

# Il Censimento della Popolazione e Nuove Metodologie

METODOLOGIE STATISTICHE E SOLUZIONI PROTOTIPALI IDONEE A  
RISOLVERE PROBLEMATICHE RIGUARDANTI LE INNOVAZIONI INTRODOTTE  
PER IL PROSSIMO CENSIMENTO DELLA POPOLAZIONE

**Carlo Trivisano**

Dipartimento di Scienze Statistiche – Università di Bologna

**NUMERI PER DECIDERE**

Biblioteca dell'Assemblea Legislativa

Regione Emilia Romagna

Bologna, 11 aprile 2011

## Sommario

**Classificazione dei censimenti**

**Le Post enumeration surveys**

**Il censimento italiano della popolazione del 2011**

L'impiego di liste provenienti da fonti amministrative

L'uso di campioni al censimento del 2011

## Classificazione dei censimenti

strategie di conduzione dei censimenti:

- 1) censimento convenzionale;
- 2) censimento basato su archivi amministrativi;
- 3) censimento basato sulla combinazione di rilevazione esaustiva e indagini campionarie;
- 4) rolling census

Abbatini D., Cassata L., Martire F., Reale A., Ruocco G., Zindato D. (2007) La progettazione dei Censimenti generali 2010-2011 – 2. Analisi comparativa di esperienze censuarie estere e valutazione di applicabilità di metodi e tecniche ai censimenti italiani, *Numero 9/2007 Documenti Istat*

## Censimento convenzionale (1)

Rilevazione esaustiva e simultanea dei dati, tramite questionario cartaceo auto-compilato, consegnato e ritirato dai rilevatori.

Al fine di rispettare il requisito della simultaneità, la rilevazione viene condotta nell'arco di un periodo di tempo limitato, immediatamente successivo alla data di riferimento.

## Censimento convenzionale (2)

censimenti register-supported: i dati amministrativi servono a costituire liste pre-censuarie e a precompilare parte dei questionari.

Censimenti register-improved: i dati amministrativi sono utilizzati per il controllo della copertura

I censimenti register-supported e quelli register-improved possono essere considerati casi particolari dei censimenti convenzionali,

## Censimento basato su archivi amministrativi

Tipologia di censimento alla quale si arriva al termine di un lungo processo, nel corso del quale si continua a condurre, almeno in parte, un censimento convenzionale.

I primi dati estratti dagli archivi sono indirizzi, dati demografici di base, informazioni di stato civile, e dati sul reddito, ma la quota di informazioni derivate dagli archivi aumenta gradualmente, censimento dopo censimento.

Vengono creati registri statistici a partire da dati di archivi amministrativi; frequentemente sono basati su una combinazione di dati provenienti da più fonti amministrative

## Censimento basato sulla combinazione di rilevazione esaustiva e indagini campionarie (1)

Si tratta di una variazione dell'approccio convenzionale.

Obiettivo: migliorare la qualità dei dati rilevati, ridurre i costi e il carico statistico sui rispondenti, rendere più tempestivo il rilascio delle informazioni.

Al conteggio esaustivo della popolazione viene abbinato il ricorso a tecniche campionarie per la rilevazione delle informazioni per le quali non sia strettamente richiesta la copertura universale.

I campioni che integrano i censimenti sono di dimensioni più ampie rispetto a quelle usate per i campioni nelle indagini correnti.

## Censimento basato sulla combinazione di rilevazione esaustiva e indagini campionarie (2)

Nei paesi che adottano questa strategia, il censimento viene condotto tramite l'uso combinato di *short form* e *long form*.

*Short form*: utilizzata per il conteggio esaustivo della popolazione e la rilevazione di poche caratteristiche socio-demografiche.

*Long form*: utilizzata su base campionaria per la rilevazione delle altre informazioni.

La decisione su quali informazioni rilevare tramite l'indagine campionaria dipende dal grado di precisione con cui è necessario produrre i risultati relativi a ciascuna variabile per piccole aree e per piccoli gruppi di popolazione.



## Rolling census

Proposto da Kish (1990) come alternativa al censimento convenzionale, ai censimenti register-based e alle indagini campionarie.

Si tratta di un'indagine cumulativa continua, che copre l'intero territorio nell'arco di un periodo di tempo anche pluriennale, consentendo di produrre informazioni dettagliate per piccole aree più tempestivamente rispetto ai censimenti tradizionali.

E' un disegno di rilevazione che consiste in una serie di  $K$  rilevazioni campionarie che si susseguono nel tempo non sovrapponendosi: ogni rilevazione può essere vista come un campione probabilistico con una frazione di campionamento pari a  $k = 1/K$  dell'intera popolazione.

## Post enumeration surveys (1)

Indagini condotte a posteriori per valutare uno dei principali aspetti della qualità della rilevazione censuaria, ovvero il grado di copertura

Prerequisiti:

tempestività dell'indagine

l'utilizzo di rilevatori diversi da quelli censuari.

Il disegno campionario dipende strettamente dalla strategia di censimento utilizzata. Spesso è utilizzato il campionamento areale a due stadi.

Se al censimento sono stati utilizzati diversi metodi di rilevazione, può essere necessario misurare l'errore di copertura in modo distinto per ciascun metodo.

Presupposto fondamentale dell'indagine di copertura è l'indipendenza dalla rilevazione censuaria.

## Post enumeration surveys (2)

In Italia l'indagine di copertura è stata effettuata per gli ultimi tre censimenti

Indagine di copertura per l'ultimo censimento: 98 comuni coinvolti (di cui 14 autorappresentativi).

1154 sezioni campione stratificate per tipo di località abitata (centri, nuclei, case sparse/località produttive).

