari da LAC:

SF e LF
parzialmente
precompilati

SF precompilati

Rilevazione integrativa di unità
non presenti in lista:

Recupero totale sul territorio

Metodo cattura-ricattura

2° Scenario: usare le liste anagrafiche comporta errori di copertura

Palermo 5 ottobre
2007

Le Anagrafi possono essere affette da:

-**Sovracopertura**: il censimento deve riuscire a depurare i propri risultati da famiglie e individui che non dimorano effettivamente nel Comune e nell'area di censimento.

-**Criticità**: verifica sul campo affidata allo stesso agente che gestisce l'anagrafe

-**Sottocopertura**: il censimento deve riuscire a recuperare famiglie e individui non iscritti in anagrafe, ma abitualmente dimoranti nel Comune e nell'area di censimento

-**Criticità**: acquisizione di informazioni da fonti terze o mediante la ricognizione sul campo

-**Quindi** è necessario dotarsi di metodi e tecniche che consentano di sanare questi errori: **ricorso a fonti ausiliarie e terze rispetto all'anagrafe**

2° Scenario: Lista anagrafica comunale di famiglie, convivenze e individui (LAC)

- L'Istat riceve dal Comune la lista con campi e formati predefiniti (nome, sesso, cod. fisc. Intestatario foglio di famiglia, indirizzo completo dell'abitazione, sez. di cens.ecc.)
- Da rendere disponibile con largo anticipo rispetto alla data della rilevazione censuaria.
- L'Istat normalizza e rinvia al Comune

Le fonti locali e centrali delle liste integrative devono essere:

- autonome rispetto alle liste anagrafiche
- riferibili a famiglie, convivenze ed individui ad una certa data prefissata, uguale a quella di riferimento dell'estrazione delle LAC
- normalizzate, validate, integrate e condivise entro tempi certi prima della rilevazione censuaria

Impiego a supporto delle attività di *back office*

- Le informazioni supplementari a quelle anagrafiche sono usabili per guidare le operazioni di recupero della sottocopertura
- Maggiore impegno sul campo per la correzione delle anagrafi e la loro integrazione con nuove unità (famiglie e individui)

2° Scenario: Lista di numeri civici [RNC]

- Riferita a numeri civici geocodificati a sez. cens., almeno quelli con un'unità immobiliare potenzialmente abitativa
- Costruita per integrazione di fonti disponibili (archivi Postali e Comunali, CP EDI 2001 e itinerari di sezione)
- Normalizzata, validata, condivisa tra Istat e Comune e integrata entro tempi certi prima della rilevazione censuaria
- Verificata sul campo da UCC con rilevazione di controllo e completamento con informazioni su edifici, abitazioni e altre unità immobiliari
- Il prodotto dovrà essere disponibile con largo anticipo rispetto alle tradizionali operazioni censuarie (es. gennaio 2011)

2° Scenario: Distribuzione e raccolta questionari da LAC (1)

- Attribuzione di modalità SF oppure LF a ciascuna famiglia in base a liste di unità campione per comune (a cura dell'Istat)
- Stampa di questionari SF e LF precompilati solo nel codice identificativo della famiglia e relative etichette d'indirizzo
- Distribuzione: *mail-out* postale per SF e LF
- Raccolta multicanale: *mail-in* postale, Web, CATI (solo SF) e da Rilevatore
- Condivisione del Sistema di Monitoraggio (ISTAT-UCC) con tracciatura di invii e restituzioni

2° Scenario: Distribuzione e raccolta questionari da LAC (2)

Solleciti mirati da parte degli UCC secondo il piano di censimento:

- 1) telefonici ove possibile
- 2) postali con nuovo invio del questionario

Recuperi mediante rilevatore mirati a non rispondenti già presenti in LAC

- 1) dopo fase di sollecito mirato
- 2) con supporto di segnali raccolti nella fase di *mail-out* e da fonti ausiliarie.
- 3) a differenza del censimento tradizionale, il rilevatore viene inviato presso specifici indirizzi

2° Scenario: Distribuzione e raccolta questionari da LAC (3)

Le criticità possibili:

- Dimensionamento dei canali di restituzione, in funzione dell'orientamento delle famiglie verso: Poste, CATI o WEB, Rilevatore
- Gestione di CATI "passiva"; il *call center* attende la chiamata della famiglia e gestisce l'agenda degli appuntamenti
- Dimensionamento delle risorse necessarie al recupero mediante rilevatore e rafforzamento del *back office* di UCC
- Certificazione alle famiglie rispondenti:
 1. motivata dall'obbligo di risposta
 2. oneri e costi differenziati secondo il canale di restituzione

2° Scenario: Distribuzione e raccolta questionari da LAC (4)

Una variante:

- Precompilazione di quesiti (SF e LF) con informazioni desunte da schede anagrafiche
- Vantaggi addizionali:
 1. conferma di SF senza variazioni mediante WEB o AVR (*Automatic Voice Responder*)
 2. ulteriore contenimento dei costi di rilevazione sul campo
- Pericoli addizionali:
 1. aumento dei costi e della complessità della stampa questionari
 2. incentivo a NON segnalare errori nei campi precompilati
 3. SF molto contenuta con rischio di raccogliere informazioni insufficienti rispetto a censimenti precedenti e/o a richieste di Regolamento europeo

2° Scenario: Rilevazione integrativa di unità non presenti in LAC (1)

- **Come si può recuperare la sottocopertura della LAC?**
- MODO A: Impiego mirato di rilevatori di UCC per il recupero totale sul territorio. Esso comporta:
 - generazione delle informazioni necessarie a ottimizzare l'attività del rilevatore in base alle liste RNC e LIFA;
 - ricerca sul territorio delle unità NON già incluse in LAC (a cura dei rilevatori, contestualmente ai recuperi delle mancate risposte a *mail-out*);
 - maggiore importanza dell'opera di *back office* di UCC per direzionare sul territorio i rilevatori di recupero.

2° Scenario: Rilevazione integrativa di unità non presenti in LAC (1)

- **Come si può recuperare la sottocopertura della LAC?**
- MODO B: Impiego del metodo “cattura e ricattura”. Esso consiste in:
 - modello statistico basato su due numerazioni indipendenti di una stessa popolazione, entrambe affette da sottotifica, con l’obiettivo di stimarne la dimensione ignota;
 - la prima rilevazione è solo su lista anagrafica precensuaria; la seconda rilevazione è campionaria areale per sez.;
 - la seconda non permette la revisione sistematica delle anagrafi perché stima i livelli di popolazione residente ma non dà completezza di informazioni individuali