

**LA VALUTAZIONE DELL'ERRORE DI CAMPIONAMENTO DELLE  
STIME DI POVERTA' RELATIVA SECONDO LA TECNICA  
*REPLICAZIONI BILANCIATE RIPETUTE***

Claudio Pauselli, Claudia Rinaldelli

*ISTAT, Servizio Condizioni economiche delle famiglie*

## **Sommario**

Il presente lavoro è stato illustrato e discusso nell'ambito del Convegno *Povertà Regionale ed Esclusione Sociale* svoltosi in Istat il 17 dicembre 2003.

La stima ufficiale della povertà viene diffusa dall'ISTAT sulla base dell'indagine campionaria sui Consumi delle Famiglie. Una famiglia è definita povera quando la sua spesa mensile per consumi è inferiore o uguale alla linea di povertà relativa, valore soglia stimato dall'indagine medesima. La povertà non è quindi rilevata direttamente sulle famiglie ma è funzione del contesto campionario osservato. Per questo motivo, non è possibile utilizzare direttamente la metodologia standard di calcolo degli errori campionari al fine di valutare la precisione delle stime di povertà. In questo lavoro, viene presentata l'applicazione della tecnica di ricampionamento Replicazioni Bilanciate Ripetute, finalizzata al calcolo degli errori campionari delle stime di povertà; i risultati sono confrontati con quelli ottenuti dalla metodologia standard applicata introducendo l'ipotesi che la povertà sia osservata direttamente sulle famiglie campione.

## **Summary**

This paper has been reported and discussed in the Conference *Povertà Regionale ed Esclusione Sociale* hosted in Istat on the 17th december 2003.

In Italy official poverty estimates are calculated by ISTAT using the data of the Household Budget sample survey. Poverty measures are disseminated by the most important features of the households and by geographical areas. Poverty is not directly observed on households, but it is function of the observed sample: actually, it is defined as poor a household whose monthly consumption expenditure is equal or below to a threshold called the relative poverty line; the threshold is estimated using the sample data. This way of calculating poverty affects the use of the standard methodology applied in ISTAT for evaluating sampling errors. In other words, the standard methodology can be applied only assuming the following hypothesis: poverty is directly measured on selected households; this simplification should be not easily accepted in the scientific world. In this paper, we report the results of the application of the resampling techniques Balanced Repeated Replication, in evaluating the sampling errors of poverty estimates against standard methodology.

## **INDICE**

1. Introduzione
2. Le tecniche di stima degli errori di campionamento
  - 2.1. La metodologia standard di calcolo degli errori di campionamento
  - 2.2. La tecnica delle Replicazioni Bilanciate Ripetute
3. Risultati e conclusioni

## **La valutazione dell'errore di campionamento delle stime di povertà relativa secondo la tecnica *Replicazioni Bilanciate Ripetute*<sup>1</sup>**

### **1. Introduzione**

In Italia le stime ufficiali di povertà sono calcolate dall'ISTAT sulla base della spesa per consumi rilevata dall'indagine sui Consumi delle Famiglie (ISTAT, 2002a; ISTAT, 2003a); la povertà è quindi misurata attraverso un'indagine campionaria. Nell'estate di ogni anno, la *Nota Rapida* "La Povertà in Italia" costituisce il mezzo di diffusione più importante sulla povertà con riferimento all'anno precedente; le misure di povertà sono analizzate secondo le principali caratteristiche delle famiglie, a livello nazionale e per ripartizione geografica (Nord, Centro, Mezzogiorno).

Entrando più nel dettaglio, la misura della povertà si basa su un approccio statistico che combina l'uso della variabile economica spesa per consumi e della *linea di povertà* (stimata come spesa media per consumi per persona in una famiglia di due persone) opportunamente modificata per tenere conto della diversa ampiezza delle famiglie. In particolare, una famiglia viene definita povera nel caso in cui la sua spesa mensile per consumi sia inferiore o uguale alla linea di povertà (ISTAT, 2002b; ISTAT, 2003b).

La metodologia appena descritta evidenzia quanto segue: la povertà non è rilevata direttamente sulle famiglie ma dipende sia dalla linea di povertà relativa sia dalla spesa per consumi delle famiglie campione.

Le stime di povertà, come tutte le stime da indagine campionaria, sono affette da variabilità campionaria.

L'ISTAT, generalmente, diffonde le stime da indagine campionaria insieme ad alcuni indicatori (errore relativo di campionamento, errore assoluto di campionamento, intervallo di confidenza) che consentono di valutare la precisione delle stime medesime. A tal fine è stata sviluppata una procedura informatica per il calcolo dei suddetti indicatori; tale procedura, che si basa sulla metodologia *standard* di calcolo degli errori campionari, è ampiamente utilizzata nelle indagini campionarie eseguite dall'ISTAT (Falorsi, Rinaldelli, 1998).

