

# Una parte per il tutto: campioni e nuovo censimento

Matteo Borioli, Ustat

## Introduzione

La novità che verrà introdotta nel censimento del 2010 consiste nel passaggio da un rilevamento decennale di tipo esaustivo ad uno annuale, solo in parte esaustivo e basato su registri amministrativi integrati da indagini a campione. L'uso integrato dei dati provenienti dai registri amministrativi con quelli ottenuti grazie ad una serie di rilevazioni per campione<sup>1</sup>, ossia l'indagine strutturale, quella tematica e la Omnibus<sup>2</sup>, completa ed arricchisce ulteriormente l'offerta di informazioni del classico censimento esaustivo.

Il ricorso a nuovi metodi di rilevamento come quello campionario non è una novità assoluta per la statistica pubblica. E' di questo tipo, ad esempio, l'annuale rilevazione sulle forze di lavoro, la RIFOS. Con la sua adozione da parte dell'"ammiraglia" della statistica federale, però, assistiamo ad una notevole innovazione: il censimento della popolazione offre una radiografia a largo spettro della nostra società, ed è tra le fonti con il più alto numero di utilizzatori, che va dal mondo delle istituzioni, ai mezzi di informazione, alla scuola, ai singoli cittadini.

Diventa perciò del tutto essenziale che questo nuovo strumento non diventi un "oggetto misterioso", un oracolo che dia responsi criptici. Si dovrà anzi promuovere un'adeguata conoscenza delle due parti in cui



si articola: dapprima la raccolta dei dati, poi – e soprattutto – la loro lettura.

Questo lavoro di promozione trova in queste pagine una sua prima realizzazione. Cercheremo in effetti, prendendo spunto dall'inchiesta strutturale, di mostrare come si procede all'esecuzione dell'indagine, quali scelte si devono compiere<sup>3</sup>, in quale forma avremo a disposizione i dati. Vorremmo in particolare rendere più concreto quest'ultimo aspetto, e lo faremo presentando fianco a fianco i dati che avremmo ottenuto con un procedimento tradizionale (il questionario a ogni abitante) con quelli che produrrà il campione. Da questo confronto risulterà con tutta evidenza che si tratta di due "linguaggi" diversi. Prima di affrontare una "full immersion" in questo nuovo universo, converrà quindi partire da uno studio della sua "grammatica".

## L'inchiesta a campione: come funziona e sue condizioni

Un'"indagine a campione" può presentarsi sotto forme diverse, nel funzionamento, come nelle condizioni di realizzazione nonché nella capacità di rispondere a specifiche necessità. Quella illustrata qui di seguito riguarda la procedura di campionamento stabilita dall'Ufficio federale di statistica (UST) per l'indagine strutturale del censimento federale.

### Metodo e condizioni di campionamento per l'indagine strutturale

Il campionamento viene utilizzato quando si vuole conoscere uno o più parametri di una popolazione, senza doverli rilevare in ogni loro elemento (nel nostro caso, dei singoli abitanti). Nel caso dell'indagine strutturale, la popolazione di base o di riferimento, ossia la popolazione all'interno della quale estrarre il campione, è costituita dalle persone che vivono in un'economia domestica privata<sup>4</sup> e che hanno 15 o più anni.

Il campione di base contiene circa il 3,5% della popolazione di riferimento, il che lo porta a contare 200.000 persone a livello nazionale, delle quali circa 8.780 risiedono in Ticino<sup>5</sup>. La stessa percentuale viene applicata anche ai livelli spaziali inferiori; nel nostro caso, si tratta dei Comuni, che vedranno quindi interpellati 3,5 abitanti su 100. Le persone incluse nel campione in

<sup>1</sup> V. l'articolo che precede.

<sup>2</sup> V. la presentazione dettagliata delle 3 indagini campionarie ancora alle pp. 61- 62.

<sup>3</sup> Sottolineiamo ancora una volta l'elemento già richiamato nell'editoriale (v. p. 1) e che verrà ripreso in modo specifico nel riquadro finale (v. p. 73): le informazioni che forniamo in questo "inserto" dedicato al Censimento del 2010 vogliono essere anche uno stimolo per delle prese di posizioni che completino la consultazione in corso in vista della decisione del Consiglio di Stato in merito a un eventuale densificazione del campione ticinese.

<sup>4</sup> Sono quindi escluse le persone che risiedono in un'economia domestica collettiva, come ad esempio i convitti, gli internati, le case di cura o di riposo, i luoghi di detenzione, ecc...

<sup>5</sup> Viene in tal modo rispettata la proporzione ticinese nel contesto nazionale, che è del 4,4%.

un determinato anno sono escluse da altre future estrazioni per una durata minima di 5 anni.

La data di riferimento del primo rilevamento è il 31 dicembre 2010, in corrispondenza dunque del rilevamento esaustivo effettuato tramite i registri delle persone e degli edifici/abitazioni. Ricordiamo che il rilevamento strutturale ha come obiettivo la raccolta di informazioni su variabili non presenti nei registri e che vanno a coprire i seguenti temi: individui-economie domestiche-alloggi, lavoro, mobilità, formazione, lingue – religioni - cultura, famiglia (per informazioni dettagliate vedasi l'articolo precedente). Infine, l'inchiesta viene realizzata per via postale tramite questionario cartaceo con la possibilità di risposta anche tramite Internet.

## Quali risultati si ottengono da un'indagine a campione

I Censimenti esaustivi realizzati fino al 2000 ci hanno fornito i valori precisi all'unità dei parametri con i quali lavorare (totali, medie, percentuali, ...)<sup>6</sup>. A partire dal 2010, con l'introduzione del campionamento, saremo confrontati con delle stime, ossia con dei valori ai quali sono associati dei gradi di fiducia (vi è l'aggiunta di un intervallo di confidenza). Ad esempio, se nel 2000 la percentuale di donne straniere italofone era del 51,2%, nel caso il rilevamento fosse stato effettuato a campione avremmo ottenuto una percentuale differente (ad esempio 50,7%), ma l'intervallo di confidenza ( $\pm 0,7\%$ ) ci avrebbe garantito (con una probabilità del 95%) di contenere il valore reale (che si stima quindi si sarebbe situato tra il 50,0% ed il 51,4%).

