

**Risultati delle misure di efficienza di
campioni di famiglie per livelli territoriali
e dettagli informativi**

Giancarlo Carbonetti

Istat - Direzione Centrale dei Censimenti Generali

XXVIII Convegno Nazionale A.N.U.S.C.A. - PalaRiccione

Impiego di metodi campionari nelle rilevazioni totali_1

Quali **opportunità** possono essere offerte dal ricorso al campionamento tramite *long form* nell'ambito della strategia complessiva per la conduzione del censimento della popolazione?

- Il successo delle innovazioni introdotte per limitare il lavoro di *front office*.
- La riduzione del fastidio statistico sui cittadini chiamati a rispondere.
- La possibilità di dare maggiore attenzione alla qualità dei risultati finali.

Impiego di metodi campionari nelle rilevazioni totali_2

Qual è il **costo** che deriva dalla possibilità di osservare l'informazione censuaria in modo non esaustivo sulla popolazione?



Presenza dell'errore di campionamento

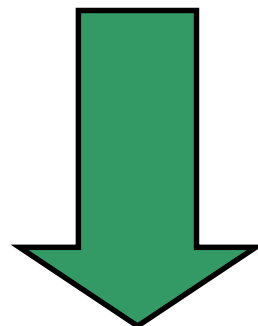


Problema?

Il maggior livello di qualità dell'informazione censuaria raccolta conduce ad una riduzione dell'errore di misura che compensa l'introduzione dell'errore campionario.

Impiego di metodi campionari nelle rilevazioni totali_3

Qual'è l'**affidabilità** del dato censuario ottenuto tramite una metodologia campionaria?



Prima esperienza in Italia

Teoria statistica

Esperienza Istat

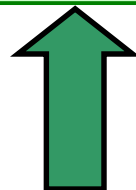
Censimenti esteri

Impiego di metodi campionari nelle rilevazioni totali_4

Quali possibili **strategie campionarie**, in termini di disegno di campionamento e di tecnica di stima, sono praticabili in un contesto censuario?



Necessità di studi e sperimentazioni



Qual'è il livello atteso dell'**accuratezza statistica** delle stime prodotte tramite l'impiego delle tecniche campionarie proposte?

Il *processo di valutazione* dell'efficienza della metodologia campionaria tramite long form: obiettivi

➤ **Obiettivi:**

- ✓ studiare l'accuratezza di stime per variabili relative a istruzione, mercato del lavoro, pendolarismo e riferite alle variabili singole o di incrocio con le informazioni demografiche;
- ✓ valutare la possibilità di estrarre campioni di famiglie nei comuni tra 5.000 e 20.000 abitanti;
- ✓ produrre stime affidabili per domini comunali e sub-comunali relativi alle *aree di censimento di centro* (aggregazioni di sezioni di censimento contigue, con dimensione tra 5.000 e 15.000 abitanti).

Il *processo di valutazione* dell'efficienza della metodologia campionaria tramite long form: percorso

➤ Percorso:

- ✓ studio di disegni di campionamento (da lista o areale, semplici o stratificati);
- ✓ sperimentazione di differenti frazioni di campionamento (10%, 20%, 33%);
- ✓ impiego di stimatori diretti (espansione, calibrati);
- ✓ confronti di efficienza tramite misure dell'errore di campionamento (*il coefficiente di variazione*);
- ✓ simulazioni di spazi campionari per valutare la variabilità delle stime.

Il *processo di valutazione* dell'efficienza della metodologia campionaria tramite long form: ambito_1

- Le **sperimentazioni** sono state condotte su 40 comuni scelti per differente 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	
Nord	4	6	6	16
Centro	2	3	3	8
Sud e Isole	4	6	6	16
Totale	10	15	15	40

Il *processo di valutazione* dell'efficienza della metodologia campionaria tramite long form: ambito_2

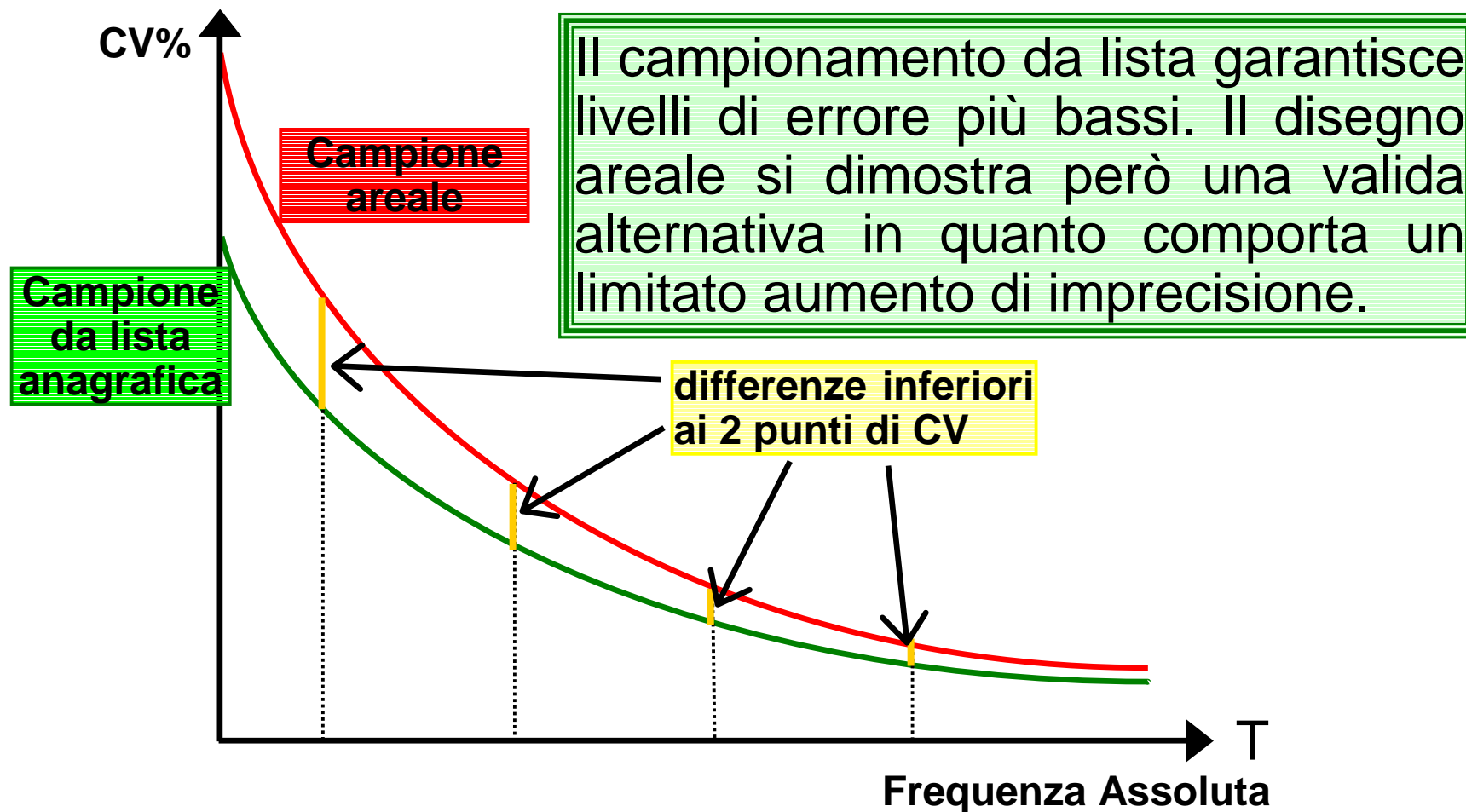
- Dimensione delle **simulazioni** in termini di aree di censimento, sezioni di censimento, famiglie e individui coinvolte.