---

<sup>1</sup> Il presente lavoro, illustrato e discusso nell'ambito del Convegno *Povertà Regionale ed Esclusione Sociale* (Istat, 17 dicembre 2003), è frutto del lavoro congiunto degli autori; Claudio Pauselli ha redatto i paragrafi 2.2 e 3; Claudia Rinaldelli ha redatto i paragrafi 1, 2 e 2.1.

Questa procedura non può essere direttamente applicata per ottenere gli errori di campionamento delle stime di povertà. La peculiarità del metodo di costruzione della stima di povertà limita pesantemente l'uso della *metodologia standard* di calcolo degli errori campionari; per meglio dire, tale metodologia può essere applicata soltanto introducendo l'ipotesi che la povertà sia rilevata direttamente sulle famiglie.

La complessità della valutazione della precisione delle stime di povertà è confermata dai pochi studi prodotti sull'argomento (Heinrich, 1996; Zheng, 2001); l'unica indicazione fornita, in ambito europeo, è quella sopra menzionata di considerare la povertà come attributo rilevato direttamente sulle famiglie (EUROSTAT, 2001).

La valutazione della precisione delle stime di povertà risulta quindi carente; tuttavia una adeguata valutazione della variabilità campionaria della povertà è divenuta sempre più importante da quando le stime sono diffuse a livello di regione geografica (Coccia et al., 2002; ISTAT, 2003c). Infatti, in linea generale, le stime che si riferiscono a livelli più disaggregati presentano errori di campionamento più elevati.

L'ISTAT si è quindi posto tra i suoi obiettivi metodologici l'individuazione di metodologie che consentano di fornire una misura più attendibile della precisione delle stime di povertà (Coccia et al., 2002).

In questo lavoro, presentiamo i risultati dell'applicazione della tecnica di ricampionamento Replicazioni Bilanciate Ripetute (BRR, Balanced Repeated Replication) proposta in letteratura nel caso di complessità funzionale delle stime di cui si intende misurare la precisione e qui utilizzata per calcolare gli errori di campionamento delle stime di povertà; tali risultati sono messi a confronto con quelli ottenuti dalla metodologia standard.

In particolare i paragrafi 2.1 e 2.2 descrivono rispettivamente l'applicazione della metodologia standard e della tecnica BRR per il calcolo degli errori campionari; infine il paragrafo 3 illustra i risultati delle applicazioni.

## 2. Le tecniche di stima degli errori di campionamento

L'errore relativo e assoluto di campionamento e l'intervallo di confidenza, utilizzati per valutare la precisione delle stime, sono funzione della varianza di campionamento delle stime medesime. Indichiamo con  ${}_d\hat{Y}$  una generica stima di interesse riferita al dominio territoriale  $d$  (ad esempio, la ripartizione geografica) e con  $\hat{V}\hat{a}r({}_d\hat{Y})$  la stima della sua varianza campionaria; la stima dell'errore di campionamento assoluto di  ${}_d\hat{Y}$  può essere scritto come:

$$\hat{\sigma}({}_d\hat{Y}) = \sqrt{\hat{V}\hat{a}r({}_d\hat{Y})} \quad (1)$$

e la stima dell'errore di campionamento relativo di  ${}_d\hat{Y}$  come:

$$\hat{\varepsilon}({}_d\hat{Y}) = \frac{\sqrt{\hat{V}\hat{a}r({}_d\hat{Y})}}{{}_d\hat{Y}} \quad (2)$$

infine gli estremi dell'intervallo di confidenza al 95% possono essere espressi come:

$$\hat{Y} - 1,96\hat{\sigma}(\hat{Y}) \quad \text{e} \quad \hat{Y} + 1,96\hat{\sigma}(\hat{Y}) \quad (3)$$

Non sempre è nota l'espressione di  $\hat{V}\hat{a}r({}_d\hat{Y})$ ; di conseguenza non è possibile calcolare le espressioni (1)-(3).

Questo si verifica anche per le stime di povertà. Come anticipato nel paragrafo 1, le stime di povertà dipendono sia dalla linea di povertà relativa, a sua volta stima campionaria, sia dalla spesa per consumi delle famiglie campione.

La varianza campionaria delle stime di povertà è quindi funzione sia della varianza campionaria della linea di povertà relativa sia della variabilità campionaria dello stato di povertà relativa delle famiglie.

Nei casi in cui non è nota l'espressione di  $\hat{V}\hat{a}r({}_d\hat{Y})$ , la letteratura propone due possibili soluzioni (Wolter, 1985; Zannella, 1989; Särndal et al., 1992):

a) ricavare, dove possibile, una espressione approssimata della varianza campionaria attraverso metodi di linearizzazione; b) applicare le tecniche di ricampionamento.