Oltre 68mila famiglie intervistate

## Post enumeration surveys (3)

Figura 1: Tassi di errore di copertura relativi ai censimenti del 2000/2001 in alcuni paesi

Paese	Anno	Errore (%)
Canada	2001	2,99
Nuova Zelanda	2001	2,20
Regno Unito	2001	2,00
<i>Londra</i>	<i>2001</i>	<i>16,00</i>
Repubblica Popolare Cinese	2000	1,81
Italia	2001	1,44
Svizzera	2000	1,41
<i>Zurigo</i>	<i>2000</i>	<i>1,46</i>
Estonia	2000	1,20
Stati Uniti d'America (USA)	2000	1,18
Australia	2001	1,08

## Evoluzione dei censimenti convenzionali (1)

A partire dagli anni 70 i paesi scandinavi hanno cominciato a pensare all'uso dei dati amministrativi a fini censuari.

Danimarca primo paese a condurre nel 1981 un censimento totalmente basato sull'uso di dati amministrativi, seguita dalla Finlandia nel 1990.

Svezia e Norvegia hanno cominciato a usare parzialmente i dati amministrativi rispettivamente nel 1980 e nel 1990, la Svezia in combinazione con un questionario breve per le persone di 16 anni e più e la Norvegia in combinazione con un'indagine campionaria

## Evoluzione dei censimenti convenzionali (2)

Su 44 paesi che hanno condotto un censimento nel 2000, 37 hanno utilizzato la strategia convenzionale (alcuni dei quali con questionari parzialmente prestampati), 5 una strategia mista (informazioni desunte da archivi amministrativi + rilevazione sul campo, esaustiva o campionaria), 2 una strategia interamente basata sull'uso dei registri.

Nel 2010: 27 i paesi che hanno continuato ad effettuare un censimento convenzionale; 17 i paesi che hanno condotto un censimento totalmente o parzialmente basato sugli archivi amministrativi

## Il censimento negli Stati uniti (1)

Dal 1940 e fino al censimento del 2000, il conteggio della popolazione è stato effettuato tramite *short form*, mentre un campione di famiglie (pari al 17% nel 2000) ha compilato la *long form*.

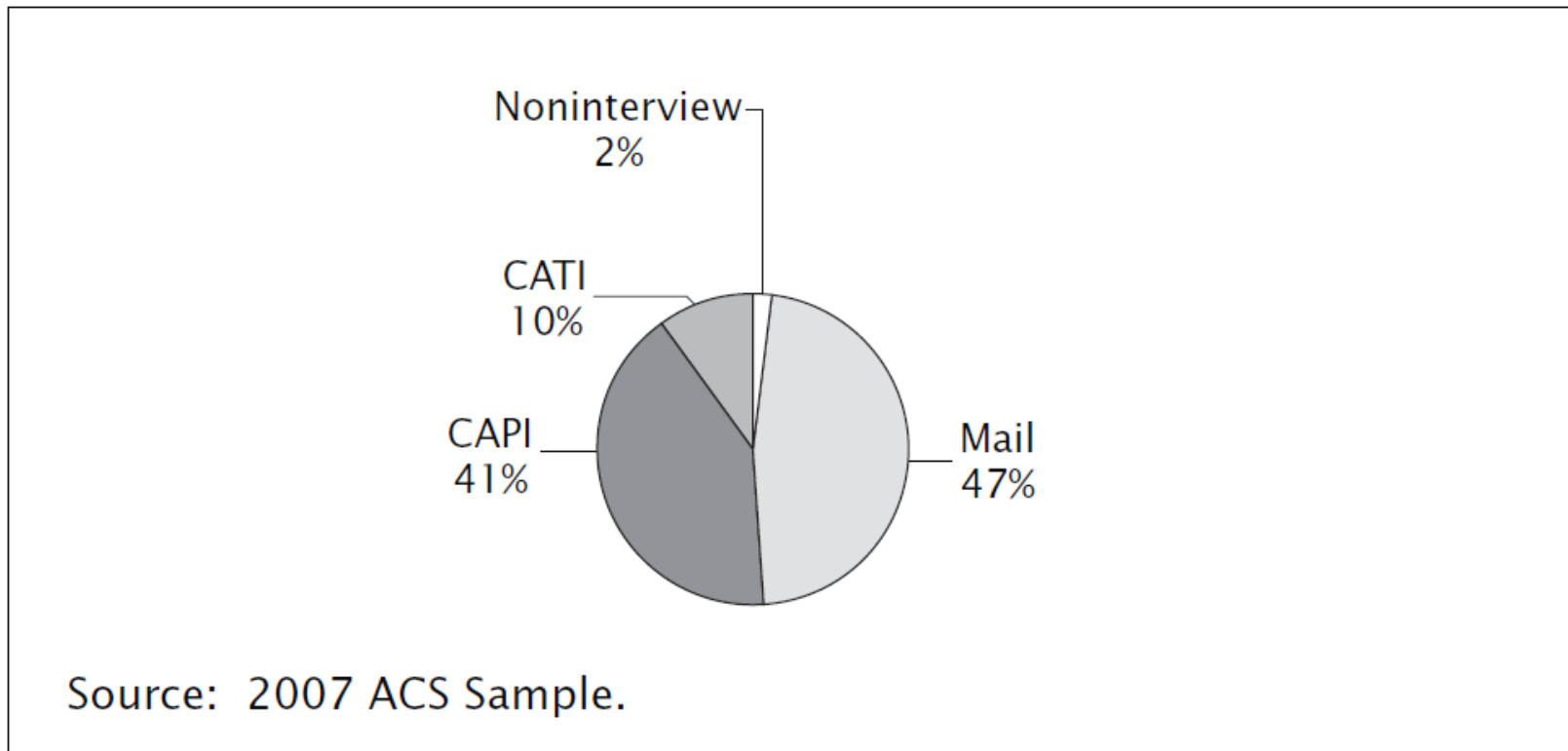
Nel 2010: *short form* ancora utilizzata per il conteggio esaustivo della popolazione e la rilevazione delle informazioni demografiche di base; mentre l'uso della *long form* è stato sostituito dall'ampliamento dell'ACS (*American Community Survey*), indagine campionaria continua di respiro nazionale che consente di fornire dati con frequenza annuale invece che decennale.