	Unità campionate	Universo	%
Aree di censimento	497	3.347(*)	14,85%
Sezioni di censimento	30.890	382.534	8,08%
Famiglie	2.243.511	21.810.676	10,29%
Individui	5.537.582	56.594.021	9,78%

(*) Numero stimato

Confronti di efficienza delle stime campionarie: risultati delle prove sperimentali_1

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).



Confronti di efficienza delle stime campionarie: risultati delle prove sperimentali_2

Disegno da
lista anagrafica

Errore percentuale atteso:
 $err\% \approx 1,96 \cdot CV\%$

Le classi di $T > 10.000$ si riferiscono unicamente
a domini superiori all'area di censimento

Classi di T	f.s. = 10%		f.s. = 20%		f.s. = 33%	
	<i>errore % medio (+/-)</i>	<i>errore % max (+/-)</i>	<i>errore % medio (+/-)</i>	<i>errore % max (+/-)</i>	<i>errore % medio (+/-)</i>	<i>errore % max (+/-)</i>
<10	280,9	376,0	198,7	242,4	130,3	187,7
10 30	148,8	166,7	94,8	107,1	66,3	75,6
30 50	101,5	112,0	62,4	72,8	45,9	50,1
50 100	75,7	80,9	43,7	55,7	34,2	37,5
100 250	49,7	55,9	30,8	38,4	22,3	25,1
250 500	31,5	35,8	20,4	24,6	14,7	15,9
500 1.000	23,2	25,1	14,7	16,1	10,3	11,5
1.000 2.500	14,7	17,5	9,3	11,5	6,6	7,7
2.500 5.000	9,5	10,6	5,9	7,1	4,0	4,9
5.000 10.000	6,3	7,5	3,9	5,0	2,6	3,7
10.000 25.000	4,0	4,9	2,8	3,2	1,7	2,3
25.000 50.000	2,4	3,2	1,5	2,0	1,1	1,4
50.000 100.000	1,6	2,0	1,1	1,1	0,8	1,0
100.000 250.000	1,1	1,3	0,6	0,6	0,5	0,6
250.000 500.000	0,8	0,8	0,3	0,3	0,3	0,3
≥ 500.000	0,3	0,3	0,2	0,2	0,2	0,2

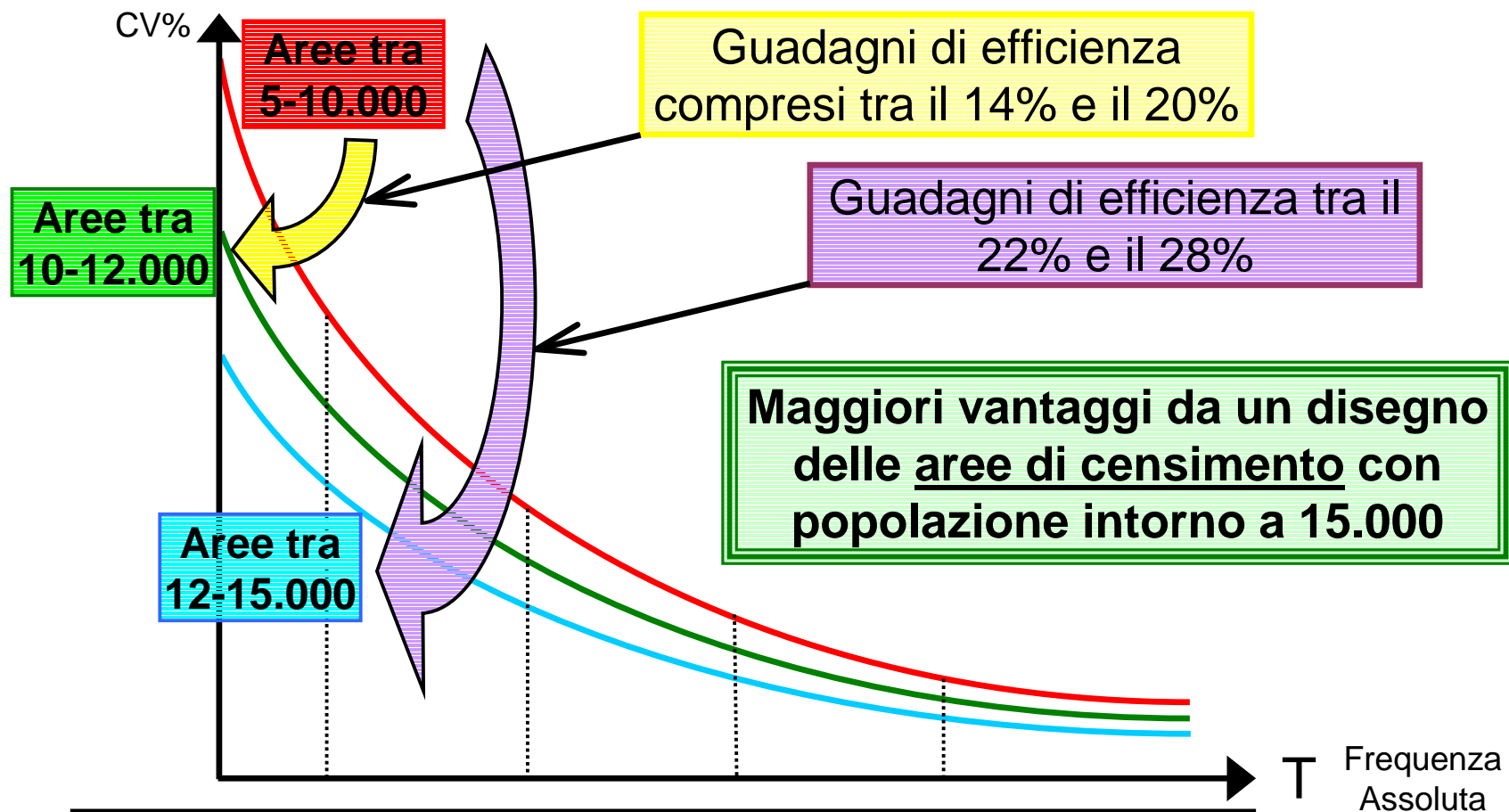
Confronti di efficienza delle stime campionarie: risultati delle prove sperimentali_3

Alcuni esempi di stime tramite Intervallo di Confidenza IC nel disegno da lista (valide sotto l'ipotesi di normalità e al livello di confidenza del 95%).