Le tecniche di ricampionamento consentono di ottenere, in maniera empirica, una stima approssimata della varianza campionaria delle stime di interesse, soprattutto in presenza di disegni di campionamento e stimatori complessi quando non è possibile ricorrere ad altri metodi (Rao, Wu, 1988; Sitter, 1992; Heinrich, 1996).

Le tecniche di ricampionamento sono, in linea generale, molto onerose da un punto di vista computazionale e spesso non si conoscono le proprietà teoriche degli stimatori della varianza campionaria proposti in questi metodi. Le proprietà teoriche di questi stimatori sono note quando la stima di interesse è di tipo lineare, ma in realtà, queste tecniche sono applicate proprio quando le stime di cui si vuole misurare la precisione non sono lineari. In questi casi, le proprietà statistiche degli stimatori della varianza campionaria sono legate a conclusioni tratte dall'evidenza empirica. La caratteristica di queste tecniche è quella di supplire alla mancanza di conoscenza teorica attraverso una intensiva elaborazione dei dati a disposizione.

In questo lavoro, presentiamo la valutazione degli errori campionari della stima "*incidenza di povertà delle famiglie*" a livello di regione, ripartizione geografica ed Italia per gli anni 1997-2002 ottenuta attraverso:

1) la metodologia standard applicata introducendo l'ipotesi che la povertà sia un attributo rilevato sulle famiglie campione; 2) la tecnica di ricampionamento Replicazioni Bilanciate Ripetute.

L'*incidenza di povertà delle famiglie* è calcolata come rapporto tra la stima del numero di famiglie povere e le famiglie residenti (totale noto esterno all'indagine):

$$\hat{I}_{\text{pov}} = \frac{\text{Stima famiglie povere}}{\text{Totale famiglie residenti}} \quad (4)$$

pertanto l'errore relativo di campionamento del *numero di famiglie povere* equivale all'errore relativo di campionamento dell'*incidenza di povertà delle famiglie*.

Si fa osservare che: l'ipotesi che la povertà sia rilevata nel campione equivale ad ipotizzare che la linea relativa di povertà non è stimata dal campione ma è esogena all'indagine e quindi non è affetta da varianza campionaria; pertanto, la varianza campionaria delle stime di povertà viene fatta dipendere esclusivamente dalla varianza campionaria dello stato di povertà relativa.

L'applicazione della tecnica di ricampionamento BRR equivale invece a considerare tutte e due le fonti di varianza campionaria delle stime di povertà ossia quella della linea di povertà relativa e quella dello stato di povertà relativa.

La descrizione delle due applicazioni è riportata rispettivamente nei paragrafi 2.1 e 2.2.

## **2.1. La metodologia standard di calcolo degli errori di campionamento**

La letteratura sulla teoria del campionamento da popolazioni finite, fornisce, per i più importanti disegni di campionamento, le formule per il calcolo della varianza degli stimatori frequentemente utilizzati nelle indagini (Kish, 1965; Cochran, 1977; Wolter, 1985; Zannella, 1989; Särndal et al., 1992).

Nel caso di stimatori lineari, vengono fornite le espressioni esatte, mentre per stimatori non lineari, ma di uso frequente, vengono messe a disposizione le espressioni approssimate. Si può quindi parlare di metodologia *standard* nel senso di metodologia notoriamente diffusa e applicata.

Le formule della varianza campionaria dipendono dal disegno di campionamento e dallo stimatore utilizzato; per procedere quindi al calcolo degli errori di campionamento è necessario predisporre di una opportuna procedura informatica.

In ISTAT, al fine di evitare la scrittura di programmi informatici ad hoc per ogni singola indagine campionaria, è stata sviluppata una procedura *generalizzata* che viene ampiamente utilizzata nelle indagini campionarie dell'Istituto (Falorsi, Rinaldelli, 1998). Tale procedura, basata sulla metodologia standard di calcolo degli errori campionari, è detta *generalizzata* poiché consente all'utente di specificare il disegno di campionamento, lo stimatore utilizzato e le stime di interesse (frequenze, totali, medie) rispetto alle quali calcolare gli errori di campionamento, senza apportare modifiche ai programmi informatici.

Gli errori di campionamento del *numero di famiglie povere* sono quindi stati calcolati attraverso la procedura informatica generalizzata introducendo l'ipotesi che la povertà sia un attributo osservato sulle famiglie campione (Rinaldelli et al., 2002). In tal modo è stato possibile usufruire della metodologia standard di calcolo degli errori campionari con riferimento al disegno di campionamento e allo stimatore utilizzati nell'indagine sui *Consumi delle famiglie*.