I tassi di campionamento annuale ai vari livelli di dettaglio territoriale variano dall'1,7% al 10% circa. Nell'arco di un periodo di cinque anni, i tassi di campionamento variano dall'8,5% al 50% circa.

Sono necessari cinque anni per produrre stime affidabili per le aree con popolazione inferiore ai 20.000 abitanti

## Il censimento negli Stati uniti (2)

Figure 7.2 **Distribution of ACS Interviews and Noninterviews**





## Un caso estremo

Per il censimento Canadese del 2011 si assumerà ancora una strategia (consolidata) di tipo *long form short form*

Fino al 2006 era prevista l'obbligatorietà della compilazione del *long form*.

La somministrazione del questionario *long form* nel 2011 è chiamata National Household Survey (NHS)

Il questionario sarà inviato a una famiglia su 3 (1/5 nel 2006)

Il questionario sarà compilato su base volontaria.

Statistics Canada prevede un tasso di risposta del 50% per la NHS

## Quadro normativo europeo

Le “*CES Recommendations for the 2010 Censuses of Population and Housing*”, preparate dall’UNECE (*United Nations Economic Commission for Europe*) in collaborazione con Eurostat (*Statistical Office of European Communities*), sono state formalmente adottate a giugno 2006.

la Commissione Europea (Eurostat) ha redatto, per la tornata censuaria del 2010-2011, un *Framework Regulation* e un *Implementing Regulation*.

Il primo rappresenta un “regolamento quadro” che costituisce la base legislativa e che contiene tutti gli *statements* fondamentali per l’effettuazione dei censimenti; il secondo include specifiche tecniche concernenti variabili, classificazioni e incroci oltre che metodologie per la valutazione della qualità dei dati prodotti.

Il Regolamento pone le condizioni per conseguire l’uniformità dell’*output* delle rilevazioni censuarie, indipendentemente dalle tecniche e dai metodi utilizzati

## Criticità emerse per il censimento italiano della popolazione del 2001

P. Grossi e S. Stoppoloni (2010). La progettazione dei censimenti generali 2010 – 2011. Informazione e formazione ai comuni sulle innovazioni di metodi e tecniche per il 15° Censimento generale della popolazione e delle abitazioni. Documenti Istat 1/2010

- elevato numero di soggetti intermedi coinvolti, ciascuno con la sua organizzazione interna; con forti differenze nella dimensione demografica e nella morfologia dei territori di competenza
- sovrapposizione di compiti e funzioni assegnate agli organi intermedi;
- costituzione, coordinamento e mantenimento di una amplissima rete di coordinatori e rilevatori (circa 108.000);
- consegna e ritiro dei questionari5 l'analisi dei dati di monitoraggio ha messo in luce forti ritardi, diversamente distribuiti, ma presenti su tutto il territorio nazionale;

## Il censimento italiano della popolazione del 2011 (1)

Le innovazioni di carattere metodologico introdotte nel prossimo censimento della popolazione vanno verso due direzioni principali:

l'impiego di liste provenienti da fonti amministrative (register-supported census)

l'introduzione di tecniche campionarie per la rilevazione delle informazioni non strettamente demografiche

Per sperimentare le nuove strategie nel 2009 è stata effettuata un'indagine pilota che ha coinvolto 82.735 famiglie di 31 comuni.

## L'impiego di liste provenienti da fonti amministrative (1)

Il prossimo censimento della popolazione e delle abitazioni prevede l'impiego delle liste anagrafiche comunali (LAC), affiancate da altre liste ausiliarie di fonte sia comunale che nazionale.

I questionari potranno essere distribuiti per posta e non più dai rilevatori e i rispondenti potranno scegliere fra diverse soluzioni per la loro restituzione: web, posta, centri di raccolta, con un aumento auspicato dei tassi di risposta.

Rinuncia alla prassi del “lembo staccabile” dal questionario di famiglia, cioè la rinuncia alla assenza di dati nominativi nei questionari inviati all'Istat

## L'impiego di liste provenienti da fonti amministrative (2)

La Lac è affetta da sottocopertura (oltre che da sovracopertura).

Un censimento assistito dai registri anagrafici deve trattare in modo opportuno gli errori commessi nei conteggi di popolazione a causa di unità non presenti in lista ma abitualmente dimoranti sul territorio comunale.

Per risolvere questo problema sono possibili due soluzioni:

- i) impiegare i rilevatori sul territorio per svolgere, con il supporto delle liste ausiliarie, un'operazione di recupero il più possibile completa (recupero sistematico);
- ii) ii) condurre una rilevazione campionaria per stimare il grado di copertura della lista (recupero statistico).

Opzione i) scelta da Istat.

## Le liste ausiliarie

Lista Integrativa da Fonti Ausiliarie centrali di individui, famiglie e convivenze non già incluse in Lac (**Lifa**). Basata su informazioni contenute nell'anagrafe tributaria, dovrebbe contenere i dati degli individui che mostrano segnali di domicilio nel comune, ma non sono registrati nella sua anagrafe.

Lista di Affiancamento basata sulla Rilevazione dei Numeri Civici (**Rnc**), mediante rilevazione precensuaria. Lista utile nella fase di recupero di individui, famiglie e convivenze che dimorano abitualmente sul territorio comunale, ma non sono registrati in anagrafe.