Da un punto di vista pratico, a dipendenza delle analisi che si desidera condurre, la disponibilità di intervalli di confidenza troppo grandi può essere un grande impedimento alla descrizione dello stato e all'interpretazione dell'evoluzione in una determinata popolazione.

E' quindi necessario cercare di migliorare la stima, ossia ridurre l'ampiezza dell'intervallo di confidenza.

## La risoluzione e il livello di aggregazione

Due concetti chiave, quello di *risoluzione* e quello di *livello di aggregazione*, forniscono una chiave di lettura più pratica per comprendere i limiti di interpretazione dei dati ottenuti da inchieste a campione.

Per **livello di aggregazione** si intende la più piccola suddivisione in sottopopolazioni (una suddivisione territoriale o socio-economica) all'interno della quale i risultati sono ancora utilizzabili<sup>7</sup>. Con il campione di base il livello di aggregazione è di circa 16.000 persone.

Il livello di aggregazione può riferirsi naturalmente a qualsiasi suddivisione della popolazione di partenza, che può corrispondere ad esempio all'insieme formato dalle donne occupate residenti a Lugano, o da quello delle persone con una formazione universitaria.

Il secondo concetto è quello di **risoluzione**, con il quale si intende il più piccolo gruppo di persone sul quale è possibile indagare nella popolazione di riferimento o in una sua sottopopolazione. Se un censimento esaustivo permette di calcolare con precisione la consistenza di un fenomeno indipendentemente dal numero di persone che presentano questa caratteristica, un'indagine a campione può spingersi solo fino ad un determinato limite. Con il campione di base questo limite si situa a 140 persone, ossia l'osservazione di fenomeni che caratterizzano gruppi con meno di 140 individui non è possibile. Ad esempio, secondo i dati del censimento del 2000, gli apprendisti svizzeri residenti a Lugano erano 102, quindi inferiori al limite dei 140 individui. Di conseguenza non saremmo stati in grado di descrivere le caratteristiche di questa categoria.

Applicando questi due concetti ad un caso pratico, nel 2010 sarà possibile approfondire l'analisi per il comune di Bellinzona, che conta circa 16.000 abitanti, distinguendo sottogruppi fino ad una dimensione minima di 140 persone (ad esempio gli stra-

nieri di origine bosniaca). Non si potrà però approfondire l'analisi per gli iberici perché il loro numero è inferiore a 140 individui. Non sarà per contro possibile analizzare ad esempio il comune di Locarno perché conta meno di 16.000 abitanti.

## Come migliorare la stima (pooling e/o densificazione)

La presenza di intervalli di confidenza (IC) troppo ampi per poter essere utilizzati è uno degli inconvenienti principali delle indagini a campione. Innanzitutto l'immagine della realtà risulta sfocata rispetto a quanto si ottiene da un rilevamento esaustivo: più l'IC è ampio, meno preciso è il quadro della situazione. Inoltre nella lettura delle serie storiche, qualora gli IC fossero grandi e dovessero sovrapporsi fra loro, l'evoluzione del fenomeno in esame risulterebbe difficilmente interpretabile.

Nasce così l'esigenza di migliorare la qualità dei dati a disposizione, risultato che si ottiene principalmente aumentando la dimensione del campione. L'UST propone due strade per raggiungere questo obiettivo, il cosiddetto *pooling* e la densificazione del campione. Un aspetto da tener presente: la prima possibilità è messa gratuitamente a disposizione dall'UST; la seconda è a pagamento.

Il *pooling*, o accumulo dei dati, permette l'uso congiunto dei risultati di più anni successivi dell'inchiesta (v. tab. 1). Il raggruppamento di vari campioni su più anni permette così di ottenere dei risultati più precisi e di conseguenza la possibilità di fornire risultati per categorie di persone più ridotte o per zone geografiche più piccole rispetto al campione di base. In questa situazione però si avranno delle stime medie sull'insieme del periodo considerato e riferite ad una popolazione media sul periodo, e non più il dato di un preciso anno. Inoltre la lettura dei dati su *pooling* presenta l'inconveniente di utilizzare il medesimo campione in più operazioni

<sup>6</sup> La popolazione residente del Ticino nel censimento del 2000 era di 306'846 abitanti: un valore preciso all'unità, ma non necessariamente esatto all'unità (qualcuno sarà senz'altro sfuggito al rilevamento).

<sup>7</sup> L'UST ha fissato dei limiti entro i quali muoversi e che sono stati definiti in base all'esperienza maturata in altri campionamenti legati alla popolazione.

Tipo	Dimensione campione (CH)	Anno di diffusione dei risultati							
		2011	2012	2013	2014	2015	2016	2017	...
Dati annuali	200.000 persone	2010	2011	2012	2013	2014	2015	2016	...
<i>Pooling</i> su 3 anni	600.000 persone	...	...	2010-2012	2011-2013	2012-2014	2013-2015	2014-2016	...
<i>Pooling</i> su 5 anni	1.000.000 persone	...	...	...	...	2010-2014	2011-2015	2012-2016	...

distinte. Il dato del 2014, ad esempio, figurerà in tre raccolte (medie) distinte (2012-2014, 2013-2015 e 2014-2016).

I dati su *pooling* sono prodotti dall'UST dopo tre e cinque anni (*pooling* triennale e *pooling* quinquennale).

Per densificazione – l'altra possibilità avanzata dall'UST – si intende l'allargamento del campione per una determinata indagine (anno). L'allargamento (la moltiplicazione del campione per 2, per 3 o altro) ha un limite teorico solo nella dimensione dello stesso universo (ad un certo punto ci ritroveremo

con un censimento esaustivo), ma nella nostra realtà deve rispettare le condizioni poste dall'UST ed esposte nel riquadro di p. 68.