A tal proposito si ricorda sinteticamente che:

a) l'indagine sui *Consumi delle famiglie* è basata su un disegno di campionamento a due stadi di selezione dove le unità di primo stadio sono i comuni e le unità di secondo stadio sono le famiglie. I comuni sono stratificati per regione ed ampiezza demografica; in particolare, i comuni che costituiscono uno strato a sé (Comuni Autorappresentativi-AR) sono selezionati con certezza nel campione mentre dai rimanenti strati vengono estratti tre comuni campione (detti Non Autorappresentativi-NAR) con probabilità proporzionale alla dimensione demografica (ISTAT, 2002a);

b) le stime sono calcolate mediante lo stimatore di *ponderazione vincolata*, basato sull'uso di variabili ausiliarie, che consente di vincolare le stime campionarie a totali noti (nella fattispecie, la distribuzione della famiglie e della popolazione per regione, la distribuzione della popolazione per sesso e classi d'età<sup>2</sup> per ripartizione territoriale) (Falorsi, Rinaldelli, 1998; ISTAT, 2002a).

Considerando il disegno di campionamento e lo stimatore utilizzati nell'indagine,  $\hat{V}ar({}_d\hat{Y})$  è stata pertanto calcolata dalla procedura informatica come (Kish, 1965; Choudhry, Lee, 1987; Falorsi, Rinaldelli, 1998):

$$\hat{V}ar({}_d\hat{Y}) = \sum_{h=1}^{H_d} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (\hat{Z}_{hi} - \hat{\bar{Z}}_h)^2 \quad (5)$$

dove

$$\hat{\bar{Z}}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{Z}_{hi} \quad (6)$$

$$\hat{Z}_{hi} = \sum_{j=1}^{m_{hi}} z_{hij} \quad (7)$$

dove  $d$  denota il dominio di riferimento delle stime (ad esempio, la ripartizione geografica),  $h$  l'indice di strato,  $H_d$  il numero di strati nel dominio  $d$ ,  $n_h$  il numero di comuni campione nello strato  $h$ ,  $i$  l'indice di comune,  $j$  indice di famiglia,  $m_{hi}$  il numero di famiglie campione del comune  $i$  nello strato  $h$ ,  $z_{hij}$  denota la trasformata di  $y_{hij}$  rispetto allo stimatore (non lineare) utilizzato nell'indagine sui Consumi delle famiglie,  $y_{hij}$  denota la *povertà* considerata come variabile osservata:

---

<sup>2</sup> Le classi d'età utilizzate sono: 0-14, 15-29, 30-59, 60 e più.

$$y_{hij} = \begin{cases} 1 & \text{se la famiglia è povera} \\ 0 & \text{altrimenti} \end{cases} \quad (8)$$

I risultati dell'applicazione sono presentati nel paragrafo 3.

## 2.2. La tecnica delle Replicazioni Bilanciate Ripetute

La tecnica BRR è stata ideata nei primi anni 60 dal *Bureau of Census*; si descriverà la tecnica applicata al caso base di un disegno campionario ad uno stadio stratificato per estenderlo poi al caso generale di più stadi (Zannella, 1989).

Si supponga che la popolazione sia suddivisa in H strati e che da ogni strato siano estratte  $k=2$  unità campionarie. Con la selezione casuale di una unità per strato si ottiene un sottocampione chiamato *replicazione* di ampiezza pari alla metà del campione originale.

Sia  $a_h$  un coefficiente che assume valore +1 se la prima unità del generico strato h appartiene alla replicazione, o -1 se invece è la seconda unità dello strato a farne parte; allora l'insieme delle possibili replicazioni potrà essere descritto da una matrice di dimensione  $2^H \times H$ . Ogni cella di questa matrice assumerà valore +1 o -1.

In questa matrice valgono le seguenti relazioni (Zannella, 1989):

$$\sum_{r=1}^{2^H} a_{rh} = 0 \quad (h=1, \dots, H) \quad (9)$$

$$\sum_{r=1}^{2^H} a_{rh} a_{rk} = 0 \quad (h=1, \dots, H) \quad (10)$$

La (9) implica che le due unità campionarie sono presenti un ugual numero di volte nell'insieme di tutte le replicazioni possibili; la (10) che le colonne della matrice sono tra loro ortogonali.

Sia  $\hat{\theta}$  la stima lineare del parametro di interesse calcolata sul campione totale e sia  $\hat{\theta}_r$  la relativa stima calcolata sulla r-esima replicazione, si dimostra che (Zannella, 1989):

$$\hat{\theta} = \frac{1}{2^H} \sum_{r=1}^{2^H} \hat{\theta}_r \quad (11)$$

$$\text{Var}(\hat{\theta}) = \frac{1}{2^H} \sum_{r=1}^{2^H} (\hat{\theta} - \hat{\theta}_h)^2 \quad (12)$$

In caso di stimatori non lineari la (11) non è valida e la (12) è una stima della varianza.