Classi di T		f.s. = 10%	f.s. = 20%	f.s. = 33%
10 30	errore %(+/-)	148,82	94,81	66,26
20	errore assoluto (+/-)	30	19	13
	IC_limite inferiore	0	1	7
	IC_limite superiore	50	39	33
100 250	errore %(+/-)	49,72	30,84	22,29
200	errore assoluto (+/-)	99	62	45
	IC_limite inferiore	101	138	155
	IC_limite superiore	299	262	245
10.000 25.000	errore %(+/-)	3,96	2,75	1,69
20.000	errore assoluto (+/-)	793	550	338
	IC_limite inferiore	19207	19450	19662
	IC_limite superiore	20793	20550	20338

Confronti di efficienza delle stime campionarie: risultati delle prove sperimentali_4

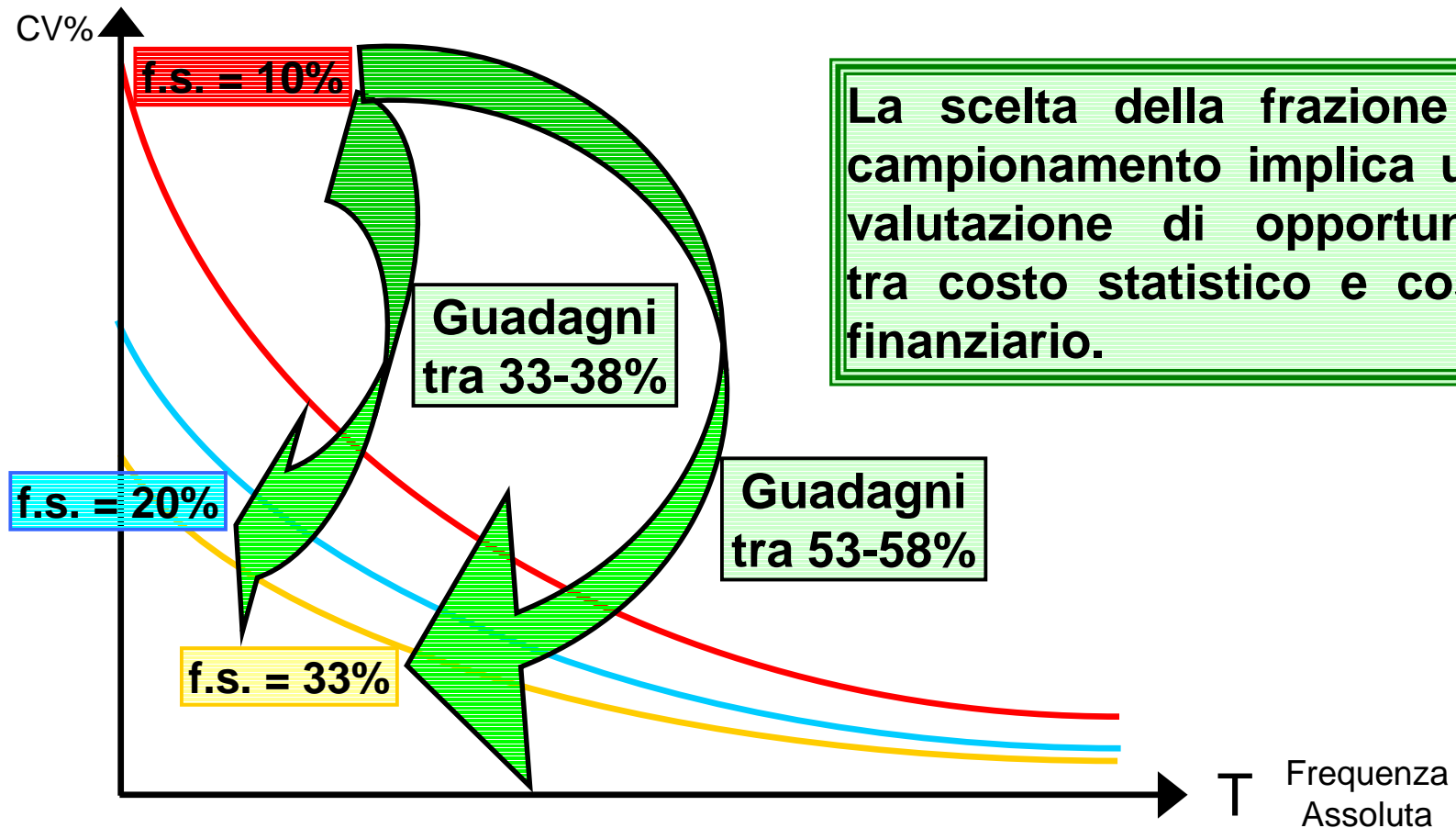
Confronto tra i livelli di errore delle stime di frequenze assolute nel disegno campionario da lista anagrafica per aree di diversa ampiezza.



Tali considerazioni sono valide per tutte le frazioni di campionamento sperimentate

Confronti di efficienza delle stime campionarie: risultati delle prove sperimentali_5

Confronto tra i livelli di errore delle stime di frequenze assolute nel disegno campionario da lista anagrafica per differenti frazioni sondate.



Confronti di efficienza delle stime campionarie: risultati delle prove sperimentali_6

Distribuzione delle stime delle frequenze assolute, oggetto della sperimentazione, in classi di valori percentuali di CV (stime riferite alle aree con più di 12.000 abit.). Confronto per 3 differenti frazioni sondate.

Classi di CV%	Frazione di campionamento		
	10%	20%	33%
< 10%	29,80	51,07	65,42
10% - 50%	54,23	40,29	29,52
> 50%	15,97	8,64	5,05

Esempio 1:

la percentuale di stime che ha un CV% inferiore del 10% aumenta dal 29,80% al 65,42% per una frazione sondata che passa dal 10% al 33%.

Esempio 2:

l'incidenza di stime che hanno un CV% superiore al 50% diminuisce dal 15,97% al 5,05% per la frazione sondata che cresce dal 10% al 33%.

Riflessi della strategia campionaria sull'accuratezza delle tabelle statistiche_1

Fino ad ora sono stati illustrati alcuni risultati dello studio di valutazione della **praticabilità** dei metodi campionari e del relativo livello di **efficienza** delle stime producibili in ambito censuario.

PASSO SUCCESSIVO

Valutare quale potrebbe essere l'**impatto qualitativo** del metodo campionario sulla produzione delle tabelle statistiche che saranno oggetto del futuro piano di diffusione dei risultati censuari.



ESEMPIO

Riflessi della strategia campionaria sull'accuratezza delle tabelle statistiche_2

Disegno da lista (f.s.=10%) → err%_soglia=25% (CV%≈12,5%) → T_s=500

sesso	classi di età	stato civile	CONDIZIONE PROFESSIONALE _ REGIONE SICILIA _ Dati Censimento2001 (estratto)											
			occupati				in cerca di prima occupazione				disoccupati			
			cittadinanza				cittadinanza				cittadinanza			
			Italia	U.E.	Altra citt.	Apolide	Italia	U.E.	Altra citt.	Apolide	Italia	U.E.	Altra citt.	Apolide
M	20-24	1	44496	22	554	0	29475	15	117	0	13537	8	61	0
		2	4221	8	109	0	968	1	18	0	1293	2	28	0
		3	0	0	0	0	0	0	0	0	0	0	0	0
		4	3	0	1	0	2	0	0	0	1	0	0	0
		5	3	0	0	0	3	0	0	0	3	0	0	0
	35-39	1	18878	26	532	0	2741	6	38	0	3898	7	98	0
		2	108808	91	2259	2	2764	4	79	0	12722	18	218	0
		3	780	2	15	0	58	0	0	0	189	0	3	0
		4	2098	4	15	0	21	1	0	0	350	1	2	0
		5	108	0	5	0	7	0	0	0	19	0	0	0
	50-54	1	5131	16	68	0	258	0	3	0	964	0	4	0
		2	98906	15	604	0	781	1	20	0	9951	5	51	0
		3	1269	2	13	0	23	1	0	0	160	1	1	0
		4	1877	3	1	0	28	0	0	0	212	0	1	0
		5	940	2	7	0	13	0	0	0	123	0	0	0

celle stimabili con qualità "critica"