Lista integrativa autonoma comunale (**Liac**). Lista di unità presenti in archivi locali e non incluse in Lac e in Lifa. Comprende informazioni acquisite dalle anagrafi su individui e famiglie iscritti dopo la data di 'congelamento' della Lac per la sua consegna all'Istat.

## Stima della sottocopertura della Lac (1)

M. Fortini e G. Gallo, (2009) Misure di sottocopertura anagrafica in base alla revisione post-censuaria del 2001

[http://www.istat.it/istat/eventi/2010/workshop\\_9nov2010/approfondimento.html](http://www.istat.it/istat/eventi/2010/workshop_9nov2010/approfondimento.html)

Per risolvere i casi di non rispondenza tra dato anagrafico e dato censuario, gli Uffici di anagrafe effettuano i procedimenti di regolarizzazione post-censuaria (revisione anagrafica). Questi provvedimenti amministrativi determinano un'iscrizione in anagrafe se le persone censite come residenti ma non iscritte si presentano presso gli Uffici anagrafici e sottoscrivono le pratiche di immigrazione da altro comune o dall'estero

Dal 21 ottobre del 2001 al 31 dicembre del 2005 sono state iscritte in anagrafe circa 236.000 persone a seguito delle risultanze del censimento del 2001. I comuni interessati a questo tipo di eventi sono 6.866.



## Stima della sottocopertura della Lac (2)

**Tabella 1 – Iscritti in anagrafe che non comportano conteggio per classe di ampiezza demografica dei comuni, cittadinanza e incidenza sulla popolazione legale del 2001. Valori assoluti e percentuali**

CLASSI DI AMPIEZZA DEMOGRAFICA DEI COMUNI	N. Comuni	Persone dimoranti abitualmente al 20.10.2001 e non iscritte in anagrafe					Popolazione legale	Non iscritti in anagrafe su 1.000 persone censite al 20.10.2001
		Valori assoluti	% sul totale	di cui:				
				% iscritti dall'estero	% cittadini stranieri			
Fino a 2.000	2.589	17.843	7,6	8,4	12,8	3.558.528	5,0	
da 2.001 a 5.000	2.052	37.051	15,7	8,8	14,1	7.032.200	5,3	
da 5.001 a 10.000	1.129	46.183	19,6	9,1	14,4	8.040.885	5,7	
da 10.001 a 20.000	629	45.348	19,2	8,9	14,9	8.669.117	5,2	
da 20.001 a 50.000	330	41.547	17,6	8,7	13,8	10.076.393	4,1	
da 50.001 a 100.000	95	17.081	7,2	10,5	15,3	6.390.014	2,7	
da 100.001 a 250.000	29	14.097	6,0	13,4	19,4	4.125.516	3,4	
da 250.001 a 500.000	7	4.922	2,1	18,0	22,8	2.133.284	2,3	
oltre 500.000	6	11.672	5,0	16,2	19,8	6.969.807	1,7	
Nord-ovest	2.492	73.729	31,3	9,8	16,4	14.938.562	4,9	
Nord-est	1.347	60.619	25,7	12,2	20,8	10.634.820	5,7	
Centro	891	36.027	15,3	11,6	18,9	10.906.626	3,3	
Sud	1.492	46.399	19,7	6,5	6,4	13.914.865	3,3	
Isole	644	18.970	8,0	6,6	5,0	6.600.871	2,9	
<b>Italia</b>	<b>6.866</b>	<b>235.744</b>	<b>100,0</b>	<b>9,8</b>	<b>15,0</b>	<b>56.995.744</b>	<b>4,1</b>	

## Stima della sottocopertura della Lac (3)

Al fine di isolare dalle determinanti della sottocopertura anagrafica gli effetti spuri dovuti alla maggiore o minore accuratezza con cui i comuni hanno effettuato l'aggiornamento delle anagrafi a seguito delle risultanze censuarie del 2001, si è fatto ricorso all'impiego dei modelli mistura.

Individuati due gruppi per ogni categoria dimensionale dei comuni :  
un primo gruppo è costituito da quei comuni che svolgono attive azioni di regolarizzazioni anagrafiche delle persone censite al 2001 come dimoranti abitualmente;  
il secondo gruppo è costituito, invece, dai comuni che non intraprendono azioni di questo tipo e al più si limitano ad effettuare le iscrizioni limitatamente agli individui che spontaneamente trasferiscono la propria residenza.

## Stima della sottocopertura della Lac (4)

L'ammontare complessivo della sottocopertura anagrafica è stato infine stimato utilizzando la componente del modello mistura che descrive i comuni più attivi nella correzione delle anagrafi per questo tipo di eventi. La somma della sottocopertura anagrafica stimata in ciascuno dei sottogruppi ammonta quindi a 406.485 individui, contro i 235.763 osservati

Si tratta quindi di una stima derivata sotto l'ipotesi che tutti i comuni abbiano svolto le operazioni di revisione con la stessa qualità del gruppo dei migliori

Con la dovuta cautela, i risultati delle analisi con i modelli mistura confermano una minore gravità del fenomeno della sottocopertura anagrafica rispetto a quello della sottocopertura censuaria

## L'uso di campioni al censimento del 2011 (1)

Nel prossimo censimento della popolazione si procederà all'osservazione di alcune informazioni socio-economiche solo su campioni di famiglie

somministrazione di *short form* a tutte le famiglie

somministrazione di *long form* a un campione di famiglie

Il campionamento interesserà solo i comuni  $\geq 20.000$  abitanti

All'interno di questi comuni interesserà le sole aree di censimento (aggregazioni di sezioni di censimento con popolazione approssimativamente pari a 15.000 individui) di centro abitato.