Appare evidente che il metodo applicato per tutte le possibili repliche diventa estremamente oneroso anche con un H piccolo (per H=30 si hanno poco più di un miliardo di repliche). Si rende necessario pertanto determinare un  $R \ll 2^H$ ; la scelta casuale delle repliche implicherebbe una stima della varianza campionaria più elevata di quella ottenuta sull'insieme di tutte le repliche.

La soluzione a questo problema è stata fornita da McCarthy (McCarthy, 1969a-1969b) con l'introduzione delle repliche bilanciate ripetute (BRR).

Un sottoinsieme di R repliche è bilanciato se vale:

$$\sum_{r=1}^R a_{rh} a_{rk} = 0 \quad (h,k=1,\dots,H; k \neq h) \quad (13)$$

Se inoltre vale:

$$\sum_{r=1}^R a_{rh} = 0 \quad (h=1,\dots,H) \quad (14)$$

si dice completamente bilanciato.

La stima della varianza basata su R repliche bilanciate ripetute produce la stessa stima basata su tutte le  $2^H$  repliche.

Per costruire un sottoinsieme di repliche bilanciate si possono utilizzare le matrici di Hadamard (Zannella, 1989 - Wolter, 1985), . Le matrici di Hadamard sono delle particolari matrici di descrizione delle repliche di ordine quadrato k. Ogni sottoinsieme di  $k' < k$  colonne della matrice soddisfa le condizioni (13) e (14) ed è quindi completamente bilanciato; l'insieme di tutte le colonne soddisfa solo la condizione (13) ed è quindi bilanciato (Zannella, 1989).

La matrice di Hadamard di ordine 2 è:

$$\begin{pmatrix} +1 & +1 \\ +1 & -1 \end{pmatrix}$$

Sia  $M$  una matrice di Hadamard di ordine  $k$ ; la procedura iterativa per la costruzione di una matrice di ordine  $2k$  è (Wolter, 1985):

$$\begin{pmatrix} M & M \\ M & -M \end{pmatrix}$$

Oltre alla stima della varianza espressa da:

$$\text{Var}_1(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R (\hat{\theta} - \hat{\theta}_r)^2 \quad (15)$$

è possibile calcolare la stima alternativa:

$$\text{Var}_2(\hat{\theta}) = \frac{1}{R} \sum_{r=1}^R (\hat{\theta} - \hat{\theta}_r^c)^2 \quad (16)$$

dove  $\hat{\theta}_r^c$  è la stima del parametro calcolata sulla parte complementare del campione alla  $r$ -esima replicazione (ovvero quella che si otterrebbe moltiplicando per  $-1$  i coefficienti della matrice di descrizione delle replicazioni).

La media tra (15) e (16) fornisce una stima più precisa della varianza campionaria.

L'indagine campionaria sui *Consumi delle Famiglie* è a due stadi con stratificazione delle unità di primo stadio (comuni); è pertanto necessario l'adattamento del metodo BRR.

L'adattamento ad un disegno complesso si svolge in quattro passi (Wolter, 1985; Zannella, 1989):

- 1) si considerano solo le unità di primo stadio (PSU); solo queste saranno incluse nella replicazione assieme alla relativa stima del totale del carattere (in questo caso il numero di famiglie povere per PSU);
- 2) nel caso in cui fosse presente una sola PSU per strato si aggrega ad uno strato contiguo; tale operazione, nota come collassamento, implica una sovrastima della varianza;
- 3) nel caso in cui le PSU fossero maggiori di due si formano in modo casuale due pseudo-PSU in cui sono divise le PSU dello strato;
- 4) negli strati autorappresentativi le famiglie campione sono considerate PSU e sono suddivise in due pseudo-PSU secondo quanto descritto in 3.

Una volta formate le pseudo-PSU e calcolato il numero di strati risultanti si applica il metodo BRR.

Si fa notare che essendo la linea di povertà funzione del campione, ad ogni replicazione si rende necessario il suo calcolo e la riattribuzione del relativo stato di povertà alle famiglie.

Data la peculiarità dell'applicazione del metodo BRR, ossia il ricalcolo della linea di povertà per ogni replicazione, non è stato possibile avvalersi di software generalizzati esistenti.

Si è pertanto reso necessario la costruzione di una procedura informatica ad hoc in codice SAS™.

Con il metodo iterativo descritto precedentemente si è ottenuta una matrice di ordine 1024 superiore al numero di strati effettivamente rilevati nel periodo 1997-2002.

Per la stima della varianza si è utilizzata la media semplice delle due stime (15) e (16).

I risultati dell'applicazione sono esposti nel paragrafo 3.