Riflessi della strategia campionaria sull'accuratezza delle tabelle statistiche_3

Disegno da lista (f.s.=33%) → err%_soglia=25% (CV%≈12,5%) → T_s=100

sesso	classi di età	stato civile	CONDIZIONE PROFESSIONALE _ REGIONE SICILIA _ Dati Censimento2001 (estratto)											
			occupati				in cerca di prima occupazione				disoccupati			
			cittadinanza				cittadinanza				cittadinanza			
			Italia	U.E.	Altra citt.	Apolide	Italia	U.E.	Altra citt.	Apolide	Italia	U.E.	Altra citt.	Apolide
M	20-24	1	44496	22	554	0	29475	15	117	0	13537	8	61	0
		2	4221	8	108	0	968	1	18	0	1293	2	28	0
		3	0	0	0	0	0	0	0	0	0	0	0	0
		4	3	0	1	0	3	0	0	0	1	0	0	0
		5	3	0	0	0	3	0	0	0	3	0	0	0
	35-39	1	18878	26	532	0	2741	6	38	0	3898	7	98	0
		2	108808	91	2259	2	2764	4	79	0	12722	18	218	0
		3	780	3	15	0	58	0	0	0	188	0	3	0
		4	2098	4	15	0	31	1	0	0	332	1	2	0
		5	108	0	5	0	7	0	0	0	19	0	0	0
50-54	1	5131	18	88	0	259	0	3	0	964	0	4	0	
	2	98906	15	604	0	781	1	30	0	9951	5	51	0	
	3	1269	2	13	0	23	1	0	0	18	1	1	0	
	4	1877	3	1	0	28	0	0	0	21	0	1	0	
	5	940	2	7	0	13	0	0	0	17	0	0	0	

celle stimabili con qualità "critica"

celle con qualità "recuperata"

Riflessi della strategia campionaria sull'accuratezza delle tabelle statistiche_4

Scelta una strategia campionaria (disegno; stimatore) e fissati un livello territoriale di riferimento e un dettaglio informativo (variabili e loro classificazione).

**Che vuol dire
qualità accettabile
per una tabella ?**

1°Esempio : meno di 1/3 delle celle
ha $CV\% > 12,5\%$.

2°Esempio : meno del 7% degli
individui ricade in celle con
 $CV\% > 12,5\%$.

Indicatori di qualità

- Percentuale di celle con errore non superiore ad una *soglia critica*.
- Percentuale di individui classificati in celle con errore non superiore ad una *soglia critica*.

Misura della qualità attesa delle tabelle statistiche

Siamo ora in grado di misurare la **qualità attesa** dei dati contenuti in tabelle statistiche definite per la diffusione dei risultati censuari.



A tal scopo sono state prese in considerazione alcune tabelle statistiche con dati del Cens2001 aventi ampio dettaglio informativo in termini di:

- ✓ livello territoriale;
- ✓ tematica di interesse.

Analisi di qualità per tabelle statistiche per *Regione*

Variabili di classificazione	Tavole richieste dal piano di diffusione Eurostat a livello regionale								
	H.B1.E0.R1	H.B1.E0.R2	H.B1.E0.R3	H.B1.E0.R4	H.B1.E0.R5	H.B1.E1.R2	H.B1.E2.R3	H.B1.E3.R4	
Sesso	X	X	X	X	X	X	X	X	
Età	X	X	X	X	X	X	X	X	
	anni di età					classi di età			
Condizione professionale	X					X	X		
Professione		X				X		X	
Attività economica			X				X	X	
Posizione nella professione				X					
Grado di istruzione					X	X	X	X	

Le variabili non demografiche sono considerate una per volta e l'incrocio con sesso ed età è al massimo dettaglio.

Le variabili non demografiche sono combinate tra loro e l'incrocio con sesso ed età è ad un minor dettaglio.

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 .

- Soglia critica T_s (relativa all'err%_soglia)
- Percentuale di celle valorizzate ($T > 0$) con frequenza assoluta $T < T_s$
- Percentuale di individui classificati in celle con frequenza assoluta $T < T_s$

Err%_massimo = 25%

	Molise			Marche			Sicilia		
Frazione sondata	Soglia critica T_s	% di celle	% di individui	Soglia critica T_s	% di celle	% di individui	Soglia critica T_s	% di celle	% di individui
f.s. = 10%	100	43,4	2,7	250	42,5	1,4	500	37,4	0,6
f.s. = 20%	50	36,0	1,1	100	33,9	0,5	250	33,9	0,4
f.s. = 33%	30	30,7	0,5	50	28,9	0,2	100	28,8	0,2

Esempio 2: Tavola Eurostat H.B1.E0.R3. Indicatori di qualità delle tavole regionali per Molise, Marche e Sicilia.

Variabili di incrocio della tavola H.B1.E0.R3: sesso (2); età (101); attività economica (17). Numero dei *possibili incroci* = 3.126 .

- Soglia critica T_s (relativa all'err%_soglia)
- Percentuale di celle valorizzate ($T > 0$) con frequenza assoluta $T < T_s$
- Percentuale di individui classificati in celle con frequenza assoluta $T < T_s$

Err%_massimo = 25%

	Molise			Marche			Sicilia		
Frazione sondata	Soglia critica T_s	% di celle	% di individui	Soglia critica T_s	% di celle	% di individui	Soglia critica T_s	% di celle	% di individui
f.s. = 10%	100	72,8	10,4	250	66,1	5,7	500	66,3	3,3
f.s. = 20%	50	61,0	5,0	100	52,2	1,7	250	57,0	1,6
f.s. = 33%	30	50,9	2,6	50	44,7	0,8	100	45,1	0,5

Esempio 3: Tavola Eurostat H.B1.E1.R3. Indicatori di qualità delle tavole regionali per Molise, Marche e Sicilia.

Variabili di incrocio della tavola H.B1.E1.R3: sesso (2); età (21); condizione professionale (6); attività economica (17); grado di istruzione (7). Numero dei *possibili incroci* = 5.574 .

- Soglia critica T_s (relativa all'err%_soglia)
- Percentuale di celle valorizzate ($T > 0$) con frequenza assoluta $T < T_s$
- Percentuale di individui classificati in celle con frequenza assoluta $T < T_s$

Err%_massimo = 25%

Frazione sondata	Molise			Marche			Sicilia		
	Soglia critica T_s	% di celle	% di individui	Soglia critica T_s	% di celle	% di individui	Soglia critica T_s	% di celle	% di individui
f.s. = 10%	100	79,2	10,7	250	78,8	6,9	500	75,9	4,2
f.s. = 20%	50	71,0	5,8	100	68,4	3,0	250	68,7	2,1
f.s. = 33%	30	63,6	3,4	50	59,8	1,5	100	59,4	1,0

Esempio 4: Tavola Eurostat H.B1.E1.R4. Indicatori di qualità delle tavole regionali per Molise, Marche e Sicilia.

