statistiche report FORMAZIONE DEGLI ADULTI

Nota metodologica

Strategia di campionamento e livello di precisione dei risultati

1. Obiettivi conoscitivi

La *popolazione di interesse* dell'indagine in oggetto, ossia l'insieme delle unità statistiche intorno alle quali si intende investigare, è costituita dalle famiglie residenti in Italia e dagli individui con età compresa tra 18 e 74 anni ad esse appartenenti, al netto dei membri permanenti delle convivenze. La famiglia è intesa come *famiglia di fatto*, ossia un insieme di persone coabitanti e legate da vincoli di matrimonio, parentela, affinità, adozione, tutela o affettivi.

I *domini di studio*, ossia gli ambiti rispetto ai quali sono riferiti i parametri di popolazione oggetto di stima, sono:

– l'intero territorio nazionale;

– le quattro ripartizioni geografiche previste dalla classificazione NUTS1 (Italia Nord-Occidentale, Italia Nord-Orientale, Italia Centrale, Italia Meridionale e Insulare);

– la tipologia comunale ottenuta suddividendo i comuni italiani in quattro classi formate in base a caratteristiche socio-economiche e demografiche:

A) *comuni appartenenti all'area metropolitana* suddivisi in:

A_1, *comuni centro dell'area metropolitana*: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari;

A_2, *comuni che gravitano intorno ai comuni centro dell'area metropolitana*;

B) *comuni non appartenenti all'area metropolitana* suddivisi in:

B_1 comuni aventi fino a 10.000 abitanti;

B_2 comuni con oltre 10.000 abitanti.

2. Strategia di campionamento

2.1. Descrizione generale del disegno di campionamento

Il disegno di campionamento è di tipo complesso e si avvale di due differenti schemi di campionamento. Nell'ambito di ognuno dei domini definiti dall'incrocio della regione geografica con le quattro aree A_1, A_2, B_1 e B_2, i comuni italiani sono suddivisi in due sottoinsiemi sulla base della popolazione residente:

- l'insieme dei comuni Auto Rappresentativi (che indicheremo d'ora in avanti come comuni AR) costituito dai comuni di maggiore dimensione demografica;
- l'insieme dei comuni Non Auto Rappresentativi (o NAR) costituito dai rimanenti comuni.

Nell'ambito dell'insieme dei comuni AR, ciascun comune viene considerato come uno strato a se stante e viene adottato un disegno noto con il nome di *campionamento a grappoli*. Le unità primarie di campionamento sono rappresentate dalle famiglie anagrafiche, estratte in modo sistematico dall'anagrafe del comune stesso; per ogni famiglia anagrafica inclusa nel campione vengono rilevate le caratteristiche oggetto di indagine di tutti i componenti di fatto appartenenti alla famiglia medesima.

Nell'ambito dei comuni NAR viene adottato un disegno a due stadi con stratificazione delle unità primarie. Le Unità Primarie (UP) sono i comuni, le Unità Secondarie sono le famiglie anagrafiche; per ogni famiglia anagrafica inclusa nel campione vengono rilevate le caratteristiche oggetto di indagine di tutti i componenti di fatto appartenenti alla famiglia medesima.

I comuni vengono selezionati con probabilità proporzionali alla loro dimensione demografica e senza reimmissione, mentre le famiglie vengono estratte con probabilità uguali e senza reimmissione.

2.2. Definizione della dimensione campionaria

La dimensione del campione teorico in termini di famiglie è stata prefissata a livello nazionale essenzialmente in base a criteri di costo ed operativi e posta pari a circa 6000 interviste

Si è anche definito che il numero di comuni campione interessati dovesse essere prossimo a 500, in modo da consentire un buon lavoro di controllo e supervisione.

L'allocazione del campione tra i domini di stima è stata effettuata sulla base della valutazione degli errori di campionamento attesi delle principali stime di interesse dell'indagine in oggetto riferite agli individui e relative alla partecipazione a corsi di formazione. Le informazioni su tali stime sono state desunte dall'indagine Multiscopo "I cittadini e il tempo libero" dell'anno 2006, a livello degli incroci tra ripartizione geografica e tipologia comunale a 4 modalità.