Il disegno delle aree di censimento deve consentire la ricostruzione, per somma di aree, delle esistenti zone sub comunali a valenza amministrativa (ad esempio; municipi, circoscrizioni, quartieri, zone di decentramento, ecc..) o a valenza funzionale (ad esempio; zone statistiche, aree elementari, zone, unità urbanistiche, zone toponomastiche, ecc..).

Indicazione a non procedere con il campionamento nelle aree di censimento sotto le 10mila unità

## L'uso di campioni al censimento del 2011 (2)

Semplificazione delle operazioni di raccolta

Incremento del tasso di risposta spontaneo

Riduzione del fastidio statistico sulle famiglie

Minor carico di lavoro per i comuni

Diminuzione dell'errore di misura (ma introduzione dell'errore di campionamento)

## L'uso di campioni al censimento del 2011 (3)

Quale schema di campionamento?

Quale frazione di campionamento?

Quale tecnica di stima?

Carbonetti G., De Vitiis C., (2007) Efficienza di stime campionarie relative ad un sottoinsieme di variabili di censimento: primi risultati delle sperimentazioni, Relazione presentata alla *Conferenza Censimenti generali 2010-2011. Criticità e Innovazioni*, Roma,

Carbonetti G., Fortini M. (2008) Sample results expected accuracy in the Italian population and housing census, *Joint UNECE/Eurostat Meeting on Population and Housing Censuses*. UN, Ginevra, Maggio 2008. ECE/CES/AC.6/2008/4

Carbonetti G., Dardanelli S., Fiorello E., Mastroluca S., Verrascina M., (2008) Ipotesi di innovazione per il censimento della popolazione del 2011: una valutazione degli effetti su un possibile piano di diffusione, Relazione presentata alla *XXIX Conferenza Italiana di Scienze Regionali*, Bari 24-26 Settembre 2008.

Carbonetti G., Fortini M., Solari F. (2008). Innovations on methods and survey process for the 2011 Italian population census. Proceedings of Q2008 European Conference on Quality in Official Statistics

## Schema di campionamento e tasso di sondaggio (1)

Sono state condotte sperimentazioni su dati del Censimento della popolazione 2001 riferiti a 40 comuni scelti per diversa dimensione demografica e differente collocazione geografica.

Area geografica	Classi di ampiezza demografica			Totale
	10.000-20.000	20.000-100.000	più di 100.000	
<b>Nord</b>	<b>4</b>	<b>6</b>	<b>6</b>	<b>16</b>
<b>Centro</b>	<b>2</b>	<b>3</b>	<b>3</b>	<b>8</b>
<b>Sud e Isole</b>	<b>4</b>	<b>6</b>	<b>6</b>	<b>16</b>
<b>Totale</b>	<b>10</b>	<b>15</b>	<b>15</b>	<b>40</b>

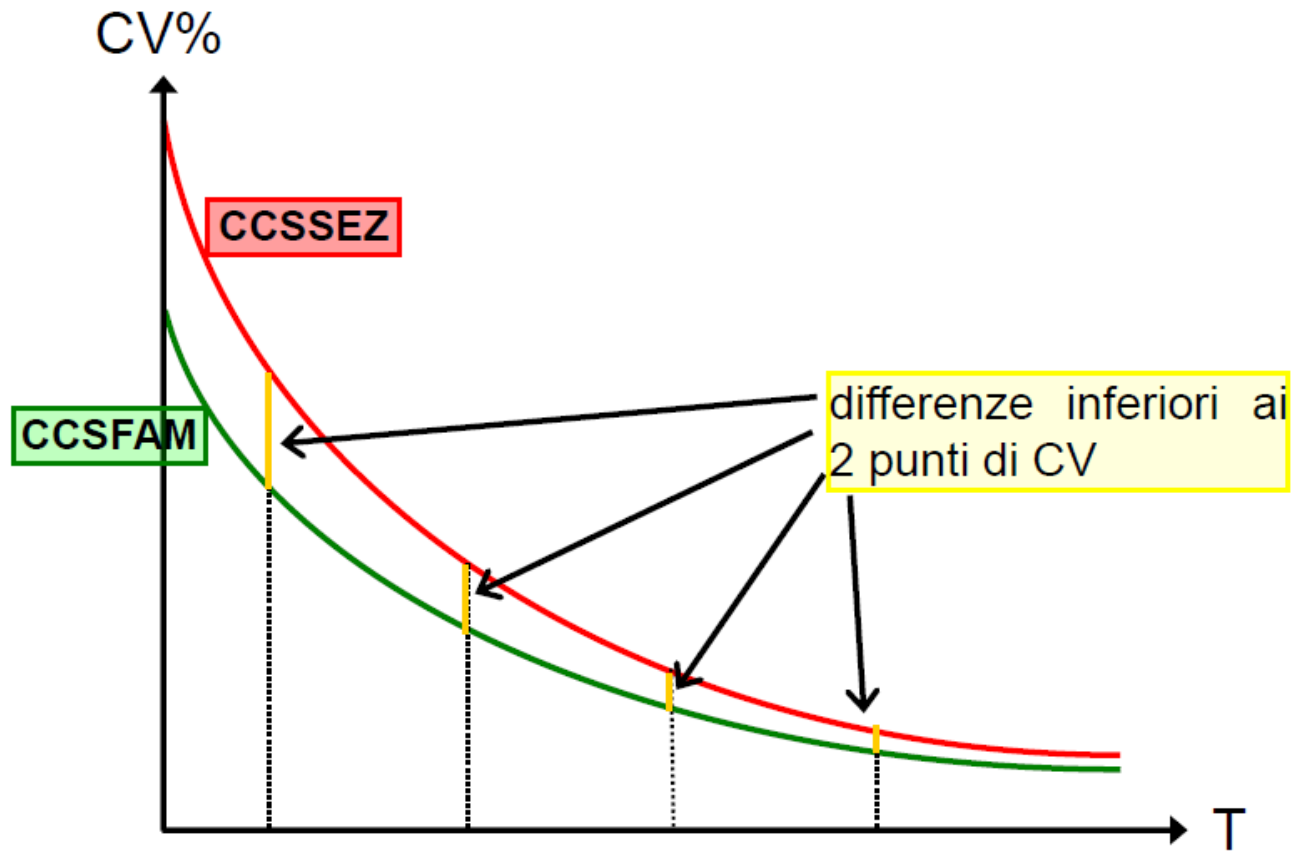
## Schema di campionamento e tasso di sondaggio (2)