Variabili di incrocio della tavola H.B1.E1.R4: sesso (2); età (13); professione (10); attività economica (17); grado di istruzione (7). Numero dei *possibili incroci* = 26.350 .

- Soglia critica T_s (relativa all'err%_soglia)
- Percentuale di celle valorizzate ($T > 0$) con frequenza assoluta $T < T_s$
- Percentuale di individui classificati in celle con frequenza assoluta $T < T_s$

Err%_massimo = 25%

	Molise			Marche			Sicilia		
Frazione sondata	Soglia critica T_s	% di celle	% di individui	Soglia critica T_s	% di celle	% di individui	Soglia critica T_s	% di celle	% di individui
f.s. = 10%	100	91,9	14,9	250	91,1	11,2	500	91,8	7,3
f.s. = 20%	50	86,5	9,3	100	84,4	6,0	250	87,6	4,5
f.s. = 33%	30	81,3	6,5	50	77,1	3,4	100	79,4	2,2

Analisi di qualità per tabelle statistiche per *Comune* e per *Area di censimento_1*

- Il passo successivo è stato quello di valutare i riflessi della strategia campionaria sulla qualità di tabelle a **maggiore dettaglio territoriale**.
- In questo caso verranno presentati alcuni risultati dell'impatto della strategia campionaria sulla stima delle frequenze assolute in alcune tabelle statistiche che presumibilmente faranno parte del piano di diffusione a livello di **comune** e di **area di censimento**.
- La analisi sono state condotte sugli stessi 40 Comuni oggetto della sperimentazione per la valutazione dell'efficienza delle stime (497 aree di censimento; poco più del 10% delle famiglie).

Analisi di qualità per tabelle statistiche per *Comune* e per *Area di censimento_2*

Tavola del "Grado di istruzione"

Popolazione residente di 6 anni e più per sesso (2 modalità) e grado di istruzione (8 modalità) per un totale di 16 incroci.

Tavola della "Attività economica"

Occupati per sesso (2 modalità), classi di età (4 modalità) ed attività economica (3 modalità) per un totale di 24 incroci.

Tavola della "Posizione nella professione"

Occupati per sesso (2 modalità), posizione nella professione (5 modalità) ed attività economica (3 modalità) per un totale di 30 incroci.

Tavola del "Settore di attività economica"

Occupati per sesso (2 modalità) e settore di attività economica (17 modalità) per un totale di 34 incroci.

Tavola della "Condizione professionale"

Popolazione residente di 15 anni e più per sesso (2 modalità) e condizione professionale (6 modalità) per un totale di 12 incroci.

Tavola del "Pendolarismo"

Popolazione residente che si sposta giornalmente per sesso (2 modalità) e per luogo di destinazione (2 modalità) per un totale di 4 incroci.

Valori mediани degli indicatori di qualità di Tavole determinate a livello di *Comune*.

Indicatori di qualità:

I1 → percentuale di celle valorizzate ($T > 0$) con frequenza assoluta $T < T_s$

I2 → percentuale di individui classificati in celle con frequenza assoluta $T < T_s$

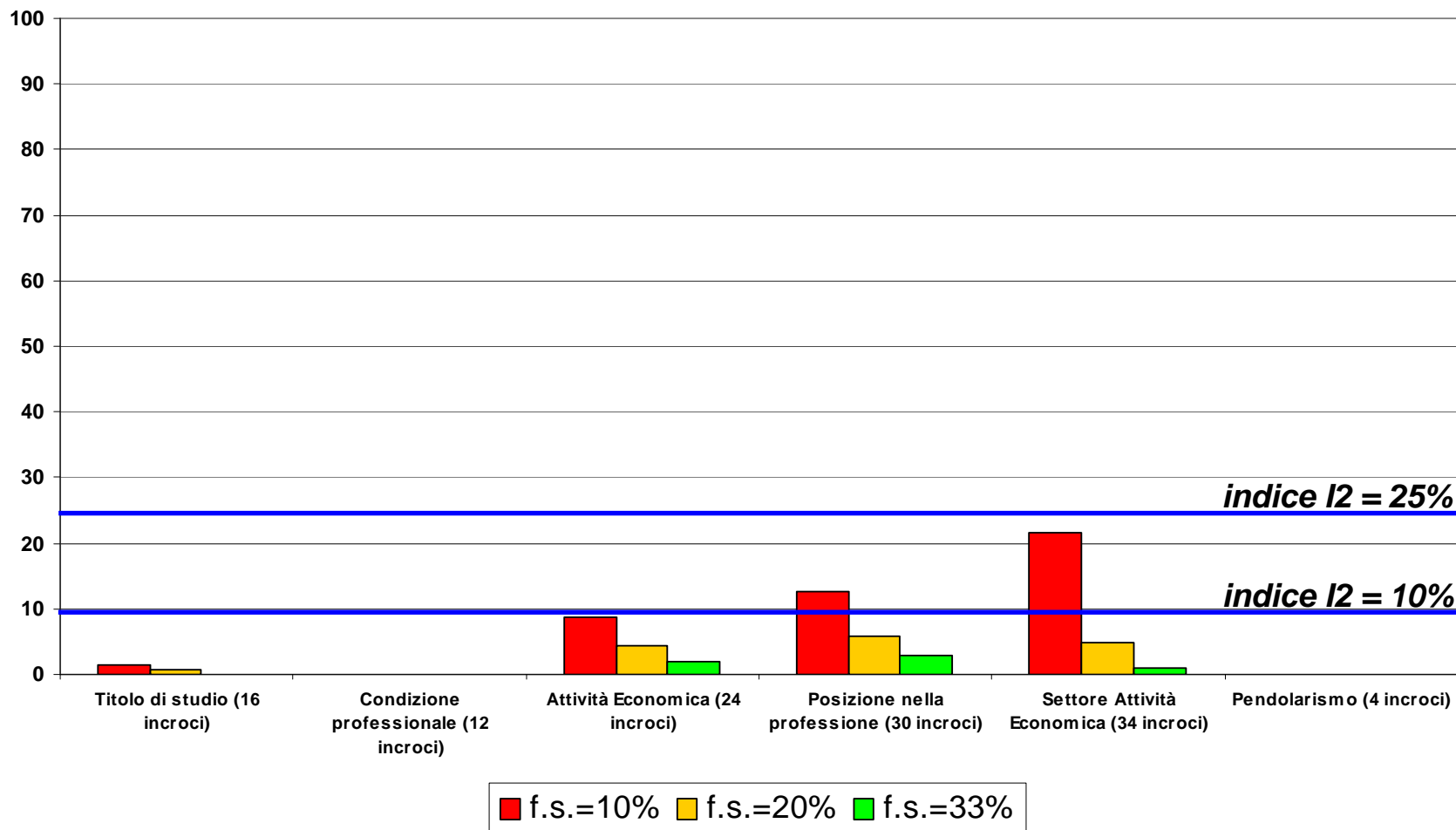
Err% massimo = 25%

Frazione Sondata	Soglia critica T_s	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo
f.s.=10%	500	I1	31,3	8,3	62,5	76,3	60,3	0,0
		I2	1,6	0,1	8,8	12,6	21,5	0,0
f.s.=20%	250	I1	18,8	8,3	53,2	66,7	36,8	0,0
		I2	0,7	0,1	4,3	5,9	4,9	0,0
f.s.=33%	100	I1	3,1	8,3	41,7	55,0	23,5	0,0
		I2	0,0	0,0	1,9	2,9	1,0	0,0

Esempio: nel caso di una strategia che preveda la f.s. del 10%, la soglia critica è pari a 500. Facendo riferimento al valore mediano della distribuzione degli indicatori calcolati per la tavola del “Titolo di Studio” sui 40 comuni: il 31,3% di celle hanno frequenza assoluta inferiore a 500 e quindi comportano un errore percentuale atteso non inferiore al 25%; in tale celle critiche è classificato l’1,6% (soltanto) degli individui relativi alla tavola.