La densificazione si riflette ovviamente sul *pooling*, incrementando ulteriormente il guadagno in termini di precisione: ad esempio, se si sceglie di raddoppiare costantemente il campione, dopo tre anni questo sarà sei volte più grande rispetto a quello dell'offerta di base (3,5% [campione di base] × 2 [raddoppio] × 3 [pooling su 3 anni] = 21% della popolazione di riferimento).

La tabella 2 mostra la progressiva riduzione degli IC rispetto ad un valore iniziale normalizzato a 1 a seconda della dimensione del campione.

Un confronto, a parità di periodo, tra il campione di base e una densificazione (raddoppio) fornisce i valori di rapporto, ossia di riduzione dell'ampiezza degli IC, riportati nella tabella 3.

Come possiamo vedere nella colonna riguardante il rapporto tra campionamento con e senza raddoppio, il guadagno che si ottiene è sempre significativo, sia che si utilizzino i dati annuali o quelli con *pooling* triennale o quinquennale (tra il 62% ed il 69% circa). La densificazione non va quindi ad escludere i benefici ottenuti grazie al *pooling* ma va a sommarsi a questi ultimi.

Ritornando ai concetti di risoluzione e di livello di aggregazione visti precedentemente, il raddoppio del campione abbassa il limi-

## Il *pooling*

### Vantaggi

- E' una tecnica sperimentata: le inchieste dell'UST sul budget delle economie domestiche (EBM) e quelle sulle condizioni di salute adottano già questo procedimento.
- Il guadagno in termini di precisione è importante e relativamente semplice da realizzare.
- La precisione del *pooling* su 3 (5) anni corrisponde circa alla precisione di un campione di 600.000 (1.000.000) persone.
- Il *pooling* dei dati del programma standard viene fornito gratuitamente.

### Svantaggi

- Si ragiona in termini di media di un periodo: l'interpretazione dei dati risulta quindi più complessa.
- Per avere a disposizione i primi dati sarà necessario attendere 3 o 5 anni.
- Il *pooling* è più efficace se i campioni sono della medesima dimensione. E' quindi raccomandabile favorire la stabilità nella forma del rilevamento.

## 2 Riduzione degli IC in seguito a densificazione e/o *pooling*

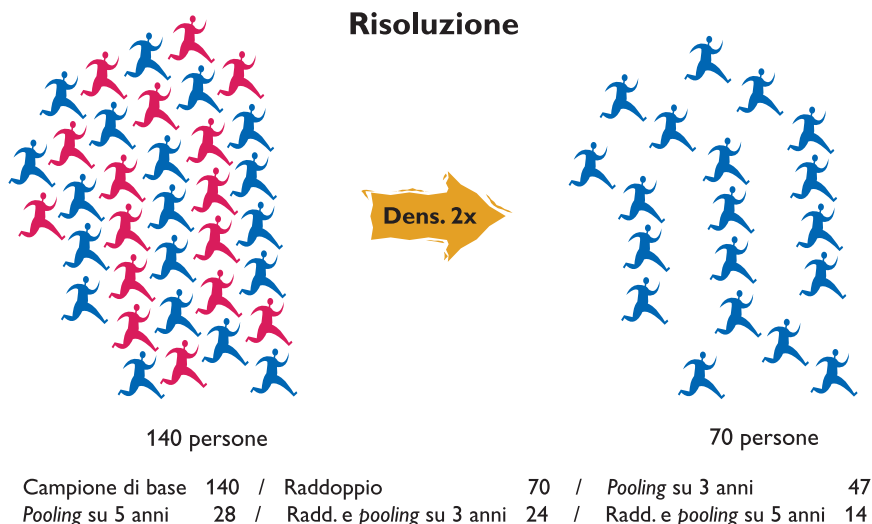
Densità campione		Ampiezza IC	Esempio
3,50%	Senza densificazione per 1 anno	1,000	(± 2,2)
7,00%	Con densificazione (2x) per 1 anno	0,694	(± 1,5)
10,50%	Senza densificazione e con <i>pooling</i> su 3 anni	0,556	(± 1,2)
17,50%	Senza densificazione e con <i>pooling</i> su 5 anni	0,414	(± 0,9)
21,00%	Con densificazione (2x) e <i>pooling</i> su 3 anni	0,369	(± 0,8)
35,00%	Con densificazione (2x) e <i>pooling</i> su 5 anni	0,260	(± 0,6)

## 3 Rapporto di riduzione degli IC con e senza raddoppio del campione

Periodo	Ampiezza IC		Rapporto con/senza raddoppio
	Senza densificazione	Con densificazione (2x)	
1 anno	1,000	0,694	0,694
3 anni	0,556	0,369	0,664
5 anni	0,414	0,260	0,627

«Con l'introduzione del campionamento, saremo confrontati con delle stime, ossia con dei valori ai quali sono associati dei gradi di fiducia (vi è l'aggiunta di un intervallo di confidenza).»