### 3. Risultati e conclusioni

Come già anticipato, la stima dell'errore di campionamento relativo del numero delle famiglie povere (e quindi dell'incidenza di povertà) è stato calcolato nei seguenti modi:

1) considerando la povertà come un attributo rilevato (quindi si è posta la linea di povertà come fissa e non dipendente dai dati campionari) secondo la procedura generalizzata per il calcolo dell'errore campionaria descritta nel paragrafo 2.1; 2) tenendo conto anche della variabilità della linea tramite la tecnica BRR descritto nel paragrafo 2.2.

Nella tabella 1 sono riportati i risultati delle applicazioni con riferimento all'indagine sui Consumi dell'anno 2002.

**Tabella 1.** Errore rel.% dell'incidenza di povertà relativa delle famiglie secondo la metodologia standard e la tecnica BRR – anno 2002

<b>Regione</b>	<b>Incidenza povertà relativa</b>	<b>Errore rel. campionario % (metodologia standard)</b>	<b>Errore rel. campionario % (tecnica BRR)</b>
Piemonte	7,0	12,0	12,2
Valle d'Aosta	7,1	18,4	14,8
Lombardia	3,7	10,5	12,1
Trentino A.A.	9,9	9,9	11,7
Veneto	3,9	12,6	15,8
Friuli Venezia G.	9,8	11,4	12,3
Liguria	4,8	14,4	18,1
Emilia-Romagna	4,5	14,0	14,9
<b>NORD</b>	<b>5,0</b>	<b>5,0</b>	<b>5,3</b>
Toscana	5,9	12,2	13,8
Umbria	6,4	17,1	21,0
Marche	4,9	12,5	15,0
Lazio	7,8	9,3	9,7
<b>CENTRO</b>	<b>6,7</b>	<b>6,4</b>	<b>6,7</b>
Abruzzo	18,0	15,0	15,5
Molise	26,2	6,4	6,9
Campania	23,5	6,1	6,8
Puglia	21,4	8,6	9,6
Basilicata	26,9	11,6	14,0
Calabria	29,8	6,6	7,4
Sicilia	21,3	5,8	5,6
Sardegna	17,1	8,8	8,7
<b>MEZZOGIORNO</b>	<b>22,4</b>	<b>3,0</b>	<b>3,1</b>
<b>ITALIA</b>	<b>11,0</b>	<b>2,4</b>	<b>2,4</b>

Le applicazioni della metodologia standard e della tecnica BRR sono state eseguite con riferimento agli anni 1997-2002 anche se soltanto a partire dall'indagine sui Consumi dell'anno 2002 sono state diffuse le stime di povertà relativa a livello regionale. I risultati sono riportati nella tabella 2.

**Tabella 2.** Errore rel.% dell'*incidenza di povertà relativa delle famiglie* secondo la metodologia standard e la tecnica BRR – anni 1997-2001

<b>Regione</b>	<b>1997</b>		<b>1998</b>	
	<b>a</b>	<b>b</b>	<b>a</b>	<b>b</b>
Piemonte	14,4	16,0	11,5	14,5
Valle d'Aosta	13,0	14,8	13,4	17,3
Lombardia	11,3	11,1	12,2	13,5
Trentino A.A.	13,8	16,5	13,7	16,5
Veneto	13,4	15,5	15,5	15,8
Friuli V.G.	20,6	21,7	11,4	10,8
Liguria	16,3	15,6	13,2	14,4
Emilia-R.	12,6	15,4	14,2	18,0
<b>NORD</b>	<b>5,8</b>	<b>6,1</b>	<b>5,3</b>	<b>6,3</b>
Toscana	14,1	14,2	13,1	13,8
Umbria	15,8	19,6	14,1	14,8
Marche	13,7	15,7	18,7	26,5
Lazio	12,8	10,7	12,3	11,2
<b>CENTRO</b>	<b>7,6</b>	<b>7,5</b>	<b>7,7</b>	<b>7,7</b>
Abruzzo	9,6	10,1	10,2	11,5
Molise	13,7	17,8	9,0	9,3
Campania	5,4	6,1	6,0	6,3
Puglia	8,7	8,5	8,3	8,0
Basilicata	9,6	11,4	9,9	10,7
Calabria	6,5	8,1	9,3	9,2
Sicilia	6,0	6,1	6,5	6,9
Sardegna	8,3	9,8	9,2	11,2
<b>MEZZOGIORNO</b>	<b>2,9</b>	<b>3,1</b>	<b>3,2</b>	<b>3,2</b>
<b>ITALIA</b>	<b>2,5</b>	<b>2,7</b>	<b>2,6</b>	<b>2,6</b>