	<b>Unità campionate</b>	<b>Universo</b>	<b>%</b>
<b>Aree di censimento</b>	<b>497</b>	<b>3.347(*)</b>	<b>14,85%</b>
<b>Sezioni di censimento</b>	<b>30,890</b>	<b>382,534</b>	<b>8,08%</b>
<b>Famiglie</b>	<b>2,243,511</b>	<b>21,810,676</b>	<b>10,29%</b>
<b>Individui</b>	<b>5,537,582</b>	<b>56,594,021</b>	<b>9,78%</b>
(*) Numero stimato			



## Schema di campionamento e tasso di sondaggio (3)

Confronto tra i livelli di errore delle stime delle frequenze assolute ottenute con disegno campionario da lista anagrafica e con disegno campionario areale (campione di sezioni di censimento).



## Schema di campionamento e tasso di sondaggio (4)

Distribuzione degli errori percentuali relativi, per classi di frequenze assolute T e per 3 differenti frazioni sondate, nel caso di disegno casuale semplice di famiglie.

Classi di T	10%	20%	33%
	<i>errore % medio</i>	<i>errore % medio</i>	<i>errore % medio</i>
< 10	<b>280,9</b>	<b>198,7</b>	<b>130,3</b>
10   30	<b>148,8</b>	<b>94,8</b>	<b>66,3</b>
30   50	<b>101,5</b>	<b>62,4</b>	<b>45,9</b>
50   100	<b>75,7</b>	<b>43,7</b>	<b>34,2</b>
100   250	<b>49,7</b>	<b>30,8</b>	<b>22,3</b>
250   500	<b>31,5</b>	<b>20,4</b>	<b>14,7</b>
500   1.000	<b>23,2</b>	<b>14,7</b>	<b>10,3</b>
1.000   2.500	<b>14,7</b>	<b>9,3</b>	<b>6,6</b>
2.500   5.000	<b>9,5</b>	<b>5,9</b>	<b>4,0</b>
5.000   10.000	<b>6,3</b>	<b>3,9</b>	<b>2,6</b>
10.000   25.000	<b>4,0</b>	<b>2,8</b>	<b>1,7</b>
25.000   50.000	<b>2,4</b>	<b>1,5</b>	<b>1,1</b>
50.000   100.000	<b>1,6</b>	<b>1,1</b>	<b>0,8</b>
100.000   250.000	<b>1,1</b>	<b>0,6</b>	<b>0,5</b>
250.000   500.000	<b>0,8</b>	<b>0,3</b>	<b>0,3</b>
≥ 500.000	<b>0,3</b>	<b>0,2</b>	<b>0,2</b>

## Schema di campionamento e tasso di sondaggio (5)

La tabella precedente è stata costruita nell'ipotesi che tutte le aree di censimento siano sottoposte a campionamento.

Nella realtà non è così.

Se il territorio (dominio)  $R$  è costituito in parte dalla presenza di aree campionabili ( $R_C$ ) e in parte da aree non campionabili ( $R_{NC}$ ) indicando con  $\gamma$  la quota di popolazione di  $R$  soggetta a campionamento per la rilevazione con *long form* ( $\gamma = \text{Pop}_{R_C} / \text{Pop}_R$ ), allora