## A Risoluzione e livello di aggregazione prima e dopo il raddoppio del campione di base



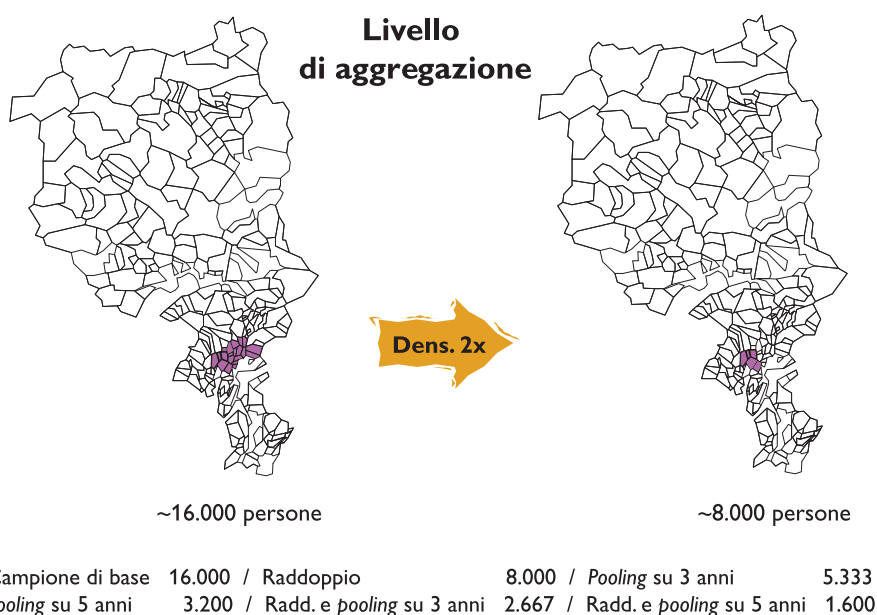
te della risoluzione da 140 a 70 individui (v. fig. A). Lo stesso vale per il livello di aggregazione, con il valore che si dimezza al raddoppio del campione di base.

In altre parole è come se utilizzassimo una lente di ingrandimento più potente, la quale ci permette di vedere i fenomeni con maggiore precisione.

### Campione di base o raddoppio? Alcuni esempi ticinesi

Come illustrato finora, il nuovo censimento federale della popolazione introduce delle novità alle quali sono associate vantaggi e limiti. Le restrizioni degli intervalli di confidenza e delle condizioni imposte dal livello di aggregazione e dalla risoluzione si percepiscono con maggiore chiarezza allorquando si analizzano i dati. E' possibile suddividere questi inconvenienti in tre grandi categorie: la perdita di informazione territoriale, la perdita di profondità dell'analisi e la perdita di precisione.

Faremo di seguito qualche esempio di ciò che si potrà trovare a partire dal 2010, esempi che illustreremo sia con delle tabelle, sia sotto forma di grafici. Negli esempi tabellari ci occuperemo del dato di un solo anno, perché ai dati del censimento 2000 (dato reale) sono accostati quelli che avremmo ottenuto se già allora avessimo fatto ricorso a un'indagine campio-



#### Come aumentare il campione di base

Qualora i Cantoni ritengano che il campione di base non risponda alle proprie esigenze, (ad esempio qualora non è possibile ottenere dei dati significativi per le principali suddivisioni territoriali), questi possono chiederne un aumento (una **densificazione**) alle seguenti condizioni:

- il campione annuale non può superare il 14% della popolazione di riferimento. Toccando questo tetto massimo, avremo una densificazione di 10,5 punti percentuali in aggiunta al 3,5% di base.
- Il campione cumulato su 10 anni non può superare il 70% della popolazione di riferimento. Ad esempio con una densificazione annuale costante del 3,5% avremmo un tas-

so di campionamento del 7,0% (3,5 + 3,5) ed il suo accumulo su 10 anni corrisponderebbe a questo limite superiore del 70%.

- I campioni densificati sono interrogati sulle stesse variabili e secondo le stesse procedure del campione di base. Non è pertanto possibile introdurre nuove domande nel questionario o modificare quelle esistenti.
- Nel rispetto dei punti precedenti può essere richiesto qualsiasi tasso di densificazione.

Ogni cantone è libero di scegliere la strategia più appropriata alle proprie esigenze. L'UST ha comunque formulato delle raccomandazioni affinché il processo di cam-

pionamento sia più efficace:

- si auspica un coordinamento tra i Cantoni riguardo sia al tipo di variante che al volume della densificazione. Questo aumenterebbe la comparabilità dei dati tra Cantoni. Inoltre per Cantoni vicini questo permetterebbe una misura migliore dei fenomeni sovracantonali.
- La variante ed il volume della densificazione dovrebbero essere stabili nel corso del tempo per ragioni organizzative e per una maggiore efficacia metodologica.
- Viene sconsigliato di procedere ad una densificazione non regolare a livello regionale, ossia senza una ripartizione proporzionale a livello spaziale.

naia (dato stimato, espresso con un IC). Negli esempi illustrati nei grafici, lo stesso tipo di confronto verrà applicato a una serie di dati annuali, quindi all'evoluzione di una realtà. La simulazione è avvenuta creando dapprima una popolazione fittizia<sup>8</sup> per ciascuno degli anni del periodo prescelto (qui, un decennio), quindi estraendone i dati individuo per individuo

(come in un censimento tradizionale), e per finire intervenendo sulla stessa popolazione ma con una rilevazione a campione. L'attenzione deve essere posta non tanto sull'evoluzione del fenomeno (si tratta di una evoluzione ipotetica), quanto sulla differenza che esiste tra il dato esaustivo e quello stimato, e sulla difficoltà nell'analizzare l'evoluzione di quest'ultimo.

#### La perdita di informazione territoriale

Il confronto tra le due parti della tabella 4 evidenzia quanto il dato del 2010 possa essere frammentario. Il limite imposto dalla soglia di 16.000 individui legata al livello di aggregazione mostra qui il suo grande peso: non solo in Ticino vi sono ben pochi comuni con una popolazione superiore a questa