Per ottenere l'allocazione migliore delle 6000 interviste prefissate tra i domini di stima per tenere sotto controllo la precisione attesa delle stime di interesse, si è fatto ricorso a una metodologia di allocazione ottima basata su un'estensione al caso multi dominio del metodo di allocazione multivariato di Bethel (1989)[1]. Poiché le stime sono riferite agli individui, si è allocato un campione di circa 8200 individui adulti, corrispondenti a 6000 famiglie.

2.3. Stratificazione e selezione delle unità campionarie

L'obiettivo della stratificazione è quello di formare gruppi (o strati) di unità caratterizzate, relativamente alle variabili oggetto d'indagine, da massima omogeneità interna agli strati e massima eterogeneità fra gli strati. Il raggiungimento di tale obiettivo si traduce in termini statistici in un guadagno nella precisione delle stime, ossia in una riduzione dell'errore campionario a parità di numerosità campionaria.

Nell'indagine *Adult Education*, i comuni vengono stratificati in base alla loro dimensione demografica e nel rispetto delle seguenti condizioni:

- autoponderazione del campione a livello regionale;
- selezione di $\bar{n}=3$ comuni campione[2] nell'ambito di ciascuno strato definito sui comuni dell'insieme NAR;
- scelta di un numero minimo di famiglie da intervistare in ciascun comune campione; per l'indagine in oggetto tale numero è stato posto pari a 11;
- formazione di strati aventi ampiezza approssimativamente costante in termini di popolazione residente.

Il procedimento di stratificazione, attuato all'interno di ogni dominio territoriale individuato dalle aree A_1, A_2, B_1, B_2 di ciascuna ripartizione territoriale, si articola nelle seguenti fasi:

[1] Bethel J. (1989). Sample Allocation in Multivariate Surveys, Survey Methodology, 15, 47-57.

[2] Si è stabilito di selezionare 3 comuni da ogni strato per consentire la suddivisione del campione sui 3 mesi del trimestre di rilevazione. Per i comuni AR, che in tal caso sono coinvolti nella rilevazione tutti e tre i mesi, il campione di famiglie è stato definito divisibile per 3.

- ordinamento dei comuni del dominio in ordine decrescente secondo la loro dimensione demografica in termini di popolazione residente;
- determinazione di una soglia di popolazione per la definizione dei comuni AR, mediante la relazione:

$$ {}_d\lambda = \frac{{}_d\bar{m}\,{}_d\delta}{{}_d f} $$

in cui per la generica regione geografica *d* si è indicato con: ${}_d\bar{m}$ il numero minimo di famiglie da intervistare in ciascun comune campione; ${}_d\delta$ il numero medio di componenti per famiglia; ${}_d f$ la frazione di campionamento a livello del dominio *d*;

- suddivisione di tutti i comuni nei due sottoinsiemi AR e NAR: i comuni di dimensione superiore o uguale a ${}_d\lambda$ sono definiti come comuni AR e i rimanenti come NAR;
- suddivisione dei comuni dell'insieme NAR in strati aventi dimensione, in termini di popolazione residente, approssimativamente costante e all'incirca pari a: ($\bar{n}$) ${}_d\lambda = 3\,{}_d\lambda$.

Effettuata la stratificazione, i comuni AR sono inclusi con certezza nel campione; per quanto riguarda, invece, i comuni NAR, nell'ambito di ogni strato viene estratto un comune campione con probabilità proporzionale alla dimensione demografica, mediante la procedura di selezione sistematica proposta da Madow[3].

La selezione delle famiglie da intervistare in ogni comune campione viene effettuata dalla lista anagrafica di ciascun comune senza reimmissione e con probabilità uguali.
La realizzazione del disegno campionario ha previsto una dimensione effettiva del campione di 6022 famiglie distribuite in 462 comuni italiani.