$$\hat{T}_R = \gamma \hat{T}_C + (1 - \gamma) T_{NC}$$

da cui si ha il

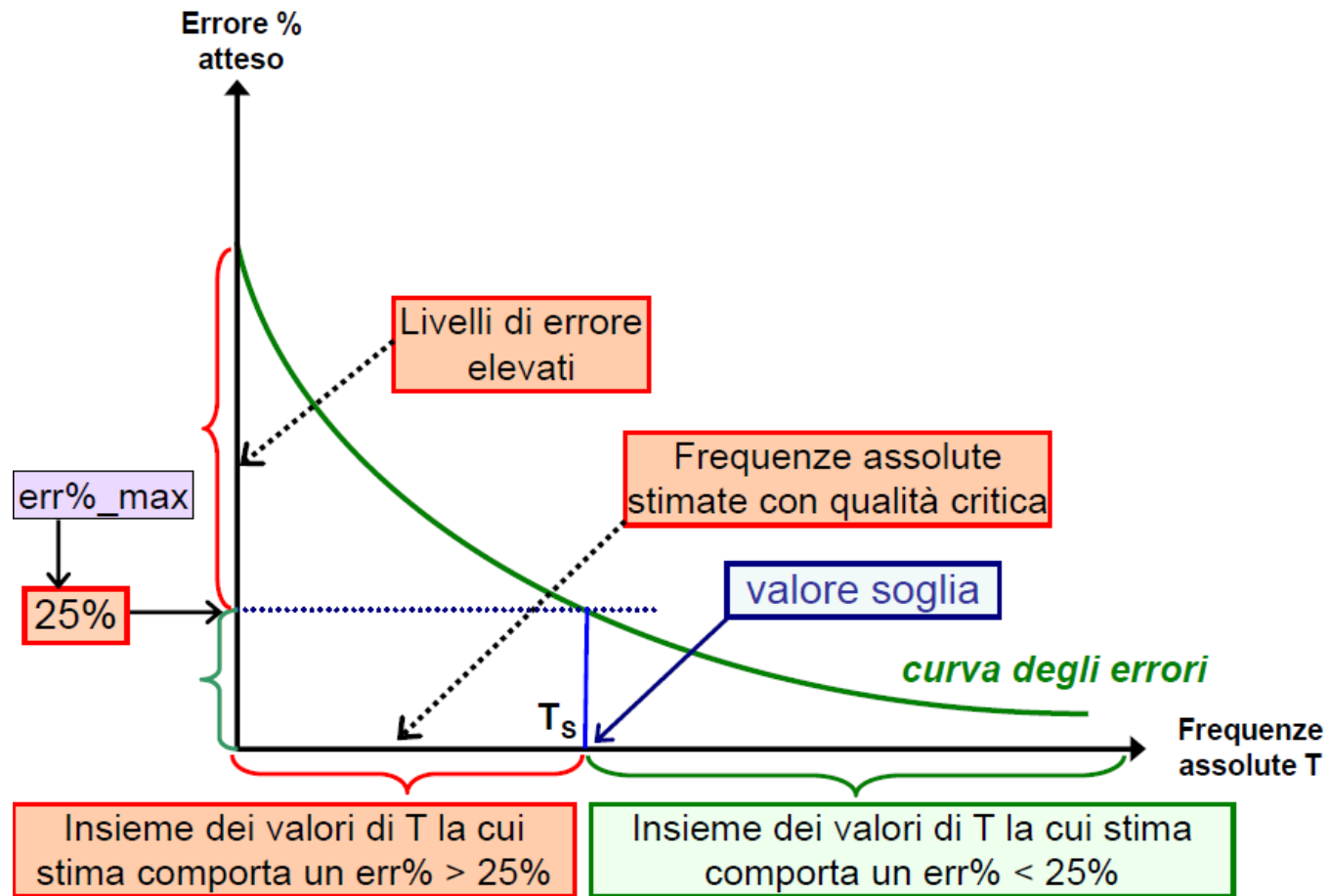
$$\text{err}(\hat{T}_R) \cong \gamma \cdot \text{err}(\hat{T}_C)$$

## Schema di campionamento e tasso di sondaggio (6)

Classi di $\gamma$	Regioni	Province	Comuni			
			20.000	100.000	100.000	250.000
$< 0,5$	3	19	23		0	0
0,5   0,6	0	15	17		1	0
0,6   0,7	7	14	23		1	0
0,7   0,8	4	21	52		4	0
0,8   0,9	5	25	88		7	2
$\geq 0,9$	1	9	228		16	11
<b>Totale</b>	20	103	431		29	13

## Schema di campionamento e tasso di sondaggio (7)

La qualità complessiva di una tabella statistica sarà tanto più elevata quanto più ridotto sarà il contenuto informativo rappresentato dalle celle con frequenza assoluta  $T$  inferiore al valore soglia  $T_s$



## Schema di campionamento e tasso di sondaggio (8)

Esempio 1: Tavola Eurostat H.B1.E0.R1. Indicatori di qualità delle tavole regionali per Molise, Marche e Sicilia.

Variabili di incrocio della tavola H.B1.E0.R1: sesso (2); età (101); condizione professionale (6). Numero dei *possibili incroci* = 1.062 .

Errore% massimo = 25%

	Valle d'Aosta ( $\gamma=0,27$ )			Marche ( $\gamma=0,67$ )			Sicilia ( $\gamma=0,86$ )		
<b>Frazione sondata</b>	<b>Soglia critica Ts</b>	<b>% di celle</b>	<b>% di individui</b>	<b>Soglia critica Ts</b>	<b>% di celle</b>	<b>% di individui</b>	<b>Soglia critica Ts</b>	<b>% di celle</b>	<b>% di individui</b>
<b>f.s. = 10%</b>	<b>50</b>	<b>81,75</b>	<b>7,09</b>	<b>250</b>	<b>84,00</b>	<b>3,20</b>	<b>500</b>	<b>82,86</b>	<b>2,22</b>
<b>f.s. = 20%</b>	<b>30</b>	<b>75,59</b>	<b>4,49</b>	<b>100</b>	<b>77,94</b>	<b>1,71</b>	<b>250</b>	<b>78,18</b>	<b>1,31</b>
<b>f.s. = 33%</b>	<b>10</b>	<b>63,16</b>	<b>2,06</b>	<b>50</b>	<b>70,95</b>	<b>0,97</b>	<b>100</b>	<b>69,99</b>	<b>0,59</b>

## Il problema della coerenza delle stime: un esempio (1)

Carbonetti (2011)

Piano di diffusione del 2001:

*“Occupati per sesso, classe di età quinquennale e posizione nella professione”*

Universo di riferimento: *popolazione residente  $\geq 15$  anni occupata*

Dominio minimo di diffusione: - *Grande Comune ( $\geq 150$  mila ab.)*

- *Provincia*

SESSO X ETA' (13 classi) X POSIZIONE NELLA PROFESSIONE (6)

Variabili rilevate in modo esaustivo sul dominio	Variabile rilevata <b>solo</b> su campioni di famiglie
---	--

## Il problema della coerenza delle stime: un esempio (2)

Tavola: Occupati per sesso, classe di età quinquennale e posizione nella professione (Censimento 2001)