## 4 Un esempio di perdita di informazione territoriale

### Popolazione residente secondo il settore economico<sup>1</sup> e il distretto, nel 2000 (in %)

	Valori effettivi			Valori ricostruiti con campione ipotetico		
	Primario	Secondario	Terziario	Primario	Secondario	Terziario
<b>Uomini</b>						
Distretto di Bellinzona	2,2	25,8	58,2	...	...	...
Distretto di Blenio	10,3	32,8	39,3	...	...	...
Distretto di Leventina	7,2	35,3	39,2	...	...	...
Distretto di Locarno	2,7	26,8	53,8	2,7 (±0,8)	26,8 (±2,6)	53,8 (±2,9)
Distretto di Lugano	1,5	18,9	64,2	1,5 (±0,5)	18,9 (±1,5)	64,2 (±1,9)
Distretto di Mendrisio	1,9	19,7	65,2	...	...	...
Distretto di Riviera	2,0	37,6	45,7	...	...	...
Distretto di Vallemaggia	5,9	34,9	40,4	...	...	...
<b>Totale</b>	<b>2,3</b>	<b>23,2</b>	<b>59,3</b>	<b>2,3 (±0,4)</b>	<b>23,2 (±1,1)</b>	<b>59,3 (±1,2)</b>
<b>Donne</b>						
Distretto di Bellinzona	0,8	8,3	75,6	...	...	...
Distretto di Blenio	4,4	9,4	54,7	...	...	...
Distretto di Leventina	2,2	10,2	64,1	...	...	...
Distretto di Locarno	1,1	10,6	71,4	...	...	...
Distretto di Lugano	0,5	7,3	76,8	0,5 (±0,5)	7,3 (±1,2)	76,8 (±1,9)
Distretto di Mendrisio	0,6	9,9	76,0	...	...	...
Distretto di Riviera	1,5	11,6	70,8	...	...	...
Distretto di Vallemaggia	3,0	8,9	66,9	...	...	...
<b>Totale</b>	<b>0,9</b>	<b>8,7</b>	<b>74,6</b>	<b>0,9 (±0,3)</b>	<b>8,7 (±0,9)</b>	<b>74,6 (±1,3)</b>
<b>Totale</b>						
Distretto di Bellinzona	1,6	18,5	65,5	1,6 (±0,6)	18,5 (±1,9)	65,5 (±2,4)
Distretto di Blenio	8,2	24,5	44,8	...	...	...
Distretto di Leventina	5,4	26,1	48,3	...	...	...
Distretto di Locarno	2,0	19,9	61,4	2,0 (±0,6)	19,9 (±1,8)	61,4 (±2,2)
Distretto di Lugano	1,1	14,0	69,6	1,1 (±0,3)	14,0 (±1,0)	69,6 (±1,3)
Distretto di Mendrisio	1,4	15,7	69,7	1,4 (±0,5)	15,7 (±1,8)	69,7 (±2,3)
Distretto di Riviera	1,8	27,9	55,1	...	...	...
Distretto di Vallemaggia	4,8	25,3	50,2	...	...	...
<b>Totale</b>	<b>1,7</b>	<b>17,2</b>	<b>65,7</b>	<b>1,7 (±0,3)</b>	<b>17,2 (±0,7)</b>	<b>65,7 (±1,0)</b>

<sup>1</sup> Data la loro irrilevanza ai fini del confronto che qui ci interessa, non abbiamo considerato i casi di "senza indicazione".

<sup>8</sup> Attraverso procedure informatiche, si sono attribuite (con procedimento casuale) le caratteristiche oggetto della simulazione a tutti gli individui che la compongono. Si è in altre parole "inventata" una popolazione.

«Dal lato degli inconvenienti, abbiamo a che fare con tre tipi di perdita: di informazione territoriale, di profondità dell'analisi e di precisione.»

soglia, ma anche facendo delle aggregazioni sia a livello di distretto sia di comprensori o di sub-regioni, la disponibilità di dati rimane pur sempre limitata.

#### La perdita di profondità dell'analisi

Qualora la sottopopolazione fosse sufficientemente grande da permettere l'acquisizione di risultati validi, la presenza degli intervalli di confidenza renderebbe tuttavia più laboriosa la lettura dei dati (v. tab. 5). Ad esempio, se una variabile dovesse presentare più modalità espresse da piccole percentuali vicine tra loro, la probabilità di vedere la loro distribuzione sovrapporsi aumenterebbe e di conseguenza la lettura ne risulterebbe troppo condizionata. In un caso simile sarà dunque necessario ridefinire i valori della variabile aggregandoli tra di loro in modo che il dato risulti più leggibile.

#### La perdita di precisione

L'interpretazione dei dati risulta facilitata adottando la rappresentazione grafica. Nel grafico B sono illustrate le coppie di valori (censimento annuale e censimento con campionatura) che si ottengono attraverso i dati annuali, i dati con *pooling* triennale ed infine quelli con *pooling* quinquennale. La linea di colore azzurro rappresenta la serie di censimenti esaustivi annuali, le barre di colore rosso raffigurano invece i censimenti a campione. L'altezza della barra riproduce l'ampiezza degli intervalli di confidenza, maggiore nei campioni annuali e sempre più ridotta nei dati su *pooling*, in virtù della maggiore dimensione del campione utilizzato per ottenerli.

Un attento esame della serie storica dei dati annuali rivela che non tutti gli intervalli di confidenza racchiudono al loro interno il valore reale (punto blu). Tale eventualità è dovuta alla probabilità, fissata al 5%, che il valore reale si collochi al di fuori dell'IC calcolato (vedi ancora gr. B), come ad esempio avviene per il campionamento del 2013.

Poiché il dato su *pooling* equivale alla media degli anni considerati, esso livella la

	Valore effettivi		Valori ricostruiti su campione ipotetico	
	Uomini	Donne	Uomini	Donne
Nessuna formazione	1,8	2,0	1,6 (±0,2)	1,8 (±0,3)
Scuola dell'obbligo	21,9	30,5	22,9 (±0,8)	31,4 (±0,8)
Scuola di cultura generale	1,5	3,3	1,4 (±0,2)	3,2 (±0,3)
Apprendistato, scuola prof. a tempo pieno	30,4	25,8	30,6 (±0,9)	25,5 (±0,8)
Liceo	6,1	7,3	5,9 (±0,4)	7,6 (±0,5)
Magistrale	1,3	3,4	1,2 (±0,2)	3,2 (±0,3)
Formazione professionale superiore	5,2	2,6	5,1 (±0,4)	2,1 (±0,3)
Scuola professionale superiore	3,2	1,7	2,9 (±0,3)	1,6 (±0,2)
Alta scuola specializzata	1,0	0,8	0,9 (±0,2)	0,9 (±0,2)
Università, politecnico	7,0	3,4	7,2 (±0,5)	3,6 (±0,3)
Età scolare	10,8	9,2	10,3 (±0,6)	8,9 (±0,5)
Età prescolare	5,2	4,6	5,3 (±0,4)	4,7 (±0,4)
Senza indicazione	4,5	5,4	4,7 (±0,4)	5,5 (±0,4)
<b>Totale</b>	<b>100,0</b>	<b>100,0</b>	<b>100,0</b> ...	<b>100,0</b> ...