Nel prospetto 1 viene riportata la distribuzione dell'universo e del campione dei comuni, delle famiglie e degli individui per ripartizione geografica.

Prospetto 1. - Distribuzione per area geografiche dei comuni, delle famiglie e delle persone nell'universo e nel campione- Anno 2011

Ripartizioni	**Universo Situazione al 31/12/2011**			**Campione realizzato**		
	Popolazione	*Comuni*	*Famiglie*	*Comuni*	*Famiglie con almeno un componente di 18-74 anni*	*Individui di 18-74 anni*
Nord-ovest	15752503	3059	7233029	137	1758	3388
Nord-est	11442262	1480	5035742	108	1345	2595
Centro	11591705	996	5033968	80	1249	2454
Sud e Isole	20607737	2557	8102924	138	1486	3156
Italia	**59394207**	**8092**	**25405663**	**463**	**5838**	**11593**

3. Le stime campionarie

L'indagine deve produrre le stime riferite al numero di individui che nella popolazione di riferimento (i 18-74enni, per questa indagine) possiedono una certa caratteristica o il livello di una quantità misurata sugli individui. Per il calcolo dei coefficienti di riporto all'universo si utilizza una procedura generalizzata di stima, basata sull'uso di una famiglia di stimatori, noti in letteratura come *calibration estimator* (stimatori di ponderazione vincolata). La metodologia alla base di tali stimatori

[3] Madow, W.G. (1949) "On the theory of systematic sampling II", Ann. Math. Stat., 20, 333-354

consente la determinazione di un coefficiente di riporto all'universo in grado di produrre stime coerenti a totali noti, desunti da fonti esterne, e correlati alle principali variabili oggetto di indagine.

La famiglia di stimatori di ponderazione vincolata coincide asintoticamente con lo stimatore di regressione generalizzato: per campioni sufficientemente grandi, quindi, tali stimatori hanno approssimativamente le stesse proprietà, ovvero sono corretti, consistenti e con la stessa varianza campionaria[4].

La strategia adottata per la costruzione dei coefficienti di riporto all'universo si sviluppa attraverso le fasi tipiche utilizzate per la costruzione degli stimatori nelle varie indagini campionarie dell'Istituto. In particolare possiamo distinguere:

- la determinazione della probabilità di inclusione di ogni unità statistica e del relativo peso diretto, pari all'inverso della probabilità di inclusione;
- calcolo dei coefficienti di correzione per mancata risposta totale;
- determinazione dei coefficienti di riporto all'universo finali vincolati ai totali noti desunti da fonti esterne all'indagine.

3.1. La probabilità di inclusione e il peso diretto

Il principio su cui è basato ogni metodo di stima campionaria è che le unità appartenenti al campione rappresentino anche le unità della popolazione che non sono incluse nel campione stesso. A tale scopo, ad ogni unità campionaria viene attribuito un peso, o coefficiente di riporto all'universo, che indica quante unità della popolazione sono rappresentate, rispettivamente, da ogni unità presente nel campione.

Senza perdere di generalità, definiamo la seguente simbologia:

U popolazione di riferimento oggetto di indagine;

y_k valore della variabile Y assunto dalla k-esima osservazione della popolazione;

y_j valore della variabile Y assunto dalla j-esima osservazione del campione;

π_j probabilità, assegnata dal disegno di campionamento, che l'unità j-esima sia inclusa nel campione S.

Il totale di una generica variabile Y, calcolato sull'intera popolazione, assume la seguente forma:

$$Y = \sum_{k \in U} y_k \qquad (1)$$

Il disegno di campionamento assegna le probabilità di inclusione ad ogni unità del campione in modo tale che

$$\hat{Y} = \sum_{j \in s} y_j \frac{1}{\pi_j} \qquad (2)$$

sia uno stimatore corretto della (1).