Valori mediани degli indicatori di qualità di Tavole determinate a livello di *Comune*. Grafico

Percentuale di individui classificati in celle con frequenza assoluta inferiore al valore soglia definito per un errore massimo del 25% (mediana su 40 comuni).




Valori medi dei indicatori di qualità di Tavole determinate a livello di *Comune* per dimensione demografica.

Frazione Sondata	Soglia critica Ts	Dimensione demografica (dati in migliaia)	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo		
f.s.=10%	500	10 20	I1	43,8	41,7	86,3	87,8	93,9	0,0		
			I2	8,8	8,0	34,6	28,6	66,8	0,0		
		20 50	I1	31,3	20,8	74,5	84,5	80,0	0,0		
			I2	1,9	2,1	16,7	20,6	40,6	0,0		
		50 150	I1	21,9	8,3	55,3	66,7	39,7	0,0		
			I2	0,9	0,1	3,3	6,2	5,6	0,0		
		≥150	I1	9,4	8,3	47,9	53,3	23,5	0,0		
			I2	0,2	0,1	1,7	2,2	0,9	0,0		
		f.s.=20%	250	10 20	I1	31,3	16,7	71,7	79,3	80,0	0,0
					I2	2,5	1,5	18,0	18,9	35,0	0,0
				20 50	I1	31,3	8,3	60,9	74,2	60,0	0,0
					I2	1,6	0,1	6,6	10,5	19,2	0,0
50 150	I1			6,3	8,3	47,8	60,0	26,5	0,0		
	I2			0,3	0,1	2,8	3,6	1,3	0,0		
≥150	I1			6,3	8,3	31,3	40,0	22,1	0,0		
	I2			0,1	0,0	0,7	1,0	0,6	0,0		
f.s.=33%	100			10 20	I1	15,6	8,3	52,2	67,2	52,3	0,0
					I2	0,5	0,1	5,8	6,9	10,7	0,0
				20 50	I1	18,8	8,3	52,2	62,1	33,3	0,0
					I2	0,4	0,1	3,2	3,8	3,1	0,0
		50 150	I1	0,0	8,3	33,3	40,0	20,6	0,0		
			I2	0,0	0,0	0,9	1,1	0,6	0,0		
		≥150	I1	0,0	0,0	18,8	28,3	13,2	0,0		
			I2	0,0	0,0	0,2	0,4	0,2	0,0		


Valori medi dei indicatori di qualità di Tavole determinate a livello di *Comune* per dimensione demografica.

Frazione Sondata	Soglia critica Ts	Dimensione demografica (dati in migliaia)	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo
f.s.=10%	500	10 20	I1	43,8	41,7	86,3	87,8	93,9	0,0
			I2	8,8	8,0	34,6	28,6	66,8	0,0
		20 50	I1	31,3	20,8	74,5	84,5	80,0	0,0
			I2	1,9	2,1	16,7	20,6	40,6	0,0
		50 150	I1	21,9	8,3	55,3	66,7	39,7	0,0
			I2	0,9	0,1	3,3	6,2	5,6	0,0
		≥150	I1	9,4	8,3	47,9	53,3	23,5	0,0
			I2	0,2	0,1	1,7	2,2	0,9	0,0
f.s.=20%	250	10 20	I1	31,3	16,7	71,7	79,3	80,0	0,0
			I2	2,5	1,5	18,0	18,9	35,0	0,0
		20 50	I1	31,3	8,3	60,9	74,2	60,0	0,0
			I2	1,6	0,1	6,6	10,5	19,2	0,0
		50 150	I1	6,3	8,3	47,8	60,0	26,5	0,0
			I2	0,3	0,1	2,8	3,6	1,3	0,0
		≥150	I1	6,3	8,3	31,3	40,0	22,1	0,0
			I2	0,1	0,0	0,7	1,0	0,6	0,0
f.s.=33%	100	10 20	I1	15,6	8,3	52,2	67,2	52,3	0,0
			I2	0,5	0,1	5,8	6,9	10,7	0,0
		20 50	I1	18,8	8,3	52,2	62,1	33,3	0,0
			I2	0,4	0,1	3,2	3,8	3,1	0,0
		50 150	I1	0,0	8,3	33,3	40,0	20,6	0,0
			I2	0,0	0,0	0,9	1,1	0,6	0,0
		≥150	I1	0,0	0,0	18,8	28,3	13,2	0,0
			I2	0,0	0,0	0,2	0,4	0,2	0,0

 indice I2 < 10%

Valori medi dei indicatori di qualità di Tavole determinate a livello di *Comune* per dimensione demografica.

Frazione Sondata	Soglia critica Ts	Dimensione demografica (dati in migliaia)	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo
f.s.=10%	500	10 20	I1	43,8	41,7	86,3	87,8	93,9	0,0
			I2	8,8	8,0	34,6	28,6	66,8	0,0
		20 50	I1	31,3	20,8	74,5	84,5	80,0	0,0
			I2	1,9	2,1	16,7	20,6	40,6	0,0
		50 150	I1	21,9	8,3	55,3	66,7	39,7	0,0
			I2	0,9	0,1	3,3	6,2	5,6	0,0
		≥150	I1	9,4	8,3	47,9	53,3	23,5	0,0
			I2	0,2	0,1	1,7	2,2	0,9	0,0
f.s.=20%	250	10 20	I1	31,3	16,7	71,7	79,3	80,0	0,0
			I2	2,5	1,5	18,0	18,9	35,0	0,0
		20 50	I1	31,3	8,3	60,9	74,2	60,0	0,0
			I2	1,6	0,1	6,6	10,5	19,2	0,0
		50 150	I1	6,3	8,3	47,8	60,0	26,5	0,0
			I2	0,3	0,1	2,8	3,6	1,3	0,0
		≥150	I1	6,3	8,3	31,3	40,0	22,1	0,0
			I2	0,1	0,0	0,7	1,0	0,6	0,0
f.s.=33%	100	10 20	I1	15,6	8,3	52,2	67,2	52,3	0,0
			I2	0,5	0,1	5,8	6,9	10,7	0,0
		20 50	I1	18,8	8,3	52,2	62,1	33,3	0,0
			I2	0,4	0,1	3,2	3,8	3,1	0,0
		50 150	I1	0,0	8,3	33,3	40,0	20,6	0,0
			I2	0,0	0,0	0,9	1,1	0,6	0,0
		≥150	I1	0,0	0,0	18,8	28,3	13,2	0,0
			I2	0,0	0,0	0,2	0,4	0,2	0,0

 indice I2 < 10%

Valori medi dei indicatori di qualità di Tavole determinate a livello di *Comune* per dimensione demografica.