a – metodologia standard; b tecnica BRR

segue Tabella 2

Regione	1999		2000		2001	
	a	b	a	b	a	b
Piemonte	14,1	16,9	12,5	13,5	11,0	12,3
Valle d'Aosta	22,6	24,9	19,2	28,9	15,8	19,1
Lombardia	14,0	15,5	11,1	12,7	11,0	11,6
Trentino A.A.	14,8	16,9	12,3	15,8	12,4	14,9
Veneto	15,6	15,5	15,9	20,8	14,5	16,3
Friuli V.G.	14,1	17,0	13,4	14,2	13,3	11,6
Liguria	20,3	21,7	12,5	10,8	15,1	15,9
Emilia-R.	15,4	17,0	14,4	17,4	17,6	17,1
<b>NORD</b>	<b>6,4</b>	<b>6,9</b>	<b>5,5</b>	<b>6,1</b>	<b>5,4</b>	<b>5,5</b>
Toscana	13,4	12,5	10,7	12,8	18,4	17,6
Umbria	15,3	18,0	8,1	13,2	14,8	16,0
Marche	14,7	15,1	11,3	14,8	16,9	18,4
Lazio	13,3	14,8	9,1	11,2	8,9	9,1
<b>CENTRO</b>	<b>8,5</b>	<b>9,1</b>	<b>6,1</b>	<b>7,6</b>	<b>7,0</b>	<b>7,2</b>
Abruzzo	10,2	10,4	11,2	14,8	9,1	9,1
Molise	10,3	11,1	9,1	8,3	10,9	12,6
Campania	6,0	6,5	6,4	6,5	5,8	6,2
Puglia	7,2	8,5	8,5	9,1	6,6	7,1
Basilicata	10,2	11,5	8,2	8,4	7,3	9,0
Calabria	7,1	7,0	7,0	5,8	7,9	7,3
Sicilia	7,1	8,7	7,4	8,9	5,8	6,2
Sardegna	9,6	11,1	8,3	8,2	9,8	10,3
<b>MEZZOGIORNO</b>	<b>3,1</b>	<b>3,2</b>	<b>3,3</b>	<b>3,4</b>	<b>2,8</b>	<b>2,9</b>
<b>ITALIA</b>	<b>2,7</b>	<b>2,7</b>	<b>2,5</b>	<b>2,8</b>	<b>2,4</b>	<b>2,3</b>

a – metodologia standard; b tecnica BRR

Al fine di confrontare gli errori di campionamento illustrati nelle tabelle 1-2, si è provveduto a calcolare il loro rapporto. La tabella 3 riporta i valori del rapporto tra l'errore relativo di campionamento secondo la metodologia standard e l'errore relativo di campionamento secondo la tecnica BRR.

**Tabella 3.** Rapporto tra errore rel.% (metodologia standard) e errore rel. % (tecnica BRR)

<b>Regione</b>	<b>1997</b>	<b>1998</b>	<b>1999</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>
Piemonte	0,90	0,79	0,84	0,93	0,89	0,98
Valle d'A.	0,88	0,78	0,91	0,66	0,83	1,24
Lombardia	1,01	0,90	0,91	0,88	0,95	0,87
Trentino A.A.	0,84	0,83	0,88	0,78	0,83	0,85
Veneto	0,86	0,98	1,01	0,76	0,89	0,80
Friuli V.G.	0,95	1,06	0,83	0,94	1,15	0,93
Liguria	1,05	0,91	0,94	1,16	0,95	0,79
Emilia-R.	0,82	0,79	0,90	0,82	1,03	0,94
<b>NORD</b>	<b>0,94</b>	<b>0,84</b>	<b>0,93</b>	<b>0,90</b>	<b>0,99</b>	<b>0,94</b>
Toscana	0,99	0,95	1,07	0,84	1,04	0,88
Umbria	0,80	0,96	0,85	0,62	0,92	0,81
Marche	0,87	0,71	0,97	0,77	0,92	0,83
Lazio	1,20	1,10	0,90	0,81	0,98	0,96
<b>CENTRO</b>	<b>1,01</b>	<b>1,01</b>	<b>0,94</b>	<b>0,80</b>	<b>0,97</b>	<b>0,95</b>
Abruzzo	0,95	0,89	0,98	0,76	1,00	0,97
Molise	0,77	0,97	0,93	1,10	0,86	0,93
Campania	0,88	0,95	0,93	0,98	0,93	0,90
Puglia	1,02	1,04	0,85	0,94	0,92	0,89
Basilicata	0,84	0,92	0,89	0,97	0,81	0,83
Calabria	0,80	1,02	1,01	1,21	1,08	0,89
Sicilia	0,98	0,94	0,82	0,84	0,93	1,03
Sardegna	0,85	0,82	0,86	1,02	0,95	1,01
<b>MEZZO-GIORNO</b>	<b>0,95</b>	<b>0,98</b>	<b>0,97</b>	<b>0,95</b>	<b>0,97</b>	<b>0,97</b>
<b>ITALIA</b>	<b>0,91</b>	<b>0,99</b>	<b>0,98</b>	<b>0,91</b>	<b>1,03</b>	<b>1,00</b>

Le due applicazioni evidenziano quanto segue:

- 1) gli errori di campionamento stimati con la tecnica BRR sono generalmente superiori a quelli stimati con la metodologia standard;
- 2) le stime dell'errore di campionamento secondo la metodologia standard sono nel complesso abbastanza prossime a quelle ottenute con la tecnica BRR;
- 3) passando dal livello regionale a quello di ripartizione geografica e nazionale, decresce la differenza tra gli errori di campionamento stimati secondo le due differenti metodologie.