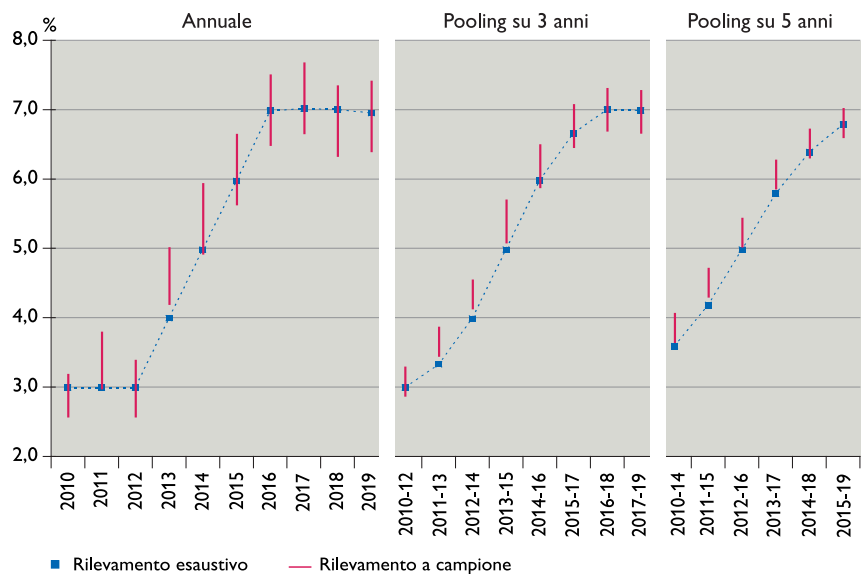
tendenza del periodo, smussando eventuali brusche variazioni. Questo effetto è chiaramente visibile nel dato su *pooling* triennale ed in quello quinquennale è ancora più accentuato: qui traspare unicamente la tendenza del decennio, ossia l'aumento del fenomeno, malgrado nei primi e negli ultimi rilevamenti del periodo la crescita sia nulla.

Considerando il contributo offerto da ogni insieme di dati, la possibilità di interpre-

tare correttamente un fenomeno e la sua evoluzione risulta adeguata allorché si dispone di almeno due *pooling* triennali. Poiché i tempi tecnici per l'allestimento delle banche dati sono di circa 8-10 mesi, ne consegue che l'attesa per avere queste informazioni può essere quantificata in almeno cinque anni dal primo rilevamento, indipendentemente che si sia scelto o meno di raddoppiare (densificare) il campione di base.

## B Confronto tra censimento esaustivo e censimento a campione

Disoccupati residenti, 2010-2019 (in %)

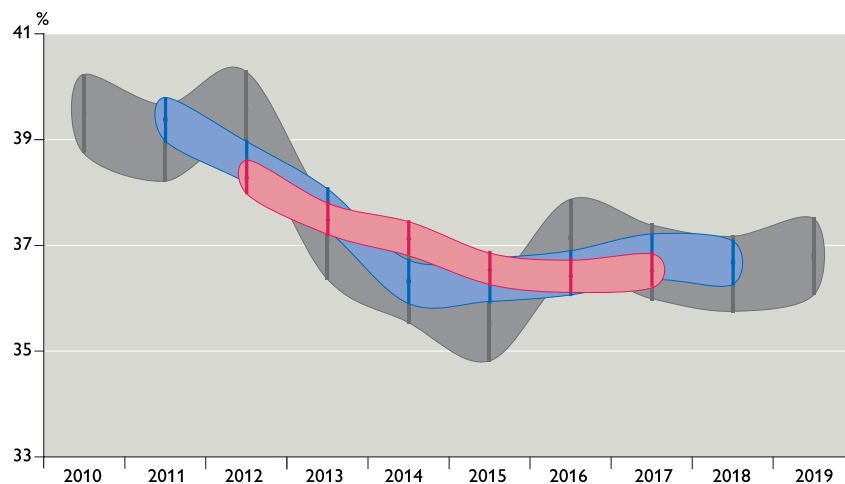


Nel grafico C sono illustrati i dati, rappresentati come delle fasce di ampiezza pari all'IC, che si potranno avere alla fine del prossimo decennio. In nero vi sono i dati annuali, in blu il *pooling* triennale ed in rosso quello quinquennale. La data di riferimento dei dati su *pooling* è quella dell'anno centrale all'intervallo considerato: ad esempio con il dato blu del 2011 si intende il valore medio del periodo 2010-2012.

L'evoluzione annuale mostra una diminuzione dei valori nel periodo centrale ed una probabile (gli IC, anche se per poco, si sovrappongono ostacolando l'interpretazione) successiva leggera ripresa. Aiutandoci con i dati su *pooling* è possibile approfondire l'analisi. In questo caso è importante sottolineare come, nel 2014, i dati su *pooling* triennale e quinquennale siano differenti (gli intervalli non si toccano). Questo sta a significare che la variazione (diminuzione) osservata con il dato annuale (fascia grigia) è molto significativa. Invece, negli anni che chiudono il decennio, i dati su *pooling* sono molto simili: questo evidenzia a sua volta una stabilità nell'evoluzione che si impone sulla maggiore altalenanza del dato annuale (fascia grigia).