Nel disegno di campionamento di questa indagine, la probabilità di inclusione di un generico individuo è data: dalla probabilità di estrazione del comune di residenza (direttamente proporzionale all'ampiezza demografica dei comuni all'interno dello strato); e dalla probabilità di

4 La metodologia è illustrata da Deville, J.C. e Särndal, C.E. in *Calibration Estimation in Survey Sampling*, Journal of the American Statistical Association, Vol. 87, n.418, 1992.

estrazione della famiglia di appartenenza tra le famiglie eleggibili del comune. Le famiglie di uno stesso comune hanno quindi tutte la medesima probabilità di inclusione.

Per una generica famiglia eleggibile *j*, nel comune *i* dello strato *h*, il peso diretto d_{hij}, inverso della probabilità di inclusione π_{hij}, assume la seguente forma:

$$d_{hij} = \frac{1}{\pi_{hij}} \qquad (3)$$

Nel corso della fase di raccolta delle informazioni presso le unità che formano il campione, come accade per tutte le indagini statistiche, alcune di queste si trovano nell'impossibilità di partecipare alle indagini. Per ovviare alla mancata partecipazione di alcune unità, sotto l'ipotesi che il comportamento dei rispondenti sia simile a quello dei non rispondenti all'interno di particolari sottogruppi di unità campionarie, il correttore per mancata risposta assume la forma dell'inverso del tasso di risposta (δ_z) all'interno del generico sottogruppo z: in questa indagine, i sottogruppi sono identificati tramite l'incrocio tra le quattro ripartizioni geografiche previste dalla classificazione NUTS1 e la tipologia comunale ottenuta suddividendo i comuni italiani in quattro classi formate in base a caratteristiche socio-economiche e demografiche, secondo le modalità descritte nel paragrafo 1. Il correttore per la mancata risposta è definito come:

$$\frac{1}{\delta_z} = \frac{m_z}{m_z^r} \qquad (4)$$

in cui m_z rappresenta il numero di famiglie che da disegno campionario avrebbero dovuto partecipare all'indagine in un generico sottogruppo z e m_z^r rappresenta il numero di famiglie effettivamente rispondenti nel generico sottogruppo z. In questa maniera, il coefficiente di riporto all'universo corretto per mancata risposta, da assegnare al campione rispondente, risulta essere una generica famiglia rispondente *j*, nel comune *i* dello strato *h*, appartenente al sottogruppo *z*:

$$k_{zhij} = d_{hij} \frac{1}{\delta_z} \qquad (5)$$

3.2. La calibrazione a fonti esterne

Per il calcolo dei coefficienti di riporto all'universo finali si adottano gli stimatori di ponderazione vincolata (*calibration estimator*). La metodologia si basa sull'utilizzo di opportune informazioni ausiliarie, sintetizzate in totali noti, che, correlate con le variabili principali oggetto di indagine, hanno la funzione di aumentare l'accuratezza delle stime. I pesi finali si ottengono risolvendo un problema di minimo vincolato, in cui la funzione da minimizzare è una funzione di distanza tra i pesi diretti corretti per la mancata risposta (*k*) e i pesi finali (*w*) degli individui del campione rispondente (S^r), e i vincoli sono proprio le condizioni di uguaglianza delle stime campionarie di alcune variabili ausiliarie con i rispettivi totali noti desunti da fonti esterne all'indagine.

$$\begin{cases} Min\left\{\sum_{j \in S^r} dist(k_j, w_j)\right. \\ \sum_{j \in S^r} \boldsymbol{x}_j * w_j = \boldsymbol{t} \end{cases}$$

dove t è il vettore dei totali noti e x_j è il vettore delle variabili ausiliarie osservate sulla j-esima unità campionaria appartenente al campione rispondente (S^r). La funzione di distanza utilizzata è la logaritmica troncata.
I totali noti introdotti come vincoli nel calcolo dei pesi finali consentono di migliorare l'accuratezza delle stime, poiché quanto più le variabili ausiliarie considerate sono correlate con le variabili oggetto d'indagine, tanto più si riduce la distorsione delle stime. Le stime campionarie sono state vincolate ai seguenti totali noti:

- popolazione residente di 18-74 anni per ripartizione territoriale, sesso e classi d'età (18-24, 25-34, 35-44, 45-54, 55-64, 65-74);
- popolazione residente di 18-74 anni per ripartizione territoriale, sesso e titolo di studio (nessun titolo o licenza elementare; diploma di scuola media; diploma di scuola secondaria superiore; titolo post-secondario).
- popolazione residente di 18-74 anni per ripartizione territoriale, sesso e condizione professionale autopercepita (occupato; disoccupato; inattivo).
- popolazione residente di 18-74 anni per ripartizione territoriale, sesso e tipologia del comune di residenza (Comuni centro di area metropolitana; Comuni periferia di area metropolitana; altri Comuni sotto i 10.000 abitanti; altri Comuni sopra i 10.000 abitanti).

La distribuzione della popolazione per titolo di studio e per condizione professionale autopercepita proviene dall'indagine sulle Forze di Lavoro.

3.3. Valutazione del livello di precisione delle stime

Le stime prodotte da un'indagine campionaria sono sempre affette da errore. Questo si distingue in errore campionario, che deriva proprio dall'incertezza derivante dall'aver osservato la variabile di interesse solo su una parte (campione) della popolazione; ed errore non campionario, che deriva essenzialmente da errori nelle liste della popolazione utilizzate per selezionare le unità del campione; mancate risposte parziali dovute a risposte mancanti o non ammissibili a causa di errori di rilevazione o di registrazione; in generale, da tutto ciò che ha a che fare con le tecniche di indagine utilizzate e i comportamenti dei rilevatori.
In questo paragrafo si descrivono le metodologie e le tecniche utilizzate per la valutazione dell'errore campionario associato alle stime prodotte.
Le principali statistiche per valutare l'errore campionario sono l'errore di campionamento assoluto e l'errore di campionamento relativo. La stima dell'errore di campionamento assoluto e relativo di una generica stima $\hat{Y}$ sono definite dalle seguenti espressioni:

$$\hat{\sigma}\left(\hat{Y}\right) = \sqrt{\widehat{Var}\left(\hat{Y}\right)}$$

$$\hat{\epsilon}\left(\hat{Y}\right) = \frac{\hat{\sigma}\left(\hat{Y}\right)}{\hat{Y}}$$

Conoscendo la stima $\hat{Y}$ di un parametro Y della popolazione e la stima dell'errore assoluto $\hat{\sigma}\left(\hat{Y}\right)$ ad essa associato, è possibile costruire un intervallo di confidenza che, con livello di fiducia α, contiene al suo interno il valore del parametro Y oggetto di stima; tale intervallo è:

$$\left\{\hat{Y} - k_\alpha \hat{\sigma}\left(\hat{Y}\right) \leq Y \leq \hat{Y} + k_\alpha \hat{\sigma}\left(\hat{Y}\right)\right\}$$

dove il valore di k_α dipende dalla forma della distribuzione campionaria dello stimatore e dal valore scelto per il livello di confidenza *α*; per grandi campioni si fa comunemente riferimento alla distribuzione normale e si ha ad esempio, per α = 0,05, che k = 1,96.

3.4. Presentazione sintetica degli errori campionari

Ad ogni stima generica stima $\hat{Y}$ corrisponde una stima dell'errore campionario relativo che consente di valutarne la precisione; pertanto, per consentire una corretta interpretazione delle stime prodotte, sarebbe necessario presentare contestualmente a ciascuna stima anche il corrispondente errore campionario stimato. Ciò, tuttavia, non è possibile quando le stime prodotte sono in numero molto elevato. Per questi motivi si ricorre frequentemente ad una presentazione sintetica delle stime degli errori campionari, basata sul metodo dei modelli regressivi. Questo metodo si basa sulla determinazione di una semplice funzione matematica che mette in relazione ciascuna stima con il proprio errore campionario relativo stimato.
Il modello utilizzato per le stime di frequenze assolute e relative riferite agli individui è il seguente:

$$\log\left(\hat{\varepsilon}^2\left(\hat{Y}\right)\right) = a + b * \log\left(\hat{Y}\right)$$

dove i parametri *a* e *b* sono stimati con il metodo dei minimi quadrati. I modelli regressivi del tipo descritto, che permettono la presentazione sintetica degli errori di campionamento, sono stati ottenuti tramite un software generalizzato messo a punto dall'Istat.