Sesso = MASCHI

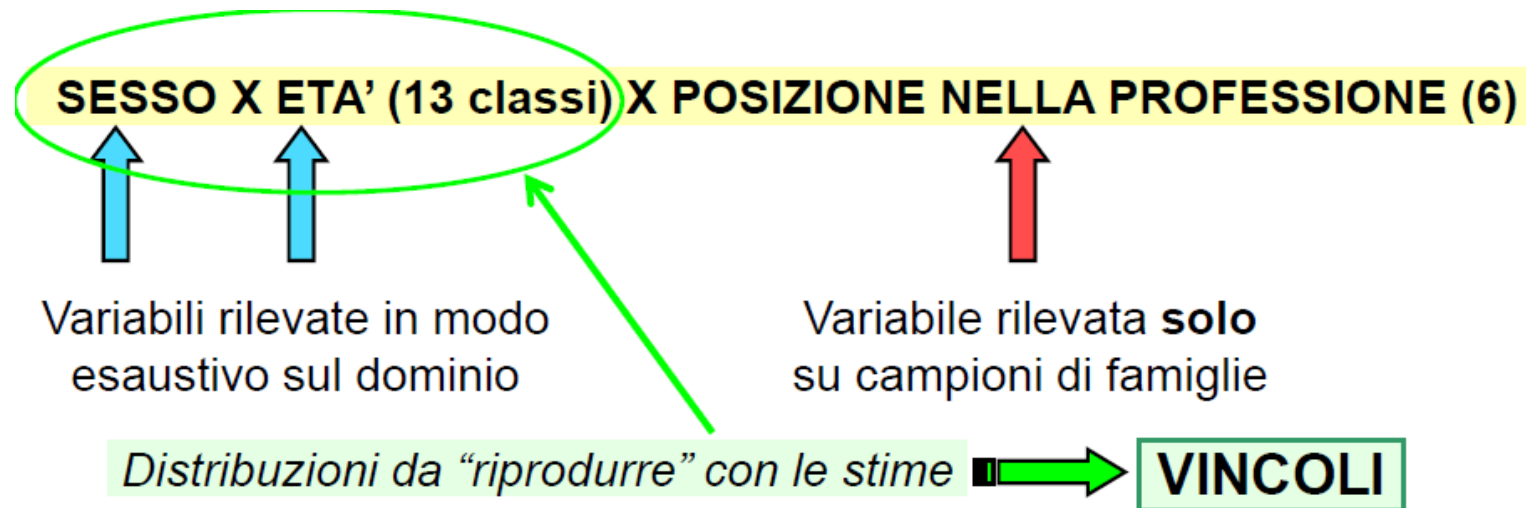
CLASSE DI ETÀ QUINQUENNALE DA 15 ANNI IN POI	POSIZIONE NELLA PROFESSIONE						Totale
	Imprenditore	Libero professionista	Lavoratore in proprio	Socio di cooperativa	Coadiuvante familiare	Dipendente o in altra posizione subordinata	
15-19	X	X	X	X	X	X	0
20-24	X	X	X	X	X	X	0
25-29	X	X	X	X	X	X	0
30-34	X	X	X	X	X	X	0
35-39	X	X	X	X	X	X	0
40-44	X	X	X	X	X	X	0
45-49	X	X	X	X	X	X	0
50-54	X	X	X	X	X	X	0
55-59	X	X	X	X	X	X	0
60-64	X	X	X	X	X	X	0
65-69	X	X	X	X	X	X	0
70-74	X	X	X	X	X	X	0
75 e più	X	X	X	X	X	X	0
Totale	X	X	X	X	X	X	0

**DATO STIMATO** (indicated by a red box and arrows pointing to the 'X' in the 'Lavoratore in proprio' column for the 50-54 and 75 e più age groups)

**CONTEGGIO ESAUSTIVO** (indicated by a blue box and arrows pointing to the '0' in the 'Totale' column for the 50-54 and 75 e più age groups)



## Il problema della coerenza delle stime: un esempio (3)



Il numero di vincoli dipende dal piano di diffusione

## L'uso di campioni al censimento del 2011 (4)

Quale schema di campionamento?

**Campionamento casuale semplice (da lista anagrafica) di famiglie nelle aree di censimento**

Quale frazione di campionamento?

**33%**

Quale tecnica di stima?

**Stimatori di ponderazione vincolata**

## L'uso di campioni al censimento del 2011 (5)

Attività	Periodo
Aggiornamento delle Basi territoriali e costituzione delle aree di censimento	Aprile 2009 – Ottobre 2009
Rilevazione pilota Censimento della Popolazione	Autunno 2009
Costituzione Uffici Comunali di Censimento	Entro Autunno 2010
Rilevazione dei numeri civici e censimento degli edifici	Autunno 2010
Rilascio liste anagrafiche (LAC)	Entro Gennaio 2011
Rilevazione Long/Short form	Contestuale alla rilevazione censuaria
Recupero della sottocopertura	
Confronto censimento-anagrafe, revisione e aggiornamento	

## E' possibile migliorare l'efficienza delle stime?

Istat ha sta conducendo ulteriori simulazioni al fine di verificare se mediante l'utilizzo ti tecniche di stima (da modello) appropriate sia possibile ottenere stime più precise a livello di piccoli domini.

Al Dipartimento di Scienze Statistiche dell'Università di Bologna stiamo conducendo esperimenti analoghi utilizzando tecniche di stima diverse. Primi risultati:

*Confronto tra approccio da disegno e approccio da modello in termini di RMSE al variare delle frequenze della tabella oggetto di stima*

Frequenze	$({}_2\text{RMSE}/{}_1\text{RMSE})$	$({}_2\text{RMSE})$
0.0001-0.0005	0.444	0.722
0.0005-0.001	0.445	0.226
0.001-0.0025	0.449	0.098
0.0025-0.005	0.616	0.059
0.005-0.01	0.776	0.037
0.01-0.025	0.908	0.019
0.025-0.05	1.091	0.008