Sembra ragionevole supporre che la metodologia standard conduca a sottostimare gli errori di campionamento delle stime di povertà poiché non tiene conto della variabilità campionaria della linea di povertà; tuttavia l'entità di tale sottostima appare comunque contenuta visto la moderata variabilità che la linea di povertà presenta (si veda Tabella 4) (ISTAT, 2002b).

**Tabella 4.** Linea di povertà relativa e corrispondente errore relativo % campionario

<b>Anno</b>	<b>Linea di povertà relativa (in euro)</b>	<b>Errore rel. di campionamento %</b>
1997	738,84	0,69
1998	762,31	0,70
1999	770,72	0,71
2000	810,21	0,72
2001	814,55	0,60
2002	823,45	0,57

Si può quindi concludere che l'uso della metodologia standard, sotto ipotesi di semplificazione (che non consente di valutare tutte le fonti di variabilità campionaria delle stime di povertà relativa), sembra non costituire nei risultati una approssimazione così severa rispetto a quanto si otterrebbe dall'applicazione di una tecnica di stima della varianza campionaria più appropriata.

## RIFERIMENTI BIBLIOGRAFICI

Coccia G., Pannuzi N., Rinaldelli C., Vignani D., 2002, Verso una misura della povertà regionale: problemi e strategie, *Sesta Conferenza Nazionale di Statistica*, Roma 6-8 Novembre 2002, in : [www.istat.it](http://www.istat.it).

Choudhry G.H., Lee H., 1987, Variance estimation for the Canadian Labor Force Survey, *Survey Methodology*, vol.13, n.2. 147-161.

Cochran W.G., 1977, *Sampling techniques*, Wiley and sons, New York.

EUROSTAT, 2001, *Sampling*, Statistics on Income and Living Conditions (EU-SILC), doc. EU-SILC 51/01.

Falorsi S., Rinaldelli C., 1998, Un software generalizzato per il calcolo delle stime e degli errori di campionamento, *Statistica Applicata*, 10, 2, 217-234.

Heinrich G.A., 1996, The prince and the pauper revisited: a bootstrap approach to poverty and income distribution analysis using the PACO data base, in: [www.ceps.lu/paco/documents](http://www.ceps.lu/paco/documents).

ISTAT, 2002a, *I consumi delle famiglie*, Anno 2001, Annuario n.8.

ISTAT, 2002b, *La stima ufficiale della povertà in Italia 1997-2000*, Argomenti n.24.

ISTAT, 2003a, *Quanto hanno speso le famiglie nel 2002*, Nota Rapida n.1.

ISTAT, 2003b, *La povertà in Italia nel 2002*, Nota Rapida n.2.

ISTAT, 2003c, *La povertà e l'esclusione sociale nelle regioni italiane*, Statistiche in Breve, 17 dicembre 2003.

Kish L., 1965, *Survey sampling*, Wiley and sons, New York.

McCarthy P.J., 1969a, Pseudoreplication: Further evaluation and application of the balanced Half-Sample Technique, *Vital and Health Statistics*, Series 2 n.31, National Center for Health Statistics, Public Health Service, Washington, D.C.

McCarthy P.J., 1969b, Pseudoreplication: Half-Samples, *Review of the International Statistical Institute*, 37, 239-264.

Rao J.N.K., Wu C.F.J., 1988, Resampling inference with complex survey data, *American Statistical Association*, 83, 401, 231-241.

Rinaldelli C., Cesarini A., Colombini S., Masi A., 2002, Indagine sui consumi delle famiglie – Calcolo e valutazione dell’attendibilità delle stime di frequenze assolute di povertà, *Documento interno ISTAT*.

Särndal C-E., Swensson B., Wretman J., 1992, *Model assisted survey sampling*, New York: Spriger-Verlag.

Sitter R.R., 1992, Comparing three bootstrap methods for survey data, *The Canadian Journal of Statistics*, 20, 2, 135-154.

Wolter K. M., 1985, *Introduction to variance estimation*, New York: Spriger-Verlag.

Zannella F., 1989, *Tecniche di stima della varianza campionaria*, Manuale di tecniche di indagine, Collana ISTAT Note e Relazioni, anno 1989, n.1, vol.5.

Zheng B., 2001, Statistical inference for poverty measures with relative poverty lines, *Journal of Econometrics*, 101, 337-356.