Nel prospetto 2 sono riportati i valori dei coefficienti *a* e *b* e del coefficiente di determinazione R^2 dei modelli stimati per l'interpolazione degli errori campionari relativi delle stime di frequenze assolute e relative per il totale Italia e per diversi domini di stima.
Utilizzando gli opportuni coefficienti è possibile calcolare una stima dell'errore campionario relativo di una generica stima di una frequenza $\hat{Y}$ applicando la seguente formula:

$$\hat{\varepsilon}\left(\hat{Y}\right) = \sqrt{\exp\left(a + b * \log\left(\hat{Y}\right)\right)}$$

Prospetto 2 – Valori dei coefficienti a, b e dell'indice di determinazione R2 (%) delle funzioni utilizzate per le interpolazioni degli errori campionari delle stime riferite alle persone per totale Italia, ripartizione geografica, tipo di comune – Anno 2012

ZONE TERRITORIALI	*a*	*b*	R^2 *(%)*
ITALIA	9.4613	-1.08932	94.0462
RIPARTIZIONE GEOGRAFICA			
Nord-ovest	9.7938	-1.13257	94.0619
Nord-est	10.0246	-1.14452	88.8633
Centro	10.9475	-1.21816	92.3105
Sud e Isole	12.4727	-1.2879	91.843
TIPO DI COMUNE			
Area urbana centro	10.4036	-1.18977	94.205
Area urbana periferia	10.5687	-1.19526	93.6828
Comuni fino a 10mila abitanti	10.9565	-1.19752	90.7354
Comuni oltre i 10mila abitanti	10.5462	-1.16389	91.9892

Infine, il prospetto 3 ha lo scopo di rendere più agevole e immediata la valutazione degli errori campionari. In fiancata sono elencati valori crescenti di stima (20.000, 30.000, 40.000, …) di frequenze assolute; in testata sono riportati i domini di riferimento delle stime; le celle interne contengono gli errori campionari relativi percentuali stimati mediante la formula precedente. Consultando queste tavole è possibile disporre di una valutazione immediata (anche se meno precisa rispetto all'applicazione della formula precedente), dell'errore campionario di una generica stima di una frequenza assoluta (o relativa), cercando nella fiancata il valore che più si avvicina alla stima di interesse e in testata il dominio di riferimento.

Prospetto 3 – Valori interpolati degli errori campionari relativi percentuali delle stime riferite alle persone per totale Italia, ripartizione geografica, tipo di comune – Anno 2012