Quali sono quindi i vantaggi del campionamento rispetto ai censimenti precedenti?

## C Confronto tra dati campionari annuali e su *pooling* Pensionati residenti svizzeri, 2010-2019 (in %)



Considerando che in passato i dati disponibili erano due, quelli di inizio e di fine periodo, con la nuova tecnica di rilevamento vi è la possibilità di rilevare con maggiore puntualità le inversioni di tendenza. Le stime, pur non indicando la reale consistenza della diminuzione, restituiscono un quadro tutto sommato chiaro della situazione, in particolar modo allorquando si impiegano le medie sul periodo con gli IC di ampiezza ridotta.

## I costi

L'inchiesta strutturale, nella sua offerta di base, non richiede nessuna spesa supplementare per i Cantoni, dato che tutte le spese sono a carico della Confederazione. Il Ticino disporrà così gratuitamente per questa indagine di un campione annuo di circa 8.700 persone e dei dati con *pooling* triennale e quinquennale.

Se si prevede invece di raddoppiare il nostro campione di base, bisognerà calcolare una spesa annua di circa 85.000 franchi, cifra ottenuta moltiplicando il costo di ogni inter-

### Valutazione dei costi per l'indagine strutturale

- Campione di base (3,5% dell'universo): **nessun costo per i cantoni.**
- Fino al raddoppio (3,5% - 7%): **CHF 9.- per persona.**
- Tra il raddoppio ed il quadruplo (7% - 14%): **CHF 15.- per persona.**
- Tutto quanto va al di là del quadruplo del campione di base, in particolare per un rilevamento esaustivo: **ca. CHF 21.- per persona.**



foto Ti-Press

«Se si prevede di raddoppiare il nostro campione di base, bisognerà calcolare una spesa annua di circa 85.000 franchi.»

vista fissato dall'UST (9 franchi) per il numero di individui intervistati (v. tab. 6).

A titolo informativo ricordiamo che i costi sono differenti a seconda della dimensione del campione dopo la densificazione. I costi supplementari sono giustificati da un maggior impegno organizzativo e logistico per realizzare il rilevamento e possono essere stimati in circa 9 franchi per individuo fino al raddoppio e in 15 franchi tra il raddoppio e il quadruplo. Per le altre modalità di densificazione rimandiamo al riquadro di p. 71.

L'UST lascia ad ogni cantone la possibilità di decidere la strategia di densificazione più appropriata alle proprie esigenze purché si rispettino

	Anno x
Popolazione	330.000
Popolazione 15 anni e più	272.377
Campione di base	9.533
Densificazione	3,5%
Campione aggiunto	9.533
Totale campione	19.066
Costo a persona in CHF	9
<b>Costo totale</b>	<b>85.799</b>

le condizioni illustrate nel riquadro di p. 68. Oltre alle due principali varianti (nessuna densificazione e raddoppio costante del campione), vi sono numerose strade percorribili a seconda che si desideri seguire con la maggiore precisione possibile un dato molto variabile nel tempo o un dato sostanzialmente stabile (vedi fig. D). Nascono

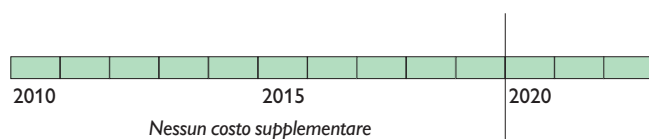
così varianti con una marcata densificazione in un determinato anno (variante E) o altre che vedono il campione densificato "spalmato" su più anni (varianti C1 e C2), o ancora una densificazione costante ma inferiore al raddoppio del campione di base, ognuna delle quali presenta dei costi complessivi differenti.

## D Esempi di possibili strategie di densificazione

La popolazione di base per il calcolo è la popolazione residente nel 2000; tra parentesi il costo con quella stimata nel 2010

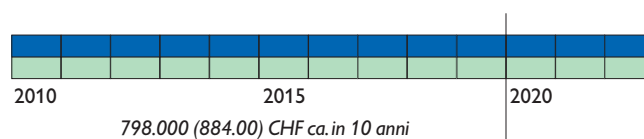
### Scenario A

#### Nessuna densificazione



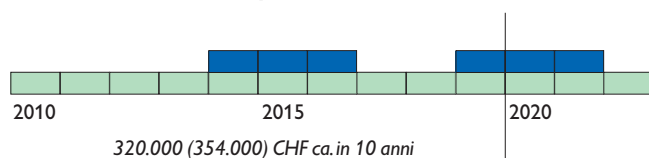
### Scenario B

#### Raddoppio costante del campione



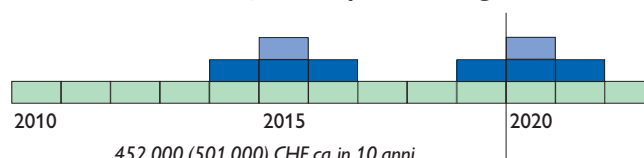
### Scenario C1

#### Raddoppio del campione solo negli anni a cavallo di quelli che terminano con le cifre 5 o 0



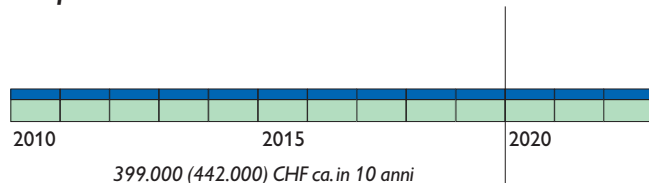
### Scenario C2

#### Densificazione del campione solo negli anni a cavallo di quelli che terminano con le cifre 5 o 0, più decisa negli anni centrali



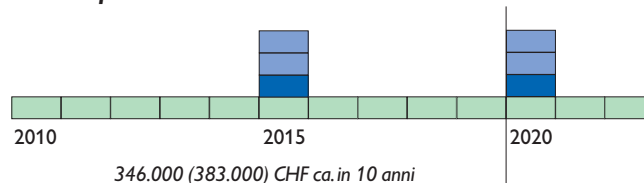
### Scenario D

#### Densificazione costante inferiore al raddoppio dal campione di base



### Scenario E

#### Densificazione molto marcata negli anni che terminano con le cifre 5 o 0



Campione di base (gratuito)