Frazione Sondata	Soglia critica Ts	Dimensione demografica (dati in migliaia)	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo
f.s.=10%	500	10 20	I1	43,8	41,7	86,3	87,8	93,9	0,0
			I2	8,8	8,0	34,6	28,6	66,8	0,0
		20 50	I1	31,3	20,8	74,5	84,5	80,0	0,0
			I2	1,9	2,1	16,7	20,6	40,6	0,0
		50 150	I1	21,9	8,3	55,3	66,7	39,7	0,0
			I2	0,9	0,1	3,3	6,2	5,6	0,0
		≥150	I1	9,4	8,3	47,9	53,3	23,5	0,0
			I2	0,2	0,1	1,7	2,2	0,9	0,0
f.s.=20%	250	10 20	I1	31,3	16,7	71,7	79,3	80,0	0,0
			I2	2,5	1,5	18,0	18,9	35,0	0,0
		20 50	I1	31,3	8,3	60,9	74,2	60,0	0,0
			I2	1,6	0,1	6,6	10,5	19,2	0,0
		50 150	I1	6,3	8,3	47,8	60,0	26,5	0,0
			I2	0,3	0,1	2,8	3,6	1,3	0,0
		≥150	I1	6,3	8,3	31,3	40,0	22,1	0,0
			I2	0,1	0,0	0,7	1,0	0,6	0,0
f.s.=33%	100	10 20	I1	15,6	8,3	52,2	67,2	52,3	0,0
			I2	0,5	0,1	5,8	6,9	10,7	0,0
		20 50	I1	18,8	8,3	52,2	62,1	33,3	0,0
			I2	0,4	0,1	3,2	3,8	3,1	0,0
		50 150	I1	0,0	8,3	33,3	40,0	20,6	0,0
			I2	0,0	0,0	0,9	1,1	0,6	0,0
		≥150	I1	0,0	0,0	18,8	28,3	13,2	0,0
			I2	0,0	0,0	0,2	0,4	0,2	0,0

○ indice I2 < 10%

Valori mediани degli indicatori di qualità di Tavole determinate a livello di area di censimento di centro.

Indicatori di qualità:

I1 → percentuale di celle valorizzate ($T > 0$) con frequenza assoluta $T < T_s$

I2 → percentuale di individui classificati in celle con frequenza assoluta $T < T_s$

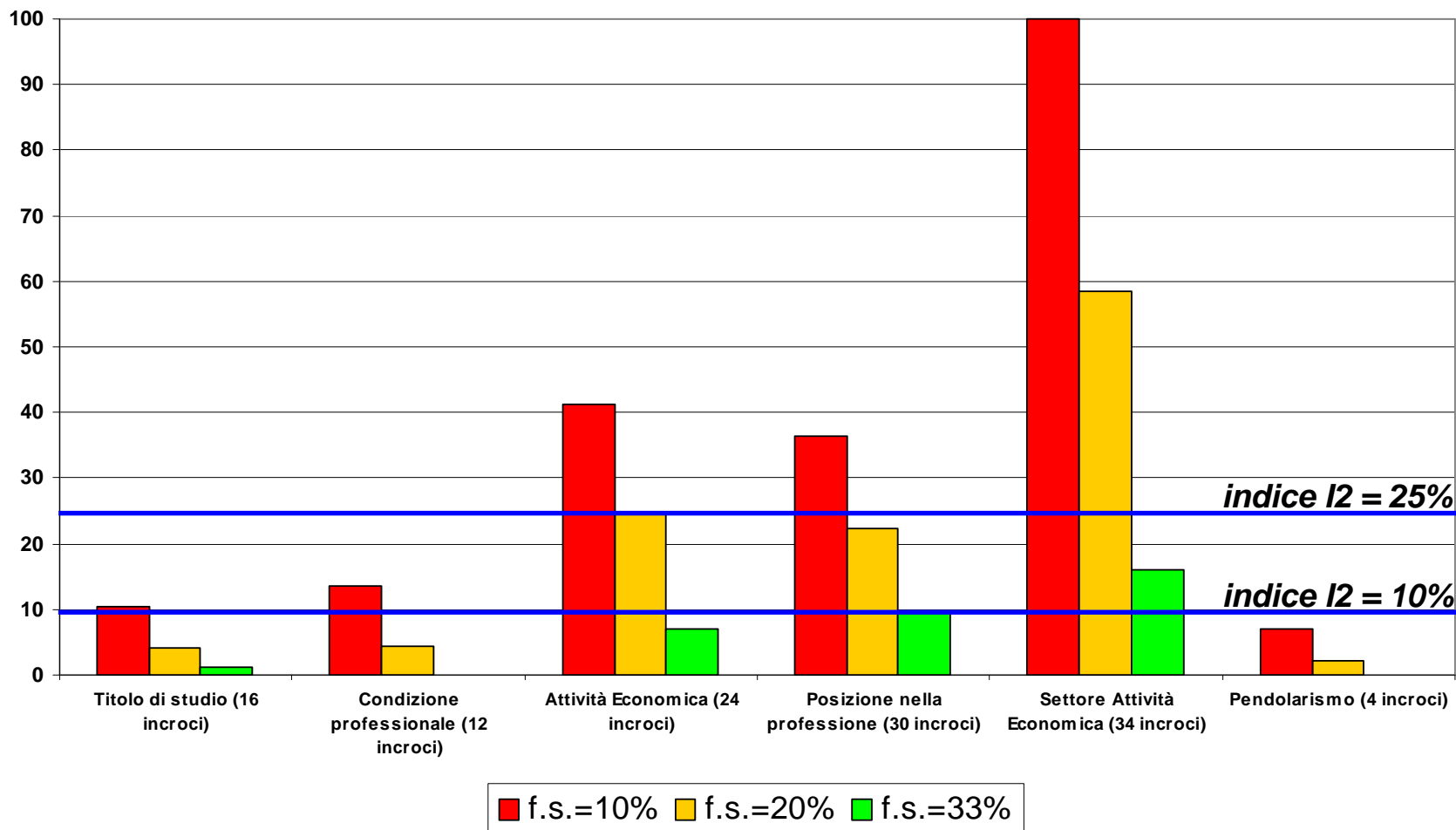
Err% massimo = 25%

Frazione Sondata	Soglia critica T_s	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo
f.s.=10%	500	I1	50,0	58,3	90,5	91,7	100,0	50,0
		I2	10,5	13,6	41,2	36,5	100,0	7,0
f.s.=20%	250	I1	37,5	25,0	81,4	85,2	87,1	25,0
		I2	4,1	4,4	24,8	22,2	58,4	2,1
f.s.=33%	100	I1	31,3	8,3	61,9	72,4	55,0	0,0
		I2	1,3	0,1	7,0	9,6	16,1	0,0

Esempio: nel caso di una strategia che preveda la f.s. del 10%, la soglia critica è pari a 500. Facendo riferimento al valore mediano della distribuzione degli indicatori calcolati per la tavola del “Titolo di Studio” sulle 497 aree: il 50,0% di celle hanno frequenza assoluta inferiore a 500 e quindi comportano un errore percentuale atteso non inferiore al 25%; in tale celle critiche è classificato il 10,5% degli individui relativi alla tavola.