		RIPARTIZIONE GEOGRAFICA				TIPO DI COMUNE			
STIME	*ITALIA*	*Nord-ovest*	*Nord-est*	*Centro*	*Sud e Isole*	*Area urbana centro*	*Area urbana periferia*	*Comuni fino a 10mila abitanti*	*Comuni oltre i 10mila abitanti*
20,000	51.5	49.1	51.9	57.2	86.8	50.2	53.0	63.7	61.3
30,000	41.3	39.0	41.2	44.7	66.9	39.4	41.6	49.9	48.4
40,000	35.3	33.2	34.9	37.5	55.6	33.2	35.0	42.0	40.9
50,000	31.3	29.2	30.7	32.7	48.1	29.1	30.7	36.8	35.9
60,000	28.3	26.4	27.7	29.3	42.8	26.1	27.5	33.0	32.3
70,000	26.0	24.2	25.4	26.7	38.8	23.8	25.1	30.1	29.5
80,000	24.2	22.4	23.5	24.6	35.6	22.0	23.2	27.8	27.3
90,000	22.7	20.9	22.0	22.9	33.0	20.5	21.6	25.9	25.5
100,000	21.4	19.7	20.7	21.5	30.8	19.3	20.3	24.3	24.0
200,000	14.7	13.3	13.9	14.1	19.7	12.8	13.4	16.0	16.0
300,000	11.8	10.6	11.0	11.0	15.2	10.0	10.5	12.6	12.7
400,000	10.1	9.0	9.4	9.2	12.6	8.4	8.9	10.6	10.7
500,000	8.9	7.9	8.2	8.1	10.9	7.4	7.7	9.3	9.4
750,000	7.2	6.3	6.5	6.3	8.4	5.8	6.1	7.3	7.4
1,000,000	6.1	5.4	5.5	5.3	7.0	4.9	5.1	6.1	6.3
2,000,000	4.2	3.6	3.7	3.5	4.5	3.2	3.4	4.0	4.2
3,000,000	3.4	2.9	3.0	2.7	3.4	2.5	2.7	3.2	3.3
4,000,000	2.9	2.4	2.5	2.3	2.9	2.1	2.2	2.7	2.8
5,000,000	2.5	2.2	2.2	2.0	2.5	1.9	2.0	2.3	2.5
7,500,000	2.0	1.7	1.7	1.5	1.9	1.5	1.5	1.8	1.9
10,000,000	1.7	1.5	1.5	1.3	1.6			1.5	1.6
15,000,000	1.4	1.2			1.2			1.2	1.3
20,000,000	1.2				1.0				1.1
25,000,000	1.1								

4. Esempi di calcolo degli errori campionari

Esempio 1

Nel 2012, il 51,5% delle persone di 18 anni e più (pari a 22 milioni e 648mila persone) ha partecipato negli ultimi 12 mesi ad attività di formazione. Si cerca il livello di stima che più si avvicina a 22.648.000 nella prima colonna del prospetto ** per il totale Italia.

L'errore relativo percentuale della stima considerata è pari a 1,2 %.

L'errore assoluto sarà: σ(22.648.000) = 0,012 x 22.648.000 = 271.776

L'intervallo di confidenza avrà come estremi : 22.648.000 - (1,96 x 271.776) = 22.115.319
22.648.000 + (1,96 x 271.776) = 23.180.681

Esempio 2

Considerando la stima precedente si possono ottenere valori più precisi dell'errore di campionamento operando mediante interpolazione lineare dei due livelli di stima consecutivi tra i quali è compreso il valore della stessa.
Tali livelli sono 20.000.000 e 25.000.000 ai quali corrispondono i valori percentuali 1,2 e 1,1.

L'errore relativo corrispondente a 22.648.000 è pari a :

$$\sigma(22.648.000) = 1,2 - (1,2 - 1,1) / (25.000.000 - 20.000.000) \times (22.648.000 - 20.000.000) =$$

$$1,14704\%$$

Il corrispondente errore assoluto è	0,0114704 x 22.648.000 = 259.782
e l'intervallo di confidenza avrà come estremi:	22.648.000 - (1,96 x 259.782) = 22.138.827
	22.648.000 +(1,96 x 259.782) = 23.157.173

Esempio 3

Il calcolo dell'errore può essere effettuato, direttamente, tramite la funzione interpolante:

$$\hat{\varepsilon}(\hat{Y}) = \sqrt{\exp\left(a + b\log(\hat{Y})\right)}$$

i cui parametri, riportati nel prospetto 2 alla riga Italia, sono i seguenti:

a = 9,4613 b = -1,08932

Per $\hat{Y}$ = 22.648.000 si ha:

$$\hat{\varepsilon}(\hat{Y}) = \sqrt{\exp\left(9,4613 - 1,08932 \times \log(22.648.000)\right)} = 0,011182$$

L'errore relativo percentuale è quindi pari al 1,1182% e il calcolo dell'errore assoluto e dell'intervallo di confidenza è del tutto analogo a quello degli esempi 1 e 2.