Fino al raddoppio del campione (9.- per individuo supplementare)

Tra il raddoppio e il quadruplo (15.- per individuo supplementare)



## Conclusioni

A partire dal 2010 l'orizzonte statistico si amplierà notevolmente; basti pensare alle informazioni sulle economie domestiche e sugli edifici ed abitazioni fino ad oggi disponibili unicamente a cadenza decennale. Nuove informazioni potranno essere utilizzate al fine di descrivere e comprendere meglio la realtà nella quale viviamo e ci muoviamo quotidianamente; nuovi percorsi di analisi potranno essere intrapresi, mentre quelli già sperimentati potranno arricchirsi di sfaccettature finora irrealizzabili.

La cadenza annuale permetterà di seguire i fenomeni con maggiore puntualità, ma la tecnica del campionamento alla quale si è dovuto ricorrere introduce anch'essa delle innovazioni che non sempre agevolano il lavoro di descrizione/analisi. Tra i principali elementi sacrificati in nome della modernizzazione del censimento vi sono, per quanto attiene alle indagini a campione, la perdita di informazioni a determinati livelli territoriali, la diminuzione della profondità interpretativa e la complessità della gestione delle stime.

In una certa misura si è in grado di compensare questi limiti aumentando il campione preso in esame, anche se questo accorgimento non sempre riesce a risolvere i problemi di precisione. Per questa ragione la decisione di procedere in tal senso deve essere accuratamente soppesata, coinvolgendo nel processo di valutazione tutti gli ambienti interessati.

Gli elementi evidenziati ed approfonditi in queste pagine mettono a disposizione degli utilizzatori dei dati dei censimenti gli strumenti necessari a comprendere il cambiamento col quale dovranno raffrontarsi, così da esprimere, qualora fosse il caso, eventuali loro esigenze. ■

Ulteriori informazioni si possono trovare nel sito web dell'Ufficio federale di statistica (<http://www.bfs.admin.ch/bfs/portal/fr/index/news/00/03.html>).

## Quale campione per quali dati? Dite la vostra

Ci basterà una 1:50.000 o sarebbe meglio partire con in mano una 1:25.000? Ci immaginiamo qui la scelta della cartina geografica da comperare per una escursione mai fatta. Perché il dilemma si pone identico a noi, in questa vigilia del nuovo censimento. La raccolta dei dati di base attraverso i registri (e non più direttamente chiedendoli ai cittadini) rappresenta in effetti solo la mezza novità del censimento 2010: cambia solo il modo, non il contenuto. La vera, la nuova novità è invece il ricorso alle inchieste presso un campione di cittadini. E, così come una cartina a scala ridotta permetterà di ritrovare più facilmente – grazie ai suoi maggiori dettagli – il sentiero giusto, un campione più numeroso ci consegnerà una conoscenza più precisa dei diversi aspetti della nostra realtà sociale.

Coprire un territorio con un cartina più dettagliata costa di più; lo stesso vale per un censimento che si affidi a un campione più numeroso. Completato con questa precisazione il nostro dilemma, non ci resta che affrontarlo.

Gli strumenti tecnici che ci devono servire nella “scelta di scala” ci sono stati presentati nei due contributi di questa sezione del “cantiere statistico”: con una data dimensione del campione si ottengono dei dati con quel dato intervallo di fiducia. Per i passi successivi basta affidarsi all’“offerta di cartine” presentata dall’Ufficio federale di statistica: i campioni si possono ingrandire (“densificare”) in diversi modi; ogni modo ha un suo prezzo.

Non è qui il caso di entrare nei dettagli di questi due aspetti. Basterà segnalare che al di là dell’“offerta di base” (campione e *pooling*; v. p. 68), la densificazione può prendere la forma di un ingrandimento (il raddoppio del campione, la sua triplicazione,...) annuale; di un ingrandimento maggiore concentrato sugli anni che finiscono in 5 o in 0 (il 2015, il 2020, il 2025,...).

La scelta che l'UST chiede ai Cantoni di compiere entro il 31 dicembre 2009 ha già fatto in Ticino alcuni passi importanti: sono stati consultati tutti i Dipartimenti dell'Amministrazione Cantonale, così come le istituzioni addette alla ricerca scientifica (Supsi e USI). Manca solo l'apporto di chi, al di fuori di questi attori, ha pure interesse a conoscere la nostra realtà sociale (per la parte che tradizionalmente i censimenti rendono “visibile”). Ci rivolgiamo perciò a questa parte di società, che si presenta qui come “lettore” della nostra rivista. Da lui non possiamo aspettarci che sappia districarsi tra intervalli di fiducia e modelli di densificazione. Vogliamo però sollecitarlo a una cosa: ripensi a fondo a quali dati dei passati censimenti ha usato, di quali non ha potuto fare a meno, pensando sia al loro contenuto specifico (la struttura familiare, la professione esercitata, ..), sia al loro dettaglio geografico (il cantone, le regioni, singoli comuni). Fatta questa ricognizione, ci comunichi i suoi bisogni attuali, ci dica cosa si aspetta di ottenere anche con il prossimo censimento. L'indirizzo mail e il numero telefonico sono quelli che compaiono in grande qui sotto, e che **vi preghiamo** di utilizzare **entro il 9 di ottobre**: le vostre prese di posizione, il contatto che stabilirete con noi ci serviranno per rendere più solida la scelta che il Ticino deve fare.

Ustat

**dfe-ustat@ti.ch**

**091 814 64 23**

«Ci rivolgiamo perciò al nostro lettore: ci comunichi i suoi bisogni attuali, ci dica cosa si aspetta di ottenere anche con il prossimo censimento.»