Valori mediани degli indicatori di qualità di Tavole determinate a livello di area di censimento di centro. Grafico.

Percentuale di individui classificati in celle con frequenza assoluta inferiore al valore soglia definito per un errore massimo del 25% (mediana su 498 aree di censimento).



Valori medi dei indicatori di qualità di Tavole determinate a livello di area di censimento di centro per dimensione demografica.

Frazione Sondata	Soglia critica Ts	Dimensione demografica (dati in migliaia)	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo
f.s.=10%	500	<10	I1	62,5	58,3	90,9	92,3	100,0	50,0
			I2	16,7	16,4	46,1	42,7	100,0	10,8
		10 12	I1	50,0	58,3	90,5	91,7	97,1	50,0
			I2	10,2	13,8	40,5	35,0	89,6	7,3
		≥12	I1	50,0	50,0	87,0	89,3	96,9	25,0
			I2	8,3	11,8	37,0	32,4	85,4	5,4
f.s.=20%	250	<10	I1	46,7	41,7	85,7	88,0	93,8	25,0
			I2	6,3	8,9	32,5	28,3	74,9	4,4
		10 12	I1	37,5	33,3	81,0	85,2	84,4	0,0
			I2	3,8	4,6	24,8	22,1	55,4	0,0
		≥12	I1	37,5	16,7	73,9	82,1	78,1	0,0
			I2	2,6	2,3	17,3	17,8	41,6	0,0
f.s.=33%	100	<10	I1	31,3	8,3	66,7	76,9	63,6	0,0
			I2	1,7	0,2	9,1	12,9	25,6	0,0
		10 12	I1	31,3	8,3	60,9	74,1	54,5	0,0
			I2	1,4	0,1	6,7	9,1	14,8	0,0
		≥12	I1	31,3	8,3	57,1	69,2	47,1	0,0
			I2	1,1	0,1	4,6	7,2	10,4	0,0

Valori medi dei indicatori di qualità di Tavole determinate a livello di area di censimento di centro per dimensione demografica.

Frazione Sondata	Soglia critica Ts	Dimensione demografica (dati in migliaia)	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo
f.s.=10%	500	<10	I1	62,5	58,3	90,9	92,3	100,0	50,0
			I2	16,7	16,4	46,1	42,7	100,0	10,8
		10 12	I1	50,0	58,3	90,5	91,7	97,1	50,0
			I2	10,2	13,8	40,5	35,0	89,6	7,3
		≥12	I1	50,0	50,0	87,0	89,3	96,9	25,0
			I2	8,3	11,8	37,0	32,4	85,4	5,4
f.s.=20%	250	<10	I1	46,7	41,7	85,7	88,0	93,8	25,0
			I2	6,3	8,9	32,5	28,3	74,9	4,4
		10 12	I1	37,5	33,3	81,0	85,2	84,4	0,0
			I2	3,8	4,6	24,8	22,1	55,4	0,0
		≥12	I1	37,5	16,7	73,9	82,1	78,1	0,0
			I2	2,6	2,3	17,3	17,8	41,6	0,0
f.s.=33%	100	<10	I1	31,3	8,3	66,7	76,9	63,6	0,0
			I2	1,7	0,2	9,1	12,9	25,6	0,0
		10 12	I1	31,3	8,3	60,9	74,1	54,5	0,0
			I2	1,4	0,1	6,7	9,1	14,8	0,0
		≥12	I1	31,3	8,3	57,1	69,2	47,1	0,0
			I2	1,1	0,1	4,6	7,2	10,4	0,0

 indice I2 < 10%

Valori medi dei indicatori di qualità di Tavole determinate a livello di area di censimento di centro per dimensione demografica.

Frazione Sondata	Soglia critica Ts	Dimensione demografica (dati in migliaia)	Indicatori	Titolo di studio	Condizione professionale	Attività Economica	Posizione nella professione	Settore Attività Economica	Pendolarismo
f.s.=10%	500	<10	I1	62,5	58,3	90,9	92,3	100,0	50,0
			I2	16,7	16,4	46,1	42,7	100,0	10,8
		10 12	I1	50,0	58,3	90,5	91,7	97,1	50,0
			I2	10,2	13,8	40,5	35,0	89,6	7,3
		≥12	I1	50,0	50,0	87,0	89,3	96,9	25,0
			I2	8,3	11,8	37,0	32,4	85,4	5,4
f.s.=20%	250	<10	I1	46,7	41,7	85,7	88,0	93,8	25,0
			I2	6,3	8,9	32,5	28,3	74,9	4,4
		10 12	I1	37,5	33,3	81,0	85,2	84,4	0,0
			I2	3,8	4,6	24,8	22,1	55,4	0,0
		≥12	I1	37,5	16,7	73,9	82,1	78,1	0,0
			I2	2,6	2,3	17,3	17,8	41,6	0,0
f.s.=33%	100	<10	I1	31,3	8,3	66,7	76,9	63,6	0,0
			I2	1,7	0,2	9,1	12,9	25,6	0,0
		10 12	I1	31,3	8,3	60,9	74,1	54,5	0,0
			I2	1,4	0,1	6,7	9,1	14,8	0,0
		≥12	I1	31,3	8,3	57,1	69,2	47,1	0,0
			I2	1,1	0,1	4,6	7,2	10,4	0,0

 indice I2 < 25%

Considerazioni finali_1

- ✓ Per la diffusione di dati qualitativamente accettabili per domini sub-comunali, la frazione di campionamento deve attestarsi sui livelli più alti (33% della popolazione residente nel dominio).
- ✓ E' più favorevole il disegno delle aree di censimento tale che le dimensioni demografiche siano intorno ai 15.000 abitanti.
- ✓ Riguardo le variabili caratterizzanti le diverse tavole, la qualità delle stime peggiora al crescere del numero delle modalità previste dalle classificazioni adottate. Si suggerisce quindi, in assenza di vincoli europei a livello comunale e sub-comunale, di adottare modalità più aggregate di classificazione.

Considerazioni finali_2

- ✓ Particolare attenzione va data al fenomeno del pendolarismo. A riguardo, sono in atto specifici approfondimenti per la valutazione della qualità delle stime sugli spostamenti quotidiani fra comuni ai fini del successivo utilizzo per la determinazione dei Sistemi Locali del Lavoro.
- ✓ I risultati ottenuti con metodi di stima diretta suggeriscono di non applicarli ai comuni tra 5.000 e 20.000 abitanti.
- ✓ La possibilità di recuperare efficienza è offerta dai metodi di stima per piccole aree.



Esperienze internazionali mostrano che l'impiego di stimatori di questo tipo può ridurre l'errore nei casi più critici. I primi risultati di alcune sperimentazioni in corso indicano miglioramenti di efficienza con una riduzione dell'errore tra il 40% e l'80% per stime di frequenze assolute tra 50 e 150 unità. Un approccio di questo tipo permetterebbe di **migliorare la qualità delle tavole prodotte**, con particolare riferimento alle **aree di censimento** e ai **comuni più piccoli**.

Grazie per la cortese attenzione.